Designing Human-Centered Algorithms for the Public Sector: A Case Study of the U.S. Child Welfare System

Devansh Saxena
Marquette University

DESIGNING HUMAN-CENTERED ALGORITHMS FOR THE PUBLIC SECTOR
A CASE STUDY OF THE U.S. CHILD WELFARE SYSTEM

by
Devansh Saxena
Social and Ethical Computing Lab
Department of Computer Science

A Dissertation submitted to the faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

Milwaukee, Wisconsin
August, 2023

ABSTRACT

DESIGNING HUMAN-CENTERED ALGORITHMS FOR THE PUBLIC SECTOR A CASE STUDY OF THE U.S. CHILD WELFARE SYSTEM

Devansh Saxena

Marquette University, 2023

Public sector agencies in the United States are increasingly seeking to emulate business models of the private sector centered in efficiency, cost reduction, and innovation through the adoption of algorithmic systems. These data-driven systems purportedly improve decision-making; however, the public sector poses its own unique challenges where policies, practices, and organizational constraints mediate all decisions. Algorithms that do not account for these pertinent aspects of professional practice frustrate practitioners, diminish the quality of human discretionary work, and amplify biases in decision-making. A human-centered research agenda can help us develop algorithms centered in social-ecological theories that support the decision-making processes of practitioners, incorporate novel sources of data, and offer a means to evaluate algorithms in their real-world contexts.

This dissertation draws upon a case study of the child-welfare system and outlines responsible pathways forward for the design of human-centered algorithms in the public sector and contributes a holistic understanding of a complex sociotechnical system through deep ethnographic work, the design of a theoretical framework for algorithmic decision-making in the public sector, and computational narrative analysis of a critical data source that can help contextualize critical factors and improve decision-making. It showcases the practical tradeoffs that need to be balanced for algorithm design - 1) at the human discretion level, I highlight different insertion points and goals of algorithms to augment practitioners' decision-making processes, 2) at the bureaucratic level, I highlight the constraints within which all decisions (human or algorithmic) must be made and how organizational resources can be leveraged to ensure the proper integration and adoption of an algorithmic system, 3) at the algorithmic level, I showcase how algorithm design can account for the uncertainties inherent within cases and support decision-making processes instead of providing predicted outcomes. This dissertation work has provided actionable steps for human-centered algorithm design to child-welfare leadership and public interest technologists that will further help ensure that decisions are centered in evidence-based practice and lead to positive outcomes for families.

ACKNOWLEDGEMENTS

Devansh Saxena

I would like to express my heartfelt gratitude to the following individuals who have played a crucial role in my academic journey and the completion of this dissertation. First and foremost, I am deeply grateful to my advisors, Dr. Shion Guha and Dr. Michael Zimmer, for their unwavering support, mentorship, and continuous belief in my abilities. Their insightful feedback, constructive criticism, and dedication to my academic growth have shaped my research and significantly influenced the outcome of this dissertation. I am truly grateful for their immense patience, support, and the countless hours they have invested in guiding me towards success.

I would also like to extend my sincere gratitude to my collaborators and mentors, Dr. Pamela J. Wisniewski, Dr. Karla Badillo-Urquiola, and Dr. Sheikh Iqbal Ahamed. Their valuable insights, thought-provoking discussions, and meticulous examination of my work have significantly improved the quality and rigor of my research and writing abilities. I deeply appreciate their time, expertise, and valuable feedback that have contributed to the overall strength of this dissertation.

I would also like to extend my deepest appreciation to my family for their unconditional love, encouragement, and sacrifices that have allowed me to succeed. To my loving wife Maggie, your unwavering support, encouragement, and understanding have been the cornerstone of my success. Your belief in me, even during the most challenging times when I did not believe in myself, has been my motivation to persevere and remain focused on my goals. The struggles and successes that we have shared together have made this journey more meaningful and rewarding and I am grateful for your presence in my life. I also want to thank my brother Piyush for his tremendous support throughout my PhD journey. His ongoing emotional, intellectual, and financial support offered me the freedom to freely learn, take risks, and make mistakes that he could not afford to make during his doctoral work. Last, but not least, I want to express my gratitude to my parents for their love and support and for motivating me to undertake this journey.

I would also like to acknowledge the tremendous support and camaraderie of my friends Nova, Romael, Eddie, and Anmol. Their friendship and encouragement have made this academic journey all the more enriching and enjoyable. Stimulating discussions that I continue to have with all of them have been invaluable in helping me shape my ideas and refine my research. I also want to thank the students who have devoted valuable time and insights to my research projects. Erina, Dahlia, Aryan, Yixin, and Charlie, I am very grateful for your support, enthusiasm, and hard work that has been vital to the success of this dissertation work.

I would like to express my gratitude to the program directors at Wellpoint Care Network and all the individuals who participated in my research studies, providing their valuable time and insights. Their contributions have been instrumental in shaping this dissertation work. And lastly, I would like to express my appreciation to all those who have contributed to my academic and personal growth, including my friends, mentors, professors, and fellow researchers. While I am unable to thank you all individually, know that I will forever be grateful for your kindness and generosity.

ACM STATEMENT ON PLAGIARISM

This dissertation includes peer-reviewed material that has been modified from first authored papers by Devansh Saxena, and has appeared in conferences and journal proceedings published by the Association for Computing Machinery (ACM). Devansh Saxena produced all the writing that has been reproduced in this dissertation.

The ACM's policy on the reuse of published materials in a dissertation is as follows: "Authors can include partial or complete papers of their own (and no fee is expected) in a dissertation as long as citations and DOI pointers to the Versions of Record in the ACM Digital Library are included."

The following list serves as a declaration of the Versions of Record for works included in this dissertation:

- Chapter 2 [397] is published in the Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems and was presented at the CHI'2020 conference. The paper is available at – https://doi.org/10.1145/3313831.3376229
- Chapter 3 [396] is published in the Proceedings of the ACM on Human-Computer Interaction (vol.5) and was presented at the CSCW'2021 conference. The paper is available at – https://doi.org/10.1145/3476089
- Chapter 4 [401] has been accepted for publication in the ACM Journal on Responsible Computing pending minor revisions.
- Chapter 5 [402] is published in the Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems and was presented at the CHI'2022 conference. The paper is available at – https://doi.org/10.1145/3491101.3519771
- Chapter 6 [404] is published in the Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems and was presented at the CHI'2022 conference. The paper is available at – https://doi.org/10.1145/3491102.3517742
- Chapter 7 [402] is published in the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems and was presented at the CHI'2023 conference. The paper is available at – https://doi.org/10.1145/3544548.3581308

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

In this chapter, I introduce the motivation for this dissertation, the methodology driving the studies, outline my research questions, and provide an overview of this dissertation by describing the structure of the chapters.

1.1 Motivation for Dissertation

Government agencies in the United States (U.S.) are increasingly looking to invest in innovative data-driven technologies that supposedly help reduce administrative costs and improve the citizens' overall experience when utilizing public services [463, 163]. The introduction of a broad range of information communication technologies (ICTs) in the public sector has allowed for digitally storing, processing, managing, and sharing of information [326]. Consequently, this digital restructuring of government agencies has led to improved data-sharing practices between different sectors of the government, promoted minimal repeated information gathering, provided targeted services to clients, and reduced bureaucratic overhead [463, 167, 465, 163]. This technologically-driven administrative transformation has been referred to as *Digital Era Governance* by public administration scholars [76]. However, despite these macro administrative changes, ICTs have generally fused onto existing micro-level work practices rather than altering them and improving them at a deeper organizational level [463]. On the other hand, using ICTs over the last three decades has allowed public entities to continually collect data about citizens during their daily operations in managing and delivering public services [326]. This data includes structured data (for e.g., quantitative assessments and records of services), unstructured data (for e.g., case narratives) as well as metadata describing different attributes of citizens' interactions with public services (for e.g., level of cooperation, involvement with care) [326].

This comprehensive cross-sector data has allowed academics, technologists, and policymakers to narrow their focus on improving decision-making by developing data-driven practices centered in algorithmic decision-making that purportedly provide consistent, transparent, unbiased, objective, and defensible decisions to citizens [397, 163, 75]. Over the past two decades, several high-stakes decision-making domains such as the child-welfare system (CWS), the criminal justice system, education, and medical services have increasingly turned towards risk assessment algorithms or predictive risk models (PRMs) as a means to standardize and improve decision-making [214, 232, 397, 163]. However, audits of these systems reveal that they are achieving worse outcomes for families and exacerbating racial biases [171, 454, 344, 230]. Here, empirical associations derived from large datasets fail to capture the underlying assumptions that go into creating such datasets, the social nuances of governance labor, the power asymmetries that vulnerable communities experience, the ethical and value-laden choices that workers must make when balancing the needs of citizens against the demands of policymakers, as well the legislative constraints within which decisions must be made and which vary across jurisdictions.

This dissertation assesses the impact of algorithmic decision-making at the macro, meso, and micro levels of governmental operations in the child-welfare system with broader implications for the changing decision-making ecosystem in government agencies. At the macro level, we assess the impact of algorithmic systems on the legislative structures within which they operate and investigate whether the value propositions of cost-effective, standardized, and objective decision-making are holding true in practice. At the meso level, we unpack the theoretical underpinnings of predictive risk models (PRMs), the impact of the empirical notion of risk as compared to the sociological understanding of risk that underscores social work practice towards informing critical decisions, and the invisible data collection processes that power "empirical risk". At the micro level (or the street-level), we assess the impact of algorithmic decision-making on caseworkers' day-to-day practices as they make critical decisions about children and families using algorithms, how the nature of human discretionary work is changing as caseworkers acquire new skills while learning to interact with algorithmic systems, and the repair work they undertake to make these systems work for their clients.

Consider, for instance, algorithms in the criminal justice system that are developed to predict the risk of recidivism and efficiently allocate resources to high-crime neighborhoods. Neighborhoods that are initially targeted by these algorithms become the subjects of increased police patrols and investigations [464]. Consequently, more data is collected from these neighborhoods which feeds the machine-learning models and draws further attention to these neighborhoods. This creates a vicious cycle where low-income neighborhoods are continually over-surveilled by police departments [464]. In addition, low-level crimes such as public intoxication and loitering are added to the datasets to improve the predictive performance of such models [79]. Taken together, predicting risk through such algorithms is inherently a function of policing practices instead of an assessment of the true underlying risk of recidivism [350]. That is, the models only find what they continually observe in the training datasets and not cases that might otherwise "slip through the cracks". Parallels can be drawn to other public agencies such as child-welfare where the majority of the cases found in training datasets come from minor instances of neglect (e.g., a child's dental hygiene), leading to the over-prediction of risk of maltreatment for low-income communities, and ongoing data collection which further increases the likelihood of these communities being over investigated by the system. This coupled with neoliberal politics in the U.S. centered in austerity and privatization has also led to government agencies increasingly looking towards data-driven decision-making both as a means to reduce costs and allocate scarce resources efficiently [167, 465, 163]. Algorithmic decision-making in the public sector has generally been adopted in the form of predictive risk models (PRMs) with their primary purpose being the preemptive recognition and mitigation of 'risk'; a core principle of this shift in governing and of neoliberal economics [22]. That is, improving productivity, accountability, and efficiency by proactively identifying clients in the riskiest circumstances and targeting services towards them using algorithms [377, 163]. This further shift in *Digital Era Governance* that embeds neoliberal politics into the principles of public administration has been called *New Public Analytics* [489]. Here, the core focus is on risk management based on predictive characteristics derived from historical administrative data while driving attention away from the needs of individual clients as well as obfuscating structural and societal problems.

Facing severely limited resources and new dilemmas in the form of burdensome workloads and high staff turnover, most human services agencies have also turned towards algorithms as they purportedly promise to reduce costs and provide greater efficiencies in public policy and social services delivery [163, 299, 400]. In addition, public services such as the child-welfare system have also been the center of public and media scrutiny because of the harm caused to children who are removed from the care of their parents [86]. On the other hand, CWS also receives severe criticism and media attention for child abuse tragedies where the system failed to remove and protect a child [220, 182]. This has further mounted the pressure on CWS in several states in the United States (U.S.) to employ structured decision-making tools (and more recently, algorithmic decision-making) to prove that they employ evidence-based, consistent, and objective decision-making processes. Decades of research in clinical psychology and medicine exhibit that statistical decision-making out-performs human experts in prediction tasks [322, 215, 14] and is often cited as a justification for introducing algorithms in the public sector. Consequently, child welfare agencies have started acquiring algorithmic systems developed by tech startups [239, 371, 447, 238] or developed through public-private partnerships [101, 433]. A nationwide survey conducted by the American Civil Liberties Union (ACLU) in 2021 found that 11 states are currently using predictive analytics within their CWS agencies while 26 other states have considered using them [391]. The underlying principle here remains that *"bureaucracy is more perfectly developed the more it is dehumanized"* [79, 474] and that *"data becomes the promise for future bureaucratic efficiencies"* [20]. That is, public administration work must be conducted with precision, without ambiguity, impartially, and as speedily as possible with algorithmic systems offering a means to accomplish this [79, 463].

This dissertation work examines these claims and draws attention to the messy reality of data collection processes, the biases that become embedded in data, and how these biases become further concealed within and obfuscated by algorithmic

systems. It also showcases how human-centered data science can help outline pathways towards the ethical and responsible design of algorithmic systems that are centered in the nature of practice and augment the quality of human discretionary work. In sum, it is imperative to understand that any algorithmic system is simply a subsystem embedded within a broader sociotechnical system within which it interacts with different stakeholders, systemic constraints, and organizational pressures. This directly impacts the underlying assumptions that go into the design of this system as well as impacts how it is used in practice [415]. Researchers have also formulated the notion of *algorithmically infused societies* [468] (i.e., societies where every fabric is co-shaped by algorithmic and human behavior) to draw attention to the need for novel methodologies that can help researchers study complex interactions within sociotechnical systems and inform the design of human-centered systems that improve practitioners' decision-making processes and lead to positive outcomes for people. Here, Human-Centered Data Science (HCDS) has emerged as an interdisciplinary field that provides the necessary research methods for studying complex sociotechnical systems.

1.2 Human-Centered Data Science (HCDS)

Traditional data science practices seek to derive generalizable information from large datasets, turn them into actionable insights, and make predictions about similar scenarios observed in the data. This methodology poses significant challenges for studying sociotechnical systems in the public sector for several reasons - 1) public agencies such as criminal justice, child welfare, and public education have a long history of racially biased decision-making that disproportionately targeted families of color and we risk further exacerbating these racial biases and concealing them within machine learning algorithms by training models using this historical data, 2) quantitative administrative data in the public sector is quite fragmented, and machine learning algorithms are inclined to learning false empirical associations that may further target some populations (e.g., learning the zip code as a proxy for race [268]), and 3) cases that may appear similar based on a broad set of statistically significant predictors may be contextually very different. In sum, datasets capture a reductive representation of society in the form of quantifiable information, and machine learning algorithms further abstract out the social context within which these systems are situated [415]. However, as previously noted, the social context has direct implications for the design and ethical use of such systems and as a result, research methods are required that draw attention to the sociotechnical system.

Human-Centered Data Science (HCDS) has emerged as a research framework to precisely address these drawbacks in traditional data science and employs methods that highlight the complex interactions between humans, digital infrastructures, human-generated data, and society [28]. It places humans (i.e., users and affected communities) at the center of the data collection,

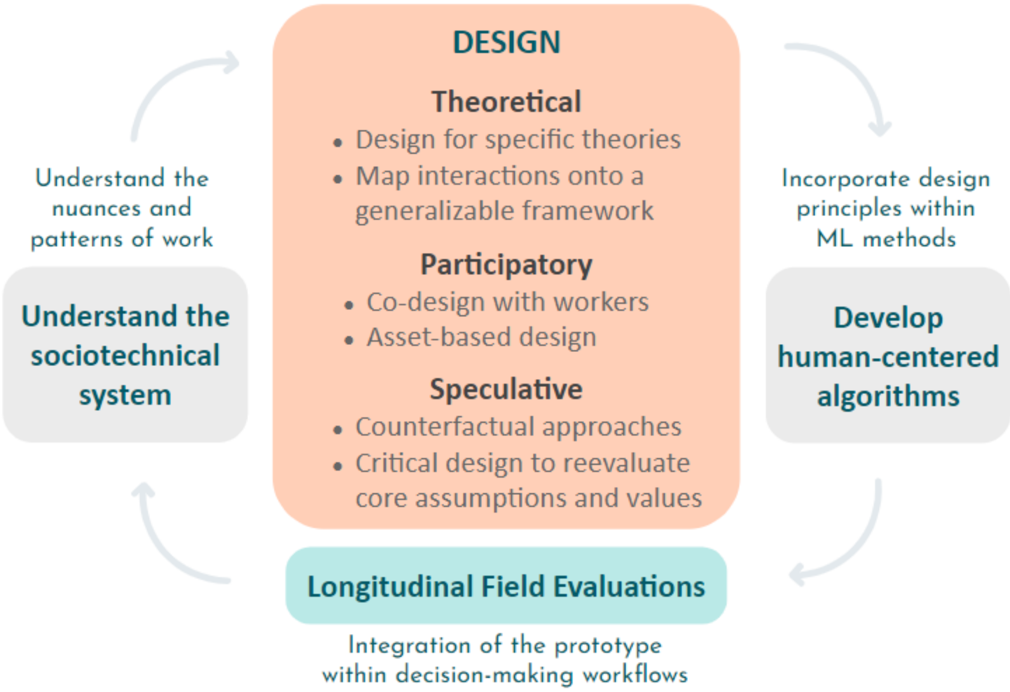


Figure 1: Human-Centered Data Science (HCDS) Research Process

analysis, and interpretation processes to ensure that the results are meaningful and actionable for the human stakeholders and not just the developers. HCDS is an interdisciplinary research framework that draws from the fields of human-computer interaction, social science, statistics, and computational techniques. It provides the theoretical scaffolding and the research and design methods that guide this dissertation work. A key characteristic of HCDS is that it incentivizes the combination of qualitative and quantitative methods. Qualitative techniques allow researchers to gain rich insights into specific phenomena, however, as it pertains to studying technology use, some of this rich information may be generalizable to other contexts. For instance, consider social workers' perspectives on predictive risk models (PRMs) and the sense-making strategies they engage in as they interact with these models. The ***humans*** (i.e., social workers with similar formal training) and the ***technology*** (i.e., PRM designed using similar machine learning techniques) are much similar across contexts and the interaction is governed by organizational workflows. Here, several generalizable aspects of practice (e.g., how social workers interpret algorithmic decisions) can help researchers design more human-centered systems and can also inform the contextual interpretation of results from quantitative data analysis. On the other hand, large-scale quantitative analysis can help researchers understand the latent patterns of practice, assess the impact of policies, and find bottlenecks in decision-making. Combining qualitative and quantitative techniques allows researchers to accomplish both - understand the depth of human behavior and understand humans at scale in regard to technology use or nonuse [28, 334, 46]. Finally, HCDS employs human-centered design techniques that facilitate the design of systems that are centered in the theory of practice and offer higher utility to practitioners. This dissertation work borrows from the three human-centered design principles described below. The HCDS research process is described in Figure 1.

- **Theoretical Design**: These design strategies seek to incorporate theories and concepts from social sciences into the design of algorithmic systems. This makes a system more human-centered because it is embedded in the practice model of workers and aligns well with their decision-making goals. In Chapter 3, we follow these strategies and introduce the first theoretical framework for algorithmic decision-making in the public sector by bringing together bodies of scholarly work in human-computer interaction (HCI), science and technology studies (STS), and public administration (PA). In Chapter 4, we also showcase the case study of a theory-driven algorithm designed using the principles of trauma-informed care that offers higher utility to caseworkers and leads to positive outcomes for families.
- **Participatory Design**: These design strategies seek to involve stakeholders in the design of algorithmic systems, interpretation of results, as well as assess the utility of novel algorithmic approaches. This ongoing engagement with stakeholders helps ensure that the designers make well-founded assumptions about the data and technology use which leads to the better integration of the prototype within organizational workflows. In Chapters 5 and 6, we engage in participatory methods with child-welfare stakeholders to understand the utility of casenotes in uncovering the invisible labor practices, systemic constraints, and power asymmetries in the child-welfare ecosystem. This ongoing engagement also helped us understand how we could algorithmically support the practitioners' cognitive environment without predicting decisions.
- **Speculative Design**: These design strategies especially focus on designing against the norm or the status quo and allow practitioners to shift their focus toward addressing underlying problems in practice and then brainstorm with designers which problems can be addressed technologically. This is especially important for innovative design because the range of what is technologically feasible continually changes. In Chapter 7, we engage in speculative design by drawing attention to the sociological understanding of risk as it occurs within child-welfare cases and impacts street-level decision-making. We show how sociological risk can be incorporated into algorithm design where the practitioners are also able to assess risks arising from procedural and systemic factors as well as the development of protective factors. This further helps us draw attention to the decision-making ecosystem, examine the temporality of different critical factors, and highlight how uncertainties arise in practice as a result of fluctuating and competing factors.

In practice, there is generally an overlap between these design strategies as researchers work closely with stakeholders and carefully deliberate over underlying assumptions and design choices. Here, HCDS also allows researchers and stakeholders better understand and attempt to bridge the sociotechnical gap [9] (i.e., the divide between what we know we must support socially and what we can support technically). This is especially important for human-centered and ethical computing because technological solutions in and of themselves may be insufficient for addressing systemic and structural problems in a sociotechnical system. It is, therefore, also necessary to draw attention to these problems and advocate for complementary policy changes. This dissertation also contributes a critical case study of a sociotechnical system (i.e., the child-welfare system) and showcases how HCDS methodologies can be used to uncover critical systemic interactions, invisible patterns of labor, power asymmetries, as well as experiences of affected communities that have direct implications for the design of human-centered algorithms.

1.3 Research Setting

The research described in this dissertation took place at Wellpoint Care Network (formerly, SaintA), a child-welfare agency in Milwaukee, Wisconsin. This agency is a non-profit private organization that is contracted by Wisconsin's Department of Children and Families (DCF) to provide child-welfare services to families that come under the attention of DCF. DCF employs its own Initial Assessment (IA) caseworkers that investigate allegations of child maltreatment and if maltreatment is substantiated, the case is officially brought into the system and referred to this agency to provide services and work towards achieving permanency (i.e., reunification, adoption, or transfer of guardianship) for the children. This agency serves about 900 families and 1300 children in the greater Milwaukee metropolitan area. Wisconsin's child-welfare system has historically suffered from racial biases where African-American children are two times more likely to be removed and Native-American children are three times more likely to be removed from the care of their parents as compared to their respective proportions in the state's child population. The state of Wisconsin has been under a federal lawsuit since 1993 brought on by the American Civil Liberties Union (ACLU) alleging that the state failed to supervise the Division of Milwaukee Child Protective Services (DMCPS) in its obligation to provide adequate child-welfare services to children and families [346]. A settlement was reached in 2002 which requires DMCPS to achieve specified performance outcomes regarding the permanency, safety, and well-being of Milwaukee County children who are placed in foster care. DMCPS has succeeded in meeting 12 of the 13 outcomes specified in the settlement but is struggling to meet a placement stability rate[1] of 90% for its foster children. To achieve this performance metric and further

[1] At least the following percentages of children in DMCPS custody within the period shall have three or fewer placements during the previous 36 calendar months of their current episode in DMCPS custody [346].

standardize decision-making DCF has implemented several algorithms throughout the process that child-welfare agencies are mandated to use.

Wellpoint Care Network has implemented a comprehensive trauma-responsive service delivery model where all caseworkers are trained in the principles are practices of trauma-informed care (TIC) [450]. The agency has fourteen case management teams each of which consists of six to eight case managers and one child-welfare supervisor. It also has specialized teams of permanency consultants and family preservation specialists. Permanency consultants assist each case management team where they are responsible for permanency planning, which includes the drafting and completion of documents for court and initiating the permanency legal process, targeted recruitment for adoption, provision of post-guardianship services, and/or post-adoption services for families. Family preservation specialists, on the other hand, help birth parents through parenting classes and in their efforts to achieve reunification with their children. The agency has developed specialized meetings (described in Chapter 3) called permanency consultations and 45-day staffings where caseworkers engage in collaborative decision-making and implement a trauma-informed protocol to assess every case from a TIC perspective. To ensure expedited permanency for foster children, this agency also employs concurrent planning such that two simultaneous plans begin when a child enters foster care – a plan for reunification with birth parents and a plan for adoption or transfer of guardianship if reunification is not possible.

This child-welfare agency served as the ethnographic field site where we conducted observations of these meetings to understand caseworkers' perspectives on algorithmic decision-making, the impact of systemic and policy-related factors on decisions, and how child-welfare staff collaboratively made decisions. We later conducted interviews with caseworkers who were frequently a part of these meetings to further contextualize findings from the observations and delve deeper into their understanding of different algorithms in use as well as the benefits and challenges as perceived by them. Caseworkers were also indispensable to the interpretation of results from the computational narrative analysis of casenotes as they were able to provide specific qualitative details and situated context about hyperlocal phenomenon observed by them in Milwaukee County and recorded in the casenotes.

1.4 Research Questions

Abebe et al. [6] highlight that much of the computational research that focuses on fairness, bias, and accountability on algorithmic systems continues to formulate "fair" technical solutions while failing to address deeper systemic and structural injustices. Through my dissertation work, we bring attention back to the sociotechnical and highlight social problems in child-welfare and how these problems become embedded in algorithmic systems. Next, we highlight how human-centered data science can help improve decision-making in child-welfare and support practitioners in their day-to-day practices. The findings of the studies discussed below have implications and are applicable to governance labor in the public sector. This dissertation assumes the dual roles of *computing as rebuttal* [6] where we highlight the technical limitations and feasibility of risk assessment algorithms, and of *computing as synecdoche* [6] by uncovering systemic complexities and social problems that directly impact families. This dissertation will also seek to make contributions at the intersection of gaps highlighted by the literature review in Chapter 2 and recommend solutions centered in strength and asset-based approaches that will improve the state of current algorithmic interventions, enhance child-welfare practice, and improve street-level decisions mediated through algorithms. Therefore, my dissertation answers the following overarching research questions:

- **RQ1**: How is algorithmic decision-making currently implemented in the United States (U.S.) Child Welfare System (CWS)?
- **RQ2**: How are algorithmic systems embedded within bureaucratic processes in child-welfare agencies and what are the implications for human-AI interaction?
  - **(a)** How do caseworkers interact with algorithms in their day-to-day work practices and how does trust and reliance on algorithms play out in practice?
  - **(b)** How do algorithmic systems impact the nature of professional practice, administration at the agency, and street-level decision-making?
- **RQ3**: How can human-centered data science (HCDS) help improve the practitioners' decision-making practices?
  - **(a)** Can theoretical signals derived from casenotes uncover patterns of invisible labor and help improve decision-making processes?
  - **(b)** How do we rethink "risk" in risk prediction and how do we incorporate sociological risk within algorithm design?

To answer these questions, I conducted six studies described below. Examining the nature of practice and street-level discretionary work as well as the impact of systemic and policy-related barriers on decision-making (human or algorithmic) allows us to develop technical solutions that operate within these constraints and augment the quality of human discretionary work.

1.5 Dissertation Overview

This dissertation fills in these gaps in research highlighted by the literature review, maps out pathways for the responsible design of algorithmic systems centered in strength- and asset-based approaches, and subsequently, makes contributions to the body of scholarly work on public interest technology, human-AI interaction, and responsible AI.

Research Questions	Contributions	Associated Study and Recognition
RQ1: How is algorithmic decision-making currently implemented in the United States (U.S.) Child Welfare System (CWS)?		
Which predictors, outcomes, and computational methods are being used to develop algorithms for CWS?	• Biases are embedded in not just the data but also the modeling choices and outcomes of algorithmic models • Human-centered design strategies can help develop theoretical algorithms, incorporate new sources of data, and refocus attention on positive outcomes	**CHI 2020** [397] • Best Paper Honorable Mention Award
RQ2: How are algorithmic systems embedded within bureaucratic processes in child-welfare agencies and what are the implications for human-AI interaction?		
(a) How do caseworkers interact with algorithms in their day-to-day work practices and how does trust and reliance on algorithms play out in practice?	• In-depth ethnography that uncovers caseworkers' interactions with a suite of algorithmic tools and the impact of systemic factors on algorithmic decisions • Collaborative decision-making ecosystem at the intersection of policies, social work practice, and algorithmic decision-making • Theoretical framework (ADMAPS) for designing public sector algorithms and conducting impact assessments	**CSCW 2021** [396] • Best Paper Honorable Mention Award • Impact Recognition Award
(b) How do algorithmic systems impact the nature of professional practice, administration at the agency, and street-level decision-making?	• Impact assessment to examine algorithmic harms caused to the nature of practice, administration, and street-level decision-making • Unpack how human factors (e.g., trust, reliance, explainability, and transparency) play out in practice	**ACM JRC** [401] (Journal on Responsible Computing)
RQ3: How can human-centered data science (HCDS) help improve the practitioners' decision-making practices?		
(a) Can theoretical signals derived from casenotes uncover patterns of invisible labor and help improve decision-making processes?	• Computational narrative analysis to highlight invisible labor, critical events, and key decision points over the life of child-welfare cases • Computational power analysis to highlight day-to-day power relationships between key personas • Identifying the right strategies to ensure timely interventions and equitable distribution of caseloads	**CHI 2022a** [405] **CHI 2022b** [404]
(b) How do we rethink "risk" in risk prediction and how do we incorporate sociological risk within algorithm design?	• Complexities in the decision-making ecosystem and street-level risk factors that impact families • Interplay between risk, protective, systemic, and procedural factors and their impact on caseworkers' decision-making • Uncertainties and confounding factors that arise in practice as a result of fluctuating and competing factors	**CHI 2023** [402] • Best Paper Award

Table 1: Overview of the Dissertation: Research Questions, Core Contributions, and Accompanying Studies and Recognition

In **Chapter 2**, I begin by first providing a comprehensive literature review of algorithms currently being used in the U.S. Child Welfare System. In this human-centered review, we critically examine the computational methods, predictors, and outcomes to uncover why current algorithmic approaches are failing to meet the desired end goals of consistent, fair, and objective decision-making (**RQ1**). The public sector poses its own challenges with respect to the ***technical*** (i.e., the quality of data), ***social and cultural*** (i.e., workers' interaction with algorithms and impact on organizational processes), ***theoretical*** (i.e., how is risk formalized), and ***societal*** (i.e., the impact of algorithms on communities and implications for public administration) implications of algorithmic decision-making.

Chapter 3 constitutes an in-depth ethnographic case study that we conducted at a child-welfare agency in Milwaukee, Wisconsin. This study contributes to the ***theoretical*** and ***social and cultural*** gaps highlighted by the literature review. Algorithms in the public sector is a domain in its own right and require a cohesive framework that explains how algorithms interact with bureaucracy and human discretion [76, 232, 72]. First, drawing on theories from human-computer interaction, science and technology studies, and public administration, we propose a theoretical framework for algorithmic decision-making for the public sector which accounts for the interdependencies between human discretion, bureaucratic processes, and algorithmic decision-making. The framework is then validated through a case study of algorithms in use at the agency. Second, the ethnography uncovers the daily algorithmic practices of caseworkers, what causes them to (dis)trust an algorithm, and how they navigate through different algorithmic systems especially when they do not account for policy/systemic barriers or resource constraints at the agency (**RQ2 (a)**). This is especially important because caseworkers are not trained in "thinking statistically" about data, algorithms, and uncertainties but are legally mandated to input data, interact with algorithms, and make critical decisions. In addition, all algorithmic decisions in the public sector must be made within the bounds of policies, current practices, and organizational constraints.

Chapter 4 further builds upon findings from the ethnography and highlights how algorithmic systems are embedded within a complex decision-making ecosystem at the intersection of the child-welfare and the judicial system and guide critical decisions throughout the child-welfare process (**RQ2 (b)**). In the previous chapter, we focused on the micro-interactions between the dimensions of human discretion, algorithmic decision-making, and bureaucratic processes to understand why algorithms failed (or succeeded) to offer utility to child-welfare staff and their impact on the quality of human discretionary work. In this chapter, we critically investigate the macro-interactions between these three elements to assess the impact of algorithmic decision-making on the nature of the practice, and the organization, as well as the interactions between human discretion and bureaucratic processes to understand how data-driven policy-making manifests itself on the street-level and whether algorithms are living up to the promises of cost-effective, consistent, and fair decision-making. This study contributes to the ***social and cultural*** and ***societal*** gaps highlighted by the literature review by highlighting how data-driven policymaking, as currently proposed, is at odds with how civil servants interpret and implement policies on the ground. We draw attention to the invisible human labor that goes into producing and maintaining datasets necessary for algorithmic systems as well as the repair work performed by caseworkers to make the algorithmic outputs work for their clients. Finally, we depict the case study of an algorithm that offers higher utility to caseworkers, however, required significant investments from the agency leadership to bring about that ecological change in decision-making where the algorithmic system plays an essential role.

Chapter 5 presents a study that conducts a quantitative deconstruction of a prominent risk assessment tool (i.e., Washington Assessment of Risk Model (WARM)) and compares it against the initial assessment caseworkers' casenotes about the same cases (**RQ3 (a)**). We found that WARM measures a parent's response to caseworkers' interventions as opposed to the efficacy of the interventions themselves. In addition, we found significant divergences between WARM risk ratings and the risks indicated in the caseworker narratives. This suggests a disconnect between quantitative and qualitative accounts of, ostensibly, the same underlying phenomenon. There are major discrepancies between the casenotes and risk assessments, i.e.- WARM scores did not mirror caseworkers' notes about family risk. This study provides empirical evidence about the biases that inadvertently become embedded in predictive risk models. We further show that critical information that can further contextualize risk factors can be derived from casenotes.

Chapter 6 presents a study that utilizes sources of information that have been hard to quantify so far, namely, caseworker narratives (**RQ3 (a)**). Child-welfare caseworkers are trained in writing detailed casenotes about their interactions with families and case progress through the life of the case. This study contributes to the ***technical*** and ***theoretical*** gaps illustrated by the literature review by deriving rich qualitative signals from case notes using natural language processing techniques such as topic modeling. We specifically analyze case notes written by the Family Preservation team that works closely with birth parents in their efforts to achieve reunification. The case notes offer a rich description of decisions, relationships, conflicts, and personas, as well as policy-related and systemic barriers. Analyzing these casenotes offers a unique lens towards understanding the workings of a team trying to achieve reunification; one of the primary policy-mandated goals of CWS. Theoretical signals derived from case notes also help contextualize the quantitative structured assessments and help assess the scope, utility, and insertion of algorithms systems that help improve decision-making practices.

Chapter 7 highlights the limitations of predictive risk models (PRMs) through a computational narrative analysis of child-welfare casenotes. It shows how there is a mismatch between how risk is quantified empirically based on administrative data versus how it is understood theoretically within the domain. This study contributes to the ***theoretical*** and ***societal*** gaps highlighted by the literature review. Algorithms model "risk" based on individual client characteristics to identify clients most in need. However, this understanding of risk is primarily based on easily quantifiable risk factors that present an incomplete and biased perspective of clients. In this study, we draw attention towards deeper systemic risk factors that are hard to quantify but directly impact families and street-level decision-making. Beyond individual risk factors, the system itself poses a significant amount of risk to families where parents are over-surveilled by caseworkers and experience a lack of agency in decision-making. we also problematize the notion of risk as a static construct by highlighting temporality and mediating effects of different risk and protective factors and show that any temporal point estimate of risk will produce biased predictions. Drawing attention to these theoretical signals embedded in casenotes can help child-welfare staff preemptively recognize risk factors that require immediate attention and lead to collaborative problem-solving and improve decision-making practices that expedite reunification for families (**RQ3 (b)**). We show how this understanding of sociological risk can be considered within algorithm design and further help caseworkers unpack complexities in decision-making that occur as a result of fluctuating and competing factors. We also draw caution against specific types of casenotes that are inappropriate for use in NLP-based algorithms by unpacking their limitations and biases embedded within them.

CHAPTER 2: LITERATURE REVIEW

ABSTRACT: The U.S. Child Welfare System (CWS) is charged with improving outcomes for foster youth; yet, they are overburdened and underfunded. To overcome this limitation, several states have turned towards algorithmic decision-making systems to reduce costs and determine better processes for improving CWS outcomes. Using a human-centered algorithmic design approach, we synthesize 50 peer-reviewed publications on computational systems used in CWS to assess how they were being developed, the common characteristics of predictors used, as well as the target outcomes. We found that most of the literature has focused on risk assessment models but does not consider theoretical approaches (e.g., child-foster parent matching) nor the perspectives of caseworkers (e.g., case notes). Therefore, future algorithms should strive to be context-aware and theoretically robust by incorporating salient factors identified by past research. We provide the HCI community with research avenues for developing human-centered algorithms that redirect attention towards more equitable outcomes for the Child Welfare System.

2.1 Introduction

As of September 2016, there were 437,465 children in the child welfare system (CWS) in the United States [348]. This is a significant (10%) rise in just 4 years since September 2012 [348], and this number is expected to keep rising unless significant efforts are made to improve youth outcomes [348]. Child abuse and neglect are severe issues that policymakers in the United States continue to battle with, and which is consistently at the foreground of public policy [117]. In recent years, CWS has been the center of public and media scrutiny [119] because of the potential damage done to the children who are removed from the care of their parents [136]. Therefore, there is significant pressure on CWS to systematize the decision-making process and show that these decisions were unbiased and evidenced-based [341]. For most policymakers, algorithmic decisions are perceived to be the epitome of being unbiased, evidence-based, and objective [455, 4]. Thus, algorithms have been developed for almost every aspect of services provided by CWS in different states. For instance, models have been developed for predicting the risk of future maltreatment event of a child [458], recommending appropriate placement settings [411], and matching children with foster parents who can meet the unique needs of every child [331]. Many of these algorithms have achieved various degrees of early success and have been shown to reduce costs [382] for CWS. However, they have also come under significant criticisms for being biased [108, 53], being opaque [455], complex and hard to explain [458, 106], being too reductive [115] and non-contextual [413] and for not incorporating factors that arise from relevant social science research literature [89].

The SIGCHI community is at the forefront of research on algorithmic bias [133, 64, 281], and has begun to examine some of the challenges of algorithmic decision-making within CWS. Brown et al. [72] studied community perspectives on algorithmic decision-making systems in CWS and found several aspects of algorithmic systems that bolstered distrust, perpetuated bias, concern over the lack of contextual understanding and 'black-box' nature of the algorithms, as well as concerns about how these algorithms may negatively impact child-welfare workers' decisions. Moreover, scholars outside of HCI have discussed how algorithms impact decision-making in CWS [86, 413, 419, 174]. Engaging in research that helps people and organizations, such as CWS, is well-suited and important for the HCI community. Therefore, a critical step in building a strategic research agenda is to synthesize the breadth of work that has already been done to identify a pathway forward. To forge this path, we posed the following research questions:

- **RQ1:** *What methods have been used to build algorithms in the child welfare system?*
- **RQ2:** *What factors (i.e., independent variables) have been shown to be salient in predicting CWS outcomes?*
- **RQ3:** *What outcomes (i.e., dependent variables) have CWS organizations been predicting?*

To answer these questions, we conducted a comprehensive literature review (n=50) of algorithms used for decision-making in CWS in the United States. We qualitatively analyzed these articles using the lens of human-centered algorithm design [46]. Overall, we found that the majority of the algorithms in CWS are empirically constructed, even though the empirical knowledge is quite fragmented [188]. Our results also revealed considerable differences in the predictors currently being used and those found salient in the literature. Finally, CWS has traditionally focused on 'risk assessment,' rather than positive outcomes that improve the lives of foster children. Based on Woobrock and Kientz's encapsulation of research contributions in HCI [485], this paper is a survey of the existing literature and makes the following contributions:

1. We apply a human-centered conceptual framework [46] to critically review the algorithms used within the U.S. child welfare system.
2. We introduce domain knowledge from the child welfare system to embed it within the SIGCHI community to allow for collaborative research between the two disciplines.
3. We identify the potential gaps in the existing literature and recommend research opportunities with careful attention to the human-centered design of algorithms to benefit CWS.

In the following sections, we discuss Human-Centered Algorithm Design and how we used this framework to inform our literature review methodology. Next, we situate our research within the SIGCHI community.

2.2 A Human-Centered Approach to Algorithm Design

As algorithms begin to permeate through every aspect of social life, HCI researchers have begun to ask, "Where is the Human?", that is, recognizing that humans are a critical, if not the central component of many domains for which Artificial Intelligence (AI) systems are being developed. A workshop organized at CHI 2019 [236], tackled this topic to identify several pertinent issues in algorithmic design, such as the opaque and isolated development of algorithms and a lack of involvement of the human stakeholders, who use these systems and are most affected by them. To address these problems, Baumer proposed Human-Centered Algorithm Design (HCAD) [46]; a conceptual framework founded in practices derived from human-centered design [237]. It incorporates human and social interpretations through both the design and evaluation phases [46]. Baumer [46] lays out three strategies that help algorithm design become more human-centered, namely, 1) theoretical, 2) speculative and 3) participatory strategies. We draw from the theoretical perspective to frame our research questions and as the qualitative lens for our analysis. Human-centered theoretical design strategy informs algorithm design as follows:

- Meaning-making: Theoretical foundations provide a much-needed scaffolding for dealing with complexity, identifying and evaluating design opportunities [365]. Designers need to study the socio-cultural domain in which they intend to situate their work.
- Design: Theoretical approaches aim to incorporate concepts and theories from social sciences into data science [46].
- Evaluation: The stakeholders' social interpretations of results can help ensure that the algorithm has higher utility and integrates well with practice.

CWS is one such domain that suffers from a complete lack of human perspectives throughout the design process. Therefore, our work focuses on how HCAD strategies can be employed to answer critical research questions in CWS.

2.3 Background

We situate our research within the SIGCHI community and provide an overview of the work that has been done to develop integrated data systems for CWS.

2.3.1 SIGCHI Research to Support the Child Welfare System

The SIGCHI community has recognized the importance of conducting research with organizations that help disadvantaged communities, such as those experiencing homelessness [441, 486] or recovering from substance abuse [309]. For example, Strohmayer, Comber, and Balaam [441] partnered with a center for people of low social stability to understand homeless young adults' perceptions of education. Similarly, Woelfer and Hendry [486] created a community technology center at a local service agency to work with homeless young people, case managers, and outreach workers. Similarly, SIGCHI researchers have started to engage with CWS to find ways to improve the lives of youth who have been displaced from their families. Some SIGCHI research has focused on foster youth and parents. For instance, Gray et al.'s [210] research with fostered and adopted children introduces a new digital memory box for creating and storing childhood memories. More recently, researchers have begun to study algorithmic decision-making systems within the child-welfare community. Badillo-Urquiola et al. [38] presented the challenges foster parents face mediating teens' technology use within the home.

Most relevant to our current work, Brown et al. [72] engaged in a participatory design effort and conducted workshops with families involved in CWS, child-welfare workers, and service providers. They found that participants were uncomfortable with algorithmic systems. Participants felt that these systems used deficit-based frameworks to make decisions and questioned the bias present within the data. Based on their findings, the investigators provide recommendations for researchers and designers to work together with public service agencies to develop systems that provide a higher comfort level to the community. Our study builds upon this related work by critically investigating the algorithms used within CWS and highlighting opportunities for future research. We provide a foundation for implementing human-centered approaches in the design and development of algorithmic systems for CWS.

2.3.2 Sociotechnical Systems for Child-Welfare

In this section, we provide the necessary background context about integrated data systems that laid the foundation for algorithmic work in CWS. In 1995, the federal government launched *SACWIS* (State Automated Child Welfare Information System) initiative to provide states with a federally funded and automated case management tool. These data systems allow states to collect and maintain data for program management and informing their decision-making [273]. States that implement *SACWIS* must also report their data to federal databases, such as *NCANDS* (National Child Abuse and Neglect Data System) [347] and AFCARS (Adoption and Foster Care Analysis and Reporting System) [348], to allow for the continual curation of comprehensive national databases. These data systems became the foundation for actuarial risk assessment tools, which have been mandated into practice, even though controversy still remains as to whether these tools should override the judgment of caseworkers who are most knowledgeable about a particular child's case [413, 412, 419, 469, 86].

Past survey papers have analyzed algorithms in CWS from a macro perspective, focusing on their reliability and validity with respect to consensus-based or clinical risk assessment models [413, 86]. Yet, they do not examine the mathematical or human-centered construction of these algorithms, that is, the techniques, the variable sets, or the outcomes predicted. This is especially important in CWS because each case of child neglect or abuse is contextually different and cannot be evaluated using the same set of significant predictors derived empirically [86]. To this end, we conducted a systematic literature review and identify the potential gaps in the literature with careful attention to the development of algorithms across time, as well as the methods and variable sets used.

2.4 Methods

We describe our scoping criteria, systematic literature search, and data analysis process.

2.4.1 Scoping Criteria: Defining Algorithms

To understand how "algorithms" are used in CWS, we first need to contextualize what we mean by algorithms. We conceptualized "algorithms" through the lens of *Street-level Algorithms*, a term recently coined by Alkhatib and Bernstein [16] in the HCI community. Street-level algorithms are algorithmically based systems that directly interact with and make on-the-ground decisions about human lives and welfare in a sociotechnical system [16]. From a more technical perspective, we use recent inclusive definitions [147, 279] for a whole suite of computational methods from statistical modeling (for e.g., generalized linear models) and machine learning. This allowed us to take a holistic view towards most forms of quantitative data analysis in CWS. Statistical modeling and machine learning are not mutually exclusive but we differentiate between them based on assumptions made about the data as specified by Breiman [68].

2.4.2 Systematic Literature Search

This study has been undertaken as a systematic literature review based on the guidelines proposed by Webster and Watson [476]. The unit of analysis for this literature review was peer-reviewed articles. We wanted to examine not just the algorithms currently being used in CWS but also newer solutions (algorithms) being proposed by researchers to better assess the current state of research. We used the following search terms to find papers at the intersection of CWS and algorithms – "child protective services," "child welfare," "foster care," "child and family services," "algorithm," "computation," "regression," "machine learning," "neural network," "data-driven," "actuarial," "computer program," "application". We used the following inclusion criteria for the articles:

- The paper was peer-reviewed, published work or a systems (or policy) report produced by a government agency.

Code	**n**	**Breakdown**
Peer reviewed	43	**40** (social science); **3** (computer science)
Agency report	7	—
Theory	5	**1** (implemented); **4** (proposed)
Psychometric scales	30	—
Actual system	27	**15** (RAs); **11** (PLs); **1** (MT)
Hypothetical system	23	**13** (RAs); **4** (PLs); **1** (MT); **5** (S-PL)
Model performance	35	—

RA: Risk Assessment model
PL: Placement Recommendation model
MT: Child-Foster parent Matching model
S-PL: Characteristics of successful placements

Table 2: Descriptive Characteristics of the Data Set

- The study (or report) engaged in a technical discussion about the computational methods, predictors, and outcomes.

Articles that did not meet these two criteria were considered irrelevant to this study and were not included in our review. We conducted a comprehensive search to identify relevant research across multiple disciplines. We searched a diverse set of digital libraries which included the ACM Digital Library, IEEE Xplore, Routledge, Elsevier, and Springer. We chose these libraries to take into account research published in multi-disciplinary conferences and journals. We then cross-referenced the citations of each article to identify additional articles or government reports that met our inclusion criteria. We did not place any constraints on our search based on the time period in which the papers were published. We identified **50** relevant articles that met our inclusion criteria.

2.4.3 Data Analysis Approach

To analyze our data, we conducted a structured qualitative analysis to answer our over-arching research questions. We used a grounded thematic process [66] to generate codes based on the data as shown in Table 3. We define theory in two ways – the system discussed in the study was developed using a theoretical framework or the system was developed theoretically based upon factors considered significant in evidence-based social work. The first author coded all of the articles, and co-authors were consulted to form a consensus around codes early in the coding process and again during coding to resolve ambiguous codes. We also coded for descriptive characteristics of the article set as shown in Table 2.

2.5 Results

In this section, we present our key findings from our review of the literature. We begin by first discussing the descriptive characteristics of our data set. Next, we organize and present the results by our three research questions, as shown in Table 3. Finally, we explore the relationship between the computational methods, predictors, and outcomes identified in our analyses.

Research Question	Dimension	Codes	n	%	Ex.
RQ1 (Computational Method)	Inferential Statistics	Generalized Linear Models	28	56%	[458]
		Discriminant Analysis/Statistical tests	6	12%	[410]
	Machine Learning	Supervised Learning	13	26%	[106]
		Unsupervised Learning	3	6%	[320]
RQ2 (Predictor variables)	Demographics	Child Demographics	20	40%	[23]
		Biological parents Demographics	10	20%	[252]
	Systemic Factors	Characteristics of Agency	2	4%	[331]
		Characteristics of Caseworker	1	2%	[331]
	Child Strengths	Child Strengths	11	22%	[104]
	Child Needs	Functioning	15	30%	[316]
		Child Behavioral/Emotional Needs	26	52%	[283]
	Child Risks	Suicide Risk	9	18%	[122]
		Child Risk Behaviors	20	40%	[382]
		Traumatic Experiences	30	60%	[39]
		Child Involvement in CWS	9	18%	[458]
	Bio-Parent Risk/Needs	Needs and Risky behavior	26	52%	[104]
	Foster Parents	Characteristics (income, occupation)	4	8%	[23]
		Preferences	2	4%	[331]
		Past performance	1	2%	[331]
		Capabilities (training/certifications)	1	2%	[331]
RQ3 (Outcome Variables)	Outcome	Risk of a future maltreatment event	28	56%	[252]
		Placement recommendation for a child	15	30%	[104]
		Matching children with foster parents	2	4%	[331]
		Characteristics of a successful placement	5	10%	[440]

Table 3: Structured Codebook: Dimensions are mapped onto their respective research questions

2.5.1 Descriptive Characteristics of the Data Set

The majority of the papers (n=40 or 80%) were published in social science venues with 3 papers (6%) published in computer science conferences, [8, 23, 106] all in 2018. We also included 7 reports (14%) from non-profit organizations, including the Children's Research Center [2]. One study discussed an algorithm that was theoretically constructed based on child-welfare research literature and four studies proposed theoretically-driven solutions. 30 papers (60%) employed psychometric scales [181] to assess the strengths, needs, and risks associated with foster children and/or biological parents. 27 papers (54%) discussed an actual algorithmic system and 23 papers (46%) proposed a new algorithmic system. Model performance was reported by 35 papers (70%).

2.5.2 Computational Methods used to Build Algorithms (RQ1)

In this section, we discuss the computational methods used to develop algorithms and organize them into the *Inferential Statistics* and *Machine Learning* dimensions.

Inferential Statistics approaches

Inferential statistics account for computational methods used in the majority of papers (68%), with 28 papers (56%) using a form of a generalized linear model (GLM). In Figure 2, we see a dramatic rise in the use GLMs after 1995, i.e., the post-*SACWIS* era. GLMs are being used to develop mostly two types of models; actuarial risk assessment and placement recommendation models. There was a general trend around the use of GLMs for developing risk assessment models [458, 86, 170]. We also identified two major concerns surrounding GLMs: their atheoretical and

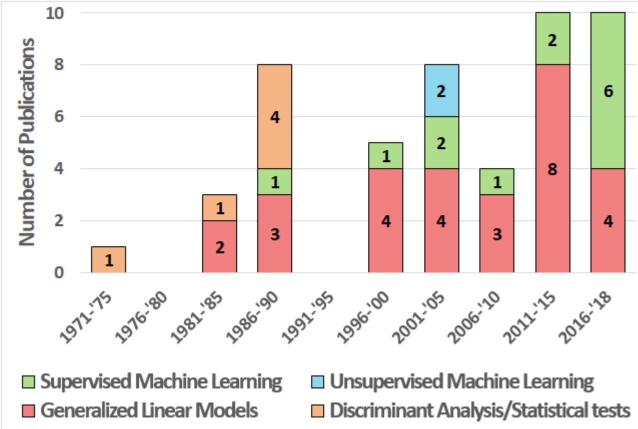


Figure 2: Methods used to build Algorithms (RQ1)

reductive nature and performance with respect to outliers.

Social scientists use validated psychometric scales [303, 84] to quantify the level of risk. GLMs have been developed using these psychometric scales, such as the CANS Algorithm [104] that only uses the most statistically significant items from the scale. This reductive and atheoretical model development has received criticism [86, 413, 419]. Each case of child neglect/abuse is contextually different and factors that are significant for one case might be peripheral to another. Moreover, GLMs do not account for the contextual factors that influence caseworker decisions leading to variable omission bias [86].

Outliers can significantly impact the performance of a regression model [353]. Traditionally, regression models seek to omit outliers as a means of improving predictive power and still account for the majority of the variance explained by significant variables [353]. However, for CWS, cases of severe abuse and neglect are the statistical outliers [65]. Regression models that are designed to predict the most moderate (average) outcomes tend to perform poorly on outliers [436]. For CWS, poor performance on outliers raises several ethical and accountability concerns [124].

Four papers (8%) used discriminant analysis to differentiate between the characteristics of foster children served by different placement settings. Figure 2 illustrates that discriminant analysis was a popular technique during 1985-1990, however, with the advent of regression techniques it gradually faded away. These papers were some of the earliest attempts at introducing algorithms to aid decision-making in CWS. However, the data was limited and its quality was questionable because of the lack of standardized data collection processes [423].

Machine Learning (ML) approaches

Machine Learning methods in CWS gained some momentum as early as 1986 with the introduction of *PLACECON*, a system designed to assist CWS with placement decisions [409]. However, with the increasing popularity of risk assessment models and limited funding available, resources were directed towards traditional regression models. Figure 2 shows a resurrection of ML methods starting in 2015 and growing interest within the computer science communities towards studying the research problems in CWS starting 2018 [8, 23, 106, 225]. Thirteen papers (26%) utilized ML methods in the form of decision trees, Bayesian networks, or inference trees. Decision-tree learning has been popular as a means of organizing large amounts of factual and empirical knowledge in the form of rules [422]. The CART (Classification and Regression Trees) algorithm has been recently used to build a child-foster parent matching system [23]. It has also been used to identify the characteristics of the most troubled children in CWS [122] as well as to study trends in child abuse and neglect data [8]. However, with such a strong emphasis on risk assessment, Children's Research Center [2] used ML methods to develop the Structured Decision-Making (SDM) model.

SDM is a decision-making framework where a risk assessment tool is used in conjunction with clinical assessment [252]. SDM utilizes an array of ML tools such as decision, value and inference trees, and Bayesian networks [213] and has been adopted by CWS in several states [61]. However, several studies have also shown that SDM produces mixed results especially when accounting for race and ethnicity [131, 134, 249]. There is also an ongoing struggle between the caseworkers' theoretical assessments and the tool's empirical judgment [413, 419]. Three papers (6%) used unsupervised ML methods in the form of neural networks [316, 320] and natural language processing (NLP) [69]. Brindley et al. [69] propose a web platform that allows foster youth to create personalized goals and talk to a chatbot that uses NLP to parse inputs and respond intelligently with recommendations about goals, finances, and housing. McDonald et al. [320] and Marshall et al. [316] propose the use of neural networks over regression techniques because their non-parametric approach performs better at modeling non-linear relationships and interactions.

One possible reason for the perpetual conflict between ML risk assessment tools and caseworkers' assessment might be at the core of Machine Learning itself and how it handles outliers. Statistical outliers in the case of child maltreatment are the most severe cases of child abuse and neglect [65]. Researchers [40] suggest that in the case of CWS, outliers are often more important for caseworkers and demand significant attention beyond the norm. Figure 2, depicts a significant dearth in the use of unsupervised learning methods with only two papers published in the early 2000s [320, 316] and one paper published in 2018 [69]. Employing neural networks in social sciences comes with its own complexities because there needs to be transparency about the proposed decisions [106]. Vaithianathan et al. [458] explored several ML methods such as Naive Bayes and Random Forests for risk assessment and achieved higher accuracies. However, they reverted to using a probit regression model because the outcomes were more explainable.

2.5.3 Predictors used in Algorithms (RQ2)

In this section, we examine the predictors that are being used in algorithms in CWS. Most algorithms are using over a hundred predictors so we systematically coded them and then grouped the emergent codes into seven dimensions (see Table 3).

Demographics and Systemic Factors

Child demographics were accounted for by 20 papers (40%) and biological parents demographics were accounted for by 10 papers (20%). Surprisingly, more than half the papers did not include child or parent demographics in their models even though racial and ethnic disparities in CWS have been recognized in social sciences [338, 372, 144]. Figure 3, illustrates that after 1990, there was a decline in the number of studies that used demographic variables in their algorithms. The *Systemic factors* dimension includes factors associated with CWS, such as characteristics of the agency and caseworkers. Two papers (4%) use variables relating to characteristics of the agency, such as location and staffing vacancies and one paper (2%) accounted for the characteristics of the caseworker, such as caseloads and the level of training. This is surprising because child-welfare literature acknowledges the impact caseworkers have on child outcomes [389, 89]. The caseworker is the child's primary contact between the biological parents, foster parents, and CWS. They navigate through the system and find services for children and families. In fact, caseworker turnover is directly associated with placement instability [89]. Factors that lead to high caseworker turnover include low salary, high caseloads, administrative burdens, low levels of training, and lack of supervisory support [89]. Systemic factors are one of the biggest reasons why children experience multiple placement moves in CWS [128]. This once again alludes to the atheoretical model construction that does not account for the salient factors well-established in evidence-based social work.

Foster-child related factors

Seven codes emerged out of the coding process and were grouped into three dimensions: child strengths, child needs, and child risks. 11 papers (22%) use variables that align with *Child Strengths*, such as interpersonal skills, coping skills, and level of optimism. Twenty-six papers (52%) took into account a child's emotional and behavioral needs and 15 papers (30%) recorded the child's day-to-day well-being and functioning, such as their school attendance and behavior, personal hygiene, and communication skills. We also coded for variables associated with risk factors that endanger child well-being. Suicide risk, risk behaviors, traumatic experiences, and child involvement with CWS were our four emergent codes that were grouped under the *Child Risks* dimension. 9 papers (18%) conducted a mental health screening to see if a child was suicidal

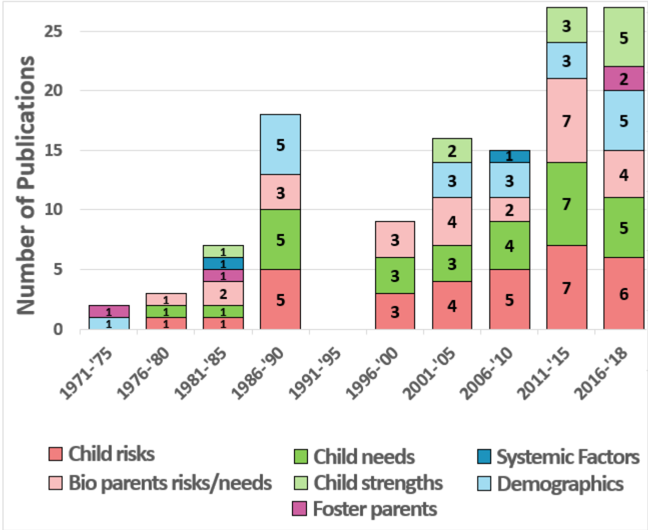


Figure 3: Predictors used in Algorithms (RQ2)

or having suicidal thoughts. 20 papers (40%) accounted for risk behaviors such as self-harm, recklessness, and social misbehavior, and 30 papers (60%) accounted for traumatic experiences such as neglect, physical/sexual abuse, history of family violence, and community violence. We noticed a trend here in that, almost all the risk assessment systems focused heavily on the *Child Risks* dimension, whereas, placement recommendation systems focused on the *Child Needs* dimension. Figure 3 depicts a rise in the number of studies that account for child strengths, child risk, and child needs since 1995, that is, the post-*SACWIS* era. This alludes to the fact that these child characteristics are well-documented by caseworkers in *SACWIS* and are being used for modeling purposes.

All the studies we reviewed accounted for foster child-related factors in terms of their needs and associated risks. However, only one study accounts for the child's interactions with other people, such as siblings, relatives, and the system itself. Moore et al. [331] account for factors such as *Placement with a sibling*, *Proximity to child's home/relatives*, and *Characteristics of the agency and caseworker*; factors well-studied in child-welfare literature [89]. Fluke et al. [170] found that placement decisions may be made as a result of interaction effects of non-case related factors such as characteristics of the agency and/or the caseworker. A study conducted in San Diego County found that 70% of the placement moves were a result of systemic or policy-related factors [244]; regardless of the child's or parent's circumstances.

Biological parents related factors

26 papers (52%) accounted for the biological parents' risk behaviors and needs, such as physical/mental health, substance abuse problems, residential stability, and knowledge of the child's

needs. We coded these variables into the *Bio-Parents Risks/Needs* dimension. Figure 3 shows that biological parent factors have been consistently used by several studies, however, we see a decline during 2005-2010. We also see a rise in the use of child-related factors during the same time period. The introduction of the CANS algorithm that focuses on the child's level of need may be a plausible explanation for this trend. Different algorithms are using biological parent-related variables differently. For example, risk assessment models quantify biological parents' risky behavior so as to discern the risk of future maltreatment. On the other hand, placement recommendation models are using this dimension to determine the level of trauma a child has experienced and recommend a placement setting based on their level of need. Factors surrounding biological parents have been studied in great detail and accounted for by most algorithms.

Foster parents related factors

Four papers (8%) that we reviewed accounted for the characteristics of the foster parents, that is, their income level, occupation, demographics, etc. Figure 3 shows that only 4 studies account for foster parent-related factors with a significant gap between 1985 and 2016 where no study accounted for these factors. Two papers (8%) look at the preferences of foster parents and one paper (2%) accounts for the foster parents' past performance and capabilities. Matching children with foster parents that are trained and prepared to meet their behavioral and emotional needs leads to increased stability for the children [89]. Matching children with foster parents that come from the same cultural background also leads to better outcomes because it leads to smoother transitions, lower stress, and a feeling of security for the children [73]. These factors are well-studied in literature [89, 378, 475], however, we see that very little research has been done from an algorithmic perspective. CWS has historically focused on ensuring safety and permanency rather than child well-being, that is, improving the quality of lives of foster children [53].

2.5.4 Target Outcomes of Algorithms (RQ3)

In this section, we examine the target outcomes of the algorithms used in CWS. Figure 4 depicts the trends in the target outcomes that algorithms have sought to model.

Risk Assessment

Predicting the risk of future maltreatment involves developing models using the empirical study of cases of child abuse/neglect [39]. The factors that show a strong association with abuse and/or neglect outcomes are selected to create an actuarial model which is then used to assess new cases of alleged abuse/neglect. Twenty-eight papers (56%) focused on predicting risk as their target outcome. Figure 4 illustrates that risk assessment has always received significantly more attention than any other outcome since the introduction of regression models in social sciences. The greatest criticism against these models is that they are not theoretically founded; these

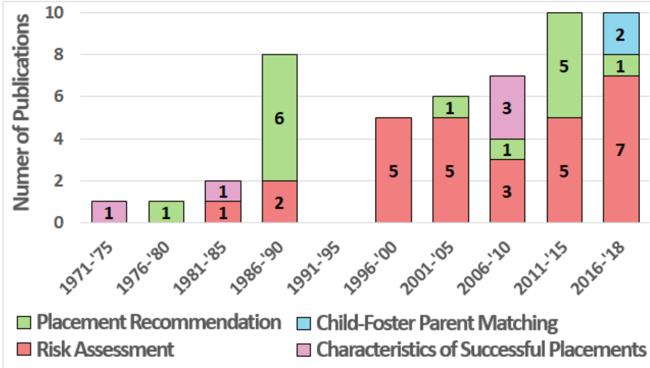


Figure 4: Target Outcomes of Algorithms (RQ3)

models are probabilistic in nature and not causal [39, 275, 413, 419]. Therefore, these models need to be empirically validated by follow-up studies to ensure their reliability. Direct comparison of any two actuarial models is a hard problem [39] and requires an in-depth understanding of the contexts in which the predictors were collected, measured, and weighted in the models.

Studies conducted on risk assessment models show that these models are more accurate at predicting target events like child maltreatment than unaided judgment, however, they lack utility [413]. Seven papers (14%) discuss the Structured Decision-Making (SDM) model, a framework that integrates predictive and contextual assessments. CWS in several states have developed their own versions of SDM, however, there are significant enough differences to treat them independently as part of our review. Even though SDM is designed to assist caseworker decisions, studies have found that there are constant disagreements between the tool (empirically-driven) and caseworker assessment (conceptually/theoretically-driven) to the point that caseworkers detest using these tools as they were intended [413]. However, caseworkers must continue to rely on these tools as a means of standardizing decisions, especially in cases of high uncertainty [419].

Placement Recommendations and Successful Placements

Models that focused on these two target outcomes were the precursors in the development of algorithms in CWS. Figure 4 depicts that these target outcomes were being studied during the time period of 1985-1990. However, no studies were published between 1991 and 2005 that focused on these target outcomes. A plausible explanation for this decline would be the increased focus on studying risk assessment during that period. 20 papers (40%) discussed recommendation systems for foster care placements.

The most prominent algorithm that determines the placement criteria based on a child's level of need is the CANS algorithm [104]. 6 papers (12%) discuss the CANS algorithm which is developed using the CANS psychometric scale [303]. CWS in a few states have developed their own versions of this algorithm, and therefore, were treated independently as part of our review. It makes a recommendation from six levels of care in the order of increasing severity – independent living, transitional living program, foster home, specialized foster care, group home, and residential treatment center. It is used in a hybrid approach in conjunction with a multi-disciplinary team which allows CWS to follow standardized admission criteria for cases with lower levels of uncertainty [104]. This is a good initial approach to ensure child safety, however, it is a minimal approach and does not seek to improve the quality of a child's life.

Child-Foster Parent Matching

This approach seeks to match the specific needs of a child with the capabilities of foster parents. That is, placing children with foster parents who are trained and certified to manage their needs. It is a proactive approach towards improving the quality of lives of children and not just minimizing the risk of maltreatment. Figure 4 shows that *Child-Foster parent matching* has only been implemented since 2015 (2 studies). This approach is different from the placement recommendation approach in that it addresses the specific needs of the child and the preferences of the caregiver. For instance, matching with respect to child temperament, parent temperament, and parental expectations leads to increased stability [378]. Placing children who have higher emotional needs with foster parents who prefer to be emotionally involved offers these children a better chance towards stability [472] than placing these children in a restrictive treatment setting. *Child-Foster parent matching* is well-studied in evidence-based social work and is known to improve stability and permanency outcomes [89, 378]. However, there is a dearth of information within CWS on how to guide this process [378]. This is a significant knowledge gap for both CWS and social scientists who seek to computationally model this approach. Moore et al. [331] recently validated a matching algorithm that was implemented by CWS in the state of Kansas for resulting in more stable placements.

		Computational methods			
		Supervised Machine Learning	Unsupervised Machine Learning	Generalized Linear models	Discriminant analysis/ statistical tests
Outcome Variables	RA	9	2	16	1
	PL	2	1	8	4
	MT	2	-	-	-
	S-PL	-	-	4	1

RA: Risk Assessment model
PL: Placement Recommendation model
MT: Child-Foster parent Matching model
S-PL: Characteristics of successful placements

Table 4: Crosstabs between Computational Methods (RQ1) and Outcomes (RQ3)

		Outcome Variables				Computational Methods			
		RA	PL	MT	S-PL	SUP	UNSUP	GLM	DAS
Predictors	Child demographics	8	6	2	4	7	1	7	5
	Bio-parents demographics	5	2	2	1	5	1	2	2
	Characteristics of Agency	-	-	-	2	-	-	2	-
	Characteristics of Caseworker	-	-	-	1	-	-	1	-
	Child Strengths	3	5	2	1	5	1	5	-
	Functioning	3	10	2	-	5	1	7	2
	Child Behavioral/Emotional Needs	7	13	2	4	7	-	13	6
	Suicide Risk	2	7	-	-	3	-	4	2
	Child Risk Behaviors	7	12	1	-	6	1	10	3
	Traumatic Experiences	15	10	2	3	9	1	16	4
	Child Involvement with CWS	2	6	1	-	3	-	3	3
	Bio-Parent Risk/Needs	15	8	-	3	7	1	16	3
	Foster parent characteristics	-	-	2	2	2	-	1	1
	Foster parent preferences	-	-	1	1	1	-	-	1
	Foster parent past performance	-	-	1	-	1	-	-	-
	Foster parent capabilities	-	-	1	-	1	-	-	-

RA: Risk Assessment model
PL: Placement Recommendation model
MT: Child-Foster parent Matching model
S-PL: Characteristics of successful placements
SUP: Supervised Machine Learning
UNSUP: Unsupervised Machine Learning
GLM: Generalized Linear models
DAS: Discriminant Analysis/Statistical tests

Table 5: Relationship between the Computational Methods (RQ1), Predictors (RQ2) and Outcomes (RQ3) used for designing child-welfare algorithms.

2.5.5 Relationship between Methods, Predictors and Outcomes

Relationship between Algorithms (RQ1) and Outcomes (RQ3)

Table 4 depicts crosstabs between the computational methods used and the outcome from all the papers in our corpus. We saw that generalized linear models have mostly been used for developing risk assessment models (16 studies) followed by placement recommendation models (8 studies). Even with the emergence of newer machine learning methods, the majority of the studies still continue to focus on risk assessment. 9 studies used supervised machine learning for risk assessment, 2 studies focused of placement recommendation, and 2 studies focused on child-foster parent matching.

Relationship between Predictors (RQ2) and Outcomes (RQ3)

Table 5 depicts the crosstabs between predictors used by computational models and the outcome they seek to predict. First, we cross-examine the risk assessment models with respect to the predictors that inform child characteristics. The majority of the models use a combination of predictors that assess *Child Behavioral/Emotional Needs* (7 studies), *Child Risk Behaviors* (7 studies), and *Traumatic Experiences* (15 studies). Several predictors coded under these dimensions (e.g., self-harm, recklessness, physical/sexual abuse) are assessed by a caseworker at an initial investigation and made available for predictive modeling. These predictors might already exist in the data if the family has previously come under the attention of CWS. This approach of aggregating the negative aspects of people's lives while ignoring the positive aspects has been criticized because of its deficit-based nature [72]. There is also an overlap between the *Traumatic Experiences* of a child and the *Bio-Parents Needs/Risk Behavior* because the same predictors (for e.g., history of physical/sexual abuse, medical trauma, parents' criminal activity) are used to conduct both *needs* assessment for a child and *risks* assessment for a parent.

Next, we cross-examine the predictors used by placement recommendation models. These models are not employed at the onset of an investigation and are used by CWS when a child needs to be placed in a permanent foster care setting. These models are generally more equitable as compared to risk assessment models because they try to weigh in the positive characteristics of a child, such as talents, interests, cultural identity, and school achievements to find an appropriate placement setting that meets their needs. Table 5 shows that placement recommendation models account for predictors around *Child Strengths* (5 studies), *Functioning* (10 studies), and *Child Behavioral/Emotional Needs* (13 studies) to weigh in the positive aspects and needs of a child and balance that with predictors around *Child Risk Behaviors* (12 studies) and *Traumatic Experiences* (10 studies) to find a suitable placement setting well equipped to meet their needs.

Relationship between Methods (RQ1) and Predictors (RQ3)

Table 5 depicts the crosstabs between predictors and computational methods. Most computational methods including both supervised machine learning and generalized linear models focused on *Child Behavioral/Emotional Needs*, *Child Risk Behaviors*, and *Traumatic Experiences* to assess the risk of a maltreatment event or the needs of a child. Some predictors that inform these three codes include traumatic events (e.g., physical/sexual abuse, medical trauma), child's conduct and anger management, and delinquent behavior. After an initial investigation is conducted by a caseworker and psychometric risk assessments are completed, these predictors become available for modeling. However, quantifying risk from such a narrow set of predictors has been criticized because it fails to account for the wide range of risk factors that arise as a result of systemic issues in CWS itself [187].

2.6 Discussion

In this section, we discuss the implications of our findings and future research directions. Our results provide implications for the human-centered design of algorithms in CWS, and more broadly for the public sector, as well as specific guidelines for developing such systems.

2.6.1 Algorithms Need to be theoretical & context-aware (RQ1)

Overall, we found a lack of theoretically derived and validated algorithms that demonstrated that they took measures to integrate knowledge from the social sciences into their designs. Only one study [331] constructed their model based on the child-welfare literature. Four studies even discussed this lack of theory and proposed solutions in the form of cumulative risk models [308], causal models [412], and revised SDM models grounded in risk and resilience theory [413]. Yet, based on the published literature, such models have yet to be consistently implemented.

This finding is problematic because it shows that these algorithms ignore many factors that affect how decisions are made in CWS. For instance, the decisions are often constrained by current policies or scarce resources [188]. Many current empirical models frustrate child-welfare workers because they do not account for such systemic factors. While some researchers have suggested [273] that empirical prediction is enough and that theory, context, or causal inferences are not always necessary in policymaking when outcomes remain desirable, we argue that this is not a desirable stance to take in child-welfare contexts because there is significant debate on how and which types of data, models, and outcomes are to be used in predictive modeling (with or without theory). Empirical knowledge related to child-welfare practice is fragmented and social science theories must be used to fill the gaps [187].

Therefore, we recommend that human-centered theoretical approaches be used to incorporate factors arising from evidence-based social work [89] and understand the causal pathways that often dictate decision-making processes. Such algorithms that are informed by appropriate causal theory would also have a greater likelihood of utilization as compared to their a-theoretical counterparts [412]. Significant work has also found disconnects between the functioning of algorithms and their social interpretations [46]. We see a similar phenomenon in CWS where the caseworkers using Structured Decision Making (SDM) model must translate information from both forms of assessments (clinical and algorithmic) leading to uncertainty and unreliable decision-making [419]. Therefore, algorithms that are meant to aid decision-making often become the source of frustration and force caseworkers to abandon their contextual judgments [413]. Human-centered theoretical approaches can help by placing the meaning-making process [365] at the center of the design process. It can help designers understand the theory of practice and uncover practitioners' sense-making processes (e.g., how they perceive quantified metrics [46]). Child-welfare workers are generally not trained in statistical thinking and make decisions based on experience, intuition, and individual heuristics [187]. Human-centered theoretical approaches can help us understand the mental models of child-welfare workers, inform feature selection (design), as well as interpret the results (evaluation).

Our results also indicate that several states adopted the SDM approach because it was supposed to integrate predictive and contextual assessments, however, it has fallen short of that goal [419, 413]. There are several factors at play in regard to any child-welfare case and it becomes critical to offer context to the case instead of focusing on a few broad factors without giving weight to important nuances [337]. For instance, understanding contextual knowledge with respect to an organization requires incorporating the organizational memory of the organization and its people [314, 11] which is inherently HCI research. Social workers are trained in writing detailed case notes by translating their context-specific experiences into text [114]. This unstructured, unanalyzed, textual data is added to *SACWIS* systems [348]. We hypothesize that valuable theoretical signals from these case notes can be considered within methodological approaches like topic modeling that can make good use of such unstructured data. Indeed, in recent years, HCI has developed a rich methodological tradition [47, 334, 95] of using signals from such unstructured data as predictors within algorithms to study complex, sociotechnical systems.

2.6.2 Going Beyond What is "Easily Quantifiable" (RQ2)

Our results suggest that the majority of the algorithms used predictors around child and parent characteristics, such as their needs, strengths, and associated risks (see Table 3). The vast majority of these predictors that are used for predictive modeling are derived from information that is easily available and readily quantifiable. For example, child-welfare workers use psychometric scales [131] to assess child and parent-associated risks and needs during an initial investigation which then becomes available for predictive modeling. Some of these predictors are found in almost every risk assessment model even though they have no predictive validity. For instance, the severity of abuse is easily quantifiable and is found in several risk assessment models even though there is little to no indication that it is related to the recurrence of abuse [85]. Moreover, several predictors (parenting skills, parent conflict, etc.) have not been properly validated [187] and can lead to unreliable predictions [187]. Such issues led the Illinois CWS (in 2017) to shut down its predictive analytics program [241]. In addition, none of the predictors account for the temporality of risk assessment. After an allegation of abuse, the assumption of escalation is the baseline for risk assessment leading to inflated risk scores and excessive CWS interventions [426].

Human-centered theoretical approaches can result in a rigorous feature selection process that relies on predictors that have been well-studied, understood, and validated in social sciences [46]. De Choudhury et al.'s [138] work in mental health is a good example, where the researchers validated constructs, focused on data biases and unobserved factors, as well as conducted sensitivity analysis. Moreover, it compels us to look towards sources of information that have been hereto hard to quantify. For instance, referring to our prior example around case notes advances in NLP [470] now allow us to quantify and make holistic inferences about all the stakeholders involved in a child-welfare case. This can address persistent issues among cases that appear similar based on the empirical data but exhibit high variation in outcomes [187].

Furthermore, human-centered participatory design [46] allows for HCI researchers to actively engage with domain experts in child-welfare to understand how risk accumulates (and how to model it), as well as engage with other stakeholders to better understand the systemic factors around policies, laws, and organizational culture [479]. Here, PD [335] can navigate the thorny, contextual differences between different legal and policy systems and the needs/values of stakeholders. Lodato and DiSalvo [299] highlight the different forms and limitations of PD as well as how PD can be conducted within such institutional constraints. Advances in CWS data systems [172] can accommodate for the collection of several new predictors concerning child well-being and systemic factors. Here, the active consideration of the needs and values of all stakeholders can help avoid the same reliability and validity pitfalls for the new predictors that exist for many of the current predictors.

2.6.3 Improve Lives and not just 'Minimize Risk'(RQ3)

One of the fundamental goals of CWS in the United States is to ensure positive outcomes for foster children [1], however, as our results confirm, the majority of the efforts in computational modeling continue to be focused on risk assessment (see Table 3). Risk assessment models only seek to minimize the risk of future harm and not improve the quality of lives of foster children. The target outcome of "risk of maltreatment" is poorly defined [496]. Federal and State law dictate how child abuse and neglect are defined and the state definitions often vary and establish the grounds for intervention by CWS [1]. Algorithms are trained on cases of substantiation, that is, cases where CWS judged maltreatment to have occurred [187]. This judgment in itself is very subjective and depends on state laws, policies, and CWS intervention criteria which are often dictated by the level of funding and caseloads [89].

Human-centered approaches can help theoretically define not only the predictors but also the target outcomes with the help of stakeholders and domain experts to ensure these key ingredients needed for algorithm design are validated and reliable. Human-centered participatory design can also unravel concerns around the social interpretations of algorithmically-based systems. For instance, Brown et al [72] investigated the community perspectives of risk assessment models in child-welfare. Child-welfare workers criticized these models because of their 'deficit-based' nature, that is, this approach only captures negative inputs to predict a negative outcome. There is growing concern that such an approach drives disproportionate negative caseworker perceptions that ultimately leads to negative actions [72]. Badillo-Urquiola et al. [38] also recognized the problems with a deficit-based framing in that it creates a sense of moral panic and diverts attention away from positive outcomes. They suggested that researchers focus on "strength-based approaches" that focus on positive factors that help improve lives.

CWS should actively focus on approaches that disrupt the status quo [224] and seek to improve the lives of foster children, such as *Child-Foster Parent Matching* [89, 378, 472]. This requires an ongoing engagement with foster parents and foster children to understand their specific values and needs as well as their cultural and parental expectations. HCI can contribute here by drawing on its rich tradition of work in action research, participatory design, and value-sensitive design to incorporate the values and needs of the stakeholders [175, 434, 32, 42, 59]. In addition, HCI researchers have developed methodological approaches that not only incorporate stakeholders into the design process but also the data analysis and interpretation processes [49, 487]. Moreover, advocating for foster children, a vulnerable and marginalized population, is inherently a social justice issue. HCI researchers have a long history of contending with social injustices and have developed theoretical and methodological approaches that seek "not so much to predict the future, but rather to imagine a radically better one [175]." Given the paucity of human-centered research into this domain and the richness of available social science literature [89, 378, 56, 73, 389, 128], this presents HCI researchers with a set of complex socio-technical challenges to study.

2.6.4 Recommendations for Future Research

Bridging AI and HCI Through Participatory Design

Our results indicate that there is a lack of theoretically-designed algorithms (see Table 2) which adds to the frustrations of child-welfare workers who are being pressured into using these algorithms as a means of standardizing decisions [413]. This situation is further exacerbated by a lack of PD leading to algorithmic systems that offer low utility [419]. Only one study in our corpus engaged with child-welfare workers to understand their concerns and needs [72]. PD [335] allows for the active inclusion of people most affected by a system. Engaging child-welfare workers in the design as well as evaluation processes ensures that their needs are met and that the system integrates well with child-welfare practices. Child-welfare workers who use algorithms on a daily basis strongly stress the need to be able to explain these models to each other and to policymakers [337]. Not only does this depend on which computational methods are used to construct an algorithm, and how they are deployed but also on how outcomes are defined and measured. This offers research pathways for HCI researchers who have increasingly started devoting attention to explaining outcomes and predictions [288, 329]. Moreover, it is imperative that researchers engage with the stakeholders because there are both, ethical and legal ramifications of using certain types of data. For instance, legal requirements might not allow a juvenile's criminal record or history of physical and/or sexual abuse to be used for modeling [421].

Algorithmic Decision Making via Speculative Design

We found that 56% of studies took a deficit-based approach to mitigate risks even though child-welfare literature has discussed the significance of equitable outcomes (e.g., child-foster parent matching). Recent studies based on newer technologies still continue to focus on risk assessment and uncritically reproduce the status quo. Designing against the status quo means setting our goals beyond risk assessment, and moving more ambitiously toward design that challenges underlying problems [224]. Speculative design [46] can allow stakeholders to shift their focus away from algorithms and be truly innovative in how they imagine problems and their underlying causes without being constrained by what might be technologically feasible. This is especially important for algorithm design where the boundaries of possibility change every day [46]. For instance, child-foster parent matching is well-documented in the literature for almost two decades but it has only recently been explored in an algorithm [331] because of advances in decision-tree learning. Similarly, innovation also creates novel avenues for studying the interactions and decision pathways resulting from different policies and programs [479].

2.7 Limitations and Future Work

We conducted a comprehensive and systematic literature review which was limited to the US-based child welfare system. We may have also missed algorithms used within CWS that are not publicly available for review. For instance, reports by non-profit organizations or state governments may have been distributed internally. Therefore, we plan to work directly with CWS agencies and conduct user interviews about the systems and algorithms being used within CWS to identify any other algorithms that have been implemented. To move towards using a human-centered approach to build new, evidence-based, and theoretically-driven algorithms, we plan to work with stakeholders in CWS to understand how different policies, practices, and programs create different decision pathways for child placements and services offered to families.

2.8 Conclusion

In conclusion, we recommend that the HCI community partners with CWS to do the following: **1)** A renewed focus on theoretically-designed algorithms with the active engagement of stakeholders through the design and evaluation phases, **2)** Develop algorithms for practice that incorporate a more comprehensive set of predictors well-studied in child-welfare literature, as well as predictors hard to quantify thus far, and **3)** Focus on equitable outcomes founded in evidence-based child-welfare and sociology research that improve the quality of lives of foster children instead of merely mitigating future risks.

CHAPTER 3: A FRAMEWORK OF HIGH-STAKES ALGORITHMIC DECISION MAKING FOR THE PUBLIC SECTOR DEVELOPED THROUGH A CASE STUDY OF CHILD-WELFARE

ABSTRACT: Algorithms have permeated throughout civil government and society, where they are being used to make high-stakes decisions about human lives. In this paper, we first develop a cohesive framework of algorithmic decision-making adapted for the public sector (ADMAPS) that reflects the complex sociotechnical interactions between human discretion, bureaucratic processes, and algorithmic decision-making by synthesizing disparate bodies of work in the fields of Human-Computer Interaction (HCI), Science and Technology Studies (STS), and Public Administration (PA). We then applied the ADMAPS framework to conduct a qualitative analysis of an in-depth, eight-month ethnographic case study of algorithms in daily use within a child-welfare agency that serves approximately 900 families and 1300 children in the mid-western United States. Overall, we found that there is a need to focus on strength-based algorithmic outcomes centered in social-ecological frameworks. In addition, algorithmic systems need to support existing bureaucratic processes and augment human discretion, rather than replace it. Finally, collective buy-in in algorithmic systems requires trust in the target outcomes at both the practitioner and bureaucratic levels. As a result of our study, we propose guidelines for the design of high-stakes algorithmic decision-making tools in the child-welfare system, and more generally, in the public sector. We empirically validate the theoretically derived ADMAPS framework to demonstrate how it can be useful for systematically making pragmatic decisions about the design of algorithms for the public sector.

3.1 Introduction

The influence of neoliberal politics and theories of New Public Management (NPM) [282] throughout most modern societies over the past two decades has sought to reform public services by emulating corporations to improve efficiency [290]. One way to achieve this goal for public sector services (e.g., child-welfare, labor, criminal justice, and public education) is through the adoption of automated processes (e.g., decision-making algorithms), as they purportedly promise to increase efficiencies, lower costs, and provide better outcomes for citizens. [382]. As such, algorithms in the public sector have become pervasive and, in turn, well-studied in recent years [307, 163, 107]. Consequently, they have also been scrutinized for achieving worse outcomes, exacerbating racial biases, and strengthening structural inequalities [163, 377, 397] within systems that are overburdened and under-resourced, yet critically needed [298, 135].

As a case in point, 423,997 children were in the U.S. Child Welfare System (CWS) in September 2019 which represents a steady increase in the past decade [349]. This number is only expected to grow in upcoming years barring major structural reforms. This has created an ever-increasing burden for CWS workers to make decisions about children that provide positive outcomes for them. Policymakers have decided that one avenue to address this issue is to implement algorithmic decision-making within CWS [4]. As such, algorithmic decision-making tools are now being used in high-stakes CWS situations, including making risk assessments of child abuse [137] and determining placement stability [331]. Brown et al. [72] conducted co-design workshops with stakeholders within the CW community (e.g., families, frontline providers, and specialists) and found that such algorithms largely bolstered distrust, perpetuated bias, and created black-boxed systems, which accelerated concerns about how these tools may negatively impact child-welfare workers' decisions. Thus, any technological solution cannot be inherently deemed 'fair' or 'just' and require complementary policy changes to affect community perceptions [72]. In a recent comprehensive review of the literature, Saxena et al. [397] highlighted the lack of human-centeredness [46] in the design and implementation of these algorithms and the need for more empirical work on how these algorithms are embedded in the daily work practices of child-welfare caseworkers. Thus, the gaps identified in these prior works led to the following overarching research questions:

- **RQ1:** *What are the high-stakes outcomes for which algorithmic decision-making is leveraged within the child-welfare system?*
- **RQ2:** *How does the implementation of a given algorithm impact algorithmic decision-making, human discretion, and bureaucratic processes?*
- **RQ3:** *What are the potential benefits and drawbacks when balancing the trade-offs between these three elements?*

To address these questions, first, we synthesized prior literature to develop a theoretical framework for Algorithmic Decision-Making Adapted for the Public Sector (ADMAPS). While SIGCHI researchers have attempted to formalize the dimensions of algorithmic decision-making in various contexts [16, 354, 232], we argue that the high-stakes decisions being made within the public sector necessitate the critical need for a distinctly unique framework for algorithmic systems that accounts for the complexities of public sector bureaucratic processes [164, 300, 178] and the delicate application of human discretion that has historically been a cornerstone of social services [297, 431]. We did this by synthesizing related works from the fields of Human-Computer Interaction (HCI), Science and Technology Studies (STS), and Public Administration (PA).

Next, we leveraged the ADMAPS framework as a theoretical lens in which to analyze the qualitative data collected from an eight-month in-depth ethnographic case study of a child-welfare agency in the mid-western United States. We attended 55 agency meetings and conducted 20 individual interviews over the course of eight months, which resulted in daily interactions with approximately 120 CWS agency employees and external consultants. To answer (**RQ1**), we first identified the algorithms used within the CW agency and the relevant data and outcomes they addressed within the system. For (**RQ2**), we assessed whether and how each algorithm affected each dimension of **human discretion** (*professional expertise, value judgments, heuristic decision-making*), **bureaucratic processes** (*resources and constraints, administration and training, laws and policies*), and **algorithmic decision-making** (*relevant data, types of decision-support, degree of uncertainty*). Finally, for (**RQ3**), we compared the four algorithms identified (i.e., CANS, 7ei, AST, and LPS) to show how ADMAPS can help balance the trade-offs in algorithmic decision-making to optimize the benefits and minimize the drawbacks associated with the high-stakes outcomes within the CWS.

Overall, we found that there is a need to refocus on strength-based outcomes centered in social-ecological frameworks [71] **(RQ1)**. We define strength-based outcomes as those that draw upon a person's assets and strengths rather than their deficits and weaknesses [38, 495]. In addition, an over-reliance on algorithmic decision-making to support bureaucratic processes can be detrimental to human discretion, however, algorithmic decision-making can support human discretionary work if they are fully supported by bureaucratic processes **(RQ2)**. Finally, algorithmic decision-making should augment human discretion (by building theory-driven algorithms centered in practice) rather than attempt to replace it; algorithm decision-support systems and bureaucratic processes need to be aligned (lack of balance creates utility issues) and collective buy-in in such systems requires trust in algorithmic outcomes at both the caseworker and bureaucratic levels **(RQ3)**. Thus, this paper makes the following unique research contributions:

- We conducted an in-depth ethnographic case study to understand the daily algorithmic decision-making practices of caseworkers in CWS.
- We go beyond existing recommendations for AI/ML to provide specific heuristic guidelines for algorithmic decision-making in CWS that can be of use to caseworkers, supervisors, government bureaucrats, and policymakers.
- We developed a theoretical framework (ADMAPS) of algorithmic decision-making in the public sector that synthesizes prior work on algorithmic decision-making in non-public sector settings with the unique challenges, limitations, and opportunities in the public sector. ADMAPS is generalizable to a wider range of public sector domains such as the criminal justice system, unemployment services, and public education.

In the sections that follow, we first highlight some of the high-stakes decisions made within the public sector. Then, we introduce our framework for **Algorithmic Decision-Making Adapted for the Public Sector (ADMAPS)**. Finally, we use ADMAPS to present an in-depth ethnographic case study of four algorithms used daily in CWS to determine high-stakes outcomes for foster children, including trauma-informed care, placement stability, and sex-trafficking risk.

3.2 The High-Stakes Decisions Made within the Public Sector

Algorithmic systems are being used to make high-stakes decisions about human lives and welfare in the public sector ranging across child-welfare, criminal justice, public education, job placement centers, welfare benefits, and housing among others. For example, the criminal justice system employs algorithms to determine the length of sentencing [214], allocate resources to neighborhoods [93], and predict the likelihood of recidivism (i.e., recommitting a crime) [151]. In the child-welfare system, decisions are being made about whether to remove a child from the care of their parents based on the risk of future maltreatment [129], who should be raising a child [331], and what types of services should be offered to families [201]. The public education system also uses algorithms to assign students to public school zones [386, 385] and determine student performance [482]. Job placement centers profile job seekers and make job placement decisions using algorithms as well [17, 232]. Algorithms are also used to establish eligibility criteria for receiving benefits and offer these benefits to families in need [163]. In short, many of the ways in which algorithms are being implemented in the public sector result in life-altering, if not life and/or death consequences.

The public sector differs uniquely from the private sector in terms of algorithmic decision-making in two distinct ways. First, outcomes in the public sector like assessing the risk of recidivism in criminal justice or assessing the need for welfare services are poorly and inconsistently defined [212, 45, 397, 163, 492]. Moreover, an individual's personal situation can (de)stabilize several times making it hard to assess what constitutes a successful outcome or intervention [232]. Second, current practices of using aggregate administrative data that is often poorly collected [118, 397, 201] and biased to predict an individual's behavior is a complex and hard task that may lead to unfairness in decision-making outcomes and is, in most western, liberal democratic systems unconstitutional and/or illegal [467]. These two factors combine to make algorithmic decision-making in the public sector, a high-stakes decision-making environment that has real repercussions for the lives and liberties of people. Therefore, the algorithmic decision-making process in the public sector needs to be scrutinized with the utmost care. Thus, there is an urgent need to develop a cohesive, yet tailored framework for algorithmic decision-making that is validated with in-depth, empirical work focusing on daily algorithmic practices.

3.2.1 Algorithmic Decision-Making within the Child-Welfare System

Algorithms in child-welfare have historically relied on a narrow set of psychometric predictors that are used to assess the risks and needs of foster children and parents. However, a more comprehensive understanding of the accumulation of risk is necessary to account for the family's social support system as well as the risk posed by the system itself [397]. Saxena et al. [397] recently conducted a systemic review of algorithms employed in the U.S. child-welfare system and uncovered several discrepancies in regard to data, computational methods, and target outcomes of these algorithms [397]. There is a need for theoretically constructed algorithms centered in the nature of the practice. Moreover, decisions must be made within the constraints of policies and systemic barriers; characteristics not accounted for by algorithms. The majority of algorithms in CWS are empirically constructed even though the empirical knowledge in child-welfare is quite fragmented and social science theories need to fill in these gaps [187]. For instance, child-welfare workers are often frustrated by algorithms because they do not account for the scarce resources in the public sector [188]. Risk assessment has also been the dominant focus of algorithms in CWS, however, there are concerns about their deficit-based nature that only seek to minimize risk but not improve the quality of children's life. This is driving attention away from strength-based frameworks [397, 72, 367, 38]. Prior work has also explored the utility of algorithms designed to aid decision-making and found that they increased uncertainty and led to unreliable decisions since caseworkers were required to translate information from both the clinical and algorithmic assessments [419, 412]. More recently, researchers have also focused on the need to uncover the politics, economics, and social implications of CWS algorithms and established the need to actively work with practitioners and domain experts to understand their perspectives about such systems as well as the systemic factors centered in policies, laws, and organizational culture that play a significant role in decision-making [400, 377, 202, 72]. This paper responds to these calls by conducting a deep ethnographic analysis of algorithms in use at a CWS agency and uncovers their social, technical, and political implications.

3.3 A Framework of Algorithmic Decision-Making Adapted for the Public Sector (ADMAPS)

As shown in Figure 1, we leveraged a socio-technical perspective of algorithmic decision-making that captures the three-way interactions between: 1) **human discretion**, 2) **bureaucratic processes**, and 3) **algorithmic decision-making**. We did this by synthesizing relevant, yet disparate, bodies of work across the fields of Public Administration (PA), Science and Technologies Studies (STS), and Human-Computer Interaction (HCI) to create a cohesive framework of Algorithmic Decision-Making Adapted for the Public Sector (ADMAPS). This framework is a

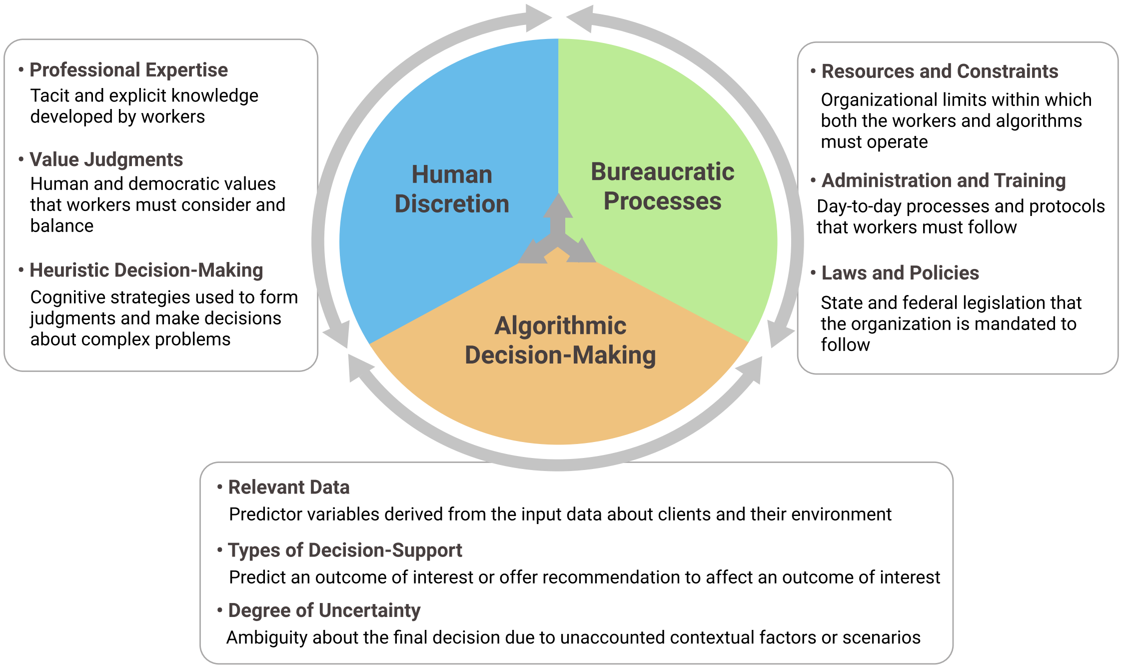


Figure 5: A Framework for Algorithmic Decision-Making Adapted for the Public Sector (ADMAPS). The three core elements of the framework are Human Discretion, Bureaucratic Processes, and Algorithmic Decision-Making.

core contribution of this paper and also served as a theoretical lens for grounding the qualitative analyses of our empirical case study within the domain of child-welfare.

Scholars within the PA field have extensively studied how **human discretion** by street-level bureaucrats [2] plays a central role in navigating **bureaucratic processes** and implementing policies. For example, bureaucrats act with a certain level of autonomy in how they interpret and apply professional standards when determining which clients must receive welfare benefits or services [431]. Public Administration scholars have also started to recognize the impact of information communication technologies (ICTs) on the nature of human discretion and bureaucracy with some recent attention paid to artificial intelligence [81, 360, 306, 75, 490]. Young et al. [490] introduced *artificial discretion* as a theoretical framework to help public managers assess the impact of AI and how it differs from human discretion with respect to task specificity and environmental complexity. However, several of these studies, rich in their understanding of human discretion and bureaucracy, continue to treat algorithms as peripheral end products; a new part of bureaucracy to which human discretion must adapt. Meanwhile, STS scholars conducting studies in the public sector have used Kitchin and Lauriault's framework of *data assemblages* [269] to deeply study the intersection of **bureaucratic processes** and **algorithmic**

[2] A street-level bureaucrat is a professional service worker (e.g., social worker, police officer, teacher) who operates in the frontline of public service provision. They interact closely with clients and make decisions about them based on how they interpret policies relating to the situations at hand [297].

decision-making by examining the politics of data systems [377, 482, 291]. Data assemblages perceive data systems as complex assemblages of human actors, artifacts, technical systems, institutions, and ideas. This framing provides a means to consider how these systems are socially, economically, and politically constructed. Similarly, STS scholars have also used Seaver's notion of *algorithms as culture* [414] to understand the social implications and values of algorithmic systems through an ethnographic analysis. These studies have made significant contributions towards the community's understanding of how algorithms shape and are shaped by cultural context, how value is inscribed to algorithms, and how power is afforded to them. However, the often macro-level perspective applied within the field of STS may at times miss some of the nuances at the bureaucratic street-level, which is where human discretion is most critical. Due to the high-stakes decisions being made within the public sector, there is an urgent need to map the complex interdependencies between the core elements of human discretion, bureaucratic processes, and algorithmic decision-making that often go unnoticed in the public sector.

The SIGCHI research community is well-positioned to do this cross-disciplinary and integrative work due to our strengths in taking a human-centered [46] and value-sensitive approach [494] to the design and development of algorithms. Further, the CSCW is a well-suited venue for this type of research due to the collaborative nature of the work being performed by teams of CWS employees and external consultants when making critical decisions about the well-being of children. In the sections below, we describe the key dimensions of our framework.

3.3.1 Human Discretion

Lipsky's theory of street-level bureaucrats [297] defines *human discretion* as an individual's ability to exercise their own judgment in implementing government policies in complex and uncertain problem spaces. Scholars across multiple disciplines have recognized the importance of human discretion in developing algorithms and interpreting and administering algorithmic decisions [46, 356, 357, 358, 149] but also in interpreting and making policy decisions [297, 94, 178, 76]. However, when HCI scholars deliberate over human discretion, it generally occurs from a design perspective. That is, how can we incorporate humans' tacit knowledge, social interpretations, and values into the design process [46, 494]. Whereas, when Public Administration scholars discuss human discretion, it refers to the decision-making latitude as well as the value-laden choices that bureaucrats must make when experiencing complexity and uncertainty [297, 81, 76]. We integrate this knowledge about human discretion and close this loop by presenting the following process model where the bureaucrats use their professional expertise, engage in value judgments, and heuristic decision-making.

Professional Expertise

The tacit and explicit knowledge developed by workers in any given domain [46, 428]. It plays an important role with respect to the workers' confidence in their own decisions [364] as well as the level of adeptness with which they navigate bureaucratic processes [178, 94], negotiate resources [76], and seek additional supervision [162]. As novice practitioners gain confidence in their skills, they become more aware of the need to pursue additional details, supervision, and other opportunities to increase their level of domain knowledge [477]. Professional expertise, however, is also domain-specific and continually evolves with time [364, 162]. It is necessary to examine the nature of professional expertise within the public sector which is rapidly evolving through the continued digitization and automation of work processes that were previously the forte of street-level bureaucrats [81, 198]. Practitioners in the public sector are continually acquiring new skills as they learn to make decisions through data and interpret algorithmic outputs [62, 75], however, these new skills are not aligned with what constitutes professional expertise [232].

Value Judgments

Practitioners must consider and balance human and democratic values when assessing cases about citizens [232, 179]. This is another key dimension of human discretion because workers must weigh the competing motivations of different clients as well as differing notions of values and incorporate them within their decision-making processes. Value judgments play a pertinent role when a practitioner is faced with ethical dilemmas and informs their decision-making ability. Practitioner's beliefs and moral values are important factors in regard to how street-level decision-making unfolds [318, 392, 300]. Moreover, practitioner's personal values are often mediated by organizational culture which subsequently yields results that can be significantly different than results centered in personal values [180]. Therefore, it is essential to understand the role of value judgments because practitioners have assumed the role of value mediators who must weigh the needs of citizens against the demands of policymakers [370].

Heuristic Decision-Making

Heuristics refer to the cognitive strategies used to form judgments, make decisions, and find solutions to complex problems [199]. Gigerenzer and Gaissmaier [199] reviewed research on heuristic decision-making in business organizations, health care, and legal institutions and established the fundamental role it plays within organizations. Heuristic decision-making can lead to more accurate decisions than complex rational models and selecting information in an adaptive manner can lead to more accurate judgments than weighing all of the information [332]. Practitioners work more effectively and efficiently when their knowledge base in well-organized and centered in heuristics since it allows them to separate relevant and irrelevant knowledge for any given context [31]. Simple heuristics can be more successful especially in uncertain and complex spaces since all the information required to make a decision might not be available as a result of uncertainty [30]. Therefore, it becomes imperative for practitioners to rely on their heuristics that are acquired through experience and practice [477]. Moreover, decision-making in organizations must involve professionals' heuristics because the ideal conditions required for rational and reductive models rarely hold true in an uncertain world [359].

3.3.2 Bureaucratic Processes

Bureaucratic processes are the critical governance characteristics essential for policy development and implementation to serving public interests [164, 165, 178]. Both HCI and Public Administration scholars have recognized the dominant role that bureaucratic processes play both with respect to establishing the role of bureaucrats (i.e., human discretion) as well as the adoption of technology [62, 76, 299, 400]. Scholars have also emphasized that policy/bureaucratic considerations must precede technology design and professional practice considerations [243]. We include three different dimensions to bureaucratic processes consistently highlighted in the literature as described below.

Resources and Constraints

Availability of essential resources (administrative, financial, personnel, political) directly impact organizational performance [289, 177, 313]. Resources can also be viewed as constraints within which the organization must operate [289]. This dimension is of special importance for the public sector which is facing severely limited resources and new dilemmas in the form of burdensome workloads, high staff turnover, and a lack of experienced workers [313, 327]. Examining how these scarce resources are allocated in public services is crucial because most agencies are experiencing a push to innovate and invest in evidence-based practices to improve performance [465], however, investing in innovation can be challenging in a resource-deficit domain [352].

Administration and Training

Protocols, workflows, and processes established at the organization that are followed by workers in their day-to-day practice and play a significant role in decision-making [297, 289, 362]. Organizational processes offer the means to understand how an agency makes decisions within policy mandates as well as how it meets diverse public needs [60]. Processes are established to improve accountability in the form of consistent, transparent, and defensible decision-making and allow the agency to effectively communicate compliance with legal mandates as well as utilize existing knowledge routines to improve reliability [362, 139]. Moreover, processes followed by practitioners in their daily lives also continually shape policy on the ground [297, 70]. This dimension also identifies the workers' training in the public sector which plays a critical role in regard to individual, team, and organizational development [264]. New assessments and tools are continually being introduced in the public sector in pursuit of creativity and innovation such that it leads to standardized and evidence-based decision-making [143, 75, 91], however, this also necessitates a need to examine if workers are being adequately trained to fully utilize these tools [460]. Training at the organization also establishes the basis for worker expertise by ensuring that the workers skillfully mediate both the nature of practice and bureaucratic processes [477].

Law and Policies

Formal actions enacted by legislatures or political executives that public administrators must comply with and implement [360]. Laws and policies establish the constraints within which all decisions (human or algorithmic) must be made as well as define the outcomes of interests themselves. For instance, the law dictates which data is available for predictive modeling and how target outcomes are defined [421, 397]. This dimension is of critical importance since it directly impacts both human discretion and algorithmic decision-making. For instance, the policy decision to expand mandated reporting in child-welfare significantly increased the number of cases referred to CWS as well as broadened the definitions of child abuse and neglect (with implications for algorithmic modeling) [324]. Prior work has acknowledged the dominant role that bureaucracy or policy plays in the public sector both in regard to decision-making as well as the adoption of technology [299, 243]. The central role of bureaucratic processes is evident from prior work conducted in public services where caseworkers pushed for algorithmic systems that could help mitigate organizational contradictions and clarify organizational processes [232].

3.3.3 Algorithmic Decision-Making

Algorithmic Decision-Making is defined through the lens of *street-level algorithms*, a term recently coined by Alkhatib and Bernstein [16] in the HCI community. Street-level algorithms directly interact with and make on-the-ground decisions about human lives and welfare in a sociotechnical system [16]. Prior work has argued that algorithms in the public sector are a domain in its own right [232, 137, 464, 397, 400] and must be characterized differently as compared to algorithms in the private sector where the decision outcomes are well-defined. Therefore, it becomes important to examine and critique the dimensions within Algorithmic Decision-Making that impact the predicted outcomes. Algorithmic Decision-Making is the most flexible element of the framework that designers can directly impact. That is, algorithms must be developed in such a way that they balance the other two elements (human discretion and bureaucratic processes). HCI methodologies such as value-sensitive algorithm design [494] and human-centered algorithm design [46] can ensure that the algorithms account for the values of stakeholders as well as theory-driven practice. Moreover, participatory design can unravel the policy mandates and institutional processes that often mediate the decision-making process and must be accounted for within systems design [299, 400].

Relevant Data

Necessary information about individuals and their environment must be collected to be able to adequately predict an outcome of interest. In several domains within the public sector, there is significant debate about which predictors are associated with which outcomes [395, 151, 232]. Moreover, the necessary information may not always be available or inconsistently available with contradicting factors [118]. For instance, risk assessment algorithms in child-welfare have traditionally only used a narrow set of predictors (child and parent characteristics) to assess risk [397]. However, a more comprehensive understanding of risk is necessary, including the risk posed by the system itself [187]. Therefore, algorithms need to be theoretically constructed with proper considerations from domain experts with respect to feature selection and modeling to ensure that the algorithm offers higher utility and complements the theory of practice [400, 46, 494].

Types of Decision-Support

Two types of algorithms are predominantly used in the public sector; predictive and prescriptive algorithms. Predictive algorithms seek to predict the likelihood of the occurrence of an outcome of interest, whereas, prescriptive algorithms act as decision aids and offer recommendations to intervene and affect the outcome of interest [141]. Examining the nature of decision-support systems is equally as important as interrogating the outcome itself. Prescriptive decision aids are often introduced as a means to improve decision-making while not shifting agency away from workers. However, prior research shows that workers allow decision-aids to supplant their own decisions when they lack confidence and/or experience [163, 397, 419]. Moreover, the calls for human-in-the-loop might be moot if there is a lack of understanding about how algorithms impact human decision-making and how the type of decision-support (i.e.- algorithm design) impacts the practical possibilities for human intervention [360, 419, 412].

Degrees of Uncertainty

Decision outcomes in the public sector are not well-defined and as previously noted, a person's life can stabilize or destabilize several times making it hard to predict what constitutes a successful outcome [232]. Prior research has also established that an irreducible degree of uncertainty exists with respect to the outcomes in the public sector [219, 137, 75] with both humans and algorithms likely to make mistakes. Pääkkönen et al. [354] further extend this argument to state that the design of algorithmic systems must identify and cultivate important sources of uncertainty because it is at these sources where the need for human discretion accumulates since ambiguity about the operation of the algorithm persists.

This framework challenges designers, practitioners, and policymakers to rethink the core assumptions and nature of their practice which are evolving in an increasingly socio-technical public sector and need to be re-examined in light of these new challenges and opportunities. It provides a structured way to think about socio-technical problems centered in algorithmic decision-making in the public sector, study the interdependencies between the dimensions, and recognize underlying causes that impact decision-making.

3.4 Methods

In this section, we describe our partnership with a child-welfare agency to address the research questions set forth in the introduction of our paper.

3.4.1 Study Overview

The goal of this study was to examine the algorithms that caseworkers use in their daily work lives and unpack the collaborative nature of how these algorithms were used in group settings. To accomplish this, we partnered with a child-welfare agency, which serves about 900 families and 1300 children in a large metropolitan area in the midwestern United States. The state's Department of Children and Families (DCF) contracts child-welfare services to this agency which must comply with all DCF standards including the use of mandated algorithms or decision tools. We conducted an eight-month long in-depth ethnographic case study at the agency from August 2019 to March 2020. Before conducting observations or recruiting participants for interviews, we obtained Institutional Review Board (IRB) approval at our mid-sized private research university to conduct our study. We then emailed the participants an IRB-approved consent form and obtained their verbal consent to participate in the study. During this time, the first author observed child-welfare team meetings and conducted semi-structured interviews with key stakeholders at the agency.

3.4.2 Meeting Observations and Interviews

The first author conducted in-person observations of meetings to gain the necessary understanding of how algorithms were used in a team setting and how caseworkers interacted with these algorithms in their daily work practices. These observations were also helpful in understanding the collaborative work of child-welfare teams that make decisions that are mediated by policies, social work practice, and algorithms. Next, we provide a detailed description of the team meetings and interviews.

45-Day Staff Meetings or Planning Meetings The 45-day staff meetings occur within the first 45 days of a case coming into the care of the agency and are attended by child-welfare team members involved at the front end of case planning. These meetings facilitate information sharing so that consensus can be reached in regard to the child's well-being and placement stability[3]. Each meeting is scheduled for 90 minutes, and we observed 15 meetings. These meetings are typically attended by the CWS employees that work in case management, permanency planning, family preservation, and licensing. Central to these meetings is the *7ei* staffing protocol that helps the child-welfare team apply principles and practices derived from trauma-informed care (TIC) [228] to each case. The 7ei Staffing protocol is accompanied by the 7ei algorithm which acts as a framework for TIC and helps track progress with respect to each case [450]. The *CANS* algorithm is also used at these meetings to establish a baseline for foster children with respect to mental health well-being. We also identified two more algorithms being used by the child-welfare teams. *Legal Permanency Status (LPS)* algorithm is used to assess the quality of the current placement and recognize systemic barriers. *Anti Sex-Trafficking (AST)* algorithm is used to assess the risk of sex trafficking for foster youth. Observing these meetings allowed us to understand how these decision-making tools were being used in practice, the benefits they offered, as well as the challenges they posed.

Permanency Consultation Meetings Permanency consultation meetings are specialized meetings designed to expedite permanency for children placed in out-of-home care by employing innovative best practices and seeking to address any systemic or policy-related barriers. These meetings are facilitated by permanency consultants and are staffed with many of the child-welfare team members that attend the planning meetings. They regularly occur at the 5, 10, and 15+ month marks for every case until the case is closed. These ongoing meetings tended to be more involved than the planning meetings as limited information is available at the onset of a case. Moreover, permanency consultations involved cases that had been with the agency for several months (if not years) and revealed the messy interaction between the complex socio-political domain of child-welfare and the algorithmic tools being used. For instance, it was interesting to observe how the child-welfare teams reached a consensus during decision-making discussions when they had to account for policy and systemic barriers, social work practice, and algorithms. Each meeting was scheduled for an hour, and we observed 40 of these meetings.

Semi-Structured Interviews Next, we used the knowledge gathered from these observations to develop our interview protocol and recruit participants who consistently attended these meetings as part of their job routines. After having first observed the child-welfare teams interact with algorithms for several months, we conducted interviews to delve deeper into the participants' understanding of these decision tools as well as the benefits and challenges as perceived by them. We asked participants a series of questions about the nature of child-welfare work and

[3]Placement stability is defined as three or fewer placement moves for a foster child during the previous 36 months.

the algorithms we observed being used at the agency. We also asked them to expand upon any interactions we had observed during the meetings, such as the participant's dislike or appreciation for a certain algorithm or feature or their frustration (or their team's) with the misuse of these decision tools. We conducted 20 semi-structured interviews with child-welfare staff members, which included permanency consultants, supervisors, program directors, ongoing case managers, data specialists, and clinical therapists. Seventeen interviews were conducted at the child-welfare agency in the participants' private offices or conference rooms.

3.4.3 Qualitative Data Analysis

The first author took detailed observational notes during each team meeting and compiled a debriefing document with his initial insights within 30 minutes of each meeting to retain as much of his thoughts as possible. The interviews were audio-recorded and transcribed verbatim for analysis. Notes, documents, and transcripts were shared among all co-authors. Our high-level research questions guided our analyses, but within those questions, we allowed for new insights to emerge and adjusted our research questions based on emergent insights. We performed thematic qualitative analyses [112] to answer RQ1 and RQ3. The first author read through the interview transcripts several times and created initial codes and consulted with co-authors to form a consensus around the codes, as well as resolve any ambiguous codes. Next, these codes were conceptually grouped into themes. However, in our results, we also use our observational notes to augment the insights we gained from the interviews and note potential discrepancies and nuances from the holistic insights gained from our site observations.

For RQ1, we used an open-coding process to identify the high-stakes decision outcomes associated with the four algorithms that are embedded in child-welfare practice, namely, Child and Adolescent Needs and Strengths (CANS), Seven Essential Ingredients (7ei), Anti Sex-Trafficking Response Tool (AST), and Legal Permanency Status Tool (LPS). In understanding how these algorithms were used in the daily practices of the CWS employees, we identified seven key purposes: 1) **Compensation Calculations:** determine the monetary value to be offered to foster parents for caring for foster children, 2) **Mental Health Assessment:** conduct a mental-health screen of foster children to assess the risks and needs, 3) **Level of Foster Care:** determine a suitable placement setting capable of meeting the needs of children, 4) **Trauma-informed Care:** a trauma-responsive service model developed through an ecological understanding of adverse events and trauma experienced by children and families, 5) **Placement Stability and Permanency:** Track outcomes from the trauma-responsive service model and assess if they are leading to better outcomes, 6) **Sex Trafficking Risk Assessment:** assess the risk of sex-trafficking for a foster child, and 7) **Quality of Placements and Systemic Barriers:** Track the current quality of placement and the systemic barriers inhibiting permanency.

For RQ2, we used the ADMAPS framework to code for how each algorithm (i.e., CANS, 7ei, AST, LPS) impacted (positively, negatively, or both) dimensions of ADMAPS. For example, we found that CANS had an overall negative impact on human discretion with 80% of the interviewees indicating that it negatively impacted professional expertise by limiting the scope for value judgments and heuristic decision-making; 75% of the interviewees said that it reduced their ability to make flexible value judgments on child outcomes; 80% of interviewees asserting that they were no longer given the discretion to make decisions on behalf of the children assigned to them because CANS made several of these decisions for them. These mappings to the ADMAPS framework allowed us to assess the role that human discretion, algorithmic decision-making, and bureaucratic processes played with respect to each of the four algorithms deployed in CWS daily practice, as well as compare the differences between them.

For RQ3, we synthesized the patterns across the four algorithms to identify emergent themes that were consistent across our analysis of RQ2 to give a bigger picture of the potential benefits and drawbacks associated when balancing the tradeoffs between human discretion, algorithmic decision-making, and bureaucratic processes.

3.5 RESULTS

In the following sections, we organize and present the results organized by our three research questions. First, we identify the high-stakes outcomes for which algorithmic decision-making is leveraged within the CWS (RQ1) and the roles of human discretion, bureaucratic processes, and algorithmic decision-making play in these decision outcomes (RQ2). We do this separately for CANS, 7ei, AST, and LPS. Next, we discuss the potential benefits and drawbacks when balancing the different dimensions of the ADMAPS framework (RQ3). The interviewees' profiles can be found in the appendix.

3.5.1 Child and Adolescent Needs and Strengths (CANS) Algorithm

The Child and Adolescent Needs and Strengths (CANS) algorithm was used at all the planning meetings (n=15) and discussed by all the interviewees (n=20). CANS algorithm is constructed using the CANS assessment; a communimetric tool designed to assess the level of need of a foster child and develop an individualized service plan [303]. With its primary purpose being communication, CANS assessment is designed based on communication theory rather than psychometric theories centered in measurement development [301]. CANS assessment was designed to support decision-making with respect to assessing a child's level of need and service planning. However, as depicted in Figure 6, the CANS algorithm has been re-appropriated to measure additional outcomes discussed below. It is conducted within the first 30 days of a child entering the child-welfare system or moving to a new placement (e.g., foster home). It is then periodically conducted every six months or per the request of caregivers if any supposed changes occur regarding the foster child's mental health.

CANS High-Stakes Decision Outcomes (RQ1)

Mental Health Assessment and Level of Foster Care. All child-welfare workers at the agency are certified in conducting CANS for mental health screenings. As such, several participants (60%, n=12) shared that the agency uses the CANS algorithm to conduct a mental health assessment of foster children based on risks posed and exhibited behaviors. Participants (60%, n=12) stated that the first CANS assessment was most useful because it helped establish a baseline for the mental health needs of a child and the level of care that the child needed. For instance, one supervisor explained:

> *"We need to have starting point.. what kind of behavioral issues does a child have? What are their needs? Because we can't naively place a high-needs kid with foster parents who are not trained and certified to manage those needs. It's a recipe for disaster"* -P9, Child Welfare Supervisor, MSW, 13 years

The first assessment is used to devise service plans for foster children (for e.g, behavioral therapy) based on exhibited behaviors. It also facilitates sharing this information with other parties such as legal parties and family preservation specialists who also play a role in case planning. Child-welfare teams in both the planning meetings (n=15) and permanency consultations (n=40) briefly discussed child needs and behaviors but then shifted to a more extensive conversation about trauma using trauma-informed care. Some participants (50%, n=10) also shared that the CANS algorithm recommends the level of foster care that the child should be placed in (see Figure 6). Higher-level foster homes are trained and certified in taking care of high-needs children. However, due to a lack of such homes, this decision often comes down to the availability of resources.

Compensation Calculations. Using the CANS algorithm for calculating foster parents' compensation was another prominent theme that emerged in 73% (n=11) of the planning meetings and 85% (n=17) of the interviews. However, it was not a dominant theme at the permanency consultations, because compensation is directly negotiated between the foster parents, case managers, and the supervisor. It does not require the input of the rest of the child-welfare team. The state reimburses foster parents for the costs associated with having foster children placed in their homes. Most of the participants (85%, n=17) quickly recognized CANS as the "rate-setting tool" even though compensation calculation was not the primary purpose of the algorithm. One supervisor explained:

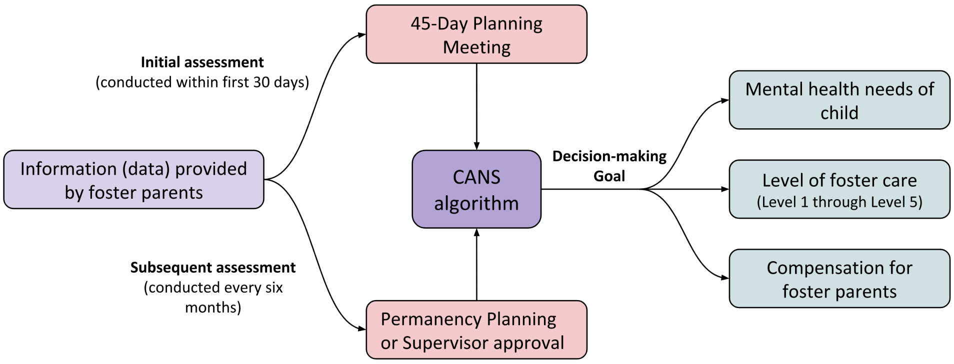


Figure 6: CANS Algorithm: Associated Decision Outcomes

> *"Foster parents who are taking in high-needs kids should be compensated for that. They have to put in significantly more time and energy in managing those behaviors, taking kids to therapy, setting healthy boundaries.. So we need this [CANS] standard to do that"* - P12, Child Welfare Supervisor, MSW, 7 years

The state needs a *metric* to be able to determine the compensation to be offered to each foster parent. The Department of Children and Families (DCF) decided to associate this compensation with the mental health needs of a child, that is, the higher the mental needs of a child, the higher the compensation offered to foster parents. Thus, even though the primary purpose of CANS was for conducting mental health assessments and for the level of foster care, it was re-appropriated to also calculate compensation associated with caring for a foster child's mental health needs. While it logically follows that a child with more mental health issues will require a higher level of care (consequently, higher cost of care), CWS employees were also well-aware that such cost-benefit analyses tied directly to something as subjective as mental health assessments were problematic.

CANS and Algorithmic Decision-Making (RQ2)

Overall, caseworkers were frustrated that CANS misses important context about the child but is still used in a mandated predictive capacity. In the sections below, we discuss how the CANS algorithm maps onto the dimensions of algorithmic decision-making (i.e., relevant data, type of decision-support, and degree of uncertainty). We use a percentage combined with an up or down arrow to denote the percentage of participants who indicated a positive or negative impact on each dimension of the ADMAPS framework for algorithmic decision-making, human discretion, and bureaucratic processes, respectively. We follow this structure throughout the remainder of our results.

70% ↓ **Relevant Data.** *CANS data does not account for trauma or social interactions.* Most of the participants (70%, n=14) shared that CANS conducted the child's assessment in an isolated manner and did not account for the quality and impact of relationships in their lives which are often more important for determining the long-term well-being of these children. CANS algorithm focuses on the child's emotional/behavioral needs (e.g., anxiety, anger control, substance use, behavioral regression) and child risk behaviors (e.g., suicide risk, self-harm, delinquent behavior, and runaway tendencies) to assess the mental health needs of the child. For instance, a supervisor shared –

> *"How do you measure empathy of others? You can't. Some foster parents are more empathetic and understanding and stand by the child no matter what. Sometimes that's all it takes. You can't put that in CANS"* -P15, Child Welfare Supervisor, 9 years

Moreover, participants noted that CANS focuses on current behaviors but not the underlying trauma or traumatic triggers. The assessment is conducted based on exhibited behaviors over the past 30 days, however, the participants explained that trauma can stay with a child for years and lead to serious emotional dysregulation from time to time.

90% ↓ **Type of Decision-Support.** *The predictive nature of CANS leaves no room for discretion.* This challenge emerged in 90% (n=18) of the interviews and 73% (n=11) of the planning meetings. The CANS algorithm is designed to predict or measure outcomes of interest once the data has been provided. Participants shared that the predictive nature of CANS leaves little room to exercise discretion and has become a great source of frustration for them. Gaming the inputs to achieve the desired outcome is the only way through which caseworkers regain agency. Participants explained that the higher the mental health needs of a child per CANS, the higher the compensation offered to foster parents. After CANS scores are entered in, the algorithm generates a monetary value to be offered to foster parents. However, if foster parents disagree with the rate and believe they should be paid more, the caseworkers manipulate the scores to produce a higher rate. Most of the participants (90%, n=18) shared their frustration regarding the inflexible and predictive nature of CANS. For instance, one supervisor explained:

> *"Case managers and even supervisors are being forced to.. and kind of pressured into scoring children higher in order to provide higher numbers. So foster parents get paid more. CANS is a manipulative tool being used to barter off children...Children are being exploited for payment"* -P10, Child Welfare Supervisor, MS, LPC, NCC, 9 years

70% ↓ **Degree of Uncertainty.** *High degree of uncertainty associated with the outcomes.* Several participants (70%, n=14) shared that CANS does not account for much of the data that they consider pertinent when assessing cases (e.g., understanding of trauma and social support system). This lack of relevant data leads to a high degree of uncertainty which is further exacerbated by the predictive and inflexible nature of CANS. Many felt that gaming the algorithm was the only viable option for caseworkers to exercise discretion. For instance, one supervisor shared:

> *"CANS has become all about the rate. Generating the right rate so foster parents are happy with little to no attention paid to mental health needs"* -P13, Child Welfare Supervisor, MSW

CANS Severely Impedes Human Discretion (RQ2)

In this section, we discuss how the CANS algorithm maps onto the ADMAPS dimensions of human discretion.

80% ↓ **Professional Expertise.** *CANS contradicts professional expertise.* All caseworkers at the agency are required to pass the CANS certification to be able to conduct the assessment with clients. However, most participants (80% n=18) shared that their training in trauma-informed care (see Section 6.3), which offers a more comprehensive understanding of trauma and the child's environment often conflicted with CANS. CANS further inhibited professional expertise because caseworkers felt that it turned them into data brokers who must collect information about children and feed it to the algorithm to make the decisions. The lack of relevant data and lack of decision-making latitude on the part of caseworkers has turned CANS into primarily a rate-setting tool to calculate compensations. For instance, one case manager shared:

> *"Caseworkers are doing CANS just to get it done.. to produce a good rate and reduce the conflict with foster parents. There is little to no attention paid to the mental health needs of kids."* -P17, Ongoing Case Manager, 8 years

75% ↓ **Value Judgments.** *CANS has introduced conflicting values.* Most of the participants (75%, n=17) shared how the re-appropriation of CANS to calculate the foster parent compensations has led to several unintended consequences. CANS is re-conducted every six months to reassess the mental health needs of children, and consequentially, compensation is re-calculated. However, with a focus on exhibited behaviors and not the underlying trauma, foster parents who are helping children cope and recover can see their compensation being lowered. A supervisor explained:

> *"It is the complete opposite of what we want it [CANS] to do. Foster parents help minimize the behaviors and offer support so that kids can develop good coping skills. They help address the mental health needs and help kids stabilize by taking them to therapy and all their activities. But then they're punished because the kid's needs go down, and so does the rate"* -P9, Child Welfare Supervisor, MSW, 13 years

Here, caseworkers are unable to prioritize properly conducting CANS because generating the adequate rate takes precedence to ensure the placement is not disrupted. Placement disruptions adversely affect foster children who develop emotional and behavioral problems and are unable to form lasting meaningful relationships with foster parents [56]. Therefore, caseworkers must prioritize supporting the current placement by any means necessary. This contradictory nature of CANS has turned caseworkers into value mediators and has left them frustrated because they are unable to adequately balance the needs of families and the demands of policymakers.

80% ↓ **Heuristic Decision-Making.** *CANS leaves no room for heuristics judgment calls.* This theme emerged in 80% (n=16) of the interviews and 73% (n=11) of the planning meetings. With a lack of relevant information pertinent to a case and a high degree of uncertainty, it becomes imperative that caseworkers are able to turn towards heuristics and make decisions with the assistance of the child-welfare team. However, CANS does not allow for heuristic decision-making with respect to the outcomes. Participants (80%, n=16) shared that every case is contextually different with salient factors that might be central to one case but peripheral to another. This was also apparent in all the planning meetings and permanency consultations where the teams adaptively focused on information pertinent to that case.

CANS and Bureaucratic Processes (RQ2)

Overall, CANS allocates resources but does not account for organizational constraints. In this section, we discuss how the CANS algorithm maps onto the ADMAPS dimensions of bureaucratic processes.

40% ↑ 70% ↓ **Resources & Constraints.** *CANS has introduced new constraints in the case planning process.* There is both a benefit and drawback to how the CANS algorithm accounts for resources and constraints. Some participants (40%, n=8) found value in the first assessment since it established the baseline for mental health and compensation for foster parents. Ideally, the algorithm offers an efficient way to allocate funds to foster parents based on the mental health needs of the child. However, it does not account for organizational constraints and its implementation ends up introducing more constraints that frustrate caseworkers. Participants (70%, n=14) shared that properly conducting CANS requires the caseworkers to interview several people, however, they are only able to interview foster parents due to high caseloads. A permanency consultant asserted:

> *"You are supposed to interview foster parents and teachers as well as others that kids interact with to get a good CANS assessment but with high caseloads, caseworkers only talk to the foster parents"* -P4, Permanency Consultant, MSW, APSW, 8 years

However, as previously discussed, CANS is conducted every six months and foster parents are motivated to exaggerate behaviors to continue to be paid consistently. This is a constraint introduced by the implementation of the algorithm itself. Participants (50%, n=10) also shared that the algorithm recommends the level of foster care that a child should be placed in, however, there is a lack of good foster homes in the system and this requirement is seldom met. Moreover, two data specialists shared that the caseworkers' yearly job performance is tied to the timeliness with which they complete and submit CANS assessments since funds need to be allocated in a timely manner. Therefore, it exacerbates the need to get the assessments completed irrespective of the mental health needs of a child, and consequentially, adds another organizational constraint.

40% ↑ 80% ↓ **Administration & Training.** *Caseworkers are trained to conduct CANS but not on managing constraints.* This theme emerged in 80% (n=16) of the interviews. CANS offers a way to support an organizational process of allocating resources. Caseworkers are trained and certified to conduct the mental health assessment, and as previously noted, some participants (40%, n=8) found value in the first CANS assessment as a means to establish a baseline for mental health needs. However, they are not trained on how to manage the systemic constraints and value conflicts introduced by the algorithm itself. Participants (50%, n=10) shared that new caseworkers socially learn from other caseworkers (and through experience) to manage these conflicts by gaming CANS.

80% ↑ **Laws & Policies.** *The use of CANS is legally mandated.* Most participants (80%, n=18) recognized (and often complained) that CANS was legally mandated by the state and offered them a convenient means to allocate resources to foster parents every six months. Participants recognized that they must comply with this policy and continue implementing CANS enough though they routinely manipulated it to generate higher compensations.

3.5.2 Seven Essential Ingredients (7ei) Algorithm

The agency uses trauma-informed care [228] as its core guiding principle, which is embedded in its practice and trauma-responsive service model [450]. 7ei is a theoretically constructed algorithm that is centered in the medical practice of trauma-informed care (TIC) and complements the agency's social work practice [228, 255]. We witnessed 7ei being utilized in all meetings (N=55). This algorithm complements specialized trainings developed by the agency for child-welfare workers and introduces the complexity of trauma, frameworks for understanding the effects of trauma, and the practices and principles of TIC [450]. 7ei acts as a guiding framework and is used to track and score each case from a TIC perspective. The algorithm is designed to be used in a team setting such that team members can offer their expertise, reach consensus decisions, and devise case plans. The team discusses and scores the child's and caregiver's wellness on seven domains as depicted in Figure 7 and brainstorms solutions on how to make progress on

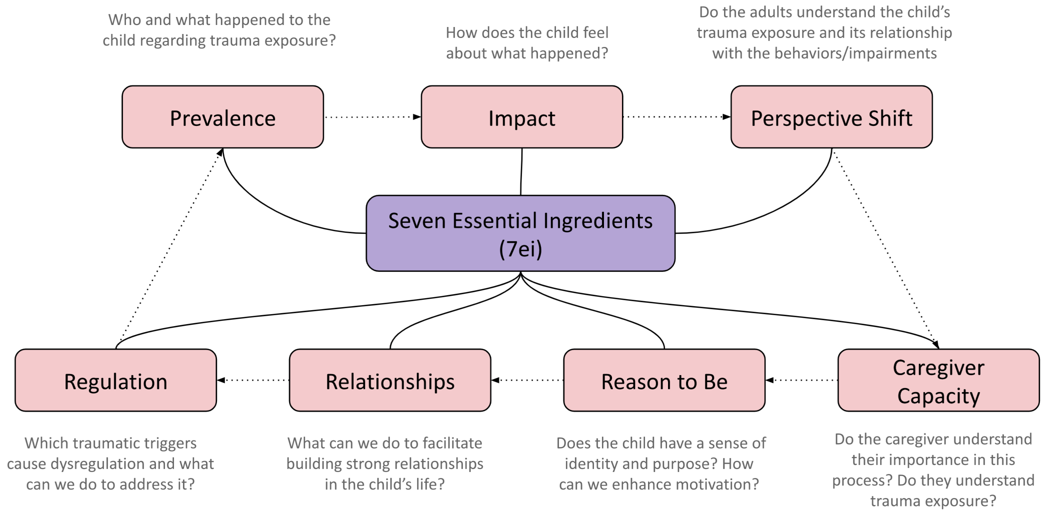


Figure 7: 7ei Algorithm: Associated Decision Outcomes

these domains. 7ei algorithm based in TIC has also been proven to improve child outcomes such as placement stability and permanence [450]. 7ei assessments are individually focused, however, unlike CANS, the results are trauma-focused and also guide family interventions [450].

7ei High-Stakes Decision Outcomes (RQ1)

The seven domains of 7ei are: *Prevalence*, *Impact*, *Perspective Shift*, *Regulation*, *Relationships*, *Reasons to Be*, and *Caregiver Capacity*. As such, 7ei is not directly tied to a specific outcome such that every time 7ei is used an outcome is recommended or predicted. Instead of predicting an outcome of interest using other factors, 7ei is used to track outcomes (i.e.- 7ei domains) over time to assess the trajectory of a child-welfare case. It is primarily used as a prescriptive tool that serves as a framework for team-based brainstorming of solutions guided by a trauma-informed care framework. As such, the seven domains of 7ei are both the input variables and the output variables of the 7ei assessment process because these are both the considerations and outcomes that the team is trying to measure and improve over time. Improvement in the 7ei domains is associated with placement stability and permanency outcomes, however, the agency leadership resisted developing a singular aggregate index that would measure this outcome. Ongoing conversations with agency leadership (program directors, quality improvement leaders) revealed that assessments are more likely to be manipulated if they are tied to singular metrics. Agency leadership also uses 7ei to assess progress within the agency with respect to trauma-informed care. For instance, it helps them understand whether conversations and meetings founded in TIC are leading to better permanency outcomes. Case-level outcomes of 7ei that are discussed at all the planning meetings and permanency consultations are depicted in Figure 7.

7ei and Algorithmic Decision-Making (RQ2)

We found that the prescriptive nature of 7ei help adaptively select data and address uncertainties. In this section, we discuss how the 7ei algorithm maps onto the ADMAPS dimensions of algorithmic decision-making.

90% ↑ **Relevant Data.** *7ei gives a comprehensive view of the child and their ecosystem.* Participants shared that the 7ei algorithm offers a comprehensive view of the foster child, caregivers, the impact of traumatic events in their life, as well as interactions in their social ecosystem. Moreover, the child-welfare teams in all the permanency consultations (n=40) and planning meetings (n=15) were able to adaptively select factors that were most pertinent to that case. Most participants (90%, n=18) emphasized that each case carried a lot of nuance and could not be addressed based on a few broad sets of predictors so they appreciated that 7ei allowed them to focus on certain factors and then brainstorm ideas on how to help the family make progress. For instance, a program director shared:

> *"We have tried the cookie cutter approach in the past. Assigning everyone to parenting classes, therapy, and other family support services. It failed and it is horrible to do to a family. So, with 7ei we focus on addressing core issues whether it's the parent's self-esteem, their own abandonment issues or child's emotional regulation that will really help this family"* -P7, Child Welfare Program Director, MSW, 20 years

Timeline of the case established which 7ei domains the child-welfare teams converged on. The teams focused more on some domains than others based on how long the child had been in the system. For instance, in the case of children who had been in the system for a few months, the team spent more time on *Prevalence* and *Impact* and then focused on *Perspective Shift* (see Figure 7). Trauma symptoms are associated with negative short-term outcomes such as placement instability [111], therefore, the team focuses on recognizing trauma early on so that proper interventions could be made that promote healing and improving outcomes. However, in cases that had been in the system for a longer period or had experienced multiple moves, the team focused more on *Regulation*, *Perspective Shift*, and *Reasons to Be*. Prior studies show that placements are often disrupted because foster parents are unprepared to manage the behaviors of children [89, 445], therefore, the team focuses on these domains to assess how to improve self-regulation for the foster child as well as expand the caregivers' understanding of trauma.

85% ↑ **Type of Decision-Support.** *The prescriptive nature of 7ei allows for brainstorming and idea generation.* Most participants (85%, n=17) appreciated that 7ei allowed for open discussions and brainstorming of solutions. 7ei is designed to be used in a team setting and acts as the TIC framework such that the child-welfare team deliberates over each domain, scores it, and formulates an action plan through a trauma-informed perspective. A case manager shared:

> *"I like 7ei because we use it as a team, and it allows us to brainstorm ideas. We look at not just the behaviors but the underlying trauma that is causing those behaviors."* -P18, Case Manager, BSW, 2 years

The 7ei domains are also the outcomes of interest that the team seeks to track and affect over time. Improvements in these domains are directly associated with the outcomes of placement stability and permanency which are tracked at the agency level.

80% ↑ **Degree of Uncertainty.** *Adaptively selecting relevant data helps address high degree of uncertainty.* Most participants (80%, n=16) felt that every case was contextually different and required their individual attention. That is, it was imperative they adequately weigh nuances and the factors pertinent to that case. 7ei allows the child-welfare team to adaptively select the domains pertinent to a case and spend significantly more time on them and brainstorm solutions. Tracking outcomes over time and not using the tool in a predictive capacity, allows the team to mitigate the high degree of uncertainty that would otherwise be associated with the predictions.

7ei Augments Human Discretion (RQ2)

In this section, we discuss how the 7ei algorithm maps onto the ADMAPS dimensions of human discretion.

85% ↑ **Professional Expertise.** *7ei helps develop professional expertise.* The algorithm is theoretically constructed and is centered in trauma-informed care. Participants (85%, n=17) believed that the algorithm allowed them to brainstorm ideas from a trauma-informed perspective and develop plans specific to that family. Continuous engagement with 7ei in trauma-informed meetings ensures that caseworkers are always thinking through TIC frameworks. A permanency consultant asserted:

> *"7ei isn't just a "thing" that we do. It is centered in everything that we do. It ensures that caseworkers are always thinking through TIC frameworks"* -P1, Permanency Consultation Supervisor, MSW, APSW, 22 years

This is especially important for CWS since the system lacks experienced caseworkers due to high turnover [89]. Continually working through 7ei under proper supervision ensures that new caseworkers are developing professional expertise.

70% ↑ **Value Judgments.** *7ei allows caseworkers to make informed value judgments.* This theme emerged in 70% (n=14) of the interviews and 72% (n=40) of the meetings. 7ei is centered in social work's core values of service, dignity, and worth of the person and allows caseworkers to prioritize these values and devise interventions that will directly help a child cope with trauma.

For instance, prioritizing the well-being of a child does not only mean sending them to therapy. It also incorporates addressing concerns within their ecosystem. A supervisor explained:

> *"Therapy only goes so far if nothing changes in the child's ecosystem and they feel continually triggered by others. With 7ei we try to address problems in this ecosystem and devise approaches that will help the family"* -P9, Supervisor, 13 years

7ei allows the child-welfare team to make value-based judgments and take steps that improve the quality of a child's relationships. This takes the form of family-level interventions or sharing information with caregivers about the impact of trauma to bring about a perspective shift.

85% ↑ **Heuristic Decision-Making.** *7ei enables heuristics judgment calls.* Most participants (85%, n=17) appreciated that 7ei offered them flexibility and autonomy in how they interact with it. This theme also emerged in all the planning meetings (n=15) and permanency consultations (n=40) where the child-welfare team adaptively selected information that was most pertinent to the case and often acted as an obstacle towards achieving permanency. For instance, in cases where the child had experienced multiple placement moves, the team focused on *Regulation* and *Reasons to Be* to devise plans that would help improve the child's emotional, behavioral, and cognitive functioning.

7ei Supports Bureaucratic Processes (RQ2)

In this section, we discuss how the 7ei algorithm maps onto the ADMAPS dimensions of bureaucratic processes.

60% ↑ **Resources & Constraints.** *7ei accounts for the resources at the organization.* Participants (60%, n=12) shared that 7ei is locally developed at the agency and accounts for the resources available at the agency in the form of supervision, expertise, and specialized trainings. Participants also shared that critical decision-making power in regard to achieving permanency sits with the legal parties (district attorneys, judges), however, 7ei operates within these constraints and allows caseworkers to help families and prepare them to be able to receive a favorable decision in court.

75%↑ 55%↓ **Administration & Training.** *7ei is embedded in daily work processes but requires additional oversight and training.* This theme emerged in 75 % (n=15) of the interviews and all the planning meetings (n=15) and permanency consultations (n=40) and there is both a benefit and a drawback associated with it. Participants (75%, n=15) recognized that continually engaging with the tool in meetings was helpful but some participants (55%, n=11) shared that the tool also added more tasks to an already heavy workload. The agency offers specialized trainings for trauma-informed care with 7ei acting as a tool that complements these trainings. Moreover, we observed that 7ei is embedded into daily work processes because it is utilized every time a case is discussed in a team setting. This continued engagement under supervision has helped earn the trust of caseworkers who learn not just from the algorithm but also from the collective expertise of the child-welfare team.

> *"I like 7ei because we use it as a team, and it allows us to brainstorm ideas. It just helps.. thinking out loud with everyone and knowing that I don't have make these decisions alone. Also, it helps guide my thought process but doesn't tell me what to do."* -P17, Case Manager, BSW, 8 years

However, 7ei has added to the workload of both the case managers and supervisors who must discuss and complete an additional tool as part of their job requirements at the agency.

65% ↓ **Laws & Policies.** *7ei is not legally mandated and only used locally at the agency.* Several participants (65%, n=13) were frustrated by the fact that 7ei was not legally mandated and that they had to continue using CANS. Even though an independent research study showed that 7ei is leading to better permanency and placement outcomes, the algorithm still lacks legitimacy at the state and federal levels.

3.5.3 Anti Sex-Trafficking (AST) Algorithm

An emerging high-stakes decision for which an algorithm is used in assessing the risk of sex trafficking for a foster child over 10 years of age. We observed this in 65% (n=13) of the interviews, 33% (n=5) of the planning meetings, and 35% (n=14) of the permanency consultations. The agency has a dedicated team called HART (Human Anti-Trafficking Response Team) that manages cases where the foster child might be at high risk of being trafficked. If a child meets the criteria for risk of sex trafficking per the AST algorithm, the case must be reported to HART. Early identification of such indicators can play a significant role in ensuring child safety.

AST High-Stakes Decision Outcomes (RQ1)

Risk indicators are divided into three domains - "At-Risk", "High-Risk", and "Confirmed". The child-welfare team must continue to closely monitor the case if they select fewer than three "At-Risk" indicators and continue to have conversations to mitigate the risks. If the team selects three or more "At-Risk" indicators or 1 or more "High-Risk" or "Confirmed" indicators, then the case must be referred to HART.

AST and Algorithmic Decision-Making (RQ2)

Overall, we found that the AST algorithm often missed the important context, which frustrated caseworkers. In this section, we discuss how the AST algorithm maps onto the ADMAPS dimensions of algorithmic decision-making.

65% ↑ 75% ↓ **Relevant Data.** *AST assesses pertinent risk indicators but misses context about the case.* Several participants (65%, n=13) shared that the algorithm offers a new perspective on sex trafficking by capturing risk indicators that they had not considered before. For instance, "possession of money, electronics, cosmetics, or clothes that are unexplained", "traveling out of the area or somewhere out of the ordinary", and "unwilling to provide information about an older partner" are some such indicators. A supervisor shared:

> *"It's helpful to have us think about it differently. I have been doing this for a very long time and when I started, we were never thinking about trafficking. So now we are more conscious of these risks"* -P8, Child Welfare Supervisor, MSW, APSW, 19 years

However, participants (75%, n=15) also shared the algorithm misses context about each case and the presence of some of these risk indicators did not mean the child was being sex trafficked.

75% ↓ **Type of Decision-Support.** *Predictive nature of AST frustrates caseworkers.* Several participants (75%) who found value in the algorithm as a guide also shared their frustrations in regard to its mandatory reporting nature. AST is used in a predictive capacity such that if certain risk indicators are selected then the case must be reported to HART. For instance, a permanency consultant shared:

> *"If a child is at risk for sex trafficking then we are having those conversations from the very beginning and taking necessary action. This decision tool is not helpful in the way it's being used and only frustrates HART"* -P5, Permanency Consultant. MSW, 12 years

Reporting to HART takes the case away from the child-welfare team who have spent a significant amount of time building a relationship with the child and their caregivers. Moreover, HART is receiving an influx of calls that do not require their expertise as a result of this algorithm.

75% ↓ **Degree of Uncertainty.** *Lack of context about the case leads to a high degree of uncertainty.* Several participants (75%, n=15) shared that even though the algorithm was useful as a guide, the presence of risk indicators did not mean the child was at risk of being trafficked. There was still a lot of pertinent information that was necessary to make such a determination. For instance, a supervisor shared that one of their foster kids had a history of sexual abuse and met a bunch of criteria on the tool, however, the supervisor's team is actively involved with the child and their caregivers, understands their needs, and did not believe the case needed to be reported to HART.

AST Impedes Human Discretion (RQ2)

In this section, we discuss how the AST algorithm maps onto the ADMAPS dimensions of human discretion.

70% ↑ **Professional Expertise.** *AST has made caseworkers more aware of risk indicators.* Most participants (70%, n=14) appreciated that this decision tool taught new caseworkers to be actively aware of indicators that are associated with sex trafficking. As stated by P5 above, AST has improved professional expertise since caseworkers are more cognizant of such risks.

75% ↓ **Value Judgments and Heuristic Decision-Making.** *AST does not allow caseworkers to make value judgments or engage in heuristic decision-making.* Several participants (75%, n=15) shared that the risk indicators were just part of a bigger picture and there was still a lot of context that the child-welfare team needed to collectively unpack to better assess the situation. This requires the child-welfare team to engage in heuristic decision-making and weigh the necessary information. For instance, a case manager shared:

> *"One of my foster teens is sexually active and has a long-term boyfriend. But looking at the tool, everyone is like... 'Is that her boyfriend or her pimp?' I know this kid and she trusts me... building trust with these kids takes time... she doesn't need to be referred to HART [Human Anti-Trafficking Response Team]"* -P17, Case Manager, BSW, 8 years

Here, the case manager emphasizes that having trusting relationships was equally important for preventing foster youth from being sex trafficked. Referring the case to HART takes this case away from the case manager and the child is assigned a new case manager from HART.

AST and Bureaucratic Processes (RQ2)

Overall, we found that AST is inadequately supported by bureaucratic processes which leads to frustrations on the part of caseworkers. In this section, we discuss how the AST algorithm maps onto the ADMAPS dimensions of bureaucratic processes.

60% ↓ **Resources & Constraints.** *AST has caused an influx of cases referred to HART.* Several participants (60%, n=12) shared that AST has added to the frustrations of HART who do not have enough resources to manage all the cases that are being reported to them because of the mandatory reporting aspect of AST. This leaves HART with significantly fewer resources to focus on the cases that need them. A supervisor on HART explained:

> *"It's become overused and is being abused. It's being used in different settings and not how it was originally intended to be used. We are getting an influx of calls that don't need to be called."* -P10, Child Welfare Supervisor, MSW, LPC, NCC, 9 years

65% ↑ 60% ↓ **Administration & Training.** *AST is conducted as an organizational process and offers some training.* Most participants (65%, n=13) shared that AST is routinely conducted for foster youth over 10 years of age. As discussed previously, it trains new caseworkers to be more conscious about risk factors but often to the detriment of HART. Participants also consider this decision tool to be peripheral to the case planning process for most cases and utilized it as a mandated requirement.

60% ↑ **Laws & Policies.** *The use of AST is mandated by law.* The Department of Children and Families (DCF) has legally mandated the use of this algorithm as a means to proactively protect foster youth as well as collect more data about associated risk factors and the number of foster youth considered at risk. Next, we discuss the Legal Permanency Status (LPS) algorithm.

3.5.4 Legal Permanency Status (LPS) Algorithm

Tracking performance metrics such as placement stability and permanency emerged as another decision outcome for which an algorithm is being used. Several participants (65%, n=13) discussed this decision tool, and we observed the tool being used at all the permanency consultations (n=40). Federal legislation has established permanency as one of the primary goals of CWS and requires agencies to meet this well-defined and measurable benchmark [481]. Legal permanency is defined as reunification with the biological family, adoption or transfer of guardianship. [275]. This tool establishes a sense of urgency and insistence towards achieving permanency since prolonged instability and multiple placement moves leads to poor well-being outcomes for foster children.

LPS High-Stakes Decision Outcomes (RQ1)

This decision tool is used to track an outcome over time instead of predicting an outcome of interest using input data. The agency uses it at permanency consultations to track the quality of the current placement and if those specialized meetings are leading to better outcomes. The tool is facilitated by the permanency consultant and upon completion, the team categorically rates the quality of the placement. The tool is also used to track systemic barriers that are getting in the way of achieving permanency. Some participants recognized (40%, n=8) the utility of this tool and a permanency consultant explained –

> *"It helps us to be actively aware of where we are at in terms of permanency. We have to follow a timeline for permanency and if parents are not showing initiative and not completing court-ordered services then we need to start exploring placement options. So, this tool is kind of an extra push"* -P5, Permanency Consultant, MSW, 12 years

LPS and Algorithmic Decision-Making (RQ2)

Overall, we found that even though LPS is used in a prescriptive capacity, it still lacks utility. In this section, we discuss how the LPS algorithm maps onto the ADMAPS dimensions of algorithmic decision-making.

65% 🡣 **Relevant Data.** *Definitions of input variables are ambiguous and frustrates caseworkers.* Several participants (65%, n=13) shared their indifference towards LPS and stated that the definitions of the input variables in terms of what constitutes a "Good", "Fair", or "Poor" placement was ambiguous. A permanency consultant who is tasked with conducting this decision tool shared:

> *"This tool is confusing in itself; the definitions are very vague. We can't necessarily put families into categories. We have to use this tool and have conversations about where they [families] might fit best, but it doesn't give a clear picture of what's really going on with the placement."* - P6, Permanency Consultant, BSW, 3 years

Participants (65%, n=13) also shared that how the tool was conducted depending on the perspective of the team. Different people might look at the same set of facts and reach different conclusions as to why the placement should be rated as good, fair, or poor.

60% 🡣 **Type of Decision-Support.** *LPS is used as a prescriptive tool, but caseworkers lack agency towards affecting outcomes.* Participants (60%, n=12) shared that the tool did not predict or recommend an outcome based on the inputs, however, the utility of the tool lies in recognizing the current state of a case with respect to permanency and addressing systemic barriers, but the child-welfare team lacks agency with respect to addressing several systemic barriers. For instance, critical decision-making power in regard to permanency decisions sits with the legal parties (district attorneys and judges) who can choose to fully disregard the child-welfare team's recommendation.

65% 🡣 **Degree of Uncertainty.** *Lack of relevant information leads to high degree of uncertainty.* Participants (65%, n=13) were most frustrated by the lack of relevant information that LPS needed to account for in order to make a proper determination about the quality of the placement. Moreover, there is high uncertainty associated with the timeliness with which some systemic barriers can be addressed making it hard to assess the quality of the placement (see P2 quote below).

LPS impedes Human Discretion (RQ2)

In this section, we discuss how the LPS algorithm maps onto the ADMAPS dimensions of human discretion and its impact on the dimension.

40% ↑ **Professional Expertise.** *LPS builds professional expertise by establishing an urgency towards permanency.* Some participants (40%, n=8) shared that the decision tool was useful in that it established a sense of urgency and insistence towards achieving permanency for foster children. It teaches new caseworkers that they needed to prioritize finding placement options even if that upsets biological parents. The child-welfare team follows a 15-month timeline where the parents must complete court order services within this timeline to achieve reunification. Towards the end of this timeline, the team must explore alternate placement options.

60% ↓ **Value judgments and Heuristic Decision-Making.** *LPS tool does not allow for value judgments or heuristics.* Participants (60%, n=12) shared that the tool does not allow for value judgments or heuristics on the part of the child-welfare team since the decision-making power about permanency decisions sits with the legal parties.

3.5.5 LPS and Bureaucratic Processes (RQ2)

Overall, LPS supports bureaucratic processes but does not account for constraints. In this section, we discuss how the LPS algorithm maps onto the ADMAPS dimensions of bureaucratic processes.

65% ↓ **Resources & Constraints.** *LPS does not account for organizational resources and imposes new constraints.* Participants shared (65%, n=13) that ambiguity around input variables and lack of agency with respect to permanency decisions has turned this decision tool into documentation that every child-welfare team must complete at the permanency consultations. Moreover, there are arbitrary constraints placed on the decision tool in regard to how a placement can be rated. For instance, a program director explained:

> *"We may have a court hearing date set and all of a sudden, the placement is "Good". We now have a "Good" rating because we have a court hearing. But it might take three years to terminate parental rights or go to a guardianship. And then you went three years without achieving permanency but somehow, we have a "Good" placement rating."* - P2, Permanency Consultant, MSW, 20 years

80% ↓ **Administration & Training.** *Caseworkers are indifferent towards how LPS is administered.* Even though the tool is expected to be central to permanency consultations, we noticed that for the majority of these meetings (87%, n=35), LPS was used towards the end as a requirement to rate the quality of the placement. Most participants (80%, n=16) were indifferent towards LPS because of the several constraints and utility issues discussed above. One supervisor shared:

> *"It's just another thing we have to do in the permanency meetings. I let the Permanency Consultants score it however they like"* - P15, CW Supervisor, MSW, 9 years

65% ↑ **Laws & Policies.** *The use of the LPS tool is legally mandated and helps track performance outcomes.* Participants (65%, n=13) shared that the Department of Children and Families (DCF) has mandated using this decision tool as a way to track the important outcomes of quality of placements and permanency. Moreover, it's important to track the systemic barriers that often impact permanency to formulate policies at the state level that address these systemic barriers.

3.5.6 Assessing the Benefits and Drawbacks of Differing Approaches (RQ3)

Table 6 offers a summary of how the four algorithms balance the dimensions of the framework. In this section, we discuss the benefits and drawbacks that arise when trying to balance the tradeoffs between human discretion, algorithmic decision-making, and bureaucratic processes.

Algorithmic Decision-Making Should Augment Human Discretion, Not Supplant it. In this section, we discuss the themes around the benefits and drawbacks that emerged when balancing the ADMAPS dimensions of human discretion and algorithmic decision-making.

When aligned, algorithms augment decision-making processes, but a lack of alignment can take away autonomy and heuristic decision-making. All of our participants emphasized that every case is contextually different and that a family's circumstances often change throughout the life of the case. For instance, a parent seeking reunification with their child might experience a lapse while trying to maintain a stable job, maintain their sobriety, or consistently attend court-ordered services. In other words, algorithms should be designed with the recognition that there will be a high degree of uncertainty associated with any relevant data and the subsequently predicted outcome. Therefore, algorithms need to not only make room for human discretion but also facilitate value judgments and heuristics in order to offer utility. Most of our participants (80%, n=16) mentioned that the 7ei algorithm augmented their decision-making processes when they were making difficult decisions. It prioritizes and enhances the value judgments and heuristic decision-making that caseworkers must engage in when devising action steps to help families. Participants (85%, n=17) especially appreciated that 7ei facilitated brainstorming and idea generation instead of predicting an outcome of interest. One case manager explained:

> *"Of all the things we have brought up, 7ei is my favorite because its helps us think differently, understand what a family has been through and then brainstorm ideas on how to help them based on this understanding of trauma"* -P3, Permanency Consultant, MSW, APSW, 9 years

Participants noted that the tool offers flexibility and autonomy in how they interact with it and which 7ei domains they focused on. Adaptively selecting information instead of analyzing all the information is a key feature of heuristic decision-making and can lead to more accurate

	Human Discretion		Algorithmic Decision-Making		Bureaucratic Processes	
CANS	Professional Exp.	80 ↓	Relevant Data	70 ↓	Resources	40 ↑ 70 ↓
	Value Judgments	75 ↓	Decision-Support	90 ↓	Admin & Training	40 ↑ 80 ↓
	Heuristic Decisions	80 ↓	Uncertainty	70 ↓	Laws & Policies	80 ↑
7ei	Professional Exp.	85 ↑	Relevant Data	90 ↑	Resources	60 ↑
	Value Judgments	70 ↑	Decision-Support	85 ↑	Admin & Training	75 ↑ 55 ↓
	Heuristic Decisions	85 ↑	Uncertainty	80 ↑	Laws & Policies	65 ↓
AST	Professional Exp.	70 ↑	Relevant Data	65 ↑ 75 ↓	Resources	60 ↓
	Value Judgments	75 ↓	Decision-Support	75 ↓	Admin & Training	65 ↑ 60 ↓
	Heuristic Decisions	75 ↓	Uncertainty	75 ↓	Laws & Policies	60 ↑
LPS	Professional Exp.	40 ↑	Relevant Data	65 ↓	Resources	65 ↓
	Value Judgments	60 ↓	Decision-Support	60 ↓	Admin & Training	80 ↓
	Heuristic Decisions	60 ↓	Uncertainty	65 ↓	Laws & Policies	65 ↑

Table 6: RQ3: Tradeoffs between balancing the dimensions within human discretion, algorithmic decision-making, and bureaucratic processes, The numbers represent percentage of participants who stated that a dimension was positively (or negatively) impacted by the algorithm.

decisions. On the other hand, tensions arise when algorithms attempt to supplant human discretion. This is the case with CANS whose predictive nature does not account for the high degree of uncertainty that accompanies each case, and consequentially, does not make room for discretionary work on the part of the caseworkers. With a lack of autonomy, gaming the algorithm is the only way caseworkers are able to exercise discretion and produce the desired outcome. One supervisor shared:

> *"CANS is all about producing a good rate so foster parents can afford the resources they need to take care of the child. I have had foster parents put in notices [to end placement] because they couldn't support the child anymore"* -P14, Child Welfare Supervisor, MSW, 30 years

This inflexible and predictive nature has shifted focus away from the primary outcome of interest (i.e., mental health screening) and towards the secondary outcome.

When aligned, algorithms can help embed important value judgments into the decision-making process. Most of the participants (80%, n=16) felt strongly about the need to support each family in a different capacity and through different practices. This was also a dominant theme in all the planning meetings (n=15) and permanency consultations (n=40) where the child-welfare team devised specific plans for each family through trauma-informed care (i.e. - using 7ei). 7ei algorithm is centered in some of the social work's core values of service, dignity, and worth of the person, and the importance of human relationships [450] and informs the child-welfare team's work processes. For instance, a supervisor explained:

> *"7ei helps us think about how we can help every family. What can we do to help mom develop her self-worth? How can we help her build relationships with relatives or people in the community, so she has more caregiver support"* -P8, Child Welfare Supervisor, MSW, CAPSW, 19 years

Several participants (65%, n=13) explained that 7ei scores translated into actionable steps that directly sought to help children and their families. In the meetings, 7ei brainstorming sessions resulted in solutions that the child-welfare team could affect directly and not simply refer children and parents to therapy. For instance, the case managers and supervisors planned activities that the family could engage in to improve child as well as family functioning, discussed information to share with foster and biological parents about impacts of trauma exposure, as well as steps caregivers could take to establish healthy boundaries with foster children and enforce positive discipline. For instance, a permanency consultant explained:

> *"It [7ei] helps you have a perspective shift on the family and the child. You don't need to refer everyone and their mother to therapy. Sometimes it's just as simple as having them do something as a family that's different than what they ever did before [picnics, sports].. and challenge them in different ways"* -P2, Permanency Consultant, MSW, 20 years

On the contrary, algorithms that do not embed human values into their design may consequentially end up minimizing them. For instance, CANS recommends the level of foster care that the child should be placed in, and based on that the child-welfare team finds foster parents who have the resources to meet those needs. However, participants explained that more financial resources do not always equate to a foster child's well-being. For instance, a permanency consultant explained:

> *"Everyone has a different compass for well-being. We had a child who was placed with well-off foster parents in a five-bedroom house and he wasn't doing well there because it was a culture shock for him. He wasn't used to a huge home, a great school, a big backyard... and he completely shut down. So, we found his aunt and we moved him down to Chicago. It's a two-bedroom house with five other kids and this child is thriving! So, he needed to be with his family, and he needed to have his own culture.* -P4, Permanency Consultant, MSW, APSW, 8 years

Here the permanency consultant emphasizes the core human value of having trusting relationships in one's life and the importance of one's culture which plays a critical role towards achieve emotional and cognitive well-being. Unlike CANS, 7ei allowed the child-welfare team to prioritize *Reasons To Be* and *Regulation* to really help children instead of simply focusing on financial resources. Caseworkers must continually negotiate values and balance the individual needs of people with the demands of policymakers. Algorithms that do not allow caseworkers to make such value judgments will most likely lead to more frustrations and limit their utility.

Algorithms replacing the need for caseworker expertise can lead to inadequate and unreliable decision-making. Another dominant theme that emerged in 70% (n=14) of the interviews, 53% (n=8) of the planning meetings, and 50% (n=20) of the permanency consultations was that the algorithms in use were diminishing the need for caseworker expertise as well as a need for family-centered care in a very contextual domain. CWS is a high-stakes domain where most caseworkers lack adequate work experience, carry high caseloads, and are under a lot of pressure from legal parties. Most of the participants (70%, n=14) expressed concerns that algorithmic systems might simply act as a safe default for most caseworkers such that questioning an algorithmic decision would add more work to their plate. A program director explained –

> *"We are not hiring people with a lot of experience. All new hires are recent graduates who don't quite know what this field looks like. They are happy to trust the machine to just get through the day"* -P7, Child Welfare Program Director, MSW, 20 years

Professional expertise in the public sector deteriorates when algorithms limit the scope for value judgments and heuristic decision-making. Value negotiations and heuristics are indispensable aspects of professional practice that workers must continually engage in to build expertise [300]. There are also growing concerns that algorithms such as CANS are leading to children being referred to unnecessary services which also shifts the focus away from family-centered care. CANS scores directly translate to actionable steps in the form of services that children are referred to. Caseworkers co-opt CANS to produce higher compensations for foster parents or the foster parents might exaggerate child behaviors, however, several participants (60%, n=12) recognized that this also meant unnecessary services being requested for children. One case manager explained:

> *"Foster parent might exaggerate behaviors just to get more money. And then we put the kid in therapy and are not addressing their needs specifically. The kid in therapy is then being asked 'why are you so sad?' and the kid is like.. 'I'm not sad.' "* -P17, Case Manager, BSW, 8 years

This is further problematic because unnecessary services are not only an added financial burden on CWS but they also add to the medical trauma of foster children who are continually being told that something is wrong with them. Finally, as previously noted, services assigned through an isolated view of a child might be less effective than family-level interventions developed through a trauma-informed perspective. Algorithms like CANS, however, limit child-welfare workers from using their expertise and developing more family-centered practices.

Algorithmic Decision-Support Systems and Bureaucratic Processes Need to be Aligned.

In this section, we discuss the themes around the benefits and drawbacks that emerged when balancing the ADMAPS dimensions of algorithmic decision-making and bureaucratic processes.

When algorithmic decision-making and bureaucratic processes are aligned, it can help train caseworkers. Most participants (70%, n=14) recognized the value of algorithms as essential training mechanisms. Continuous engagement with 7ei in trauma-informed meetings under proper supervision ensures that caseworkers are continually having conversations centered in trauma and are coming up with solutions founded in TIC. This is essential because CWS suffers from a high turnover with lack of adequate training and supervision being some of the leading causes [89]. During the observations, experienced members of the child-welfare team such as the supervisors and permanency consultants brainstormed ideas with the caseworkers using 7ei and shared practices and approaches that had worked in the past with other families. A program director explained:

> *"It [7ei] makes us think differently and in the moment think through TIC [trauma-informed care].. what is the impact of that [incident]? How can we help support parents and children and help with emotional regulation? What can we do to build their relationships and how do we support the people in their life so that they show better caregiver capacity to manage what this child is going through"* -P7, Child Welfare Program Director, MSW, 20 years

7ei meetings are also attended by Caregiver Support Specialists and Family Preservation Specialists who based on case circumstances also offer their expertise to the caseworkers on how to proceed. Several participants (70%, n=14) also recognized the Anti Sex-Trafficking algorithm as a good training tool such that caseworkers are always aware of and looking for red flags associated with sex trafficking. This tool is also used in a team setting and allows for the less experienced caseworkers to learn from the seasoned members of the team in how to perceive certain risks as well as how to follow up on them without confronting the foster youth. One supervisor explained:

> *"Its a whole change of mindset and and we're now more cognizant of some of those red flags. So I appreciate that. So, anytime in supervision, if they [caseworkers] start talking about some of these things, we pull up the tool and go through it and start discussing how to proceed"* -P15, Child Welfare Supervisor, 9 years

However, participants who found great value in the Anti Sex-Trafficking algorithm as a training mechanism were also equally frustrated by the mandatory reporting nature of it.

When the algorithmic decision-support system is not aligned with the constraints of the bureaucratic system, it leads to utility issues. The utility of algorithmic decision-support systems is severely limited when they do not account for the organizational constraints within which they must operate. This challenge and frustration on the part of caseworkers was observed in 80% (n=16) of the interviews, 73% (n=11) of the planning meetings, and 82% (n=32) of the permanency consultations. Most of the participants (80%, n=16) stated that they were unable to use algorithms as intended because of several organizational constraints such as limited time availability due to high caseloads. For instance, properly conducting CANS requires the caseworkers to interview several individuals such as the foster parents, relatives, and teachers to be able to get consistent information. However, most caseworkers have high caseloads and do not have enough time to devote to properly conducting each assessment. A permanency consultant explained:

> *"It can take hours to properly do CANS. Everyone is stretched too thin.. we literally do not have the time to do that."* -P2, Permanency Consultant, MSW, 20 years

Moreover, the bureaucratic policy that requires CANS to be conducted every six months makes it unfeasible for caseworkers to interview anyone beyond the foster parents. However, as previously noted, the assessment is also tied to the compensation offered to foster parents which further leads to foster parents exaggerating child behaviors. A supervisor explained that the algorithm did not need to be manipulated and caseworkers could request an exceptional rate to account for several different factors such as transportation services, therapy, school activities, etc. However, she also explained that high caseloads make it hard to devote any extra time to each assessment:

> *"The supervisor or worker can take a little extra time and put an exceptional amount in. But again, everybody has too much to get done and it's unlikely for somebody to even think about that until they are in the middle of submitting CANS and its due the next day."* -P8, Child Welfare Supervisor, MSW, CAPSW, 19 years

Caseworkers who continue to feel disempowered and frustrated by CANS find it easier to manipulate CANS scores to produce a higher rate than request the addition of an exceptional rate which needs to be approved by the supervisor. Most participants (75%, n=15) also shared significant concerns about the fidelity of data being used to develop algorithms. Administrative data curated through bureaucratic processes can not be uncritically used to develop algorithms. Participants shared that the data about families is collected by caseworkers and it was hard to acquire accurate information. Oftentimes, caseworkers might receive conflicting information from parents, relatives, and neighbors. A supervisor explained:

> *"People aren't willing to happily share personal information about themselves. And understandably so. It's really hard to be able to get consistent information to be able to put in some data system to bust out a decision. So I find this really difficult to comprehend how we would even consider that"* - -P13, Child Welfare Supervisor, MSW, 12 years

Families view caseworkers as representatives of the state and are unwilling to trust them or share any information that they might consider incriminating. Therefore, much of the information is based on the caseworker's perception of the family. This problem is further exacerbated by the fact that most caseworkers are new graduates and lack adequate experience. A program director explained that during investigations, caseworkers need to ask the right questions, read situations, and follow up to be able to derive meaningful information. However, caseworkers acquire these skills through years of experience.

If the algorithms are not explained well to the professionals, they don't trust them. Most of the participants (85%, n=17) were unaware of the algorithms being used. They did not recognize CANS or 7ei as algorithmic systems, even though they interact with them on a daily basis. This finding is especially interesting because the CANS training material recognizes it as an algorithm and explicitly lays out the purposes (i.e. - outcomes) that the algorithm is designed to accomplish. CWS started using algorithms as a means to track important performance metrics such as permanency and placement stability as mandated by federal legislation [483]. Moreover, several psychometric assessments that are routinely used in child-welfare to assess the risks and needs of children and parents have adopted algorithmic analogues in the form of risk assessment algorithms. They are now being used as data collectors for all cases as well as decision arbiters for future cases. The CANS algorithm is a case in point of this scenario which is an algorithmic version of the CANS communimetric assessment [301] and has been re-purposed for other outcomes. Interestingly, caseworkers were well aware of assessments used at the agency but did not actively recognize the algorithmic components unless the interviewer nudged them to think about some of the automated aspects. For instance, when explicitly discussing CANS, all the interviewees realized it to be an algorithm that plays a pertinent role in decision-making. We asked a permanency consultant how she thought CANS worked and she exclaimed:

> *"Yes, okay! Right! Because we put in the scores, and it generates the rate. Oh and I hate the CANS! I, I've expressed it a lot of times that CANS should not be tied to money."* -P6, Permanency Consultant, BSW, 3 years

As soon as the participants recognized the automated decision-making aspects of these decision tools, they started to share some other ambiguous aspects. For instance, they had trouble recognizing who managed these decision tools and shared that different tasks associated with these tools were distributed across the agency. Most of the participants (85%, n=17) were unable to isolate CANS to a particular role or "place". CANS is embedded in several aspects of the case planning process with different departments involved. The case management team focuses on assessing risks and needs whereas the data specialists track scores, and the fiscal liaisons reviews and approve foster parent compensations. Participants (70%, n=14) also alluded to the fact that algorithms meant different things to people based on their intended goals. For instance, one permanency consultant asserted:

> *"Case management try to do a good job with risks and needs but of-course foster parents only care about the rate. I know X's team [data specialists] look at scores to see if there is improvement. But there is also Y [fiscal liaison] who just wants the scores to be turned in on time so that he doesn't have to chase people down"* -P1, Permanency Consultation Supervisor, MSW, APSW, 22 years

Several participants (75%, n=15) recognized that CANS, AST, and LPS decision tools meant different things in different contexts. Lack of awareness about algorithms at the organizational level and their distributed nature augmented caseworkers' distrust.

Collective Buy-in Requires Trust in Outcomes at Both the Caseworker and Bureaucratic Levels of the Organization.

In this section, we discuss the themes around the benefits and drawbacks that emerged when balancing the ADMAPS dimension of human discretion and bureaucratic processes with respect to decision outcomes that the CWS agency implements.

When the algorithmic outcomes miss important context, people don't trust them. This challenge emerged in 70% (n=14) of the interviews, 67% (n=10) of the planning meetings and 75% (n=30) of the permanency consultations. Algorithms attempt to generalize child and family characteristics and place them in certain categories to be able to make a determination. However, this inadvertently leads to a loss of information with respect to the final outcome since all the information cannot be accounted for by these algorithms. For instance, the CANS algorithmic assessment focuses on child behaviors, risks, and needs based on the last 30 days and does not account for traumatic triggers that can lead to serious emotional dysregulation from time to time. A permanency consultant explained:

> *"CANS does not account for trauma in the way 7ei does. We have a child that goes into manic depression every year around holiday season and needs to be cared for 24/7. So, one of the foster parents has to quit their job. However, there is no way to account*

for that in CANS. This child is doing fine right now but we know that traumatic trigger is coming." -P3, Permanency Consultant, MSW, APSW, 9 years

Participants shared similar concerns about the Anti Sex-Trafficking tool. Participants (65%, n=13) shared that the tool was useful to identify risk factors, however, the presence of risk factors does not necessarily mean that the foster youth is being trafficked. A supervisor explained:

"We have a 12 year old boy who suffers from a lot of sexual trauma. He met a bunch of criteria on the tool and we had to report to HART. We have been with this boy for many years.. we know him and what his behaviors are connected to. Reporting to HART felt very odd" -P11, Child Welfare Supervisor, MSW, APSW, 9 years

Here, the supervisor alludes to the fact that the presence of risk factors did not capture the full picture and missed important context about what is going on with this child. As previously discussed, child-welfare teams are having ongoing conversations about the risk of sex trafficking from the onset of a case and are able to investigate concerns without reporting to HART. Continued interactions illustrated by these examples lead to an accumulation of distrust towards algorithms that often miss important nuances and context.

There is collective buy-in when people feel like it leads to better outcomes. This theme emerged in 85% (n=17) of the interviews, 67% (n=10) of the planning meetings, and 62.5% (n=25) of the permanency consultations. Participants appreciated 7ei and trauma-informed care because it allowed them to directly help families by developing specific plans for them. Moreover, they were aware that trauma-responsive services developed through TIC lead to better permanency and placement outcomes for children. When frustrated by algorithms such as CANS, caseworkers often asked why they were still expected to use CANS when 7ei was leading to better outcomes.

The agency developed a comprehensive four-part evidence-based service program centered in trauma-informed care of which the algorithmic tool is just one component. The first component of this program involves extensive ongoing TIC trainings. The second component introduced child-welfare staff to trauma-informed assessments that resulted in family-level interventions. The third component was the availability of specialized supervision and consultations provided by a clinical supervisor, TIC program administrator, medical expert, and caregiver support specialist. Finally, the fourth component of this program is the 7ei algorithm which allows the child-welfare team to break down each case into seven domains and apply TIC principles and practices to it. In sum, the agency significantly invested in resources to ensure that TIC practice and 7ei were fully supported by bureaucratic processes. A better understanding of what a family might be going through at a psychological level and knowing exactly how to help them has led to the collective buy-in from caseworkers. Most participants (85%, n=17) found great value in

TIC and appreciated 7ei as a tool designed to be a guiding framework for TIC. One permanency consultant explained:

> *"Compassion fatigue in child welfare is very real. This tool has helped me truly understand what trauma can do to the brain. So I think now that we have a better understanding of trauma... we also have more empathy, and we go the extra mile to help families."* -P1, Permanency Consultation Supervisor, MSW, APSW, 22 years

This was also evident at the meetings where the child-welfare staff urged case managers to ask and think in terms of *"What happened to you and how can I help you?"* instead of *"What is wrong with you?"* when working with children and families. That is, having a perspective shift and thinking in terms of the impact of trauma and not just exhibited behaviors. Moreover, a study published by independent researchers showed that child-welfare teams who implemented TIC practices using 7ei exhibited improved permanency and placement stability outcomes for their cases. This has further deepened the collective buy-in from caseworkers who often brought up the "study" to state that their practice centered in TIC worked and led to better outcomes, and therefore, the legal parties needed to trust their judgments more. However, interestingly, none of the participants except the program director were able to locate the study and share it with us.

Lack of trust in algorithmic outcomes, leads to concerns about unethical and unsound decision-making. This theme emerged in 80% (n=16) of the interviews, 40% (n=6) of the planning meetings, and 45% (n=18) of the permanency consultations where participants recognized that algorithms in use sometimes led them to make unethical or unsound decisions. Surprisingly, even though 90% (n=18) of the participants recognized that caseworkers were gaming the system to produce higher compensations, some participants (40%, n=8) did not consider the decisions made to be unethical or unsound. To them, this is how the system was set up to be used. With contradictory incentive structures and conflicting values, caseworkers are often put in a position where they are forced to make such decisions. As previously discussed, both caseworkers and foster parents are co-opting CANS to produce a higher compensation, however, most participants (90%, n=18) recognized that the base compensation offered to foster parents is pretty low and gaming the algorithm was the only convenient way to produce an adequate rate. A supervisor explained:

> *"There is a lack of good foster homes and there is no financial incentive to be doing this work... So caseworkers do whatever they can to get them [foster parents] the money they need or we risk disrupting a placement...Several foster parents have put in notices in the past because they can't financially sustain the placement"* -P9, Child Welfare

Supervisor, MSW, 13 years

Another challenging aspect of using CANS which leads to unsound decision-making is that it is now being used to track progress with respect to mental health; an unintended outcome of the continuous data collection. Consequently, an improvement (or deterioration) in mental health is impacting placement decisions. A clinical therapist explained:

> *"Sometimes a kid isn't doing well because they are working through things. For example, when a kid starts therapy sometimes their behaviors gets worse, but it might actually be that they're working through some things in therapy. And that's not captured by the tool. When there is no context, I just can't interpret whether or not that worsening is actually a bad thing."* -P16, Clinical Therapist, LCSW, 5 years

These frustrations were also consistent with the fact that CANS is tied to service-planning where CANS recommends unnecessary child-level services when family-level interventions are often more effective in helping the family heal and cope with trauma. Prior studies have found that caseworkers distrust risk assessment algorithms such as CANS because of their deficit-based nature [72]. Interestingly, participants who were averse to this deficit-based nature did find value in risk assessment in regard to understanding a family's history. One permanency consultant explained:

> *"The data gives you a bigger picture of what the family is going through. Because the families are not always honest with us about what they need. And so if you can see that they were in drug and alcohol services or housing authority... that would be helpful in child welfare. Not necessarily giving them a score. I don't even know what to do with the score. It tells me nothing."* -P6, Permanency Consultant, BSW, 3 years

Understanding a family's past helps caseworkers assess the type of services that would be most beneficial for them. Looking at this data from a trauma-informed perspective allows caseworkers better understand the underlying trauma in the family that might be leading to exhibited behaviors and subsequent interactions with the child-welfare system. This points to a change in mindset in regard to the intended goal of risk assessment of (i.e. - predicting the risk of future harm) towards a strength-based outcome of providing the right services to families and not necessarily labeling them as high or low risk. The permanency consultant here also invokes social work's core value of service and helping families and sees the utility of risk assessment through that perspective.

3.6 Discussion

Our results provide implications for algorithmic decision-making in CWS, more broadly for the public sector, as well as specific design guidelines for developing such systems in the public sector.

3.6.1 Implications for Algorithmic Decision-Making in the Child-Welfare System

Identifying Gaps and Opportunities for Algorithmic Decision-Making for Child-Welfare through the ADMAPS Framework

Our results suggest that caseworkers have differing perceptions of the design, outcomes, and intended uses of the various algorithms. ADMAPS dimensions helped uncover the caseworker perspectives with respect to different aspects of algorithms. For instance, examining the bureaucratic processes at the agency revealed that the most frustrating part about CANS was that it was being used to calculate financial resources for foster parents. Furthermore, deliberating over the relevant data and degree of uncertainty with the participants revealed that CANS did not account for traumatic triggers in a child's ecosystem or the lack of interpretation regarding worsening behaviors. That is, ADMAPS uncovered the multiple and conflicting roles of CANS as a compensation calculator as well as a mental health assessment tool. On the other hand, participants considered 7ei to be analogous to the trauma-informed care framework but not an algorithm that was continually collecting data and tracking outcomes over the life of a case. Differing socio-technical imaginations of algorithms in the public sector are aligned well with existing literature [72, 397, 414, 232, 385]. Moreover, while many of our participants were able to connect how CANS risk scores were used in making decisions about a child based on state guidelines, yet others were unable to make this connection since CANS is deeply embedded within their daily culture, where they did not recognize it as an algorithm anymore. This aligns with prior literature on thinking about algorithms as part of culture [414] and points to a need in raising more awareness among caseworkers about the algorithms that they use daily. One way to improve this is to support explanations of algorithms in daily use; not just interpretations of outcomes but also designing around other interactions (e.g. data collection, input, visualization, etc). Recent work on explainable AI has found support for focusing on explanations of daily interactions of users with algorithms [293, 333]. In addition, transparency of data, methods, and outcomes can support collaborative algorithmic decision-making processes [407, 231]. Finally, assessing the scope of bureaucratic processes and how algorithms must function within these constraints allows for better practitioner engagement and buy-in.

Balancing Strength-based and Deficit-based Approaches

Our results show that there are differing approaches in algorithmic implementation that is mandated by the state as opposed to the non-profit agency that is contracted by the state, to provide child-welfare services. For instance, the state mandates the use of CANS, which the caseworkers are not very amenable to as they can see the deficiencies in using this algorithm to provide good services for children in reality. In response, the agency developed 7ei using best practices from health services and social work (Trauma-Informed Care) [450, 292], and continues to use this algorithm in parallel with CANS to provide a more holistic perspective for the best outcome for a child and family. As outlined in our results, CANS and 7ei affect the three dimensions of ADMAPS differently because fundamentally, CANS is a deficit-based assessment that focuses on risk mitigation and resource allocation; on the other hand, 7ei is a strength-based assessment that focuses on improving the outcomes for children. This is not to imply that CANS is always employed in a deficit-based context while 7ei only focuses on building strengths. These kinds of assessments can be commonly found in public services beyond child-welfare. For instance, within criminal justice systems, deficit-based risk assessments are the norm [437] but are heavily criticized for being unfair [54]. An important takeaway is that we need both strength and deficit-based approaches to make better decisions in high-stakes environments. An initial assessment of risk is necessary but a transition towards helping people through strength-based approaches is equally important to prevent referrals and future interactions with the system [37, 35, 38, 36]. Moreover, regardless of the type of outcome, street-level bureaucrats in the public sector must still exercise discretion and contextualize the algorithmic results for each case and act within the constraints posed by bureaucratic processes as explained by the ADMAPS framework [16, 354, 232].

3.6.2 Implications for Algorithmic Decision-Making in the Public Sector

ADMAPS Emphasizes Managing Interdependencies and Trade-offs in Algorithmic Decision-Making Within the Public Sector

The purpose of ADMAPS is twofold: 1) to interrogate algorithmic interventions in the public sector in a way that ensures that they balance the dimensions of *human discretion* and *bureaucratic process* with *algorithmic decision-making*, and 2) to offer practical guidelines for developing algorithms that offer higher utility to practitioners. A key implication of using ADMAPS is that better algorithmic decisions are made when algorithms account for and balance the complex interdependencies within socio-technical systems, rather than operate in isolation. We found that when one aspect of the framework was optimized, the other aspects of the framework often suffered. For instance, CANS demonstrated an over-reliance on predictions to support mandated bureaucratic processes; therefore, it minimized human discretion to the point that negated collective buy-in and created ethical dilemmas where caseworkers felt pressured to manipulate the system. Additionally, each dimension of the framework had cross-dependencies with others, demonstrating practical trade-offs at the level of the model. For example, predictive versus prescriptive types of decision support affected administration and training in different ways. The predictive nature of CANS tended to replace the need for training for interpreting outcomes (i.e., giving the answer), while the prescriptive nature of 7ei augmented training by helping new CWS employees learn through the collaborative brainstorming process. Thus, a core contribution of this paper is that we demonstrate how the theoretically derived ADMAPS framework (empirically validated through a case study in child-welfare) can be used within other public sector domains (e.g., criminal justice [222, 221]) to assess the strengths and weaknesses of algorithms making high-stakes decisions in the lives of people. For instance, ADMAPS can help develop algorithms for judicial decision-making designed to aid judges. Judges must make decisions based on legal justification, interpretation, and application based on relevant laws and precedents [159]. The fluid nature of legal reasoning is often at odds with the discrete predictive nature of algorithms [212]. Here, similar to the 7ei algorithm, ADMAPS can help develop analogues that augment human discretion instead of curtailing it. Similarly, ADMAPS can help design algorithms for job placement centers where the caseworkers must exercise discretion in applying complex legal frameworks, assessing resource constraints, as well as resolving organizational contradictions to extend unemployment benefits to citizens [232].

Holistic Assessments, Not Deterministic Scoring Lead to Improved Decisions

Through the lens of ADMAPS, the tension between all three dimensions implies that policymakers do not get an opportunity to understand and appreciate the value of holistic algorithmic assessments that can augment the current mandates of univariate risk scores. These mandates exist based on state legislation and are meant to provide a legal justification for using algorithms in child-welfare. While the legal implications are out of scope for this paper, our main takeaway is that the current tension that exists between the state and the agency needs to be redressed in order to improve outcomes for children. This is not to say that we must mandate decision-making from both CANS and 7ei but on the other hand, reduce the dependency on singular metrics and increase the dependency on understanding the underlying trauma and behaviors of a particular child. Moreover, as clearly depicted through our results, using singular metrics from a risk assessment to make subsequent decisions about unrelated determinants (for e.g., calculating compensation for foster parents) only incentivizes gaming the system and results in a vicious cycle for children in foster care. The benefits of strength-based approaches have been supported in prior literature [38, 367, 397, 450] as well as the pitfalls for reappropriating algorithmic outcomes for different purposes [395, 303]. More importantly, algorithmic decision-making should not be treated as inevitable; knowing when not to design [48], not to deploy [44], and to resist [399] is equally important. If systems designers and policymakers are unable to balance the dimensions of ADMAPS in a way that it serves on-the-ground practitioners then alternative, non-algorithmic approaches such as collaborative assessments or processes should be developed that support and/or streamline bureaucratic processes and augment the quality of human discretionary work.

3.6.3 Design Guidelines for Algorithmic Systems in the Public Sector

Based on our findings we provide the following design guidelines for developing algorithmic decision support systems in the public sector. Our guidelines highlight the need to support the complex interdependencies between the three dimensions of ADMAPS.

- Consider making algorithmic outputs multidimensional, rather than a singular metric. This allows for flexibility in interpreting the output through the use of human discretion.
- Make algorithmic metrics suggestions, rather than mandated decision outcomes. Creating flexibility in the bureaucratic process will reduce overhead and prevent the system from getting overburdened.
- Account for the degree of uncertainty in the data and the associated outcomes and make room for value judgments and heuristic decision-making.
- Design algorithmic systems to be used collaboratively so that joint oversight and expertise can provide fairness, transparency, and accountability.
- Consult key stakeholders to form a consensus on what data should and should not be collected to inform high-stakes decisions.
- Design the system to learn and adapt from expert users by being able to identify exemplar cases and the reasons why they were successful. This may lead to serendipitous data points that were not previously captured formally by the system.
- Account for the organizational resources and constraints within which all decisions (human and algorithmic) must be made and incorporate this into algorithm design.
- Avoid direct trade-offs between input variables that create ethical dilemmas. Create safeguards in the system that check for gaming behaviors, such as "what-if" analyses.
- If tradeoffs cannot be properly managed, consider alternative, non-algorithmic approaches that streamline and/or support bureaucratic processes and augment human discretion.

3.7 Limitations and Future Work

Our study contributes both a generalizable framework of Algorithmic Decision-Making Adapted for the Public Sector (ADMAPS) and an in-depth ethnographic case study of a sociotechnical system used in the domain of child-welfare. However, there are several limitations that introduce opportunities for scholars to expand upon this research. First, this study solely focuses on the perspective of CWS caseworkers on algorithmic decision-making systems. We believe it is important to also include the perspective of CWS children and families as they are the affected communities of concern. Future research should focus on understanding the perspectives of families and their (lack of) agency with respect to the decisions made about them through the use of algorithmic systems. Second, this study only focused on the algorithms being used in collaborative team settings; however, there might be other decision tools that caseworkers might use independently in their daily work. For instance, all CWS employees are required to use a comprehensive state-mandated data system that has several data-driven visualizations and decision tools built into them. But as our results indicate, caseworkers lack an adequate understanding of algorithms or decision tools and only recognize the automated decision-making aspects when explicitly asked to think in those terms. Therefore, future research should focus on studying the state-mandated data systems, in-built decision tools, and their impact on caseworkers' decisions more holistically.

3.8 Conclusion

We conducted an in-depth ethnographic study to understand the daily algorithmic practices of caseworkers at a child-welfare agency. Concurrently, we also developed a cohesive framework of Algorithmic Decision-Making Adapted for the Public Sector (ADMAPS) by systematically reviewing and synthesizing prior literature in HCI, STS, and PA. We qualitatively coded our data from the ethnography to the dimensions of ADMAPS to reveal the complex interdependencies between human discretion, algorithmic decision-making, and bureaucratic processes. Our findings show that there is a need to invest in strength-based approaches centered in ecological frameworks. This approach not only seeks to improve the lives of people but also builds collective trust in the outcome itself, and subsequently, leads to collective buy-in at both the caseworker and bureaucratic levels. Moreover, algorithms need to be designed such that they augment human discretion by allowing practitioners to engage in value judgments and heuristic decision-making. In addition, algorithms need to be fully supported by bureaucratic processes by allocating necessary resources and accounting for organizational constraints. As a result of this study, we also propose heuristic guidelines for the design of high-stakes algorithmic decision-making tools in the public sector.

CHAPTER 4: ALGORITHMIC IMPACTS IN THE CHILD WELFARE SYSTEM: UNCERTAINTIES IN PRACTICE, ADMINISTRATION, AND STREET-LEVEL DECISION MAKING

ABSTRACT: Algorithms in public services such as child-welfare, criminal justice, and education are increasingly being used to make high-stakes decisions about human lives. Drawing upon findings from a two-year ethnography conducted at a child-welfare agency, we highlight how algorithmic systems are embedded within a complex decision-making ecosystem at critical points of the child-welfare process. Caseworkers interact with algorithms in their daily lives where they must collect information about families and feed it to algorithms to make critical decisions about child safety and well-being. Algorithmic decision-making is causing real harm to the nature of social work practice where caseworkers consider their role being diminished from street-level bureaucrats who focused on understanding the individual needs of clients and helping them into mere data brokers. We also highlight how data-driven policy, as currently proposed, is at odds with how street-level bureaucrats interpret and implement policies on the ground. Our work also problematizes the popular narrative on transparency and explainability of algorithms by illustrating that different stakeholders held different notions of these human factors based on their end goals. Finally, we show how a simple decision-tool designed to support decision-making processes offered higher utility and gained support from caseworkers.

4.1 Introduction

Decades of neoliberal politics in the United States (U.S.) centered in austerity and privatization have led to public sector agencies increasingly looking towards digitization and automation both as a means to reduce costs as well as provide greater efficiencies in public service delivery [167, 465, 163]. Principles of New Public Management (NPM) have been re-positioned in the public sector to emulate the business model of private corporations centered in efficiency, cost reduction, and innovation [290]. Theories of NPM coupled with the introduction of digital technologies seek to improve data sharing practices between different sectors of the government, promote minimal repeated information gathering, provide targeted services, reduce bureaucratic overhead, and improve decision-making [463]. Consequently, most decisions about citizens are now being made from behind the screens instead of interacting with them at the street-level [62]. This transformation from 'street-level' bureaucracy to 'screen-level' bureaucracy has been characterized as *Digital Era Governance* by public administration scholars where digital technologies played a central role [76]. However, despite promises of significant improvements in efficiencies and cost-effectiveness, several digital tools have fused onto existing street-level discretionary practices without altering public management at a deeper organizational behavior level [463, 315].

However, digitization has allowed public services to collect a vast amount of data about citizens in the course of their daily operations in managing and delivering public services [326]. This comprehensive cross-sector data has allowed policymakers, tech companies, and academics to narrow their focus on improving high-stakes decision-making by developing data-driven practices that purportedly provide consistent, transparent, objective, and defensible decisions to citizens [397, 163, 75]. Algorithmic decision-making in the public sector has generally been adopted in the form of risk assessment algorithms with their primary purpose being the preemptive recognition and mitigation of 'risk'; a core organizing concept of this shift in governance and of neoliberal economics [22]. That is, improving productivity, accountability, and efficiency by proactively identifying clients in the riskiest circumstances and targeting services towards them using algorithms [377, 163]. This further shift in *Digital Era Governance* that embeds neoliberal politics into the principles of New Public Management has been called *New Public Analytics* [489] and focuses more on risk management based on individual client characteristics while driving attention away from structural and societal problems [377, 258].

Over the past two decades, several high-stakes decision-making domains such as the child-welfare system (CWS), the criminal justice system, education, and medical services have increasingly turned towards risk assessment algorithms as a means to standardize and improve decision-making [214, 232, 397, 163]. Facing severely limited resources, burdensome workloads, and high staff turnover, most public sector agencies have also turned towards algorithms as they promise to allocate resources more efficiently and fairly [163, 299, 400]. The Child-Welfare System has also been the center of public and media scrutiny because of harm caused to children who are removed from the care of their parents [86]. On the other hand, CWS also receives severe criticism and media attention for child abuse tragedies where the system failed to remove and protect a child [220, 182]. This has further mounted pressure on CWS in several states in the United States (U.S.) to employ structured decision-making tools (and more recently, algorithmic decision-making) to prove that they are employing evidence-based, consistent, and objective decision-making processes.

Consequently, critics of algorithmic systems in the public sector have continued to focus on the outcomes predicted by these systems and their disparate impact on affected communities [432, 72, 377]. In this study, we deepen this discussion about algorithmic impacts by examining the impact of algorithmic tools on professional practices, bureaucracy, and street-level decision-making in child-welfare. Digital Era Governance witnessed digital technologies assimilating into existing work practices without altering or improving them at a deeper level [463]. A similar exploration needs to take place to assess the impact of algorithmic decision-making on the nature of practice (i.e., street-level human discretionary work) and organizations (i.e., bureaucratic processes) as well as assess how street-level decision-making is changing and whether algorithms are living up to the promises of cost-effective, consistent, and fair decision-making.

4.2 Related Work

The breadth of research at the intersection of technology and governance has been wide-ranging, including studies that examine how emergent technologies are shaping collaborative work [58, 325], designing technologies that empower affected communities [72, 146], studying issues concerning civic engagement [150, 206], and social and ethical implications of datafied public services [377, 232, 374]. For instance, researchers have unpacked the forms, limits, and complexities of participatory design within the public sector where newer technologies are being designed for the governance of smart cities [478, 416, 227]. Several of these technologies are being developed through public-private partnerships [150, 299] where the expectation is that these entities are able to transform data into knowledge and inform decisions that are centered in efficient allocation of resources [232]. As Holten Møller et al. [232] note, *"here, data becomes the promise of future bureaucratic efficiencies"*.

Specific to affairs of public administration, researchers have studied how the shifting decision-making latitude has impacted the work of street-level bureaucrats. [4] Studies have found that value conflicts arise when the logics embedded within the government's digital platforms do not align with street-level bureaucrats' discretion when they tried enacting the same shared values in practice [256, 466, 146, 232]. Research within public administration, science and technology studies, and human-computer interaction have recently drawn attention to how the digitization of public services is leading to distinct changes in street-level bureaucrats' discretion [80], power asymmetries between public officials and citizens [432, 405, 404], need for re-skilling of public officials [256, 232], and actor transparency in government decision-making [198]. Lindgren et al. [296] go a step further and argue that "public officials can no longer be understood as merely human". They call for a reinterpretation of citizens' trust in their government in regard to the legitimacy and accountability of e-governments. Busch et al. [81] explore public service workers' digitized discretionary practices as they balanced conflicting demands of market-oriented goals and norms of professional practice and found that workers responded positively to digitization when it supported professional aspects of their work. In a similar vein of work, Giest et al. [198] highlight that a disconnect between bureaucratic processes and digital tools

[4] A street-level bureaucrat is a professional service worker (e.g., social worker, police officer, teacher) who operates in the frontline of public service provision. They interact closely with clients and make decisions about them based on how they interpret policies relating to the situations at hand [297].

magnified in street-level decision-making where workers' discretionary power was obfuscated and led to more complications and time-consumption in accomplishing daily tasks. In the context of smart cities, Meijer [323] further argues that building technologies for governance is centered in political, strategic, and value-laden choices made between three key actors: state, market, and civil society. This requires a re-conceptualization of sociotechnical structures that result from their interactions. Public administration scholars have also foreboded a "digital sclerosis" characterized by the stiffening of governmental processes and lowering of innovation feedback from workers [21]. They predict "decreased bargaining and discretionary power of governmental workers" as one of the early warning signs of this phenomenon. These changes in professional work practices through the adoption of digital tools have similar, yet more serious implications for the adoption of algorithmic tools which further shift discretion away from public service workers.

As a result, researchers have started investigating the intersection of human discretionary work conducted at the street-level and algorithmic decision-making in public services [16, 354, 232, 385]. Alkhatib and Bernstein introduced the theory of *street-level algorithms* to distinctly highlight the gaps in algorithmic decision-making that human discretion needed to address [16]. Unlike street-level bureaucrats who used discretion to reflexively make decisions about novel cases, street-level algorithms produced illogical decisions that offered no recourse and could only be addressed by 'learning' through new data in the future. Pääkkönen et al. expand upon this theory to highlight that algorithm design needed to identify and cultivate important sources of uncertainty because it was at these locations that human discretion was most needed [354]. Additionally, recent work [20, 396] has highlighted the collaborative nature of caseworkers' decision-making processes and the impact of bureaucratic structures that algorithm design needed to account for. In sum, researchers have reached a general consensus that any algorithmic interventions in the public sector needed to understand the complexities of human discretion carried out at the street-level when implementing day-to-day bureaucratic processes and legislative policies. Recently, Saxena et al. [396] synthesized prior work conducted on algorithmic governance systems into a framework for algorithmic decision-making for the public sector which accounts for the complex interdependencies between human discretion, algorithmic decision-making, and bureaucratic processes. Next, we offer a brief overview of this framework and explicate its utility in unpacking the impact of algorithmic decision-making on the nature of the practice, organization, and street-level decision-making.

4.2.1 Algorithmic Decision-Making Adapted for the Public Sector (ADMAPS) Framework

Saxena et al. (2021) [396] developed a theoretical framework for algorithmic decision-making in the public sector that accounts for the complex inter-dependencies between *human discretion*, *bureaucratic processes*, and *algorithmic decision-making*. Below, we provide a brief summary and explain the three high-level elements of the framework and their utility for our analysis -

Human Discretion refers to the decision-making process model that practitioners in the public sector engage in where they use their professional expertise, make value judgments, and engage in heuristic decision-making based on the available facts. This is especially important in the public sector because government officials must make decisions within the bounds of policies and organizational constraints. Moreover, they must use discretion in resolving missing or conflicting sources of information.

Bureaucratic Processes refer to critical governance characteristics that include the systemic constraints within which all decisions must be made, day-to-day protocols, and the legislation that the organization is legally mandated to follow. Bureaucratic processes directly impact practitioners' training and nature of the practice (i.e., human discretion) at the agency as well as how well an algorithm is integrated into the day-to-day workflows and decision-making processes.

Algorithmic Decision-Making adopts *street-level algorithms* [16] as a theoretical lens to identify algorithmic systems that are used to make on-the-ground decisions about clients and welfare in the public sector. It draws attention to the relevant data needed for decision-making, the degree of uncertainty associated with predicted decisions, as well as the decision-making latitude (i.e., predictive or prescriptive) that is allocated to algorithmic systems. Authors refer to *algorithmic decision-making* as the most flexible element of the framework that researchers can directly affect by designing systems that balance the other two elements (i.e., human discretion and bureaucratic processes). In addition, the authors further highlight the need to understand human discretionary work because algorithm design needs to make space for (and be preceded by) human discretion to fill in the gaps in data as well as make sense of organizational and legislative protocols.

In sum, the ADMAPS framework showcases how practical trade-offs must be made to manage the cross-dependencies at both the macro- and micro-levels of the algorithmic model to offer autonomy to practitioners and improve human discretionary work. The framework also draws attention to the high degree of uncertainty inherent in the administrative data which consequently means unreliable predictions. Therefore, the goal of algorithms in the public sector must be re-evaluated to support the decision-making processes of stakeholders instead of providing predicted outcomes. In their study, Saxena et al. [396] focused on the micro-interactions between the dimensions of these three elements to understand why each algorithm failed (or succeeded) to offer utility to child-welfare staff and their impact on human discretionary work. However, in this study, instead of focusing on singular algorithmic tools and their impact on human discretion, we draw attention to the broader decision-making ecosystem and critically investigate the macro-interactions between these three elements to assess the impact of algorithmic systems on the nature of the practice (i.e., the interaction between human discretion and algorithmic decision-making), the organization (i.e., the interaction between bureaucratic processes and algorithmic decision-making), as well as three-way interactions between the three elements to understand how the nature of street-level decision-making is changing in child-welfare.

4.2.2 Current Landscape of Algorithms Used in the U.S. Child Welfare System

Child-welfare (CW) agencies in the United States have increasingly adopted algorithms for high-stakes decisions as they promise to improve decision-making, lower costs, and provide better outcomes to citizens [382, 163]. A nationwide survey on predictive analytics in child-welfare conducted by the American Civil Liberties Union (ACLU) in 2021 revealed that 26 states have considered employing predictive analytics in child-welfare [391]. Of these 26 states, 11 are currently using them [391]. Several states in the United States continue to experiment with predictive analytics, however, audits of these systems reveal that they are achieving worse outcomes for families and exacerbating racial biases [171, 454, 344, 230]. In the past, Los Angeles County and the state of Illinois have shut down their predictive analytics programs for these reasons [171, 454] with Oregon recently joining their ranks in June 2022 [344]. A recent study conducted by Cheng and Stapleton et al. [99] on the Allegheny Family Screening Tool (AFST) found that AFST-predicted decisions were racially biased and workers reduced these biases by overriding erroneous decisions. AFST algorithm was designed to mitigate call screeners' biases and subjective decisions and augment decision-making by making it more objective through data. Ironically, AFST has introduced more complexities in decision-making and the call screeners are the ones mitigating algorithmic biases. A comprehensive literature review of algorithms in CWS revealed other sources of biases embedded in the predictors, outcomes, and computational methods being used to develop these systems [397].

Federal initiatives such as improved data infrastructures for CWS [225] have paved the way for tech startups to develop and pitch algorithmic systems to human services agencies across different states [239, 371, 447, 238]. However, there is a need to critically examine the current points of failures in algorithm design as well as understand how workers engage with algorithms as they make critical high-stakes decisions about families. Critical to the conversation about predictive analytics or predictive risk models (PRMs) is also the underlying principle of "risk" and how its understanding has shifted in response to the restructuring of public services to be economically efficient, productive, and accountable [83, 377, 22]. Traditionally, child-welfare services have focused on risks and protective factors within families to be able to provide them with individualized care. However, with a shift towards an economic understanding of risk and the introduction of PRMs, risk has now become a function of client characteristics as existing in prior cases (and not individual family circumstances) and their impact on a predictive outcome (i.e., risk of maltreatment). That is, the risk is estimated based on historical administrative data and is being used to identify the "deserving poor" who pose the most risk to governmental apparatus [163]. Redden et al. [377] refer to this as the embedded logic of actuarialism that also obfuscates and drives attention away from social and structural issues that bring poor and vulnerable communities under the attention of public services such as child-welfare, housing authority, and public assistance [258].

4.3 Methods

We partnered with a child welfare agency that serves a metropolitan area in the midwestern United States. This private, non-profit agency is contracted by the state's Department of Children and Families (DCF) to provide child welfare and family services and must comply with all DCF standards, including the use of mandated decision-making algorithms. DCF's Initial Assessment (IA) workers investigate allegations of child maltreatment, and if abuse/neglect is substantiated, the child(ren) may be removed from the care of their parents or an in-home safety plan might be developed. At this point, the case is referred to the agency to provide services. These services are negotiated between the parents' attorney, the district attorney's office, and the judge after caseworkers have conducted initial structured assessments and provided their recommendations to the court. A case manager is assigned to each case and is supported by a multi-disciplinary child-welfare team that brings in domain expertise from social work, family psychology, medicine, and law. The agency is mandated to use some algorithms at different stages of a child-welfare case per DCF standards but has also developed an algorithm in-house to improve its decision-making processes.

4.3.1 Study Overview

We conducted an extensive ethnography to understand how caseworkers interacted with algorithmic systems, and their perspectives on these systems, as well as unpack how decisions were made at the intersection of child-welfare practice, regulations, and algorithmic decision-making. This study is conducted within the same broader child-welfare system that was also the ethnographic site for Saxena et al. [396]. However in this study, instead of focusing on singular algorithmic tools and assessing their impact on human discretion, we draw attention to the broader decision-making ecosystem and assess the impact of algorithms on the nature of the practice, administration at the organization, and the changing nature of street-level decision-making. Before conducting observations or recruiting participants for interviews, we obtained Institutional Review Board (IRB) approval at our research institution to conduct our study. Before the interviews, we emailed the participants an IRB-approved consent form and obtained their verbal consent to participate in the study before beginning the interviews. The first author observed child-welfare team meetings and then conducted semi-structured interviews with key stakeholders at the agency to better contextualize interactions that occurred at these meetings and gather caseworkers' perspectives. Below, we first highlight our scoping criteria for algorithms followed by a description of observations and interviews, and the qualitative data analysis process.

4.3.2 What is an "Algorithm"?

Different definitions of algorithms have been adopted across different disciplines due to the intersection between statistical modeling and machine learning as well as ongoing innovations in advanced neural networks [229, 456, 414]. From a social perspective, we define algorithms through the lens of *street-level algorithms* [16] - computational predictive tools used to make on-the-ground decisions about human lives and welfare. That is, algorithms that directly affect families utilizing public services and not second or third-removed algorithms that might be internally used by government analytics teams. Technically, we define an algorithm as a computational system that takes in input data, computationally processes data, and produces an output. This output can be a predicted outcome of interest or change in output over time when some statistical property of variables change as is the case with change point detection algorithms [19].

4.3.3 45-Day Staff Meetings, Permanency Consultations, and Concurrent Planning

45-day staff meetings occur within the first 45 days of a case coming into the care of the agency and are attended by child-welfare staff involved at the front end of case management and planning. These meetings facilitate information sharing such that consensus decisions can be made in regard to child well-being. Each meeting is scheduled for 90 minutes and the first author observed 25 such meetings. These meetings are typically attended by child-welfare staff that works in case management, permanency planning, family preservation, and licensing. The goal of these meetings is to develop a deeper understanding of a family's circumstances from a trauma-informed perspective, identify parents' support system that can help with child care, identify systemic barriers that may inhibit reunification, and develop action steps for child-welfare staff. These meetings facilitate collaborative decision-making and ensure congruence in case planning.

Permanency consultation meetings are specialized meetings designed to expedite permanency for children placed in out-of-home care by employing collaborative practices as well as actively addressing any systemic or policy-related barriers. These meetings are facilitated by permanency consultants and are staffed with many of the child-welfare team members that attend 45-day planning meetings. These meetings regularly occur at the 5, 10, and 15+ month marks for every case until the case is closed. These ongoing meetings tended to be more informative than the 45-day planning meetings because not enough information is available at the onset of a case. Moreover, permanency consultations involved cases that had been with the agency for several months (if not years) and revealed the complicated interactions between policies, systemic barriers (legal, resource, administrative), social work practice, and algorithmic systems being used at the agency. The first author attended 55 permanency consultations.

Critical to a discussion about collaborative meetings and the decision-making ecosystem is the *Adoption and Safe Families Act (1997)* which introduced some of the most sweeping changes to the child-welfare system and shifted the focus primarily towards child safety concerns and away from the policy of reuniting children with parents regardless of prior neglect/abuse. The Act introduced federal funding to assist states with foster care, adoption, and guardianship assistance and expanded family preservation services. In addition, it also introduced a **15-month timeline** where the State must proceed with the termination of parental rights if the child has been in foster care for 15 out of the last 22 months [209]. This speedy termination of parental rights has received widespread criticism but still establishes the constraints within which caseworkers must conduct their work [216, 438, 368]. To ensure expedited permanency[5] for foster children, the agency employs concurrent planning such that two simultaneous plans begin when a child enters foster care: a plan for reunification with the birth parents and a plan for guardianship and/or adoption if reunification is not possible.

4.3.4 Semi-Structured Interviews

Next, we used the knowledge gathered from these observations to develop our interview protocol and recruit participants who consistently attended these meetings. We conducted these interviews to delve deeper into caseworkers' understanding of these algorithms as well as the benefits and challenges as perceived by them. We asked participants a series of questions about the nature of child-welfare work, how algorithms in use impacted their practice, and the organizational support made available to them to mitigate conflicts or challenges associated with these systems. We also asked them to expand upon any interactions we had observed during the meetings such as the participant's dislike or appreciation for a certain algorithm or feature or their frustration with the misuse of algorithmic tools. During observations, we also noted any benefits or frustrations that the participant (or their team) experienced in regard to algorithmic decision tools and brought them up during the interview to further expand upon and better

[5]Permanency is defined as reunification with birth parents, adoption or legal guardianship.

understand their outlook. We conducted 20 interviews with participants that included program directors, child-welfare supervisors, permanency consultants, caseworkers, data specialists, and clinical therapists.

4.3.5 Study Participants

We interviewed child-welfare workers who attended the collaborative meetings mentioned above and had more experience working with algorithmic tools and assessments in their day-to-day work. The case manager position experiences high turnover with most case managers quitting within the first two years. Therefore, we focused our attention on the more experienced members of the staff who could provide deeper insights into the workings of the system as well as systemic barriers that impede their work. Below, we provide job descriptions of various positions and participant information in Table 1.

Program Directors (n=1): Agency leadership responsible for professional development programs, trainings, research and policy initiatives, and grant writing. They supervise child-welfare supervisors and the agency's trauma-responsive service model.

Child-Welfare Supervisors (n=8): Supervisors manage a case management team comprising of 6-8 case managers and oversee about 140 cases. They are responsible for the professional development of case managers and provide additional support when interacting with families and legal parties. They also facilitate 45-day staffing meetings.

Child-Welfare Case Managers (n=2): Frontline workers that directly interact with families and act as mediators between parents, foster parents, relatives, legal parties, and child-welfare staff. They conduct home visits, safety assessments, and psychometric assessments, and transport foster children to supervised visits and medical appointments. On average, they manage about 20 cases.

Permanency Consultation Supervisor (n=1): The supervisor in charge of the permanency consultation program is designed to address systemic barriers that pose obstacles in the way of achieving permanency for foster children. They manage and supervise permanency consultants.

Permanency Consultants (n=5): They are responsible for managing the legal process underscoring permanency where they prepare documentation for court, focus on recruitment and licensing of foster homes, and manage post-guardianship and post-adoption services, among other tasks that help expedite permanency for foster children. They provide consultations on about 150 cases each.

Data Specialists (n=2): Responsible for tracking federally-mandated performance benchmarks for the agency as well as case-level data. They also analyze data and manage algorithmic systems implemented at the agency and present results to agency leadership.

Clinical Therapist (n=1): A licensed clinical social worker who conducts mental health as-

Participant	Sex	Job Title	Experience (years)
P1	F	Permanency Consultation Supervisor	22
P2	F	Permanency Consultant	20
P3	F	Permanency Consultant	9
P4	F	Permanency Consultant	8
P5	F	Permanency Consultant	12
P6	F	Permanency Consultant	3
P7	F	Child Welfare Program Director	20
P8	F	Child Welfare Supervisor	19
P9	F	Child Welfare Supervisor	13
P10	F	Child Welfare Supervisor	9
P11	F	Child Welfare Supervisor	9
P12	F	Child Welfare Supervisor	7
P13	M	Child Welfare Supervisor	12
P14	M	Child Welfare Supervisor	30
P15	F	Child Welfare Supervisor	9
P16	F	Clinical Therapist	5
P17	F	Child Welfare Case Manager	8
P18	M	Child Welfare Case Manager	2
P19	F	Data Specialist (Program Director)	17
P20	M	Data Specialist	17

Table 7: Interview Study Participants

sessments and has a deeper understanding of psychometric assessments used in child-welfare.

4.3.6 Qualitative Data Analysis

The first author took detailed observational notes during each team meeting and compiled a debriefing document with their initial insights. They also noted questions that needed further clarification and conferred with team leaders (i.e., supervisors or permanency consultants) after the meetings and noted their responses. All authors read through the observational notes and collectively discussed meeting notes that uncovered pertinent aspects of the decision-making ecosystem in regard to the nature of the practice, policies and regulations, and/or the use of algorithmic tools. Interviews were audio-recorded and transcribed verbatim for analysis. After carefully reading through the transcripts, we conducted several rounds of iterative coding to identify patterns and converge on appropriate themes associated with the three elements of the ADMAPS framework (i.e., human discretion, bureaucratic processes, and algorithmic decision-making). We performed thematic analysis [112] to create these initial codes, formed a consensus around the codes, as well as resolved any ambiguous codes. In our results, we also use our observational notes to augment the insights we gained from the interviews and note potential discrepancies and nuances from the holistic insights gained from our site observations. These initial codes were grouped into three high-level themes of *impact on practice*, *impact on organization*, and *impact on street-level decision-making*. Codes that highlighted pertinent issues regarding how human discretionary work has changed under algorithmic systems were grouped under *impact on practice*. For instance, 80% of the participants shared that the CANS algorithm had negatively impacted social work practice due to its lack of understanding of trauma. Similarly, codes that highlighted issues such as an algorithmic system's inability to account for organizational constraints or legislation were grouped under *impact on organization*. For instance, 75% of the participants shared that all decision-making processes were first situated in the 7ei algorithm which is central to the agency's trauma-responsive service model. Finally, codes that highlighted how street-level decision-making is changing on the ground where bureaucrats are mandated to use algorithmic decisions (but do so while employing their professional expertise and operating under organizational and legislative constraints) were grouped under *impact on street-level decision-making*. For instance, 80% of the participants shared that they were mandated to use the foster care placement as recommended by CANS, however, foster children were more likely to achieve stability and well-being when placed with relatives rather than a more restrictive foster care setting as recommended by CANS. In addition, CANS did not account for the limited number of available foster homes; an organizational constraint that rendered the algorithmic recommendation nonsensical.

4.4 Results

In the following subsections, we first share some overarching findings regarding the decision-making ecosystem in child welfare that sits at the intersection of policies, social work practice, and algorithmic systems. Next, we discuss findings from two algorithms that are in use at the child-welfare agency and how their adoption has impacted the nature of the practice, administration at the organization, and street-level decision-making.

4.4.1 Decision-making Ecosystem in Wisconsin's Child-Welfare System

In order to understand the role of algorithms in decision-making, it is necessary to map out the complexities within the broader decision-making ecosystem to be able to assess the utility and scope of algorithmic tools. In addition, it is imperative to understand the systemic constraints within which such systems must operate. This helps us to better contextualize how caseworkers interact with algorithms, their perspectives on these systems, as well as how decisions are made at the intersection of policies, social work practice, and algorithms, i.e. - how algorithmic decisions were used by child-welfare staff working under legislative and organizational pressures. Figure 1 describes **"Life of a Case"** in child-welfare and highlights all the critical decision-making steps and algorithms (red boxes) that are embedded throughout the child-welfare process. We co-designed this diagram with caseworkers. The yellow section of the diagram represents the initial allegation of maltreatment and investigation that is conducted by Initial Assessment (IA) caseworkers at the Department of Children and Families (DCF). If maltreatment is substantiated, the case is officially opened and is referred to the child-welfare agency to provide ongoing services to the family. This is represented by the orange section of the figure. Finally, the blue section of the figure represents the court system. Contrary to popular belief, critical decision-making

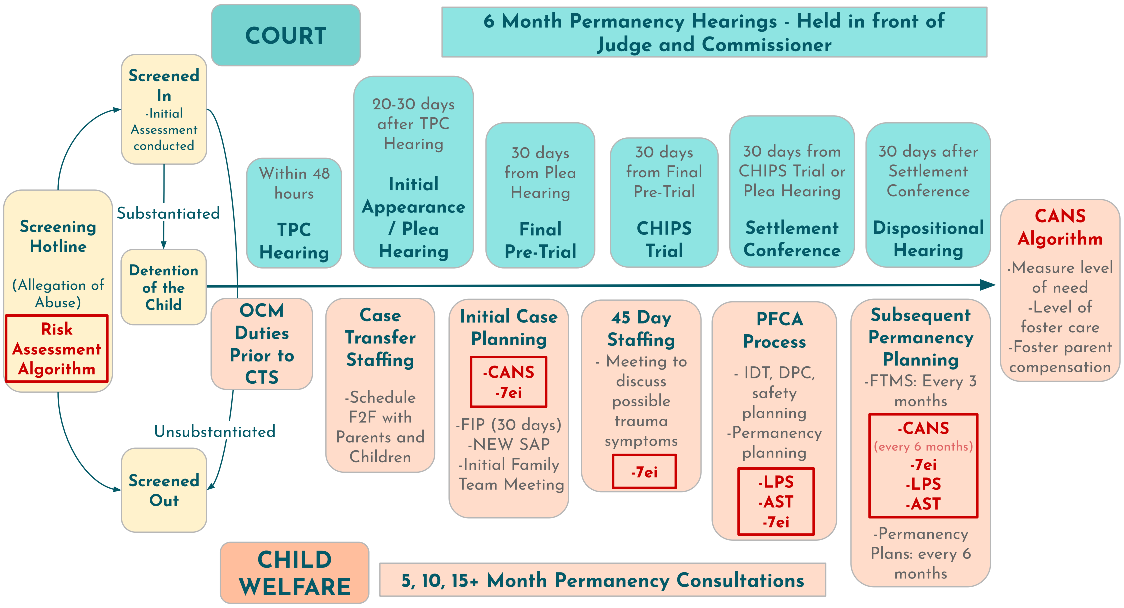


Figure 8: Life of a Child-Welfare Case in [State]

power in regard to permanency (i.e., reunification, adoption, termination of parental rights) for foster children sits with the court system and child-welfare staff only makes recommendations to the district attorney's office. Tensions between child-welfare staff and the court system are well captured in social work literature [158, 154, 90]. This ethnography is conducted within the orange section of the figure (i.e., the child-welfare agency) where algorithmic tools are being used by caseworkers to make several day-to-day street-level decisions (as opposed to critical permanency decisions made in courts) about children and families. Underscoring the use of algorithmic decision-making in child-welfare are two dominant concerns that frequently arose in this ethnography -

- Caseworkers are not trained in "thinking statistically" about data, algorithms, and uncertainties but are legally mandated to input data, interact with algorithms, and make critical decisions.
- All algorithmic decisions in the public sector must be made within the bounds of policies, current practice, and organizational constraints.

As depicted in Figure 1, when an allegation of abuse is made at the hotline, a risk assessment algorithm helps call screeners to decide if the call should be screened for an investigation. We did not have access to this decision step since it is conducted at DCF before the case is referred to the agency where the ethnography was conducted. If the call is screened in and the investigation substantiates abuse and/or neglect, the child(ren) is removed from the care of their parent(s), or an in-home safety plan is put in place. The case is then referred to this child-welfare agency.

Throughout the life of the case, the CANS algorithm is used to assess the mental health needs of a child, the level of foster care (level 1 to 5) the child should be placed in, as well as calculate the compensation that foster parents should be paid by the state. 7ei (Seven essential ingredients) algorithm is used to develop trauma-responsive services. LPS (Legal Permanency Status) algorithm is used to track federal benchmarks such as placement stability, as well as policy and systemic barriers. And finally, AST (Anti Sex-Trafficking Response) algorithm is used to assess the risk of sex trafficking for a foster child. These algorithms are being used to make everyday decisions about foster kids but there are also other algorithms that are being used internally by the Department of Children and Families. For instance, an algorithm is used to assess the risk of re-entry into the system for every case. For the purpose of this study, we only focus on CANS and 7ei algorithms.

4.4.2 Caseworkers' Perspectives: Assessments or Algorithms?

Through the observations of child-welfare meetings, we learned about several different sources of data that are collected by the agency as a means to collect consistent information about all cases. The agency uses several psychometric assessments for this purpose (see for e.g., [304, 379, 305]) as they provide a structured framework to conduct risk assessments. Child welfare initially adopted psychometric risk assessment instruments (RAI) as a means to standardize the process of assessing children's and parents' risks and needs and allow for a more consistent decision-making process. Consequently, RAIs have facilitated the collection of data about children and families for the past three decades. Over the last decade, psychometric and administrative data from prior substantiated cases of maltreatment is now being used to train algorithms to make predictions about current cases [397]. The algorithms depicted in Figure 1 have all been developed using their RAI counterparts. RAIs have been widely adopted in social work practice, and as a result, all participants recognized these assessments but did not actively recognize their algorithmic components (i.e., automated aspects) that routinely frustrated them. However, once we focused more on these frustrations and expectation violations, we uncovered several pertinent issues at the intersection of human discretion, bureaucracy, and algorithmic decision-making. In the following sections, we discuss the CANS and the 7ei algorithm and their impact on the nature of social work practice, impact on the organization, and street-level decision-making.

4.4.3 CANS Algorithm

CANS (Child and Adolescent Needs and Strengths) algorithm is constructed using the CANS communimetric assessment that consists of 104 psychometric items organized across eight domains that address child needs and strengths (see [105] for more details). It was designed to assess the *level of need* of a foster child and utilize this assessment to develop an individualized care plan. CANS offers the child welfare team a structured way to assess a case as well as share this information with other parties. With its primary purpose being communication, CANS is designed based on communication theory rather than psychometric theories centered in measurement development [2]. That is, the tool was not designed to explicitly measure any variables or predict outcomes based on these variables. CANS was designed to be the expected outcome such that it facilitated the linkage between the assessment process and the design of individualized service plans. That is, CANS was designed to support care planning, facilitate quality improvement initiatives, and monitor the outcomes of new evidence-based practices.

CANS algorithm, on the other hand, uses the risks and needs data about children collected by the assessment and has been re-purposed to explicitly measure three outcomes - **1)** mental health services to be offered to foster children, **2)** level of foster care they should be placed in, and **3)** generate subsidized guardianship rate offered to foster parents. Based on a child's risks and needs, the algorithm recommends mental health services that must be offered. Next, based on this level of need, it predicts the level of foster care the child should be placed in. Foster homes in this state range from Level 1 to Level 5. Higher-level foster parents are more trained and certified to take care of higher-needs children. Therefore, the higher the level of need, the higher the level of foster care that the child should be placed in. Finally, based on the level of need and foster home setting, CANS generates the subsidized guardianship rate that foster parents are paid by the state for the costs associated with having a child placed in their care. CANS is conducted within the first 30 days of a child entering the child-welfare system or moving to a new placement. Subsequently, it is then conducted every six months. CANS algorithmic assessment is completed by caseworkers with information (about child behaviors and needs) provided by foster parents.

This re-appropriation of CANS (and therefore, CANS data) to predict these outcomes has led to several unintended consequences that frustrate caseworkers and impede theory-driven practice centered in trauma-informed care. All caseworkers at the agency are trained and certified in conducting CANS assessment, however, they are not trained in managing conflicts that arise due to the re-appropriation of CANS. This further obfuscates the boundary between the role of the **assessment** versus the **algorithm**. Below, we discuss some unintended consequences on the nature of the practice, processes at the organization, and street-level decision-making.

Impact on Nature of Social Work Practice.

CANS contradicts trauma-informed care: Caseworkers at the agency are trained in trauma-informed care (TIC) and apply principles and practices from TIC while employing the agency's trauma-responsive service model to cases. However, the CANS algorithm only offers a live snapshot of the child's mental health based on exhibited behaviors (over the past 30 days) and does not account for underlying trauma (or traumatic triggers) that can cause serious emotional dysregulation from time to time. Here, caseworkers co-opt CANS to account for trauma-informed care and anticipate the child's needs. One permanency consultant explained -

> "We have a child that goes into manic depression every year around the holiday season. At this point, one of the foster parents has to quit their job and care for the child full-time. We know that this traumatic trigger is coming up but there is no way to account for that in CANS. So we edit scores in anticipation for these upcoming needs" – P3, Permanency Consultant

CANS Misses Important Context about each Child: Algorithms attempt to generalize child and family characteristics and place them in certain categories to be able to make a determination. However, this inadvertently leads to a loss of information or context since all the information cannot be accounted for. CANS focuses on child behaviors, risks, and needs but it does not account for the child's interactions and relationships with other people. Caseworkers shared that CANS conducted the child's assessment in an isolated manner and did not account for the quality and impact of relationships in their lives which are often more important in determining their long-term well-being. For instance, a permanency consultant shared -

> "A child's behaviors are often a result of what is happening in their environment. What or who is really triggering them and where is that coming from.. there is no way to put that in CANS" – P4, Permanency Consultant

CANS offers no context regarding worsening behaviors. A clinical therapist shared that it was more important for them to be able to understand the context around worsening behaviors. They shared that worsening behaviors were not always a bad thing since children's behaviors initially get worse (before getting better) when they started therapy. This is simply because children are finally starting to address the underlying trauma in their lives in therapy which might lead to worsened behaviors in the short term but is necessary for their mental health and well-being in the long run. The clinical therapist also shared the following concern about the short-term focus on mental health -

> "If CANS is used as a communication tool, that is, just as a structured way to talk about children and families and then to make decisions on it, then that's fine! But that's not how it's being used.. A lot of it is just entering [data] into a computer and seeing whether or not there's improvement from one point to another. And if there's no improvement then making a decision." – P16, Clinical Therapist

These concerns regarding the re-appropriation of CANS were inconsistently shared by child-welfare workers. That is, not all caseworkers were equally aware of other concerns and only shared issues that affected their day-to-day practice. For instance, caseworkers were unaware of the clinical therapist's concerns regarding *worsening behaviors.* A supervisor shared that children can have severe underlying trauma that is not captured by CANS because of its focus on exhibited behaviors. At this point, we shared the clinical therapist's concerns with the supervisor and they further elaborated by saying that underlying trauma did not always manifest in terms of exhibit behaviors and that CANS may be measuring wrong indicators altogether. This also raises questions regarding which explanations associated with the tool are deemed more useful - caseworkers were more invested in explanations of underlying trauma (based on their training in trauma-informed care), whereas the clinical therapist wanted better explanations about the context surrounding worsening behaviors. That is, different stakeholders have different needs in regard to explainability and a "one-size-fits-all" approach with respect to algorithm design may not be feasible here.

In sum, such interactions with the algorithm over time led to cumulative distrust in algorithmic decision-making. The implementation of CANS also further obfuscated the difference between the algorithm and the assessment. For instance, caseworkers generally score CANS on the paper-based assessment when speaking with foster parents. However, at the agency, they input this data into the CANS algorithm to predict outcomes. Re-appropriation of CANS is also a case in point regarding why it is imperative to understand the context within which (and how) data is collected about clients in the public sector.

Impact on the Child-Welfare Agency.

Below we discuss the impact of the CANS algorithm and its re-appropriation on the day-to-day bureaucratic processes carried out at the agency.

Subsidized guardianship rate has become the primary outcome of interest. Both caseworkers and foster parents are actively aware that the subsidized guardianship rate is directly tied to mental health needs. This leads to the gamification of the algorithm for three main reasons -

- Maintaining a stable placement - There is a lack of good foster homes in the system and caseworkers try to support foster parents by any means necessary to ensure that the placement is not disrupted and the child does not need to be placed elsewhere. Multiple placement moves for foster children are associated with poor well-being outcomes where they are unable to develop any meaningful relationships with foster parents or other caregivers [56]. Therefore, gaming the algorithm is the only way in which placements can be continually supported.

- Base rate is low - Caseworkers shared that the subsidized rate for most foster placements was too low and there was no financial incentive for foster parents to be doing this work.

Here, gaming the algorithm is a convenient way to generate a higher compensation by exaggerating the mental health needs on the assessment.

- Maximizing financial incentive - Caseworkers also shared (see quote below) that even though most foster parents did their best to care for foster children, there were still other foster parents who accepted multiple placements in their home and ran foster care like a business. They routinely exaggerated child behaviors to receive higher compensations and renegotiated to what they considered to be their "standard" rate. One supervisor explained -

> "Case managers are being pressured into scoring children higher. Foster parents will match [the rate] with the previous kids. They think that their previous kid got 1500 dollars, so now that is their standard rate, and will demand 1500 dollars for the next kid. They don't have an understanding of CANS or the child's strengths or needs. Money is the key part of these decisions." –P13, Child Welfare Supervisor

Impact on Street-level Decision-Making.

Below we discuss the impact of the CANS algorithm on street-level decisions that caseworkers make in their work relationships with foster parents.

CANS punishes good foster parents. Even if there were safeguards in place that prevented the gamification of CANS, the core incentive structure of the algorithm is problematic. Foster parents offer stability and support so that kids can develop good coping skills. They also help address mental health needs and help kids stabilize by taking them to therapy and all their activities. However, CANS scores are recalculated every six months and as the child supposedly exhibits less overt behaviors, their needs are lowered (per CANS assessment) and so is the compensation offered to foster parents. One supervisor explained (see quote below) that by lowering the rate, foster parents were being punished for being actively involved and caring for foster children. CANS algorithm is set up to incentivize worsening behaviors.

> "There are an enormous amount of foster parents that do an exceptional job. And so, it is hard, how do you reward them? Because in my opinion it needs to be a rewarding system.. like you're putting in the energy.. you're doing what you're supposed to do.. There's foster parents getting their kid to all these activities and making sure this kid has a "normal social life", like normalcy, just part of parenting law. And they are getting half the rate compared to foster parents who are putting in no energy." – P9, Child Welfare Supervisor

Inconsistent placement decisions due to a lack of foster homes. All participants shared that the majority of placement decisions (i.e., where to house a foster child) come down to the availability of good foster homes. For instance, a child might have severe mental health or medical needs and the algorithm might recommend placing the child in a residential care center. However, most of these centers have limited openings, and the child-welfare staff are forced to manipulate CANS and place the child in a group home or foster care setting that is ill-equipped to manage their needs. In addition, policy dictates placement decisions regarding relatives willing to accept guardianship for a child. For instance, a supervisor shared that relatives' homes must meet all legal and safety requirements and go through a licensing process to become foster parents (see quote below). However, if relatives fail to meet these requirements, CANS is used again to recommend a placement setting where the foster child can at least be temporarily placed until better placements become available.

> Houses need to be legally compliant with all safety codes before a child can be placed there. These rules make sense on paper but in practice we lose so many good placement options because relatives can't afford to move or fix everything in the house to be legally compliant"
> – P10, Child Welfare Supervisor

The reappropriation of CANS has become a source of frustration for child-welfare staff due to inconsistent street-level decisions where they must continually find ways to address policy and systemic barriers as well as ways to make the algorithm work for their clients.

4.4.4 7ei (Seven Essential Ingredients) Algorithm

This child-welfare agency's core service delivery model is centered in trauma-informed care (TIC) where the staff assesses every case from a trauma-informed perspective and develops trauma-responsive interventions. However, program directors at the agency were concerned that decisions were still being made in an arbitrary manner where only some child-welfare teams were employing TIC practices. To address these gaps in practice, agency leadership has made significant investments towards developing a comprehensive four-part evidence-based program. The program is based on core concepts of the Neurosequential Model of Therapeutics [361]. Recently, principles of trauma-informed care have also been adopted within computational research where academics have proposed a framework for trauma-informed computing [98].

The 7ei algorithm (commonly referred to as the 7ei tool) is one of the four parts of this program. It is designed using TIC principles where the child-welfare team discusses and scores variables across seven domains. We provide some high-level explanations of the seven domains in Table 1 but do not have the agency leadership's permission to share the complete tool. Instead of predicting a singular outcome of interest, 7ei is used to assess the trajectory of child-welfare cases, i.e. - *change in 7ei domains over time*. Supervisors and program directors use 7ei to monitor trends in the seven domains to ensure progress is being made in cases using a seven-pronged TIC approach. For instance, if supervisors notice a downward trend for *Regulation* for a given

7ei Domain	Explanation
Prevalence	Exposure to and difficulty adjusting to adverse life experiences, i.e. - what happened to the child regarding trauma exposure?
Impact	Trauma occurs when a person's ability to cope with an adverse event is overwhelmed and contributes to difficulties in functioning
Perspective Shift	Caregivers' understanding of child's trauma exposure and its relationship with child's behaviors and/or impairments
Regulation	Traumatic triggers that cause serious dysregulation and action steps to address it
Relationships	Current progress towards facilitating strong relationships in a child's life. Strong relationships help create resilience and mitigate the effects of trauma
Reason to Be	Child's sense of identity and purpose and their sense of connectedness to their family, community, and culture
Caregiver Capacity	Caregivers' understanding of their importance to this process and whether they have additional social support

Table 8: 7ei Algorithm: Explanation of Seven Domains.

case, it triggers consultations with a clinical supervisor and consequent action steps that must be followed. This integrated approach allows the team to also focus on the child's and parent's ecosystem and their social support system such that parents have more caregiver support in the future. In addition, casenotes from the child-welfare team are also collected and linked to the 7ei quantitative scores which provide program directors with more contextual information about the scoring of variables at any point in time. This has allowed the agency to collect pertinent data about their practices in TIC and informs future improvements towards their TIC service delivery model. Below, we discuss how 7ei has impacted the nature of social work practice, its impact on the agency, and street-level decision-making.

Impact on Nature of Social Work Practice.

The first and second components of the four-part program involve extensive ongoing TIC trainings where the child-welfare staff (e.g., caseworkers, supervisors, family preservation team, permanency consultants) is introduced to the complexity of trauma, frameworks for understanding the effects of trauma, and practices and principles of TIC. This is accompanied by specialized supervision and consultation with a clinical supervisor, caregiver support specialist, program administrator, and a national expert (Dr. Bruce Perry) who provide varying degrees of support to frontline caseworkers based on family circumstances. These two components ensure that the nature of practice is centered in TIC model and allows for deeper engagement with 7ei. One supervisor explained -

> "Every case is centered in the trauma-responsive model using 7ei. It ensures that caseworkers are always thinking through TIC" – P14, Child Welfare Supervisor

Even though most participants shared that they appreciated 7ei and the trauma-responsive model, some participants drew attention to deeper systemic issues in child-welfare and the nature of social work practice that need redressing for tools like 7ei to offer more utility. One program director explained -

> "During home visits, caseworkers need to ask the right questions and read situations to be able to derive meaningful information. But we are not hiring people who have been doing this for 5-10 years or are highly qualified. The only new hires available are newly graduated social workers." –P7, Program Director

7ei was not designed to address such issues in child-welfare, however, it is pertinent to note that any algorithmic tool can only offer limited utility when the workers (and social work practice) are inherently impacted by deeper systemic and structural issues.

Impact on the Child-Welfare Agency.

The third component introduces a staffing protocol in the form of specialized collaborative meetings (commonly referred to as *7ei meetings*) where experienced members of child-welfare staff (e.g., family preservation, program director, permanency consultants) share their expertise and provide support to frontline staff members (i.e., caseworkers and supervisors). One supervisor explained -

> "I get the most out of having those conversations. What usually happens is that we end up talking about other things, but then it [7ei] brings us back around then too. So, we may be talking about *Prevalence* and *Impact* but when we are on *Relationships* or *Regulation*, we can start tying those to *Prevalence* and how we may be able to help with an intervention or explain why *Regulation* is off. So, I like connecting the dots through TIC and having conversations and processing it [7ei] with my staff." – P8, Child Welfare Supervisor

Implementation of 7ei (i.e., the fourth component) has also helped the agency address some problems in CWS. Lack of supervisory support is one of the main reasons for high turnover in child-welfare [89]. 7ei meetings ensure that new caseworkers are receiving adequate supervision and support from experienced members and they are also consistently using the trauma-informed framework. Agency also decided to train only one member of the child-welfare team (permanency consultant or supervisor) on the definitions and scoring criteria for variables where they facilitate the meetings and utilization of the tool. This ensures that the rest of the team is able to freely brainstorm as they work through TIC principles.

Impact on Street-level Decision-Making.

7ei algorithm is constructed using TIC principles and is embedded within the agency's service delivery model that child-welfare staff must follow. One supervisor shared that the caseworker position experienced high turnover such that the agency was always understaffed. Here, it becomes imperative that the agency is creating mechanisms to train new caseworkers where they are employing evidence-based best practices in street-level decision-making when they work with families. 7ei algorithm and 7ei meetings have helped created this training process -

> "7ei is embedded in the agency. Everything is based in the trauma-responsive model and often caseworkers are working through 7ei domains without even realizing it" – P9, Child Welfare Supervisor

However, participants also shared that even though 7ei is leading to better outcomes for some cases, it also ends up being impractical for cases where critical decisions come down to systemic or legal barriers. For instance, a family may have completed trauma-informed interventions and made changes to their household to provide a safe environment for children, however, the district attorney's office might still advocate against reunification because of the family's past (e.g., criminal history, domestic violence, drug use). One supervisor explained -

> "7ei is useful but so many decisions come down to systemic problems. Like..for example..the biggest barriers can come from legal parties. They are not in the [parent's] home every month. They are not talking to these parents, day in and day out. So, sometimes it can be challenging to try and fight for reunification." – P12, Child Welfare Supervisor

Here, the supervisor alludes to ongoing systemic tensions between child-welfare staff and the court system that are well-documented in social work literature [158, 154, 90]. 7ei was not designed to address these systemic issues, however, it still adds to the frustrations of caseworkers who continually employ a trauma-informed perspective using 7ei but are unable to receive a favorable decision in court for their clients.

4.5 Discussion

In this section, we discuss how algorithmic tools at the agency are inadvertently causing harm to the nature of the practice (i.e., *human discretion*), harm to the organization (i.e., *bureaucratic processes*), as well as uncertainties in street-level decision-making. Next, we discuss the implications for the design of responsible data science practices in the public sector.

4.5.1 Algorithms Harms to Social Work Practice

The child-welfare system has traditionally suffered from inconsistent decision-making with respect to child safety and family maintenance [86]. The problem of inconsistent decision-making is further aggravated by the fact that CWS experiences chronic turnover with the majority of caseworkers quitting within the first two years [418, 41]. However, research in evidence-based social work suggests that it takes caseworkers about two years to learn how to do the job and adeptly navigate and interact with the legal parties, service providers, children and families, and the nature of child-welfare practice [157]. Consequently, inexperienced caseworkers who use risk assessment algorithms such as CANS are led to believe that they are acting in an unbiased and objective manner. But as our results in this study and prior work [405] indicate, bias can be embedded within these assessments and the underlying data collected as a result of the practices of similarly inexperienced caseworkers. For instance, two of the most significant predictors of risk of maltreatment per the WARM risk assessment are *parents' cooperation with the agency* and *stress of caretaker* [405]. These variables are singly scored by caseworkers and encapsulate their impressions of the family with no input from the parents themselves. In addition, such variables capture a parent's response to the agency's intervention in their lives rather than the effectiveness or means of the intervention itself (i.e., how the caseworker approached and engaged with the family). Child-welfare is supposed to transition towards a "families as partners" model, however, algorithmic tools are creating a layer of obfuscation where such power asymmetries can become embedded within these tools. Ironically, risk assessments have worsened the turnover problem in CWS. Caseworkers leave due to their frustrations with practices and working conditions that have been exacerbated by the adversarial nature of risk assessments [83]. In addition, what remains missing from the conversations about algorithms in child-welfare, and the public sector in general, is the significant amount of data labor that is required of caseworkers as well as the repair work [110, 260] necessary to bring these systems to work for families and not just the governmental apparatus.

In addition, caseworkers at the agency were frustrated that the CANS algorithm required them to provide data labor where they must collect information about children and feed it to the system, however, the system also stripped them of discretionary power in regard to these decisions. As highlighted by recent work [233], predictive systems in high-stakes domains are *extractive by design* where they lead to a systemic extraction of discretionary power such that probabilistic outcomes are being used to supplant workers' contextual knowledge. Moreover, this interaction between caseworkers and algorithmic tools is further problematic because *human-in-the-loop* is the often proposed solution to erroneous decisions made by algorithms [137]. However, in this scenario, humans may be just as likely to make mistakes. Power asymmetries and incentive structures also directly impact how data is collected about children and how these assessments are scored. For instance, the CANS algorithm is conducted by both clinical therapists and caseworkers. However, a prior study [303] conducted on CANS found that clinical therapists were more likely to detect mental health needs because they are medically trained to detect these needs but also because they provide services when needs are detected, and consequently, they are paid for providing these services. That is, there was a clear financial incentive to exaggerate CANS scores. On the contrary, caseworkers very significantly less likely to detect mental health needs because 'detecting needs' inadvertently created more work for them. Once needs are detected, caseworkers must reach out to service providers and secure appointments for their clients.

4.5.2 Algorithmic Harms to the Child-Welfare Agency

Child-welfare agencies in several states in the United States have continued to rely on counsel from the federal government in the form of initiatives, regulations, and evidence-based approaches to improve their practice model. As previously noted, recent federal initiatives have laid the groundwork for more algorithmic interventions in CWS. However, federal directives have continually focused on the need for CWS agencies to adopt data-driven practices without providing adequate guidelines that focus on the ***why*** and ***how*** CWS agencies could employ evidence-based data-driven practices within their day-to-day processes in addition to training caseworkers on these practices. Consequently, CWS agencies in several states have rushed to adopt "something" in order to prove that they are employing scientific and evidence-based practices without ensuring that child welfare stakeholders have a strong understanding of how the model works, how to assure fidelity, and how to assess the model for issues of ethics and equity. The Allegheny Family Screening Tool (AFST) is a case in point of this scenario. A recent study conducted by Kawakami et al. [256] found that call screeners were offered minimal information about the working of AFST, considered it to be unreliable due to unexpected model behaviors, and even engaged in a collaborative "game" to learn more about the tool. Moreover, Cheng and Stapleton et al. [99] found that AFST-only decisions were racially biased and workers mitigated these biases by overriding erroneous decisions. Yet, these (sometimes hastily adopted) data-driven risk assessment models have become a central activity in many child welfare organizations [83]. As agencies across the United States begin to adopt the new CCWIS data model [225], it has also paved the way for tech startups [239, 371, 447, 238] to start pitching CCWIS-based algorithmic tools to agencies to help them meet their accountability requirements.

At this agency, there are serious data provenance concerns about data collected about children through the CANS algorithm since the data is so heavily manipulated by both caseworkers and foster parents. CANS was reappropriated to calculate foster parent compensations because policymakers believed that it would offer a fair and unbiased means for allocating resources. In addition, it would reduce costs over time since the improvement in care would lower the mental health needs and also the resources required to provide care. However, data specialists at the agency shared with us several cases where the compensation continually increased over time because the algorithm was being gamed. Another unintended consequence of exaggerating CANS scores is that foster children are now being sent to services (e.g., individual therapy) that they don't necessarily need. This is an added financial burden on an underfunded system that must pay for these unnecessary services. Here, CANS has added more barriers to consistent and evidence-based decision-making and introduced more constraints within the process. Moreover, caseworkers are mandated to provide the data labor that runs AI systems, however, they have no agency over these data production or future data utilization processes. This non-profit agency is contracted by the state's Department of Children and Families to provide child-welfare services and the agency must use DCF's centralized data system (i.e., eSACWIS) to record all case data. However, they are unable to download this data from their cases for the purpose of critiquing and improving their own practices. Skeptics of AI have invoked ethical frames where engineers and designers of algorithms must refuse to design systems that may raise social and ethical concerns [44]. However, such concerns might not arise at the onset, and as Selbst et al. note - "repurposing algorithmic solutions designed for one social context may do harm when applied to a different context" [415]. The CANS algorithm is a case in point of this scenario which was developed by academics with all the best intentions and as a medium to facilitate sharing of information about children's needs. Refusal to design may also not be an option for most agencies who must continually prove that they are employing innovative data-driven practices and may also be under political pressure to adopt algorithmic tools that seek to automate public service delivery.

4.5.3 Algorithmic Harms to Street-Level Decision-Making

> "All government policy and regulation is contradictory. That's why we exist, to fill in these gaps and make sense of policy in implementing it. This CANS is basically policy, right? So we interpret it and bend it however necessary to make it work for the people. I don't think we are gaming the algorithm. Because that would mean we are gaming all policy."
>
> – P7, Program Director

With inadequate federal guidelines on how to adopt, implement, and use algorithmic tools, there is significant misunderstanding regarding the roles of these tools in implementing rules or standards. Rules can be thought of as triggering criteria, i.e. - authoritative conditions that can be algorithmically coded for. On the other hand, traditionally street-level bureaucrats have exercised a significant amount of autonomy and flexibility in applying professional standards. CANS algorithm offers a case in point where it is expected to be implemented as a rule, however, caseworkers engage with it as if it were a standard. In addition, the program director in the above quote alludes to the *repair work* [110, 260] that caseworkers undertake in order to make the algorithm work for their clients. Caseworkers are expected to use algorithmic tools that shift discretion away from them, however, they are also expected to assume responsibility when automated decisions lead to poor outcomes for families. Recently, public sector algorithms have received criticism when they lead to poor outcomes and exacerbate racial disparities [473, 459, 163]. However, finding the sources of harm can be difficult when we witness distributed use of smaller algorithmic tools as opposed to a larger and more visible algorithmic system. Distributed

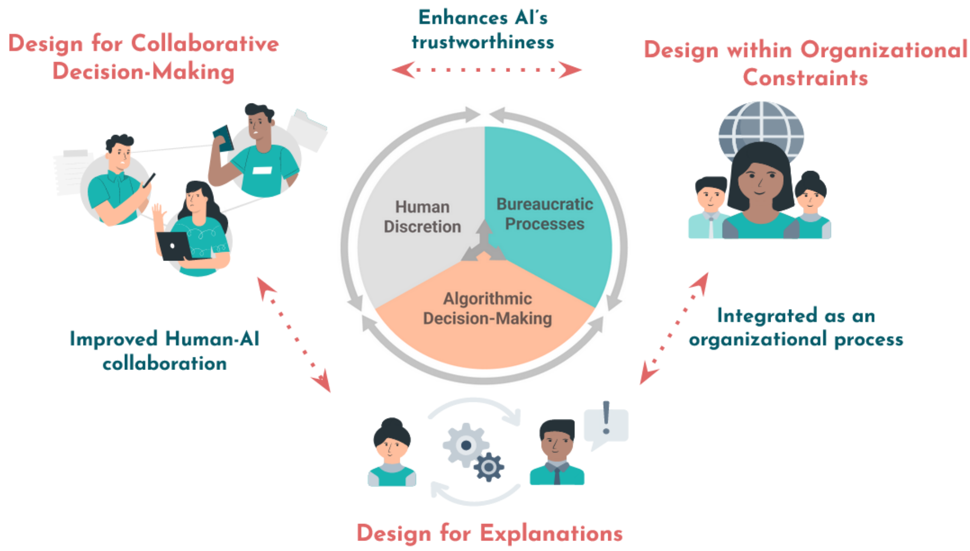


Figure 9: Human-Centered Design of Algorithms using the ADMAPS Framework

nature of several of these tools also shifts accountability from one group to another within the agency, and therefore, can be hard to assess. For instance, as highlighted in prior work [396], the Anti Sex-Trafficking algorithm refers cases to the Human Anti-Trafficking Response Team (HART) based on a set of risk predictors. However, HART is now receiving an influx of cases with no added resources provided to them. Consequently, cases that really need their expertise are receiving inadequate attention; indirect harm that is hard to measure. As illustrated in this case study, actuarial tools such as CANS are transformed into algorithmic decision supports which are now further being legally mandated to ensure caseworkers' compliance. AFST algorithm implemented at a child-welfare agency in Allegheny County, Pennsylvania is another example of this scenario where it was initially implemented in a voluntary capacity, however, its use was mandated when it did not receive enough engagement from caseworkers. AFST is still characterized as an algorithmic decision-support but offers an interesting example of how such tools are increasingly limiting human discretionary work instead of findings ways to improve it.

4.6 Implications for Responsible Data Science in the Public Sector

Crucial to the discussion of responsible data science practices in the public sector are the following points. First, it is imperative to recognize that decision-making is a complex process where information needs to be shared among several parties (e.g., child-welfare staff, district attorney's office, parents' attorneys, service providers, and judges) and decisions are collaboratively made. As illustrated by Figure 2, **algorithmic tools must be designed for collaborative use** for them to be able to offer higher utility to practitioners and improve decision-making processes. In addition, practitioners must be able to explain decisions made to other involved parties.

Therefore, it is necessary to **design algorithms that facilitate explanations**. Designing for collaborative use and explanations will aid deeper integration between human discretion and algorithmic decision-making, and consequently, lead to improved Human-AI collaboration over time. In addition, **algorithm design must occur within organizational constraints**, i.e.- it is imperative to account for the resource and legislative constraints within which all decisions (human or algorithmic) must be made. In the absence of these design requirements, algorithms frustrate practitioners who are unable to follow the 'ideal algorithmic decision'. For instance, CANS might recommend that a foster child be placed in a Level 4 foster home, however, with a lack of good foster homes in the system, the caseworker must manipulate data to produce an outcome that points to an available placement [395]. Consequently, algorithmic tools that account for such constraints are more likely to be integrated as an organizational process as well as enhance the tool's trustworthiness among practitioners who must interact with it collaboratively.

Second, as highlighted by the adoption and use of the 7ei algorithm, agency leaders had to invest in a significant amount of resources in terms of trainings, specialized consultations, hiring experts and creating the time and space (in terms of collaborative 7ei meetings) to ensure proper utilization of the algorithmic tool. That is, there is a significant amount of human labor that went into the integration of 7ei into daily work routines. It is imperative to note that these investments must be made in order to rebuild and improve decision-making processes that utilize algorithmic systems.

Third, as highlighted by the development of 7ei, the tool was designed with the intent to decompose the algorithm and turn it into an open-ended and transparent process such that it tracked outcomes over time instead of predicting an outcome of interest. As highlighted by prior work, there is an irreducible degree of uncertainty associated with each predicted outcome [137, 354] and this problem is further exacerbated in the public sector where all the relevant information is not always available and there may also be contradicting sources of information [75, 404]. In such scenarios, predicted decisions led to frustrations on part of caseworkers and poor decisions for families. In sum, there is a need to prescribe away from the black-box model of ***input-computation-output*** and build tools that support the decision-making processes of practitioners. Designing to support decision-making processes will also lead to fair, transparent, and accountable decisions for clients as well as responsible AI practices from the data science community.

Furthermore, as highlighted by prior work [462], there is no need to over-engineer solutions using complex neural networks when simple models have been shown to be just as accurate in high-stakes domains [462, 253, 388]. Vaughan and Wallach [462] argue that we must begin by considering the needs of relevant stakeholders and then design for intelligibility techniques that support these needs. Specifically for child-welfare stakeholders in this case study, intelligibility encompasses three things - 1) provide relevant information to the child-welfare team, 2) provide explanations for recommended trauma-responsive services, and 3) demonstrate compliance with evidence-based practice. In addition, caseworkers' frustrations with CANS and the repair work they conducted highlighted key aspects of decision-making that the algorithm failed to capture. This further highlights the need to conduct extensive ethnographic work as well as the need to co-design with stakeholders such that their needs and concerns are addressed.

4.7 Limitations

We conducted an extensive ethnographic study that highlights the messy interactions between social work practice, policies, and algorithmic decision-making at a child-welfare agency. However, this study has some limitations that create opportunities for researchers to further expand upon this body of work. First, this study only focuses on the perspectives of on-the-ground caseworkers and how their interactions with algorithmic systems. It is important to understand the perspective of affected communities (i.e., foster children, parents, and foster parents) about whom decisions are being made through algorithms. For instance, a recent study conducted by Stapleton et al. [432] found that parents considered CWS algorithms to be punitive and unsupportive. Parents instead wanted systems that would help them fight against CPS as well as evaluate CPS and the caseworkers themselves. Future research should continue to focus on uncovering street-level complexities within this complicated sociotechnical environment. For instance, recent work [404, 405] used computational methods to uncover patterns of invisible labor, systemic constraints, and power asymmetries that impact both families and workers. Finally, this ethnographic study only uncovers complexities at one child-welfare agency, however, child-welfare practice and policies can significantly vary from one state to another. Therefore, we recommend that researchers conduct similar qualitative explorations in other jurisdictions.

4.8 Conclusion

We conducted an in-depth ethnographic study to understand the daily algorithmic practices of caseworkers at a child-welfare agency. We qualitatively coded our data from the ethnography to the dimensions of the ADMAPS framework to reveal the complex interdependencies between human discretion, algorithmic decision-making, and bureaucratic processes. We focused on the macro-interactions between the three core elements of ADMAPS to highlight how algorithms in use at the agency were impacting the nature of social work practice, bureaucratic processes at the agency, as well as the nature of street-level decision-making. Our findings highlight that there is a need to focus on the proper implementation and integration of algorithmic tools into decision-making processes and not just the initial development and deployment of the algorithmic model.

That is, there is a need to rethink the amount of investment required to ensure the proper adoption of algorithms in complex sociotechnical environments. In addition, we show how a simple algorithmic tool that tracks variables over time instead of predicting an outcome offered higher utility to caseworkers. Moreover, algorithms need to be designed to support explanations and collaborative use such that they augment human discretionary work. In addition, algorithmic tools need to be fully supported by bureaucratic processes by allocating necessary resources and accounting for organizational constraints to ensure that they are integrated as an organizational process. As a result of this study, we also propose responsible data science practices for algorithm design in the public sector.

CHAPTER 5: HOW TO TRAIN A (BAD) ALGORITHMIC CASEWORKER: A QUANTITATIVE DECONSTRUCTION OF RISK ASSESSMENTS IN THE CHILD WELFARE SYSTEM

ABSTRACT: Child welfare (CW) agencies use risk assessment tools as a means to achieve evidence-based, consistent, and unbiased decision-making. These risk assessments act as data collection mechanisms and have been further developed into algorithmic systems in recent years. Moreover, several of these algorithms have reinforced biased theoretical constructs and predictors because of the easy availability of structured assessment data. In this study, we critically examine the Washington Assessment of Risk Model (WARM), a prominent risk assessment tool that has been adopted by over 30 states in the United States and has been repurposed into more complex algorithmic systems. We compared WARM against the narrative coding of casenotes written by caseworkers who used WARM. We found significant discrepancies between the casenotes and WARM data where WARM scores did not mirror caseworkers' notes about family risk. We provide the SIGCHI community with some initial findings from the quantitative de-construction of this risk assessment algorithm.

5.1 Introduction

Child welfare (CW) agencies began using risk assessment tools in the 1980s to reduce bias and standardize decision-making in cases of possible child abuse and neglect [205, 148]. Today, risk assessment tools that were designed to be used in conjunction with caseworkers' clinical decision-making are often fed into more complex algorithms to support decision-making. Algorithmic risk scores are often viewed as more neutral than a worker's clinical impression of risk, which may be biased. However, worker bias may be embedded in algorithms themselves, offering a veneer of standardization that disguises the degree to which algorithmic risk scores still represent potentially faulty risk assessment. In some cases, a worker's opinion of risk may account for most of the variance in whether children are removed from their homes, even when it comes to quantitative tools [161, 284]. For a variety of systemic reasons, child-welfare caseworkers nationally have an average of fewer than two years of work experience in their positions [157]. Caseworkers vary widely on their impressions of family risk, which is known to be influenced by demographic factors such as race and gender of the worker and the families [144]. This raises additional concerns about the possibility that the data being embedded in algorithmic models further amplifies bias in risk assessment.

When an allegation of abuse is made at the screening hotline, a frontline worker assesses the case based on prior referrals and family history and makes a determination to screen in (or out) the case for an investigation [72]. Caseworkers conducting these investigations (e.g., home visits, interviewing referent, parents, relatives, neighbors) typically record information in two forms: **1) quantitative risk assessments** and **2) unstructured narrative casenotes** that are electronically stored in the case record. These risk assessments were designed to provide a consistent assessment of risk based on specific risk factors that are believed to predict future abuse/neglect. One such tool, the Washington Assessment of Risk Model (WARM), is scored at the investigation alongside an outcome of maltreatment (i.e., founded, unfounded, inconclusive). Concurrently, caseworkers also write detailed casenotes that contain more contextual information about the family such as a text summary of abuse/neglect, a discussion of major risk factors, and explanations of risk factor ratings on the risk assessment. However, over the past decade, artificial intelligence (AI) research has grown exponentially with significant attention being paid to developing algorithms using quantitative data collected from risk assessments while overlooking casenotes as a data source that contain more contextual signals about case circumstances.

With significant growth in AI research, SIGCHI researchers have become very engaged in understanding how fair [415, 15, 155], accountable [480, 223], transparent [271, 261], and explainable [102, 278] algorithms may be developed using a variety of design methods. More specifically, researchers have been very interested in how algorithmic decision-making is carried out in the public sector [109, 232, 385, 398, 397]. In this study, we critically examine the WARM risk assessment and predictors from WARM which have been adopted into newer algorithmic models. Specifically, we investigate the congruence between WARM and qualitative casenotes. We also examine the degree to which subjective variables in WARM (e.g., parents' cooperation with the agency) impact caseworkers' decision-making. Therefore, in this study, we ask the following over-arching research questions:

- **RQ1:** *Where do qualitative caseworker narratives align with (or diverge from) quantitative structured decision-making WARM assessments in regard to risk factors?*
- **RQ2:** *How do caseworkers' biases and perceptions of families become embedded into quantitative structured decision-making assessments?*

In addressing these research questions, our specific contributions are as follows:

- We explore a quantitative deconstruction of WARM predictors and theoretical constructs using public child-welfare data. Our findings show how caseworkers' biases become embedded in structured risk assessments as well as the core discrepancies between the purpose of risk assessments and how they are currently being used.

Physical Abuse
Severity 1 = No marks indicated
Severity 2 = Minor marks
Severity 3 = Numerous or non-minor marks
Severity 4 = Emergency room or medical treatment
Severity 5 = Hospitalization for more than 24 hours
Severity 6 = Permanent disability or death

Table 9: Physical abuse to the head, neck, or facial region. Severity is coded on a scale of 1 (low) through 6 (high).

- Specifically, we find that WARM measures a parent's response to a caseworker intervention as opposed to the efficacy of the interventions themselves. In addition, we find significant divergences between WARM risk ratings and the risks indicated in the caseworker narratives. This suggests a disconnect between quantitative and qualitative accounts of, ostensibly, the same underlying phenomenon.

These initial findings are part of a larger work-in-progress research project on the quantitative deconstruction of algorithms employed within the U.S. Child-Welfare System (CWS).

5.2 Methods

The Dataset. The secondary data used in this study comes from a dataset, *Factors that Influence the Decision Not to Substantiate a CPS Referral*, housed at the National Data Archive on Child Abuse and Neglect (NDACAN). The federally-funded project sought to assess child-welfare system decision-making in the state of Washington [160], specifically, 1) to identify the factors that influence the decision ***not*** to substantiate (i.e., find neglect/abuse) a CWS referral; and 2) to identify the characteristics of CWS referrals that are more likely to be unsubstantiated compared to those that are substantiated. For that project, researchers independently examined 2000 cases, exploring two sources of data: **1) qualitative casenotes** from narrative portions of case records, which were coded and quantified based on narrative risk factors, and **2) WARM quantitative data** for the same cases. Table 1 depicts one of the coding schemes used specifically for coding physical abuse (as depicted in casenotes) into numeric data. Please refer to English et al. [160] for more details about the initial study's coding criteria. They conducted multivariate analysis on WARM data to determine predictors related to unsubstantiation of maltreatment in each dataset in order to ascertain protective factors.

Analysis. For the purpose of our study we re-examine this dataset. However, instead of treating narrative data and quantitative data as two independent sources of information, we compare across the two to assess how much of the pertinent information is captured by each. As previously noted, risk assessments were designed to be used in conjunction with caseworkers' contextual and clinical judgments [419, 413]. However, over the past decade, the decision-making latitude has transitioned towards algorithmic decision-making [397]. Therefore, motivated by concerns about how public services decision-making has shifted significantly towards the use of quantitative data from risk assessments to recommend/predict an outcome of interest, we compared these information sources to better understand how structured assessments and algorithms might influence risk analysis.

Therefore, to answer **RQ1**, we compared the narrative codes against WARM risk factors to assess how much of the pertinent information needed for decision-making is accounted for by each of these two sources. First, to assess congruence between quantitative (WARM) risk factors and qualitative (case noted) risks, we compared risk categories extracted from each based on the primary researcher's analysis of risk rating to assess if they were measuring similar items. The shared categories between the two were then used to calculate the degree to which the cases demonstrated incongruities between risk factors identified in casenotes versus the WARM assessment. Overall risk on WARM is depicted on a scale of 0-5 (0=no risk, 1=low risk, 2=moderately low risk, 3=moderate risk, 4=moderately high risk, 5=high risk). Cases that received a 0-2 risk rating receive a *low standard investigation*[6], referred to community-based services, and quickly closed. Whereas, cases with risk ratings of 3-5 were assigned a *high standard investigation* [7]. This distinction between CW cases, as established by child-welfare in Washington, allowed us to compare across casenotes and WARM data. A case was considered incongruent if the narrative notes explicitly mentioned risk (rated 3-5) but a "No Risk" rating on WARM was entered, and conversely, if the narrative notes explicitly mentioned no risk (rated 0-2) but WARM categorized the risk as "Moderate", "Moderately High", or "High" risk. Additionally, we examined the correlations between WARM risk factors to see if there were any surprising or unusual relationships.

For **RQ2**, we critically examined the data being recorded by WARM itself and how biases may be embedded within this quantitative assessment. Data was randomly organized into an 80/20 training/testing split to perform backward and forward feature selection to mirror the ways this type of data might be used in predictive decision-making. This informed variable selection for building a multinomial logistic regression model (using 10-fold cross validation) to classify overall risk to the child. Variables' significance to the model was quantified and used to further understand whether caseworkers' perceptions of families potentially impacted how risk was measured.

[6]**Low standard investigations** are defined as a review of prior CPS involvement and collateral contacts to determine if a further investigation should occur. They do not require face-to-face contact with the child or caregiver.

[7]**High standard investigations** includes a review of prior CPS involvement, collateral contacts, face-to-face interview with child and caretaker, and additional assessments required to determine whether or not abuse/neglect occurred

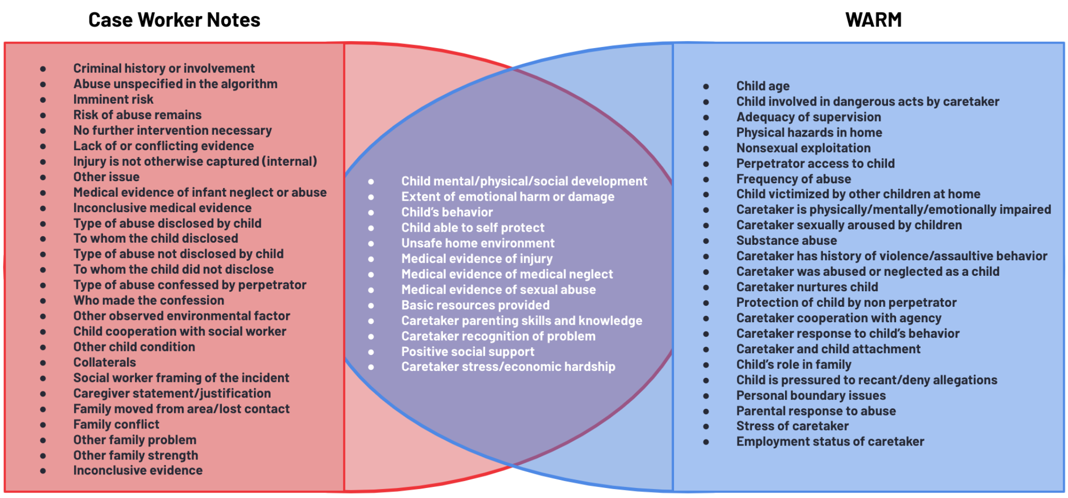


Figure 10: Narrative coding variables were compared to WARM categories; while there are some common risk factors, both contain risk factors that the other does not address.

5.3 Results

5.3.1 Discrepancies between Narrative Coding and WARM Risk Factors (RQ1)

As depicted in Figure 1, several differences exist between the risk factors mentioned in narrative coding versus the WARM assessment. This could be for several reasons - 1) caseworkers felt constrained by WARM risk factors because they do not appropriately capture safety concerns, 2) different risk/safety priorities emerged during the investigations but were not captured by WARM, 3) depending on when casenotes were completed, the caseworkers' memory might be erroneous or focused on different factors, or overall risk may have shifted over the course of the assessment. However, as depicted in Table 2, caseworkers' casenotes and WARM were aligned in capturing several risk factors, however, there are significant discrepancies in how these factors are scored.

The risk factors common to both the narrative coding and WARM assessment were used to

Shared Variable	False Positive	False Negative
Child has Bad or Difficult Behavior	7.69%	13.18%
Caregiver Doesn't Recognize Problem	11.25%	13.07%
Economic Stress or Hardship	-	27.22%
Medical Evidence of Medical Neglect	-	3.70%
Medical Evidence of Physical Injury	-	9.52%
Medical Evidence of Sexual Abuse	-	25%
Negative Emotional Condition of Child	8.41%	17.78%
No Basic Resources	3.66%	22.64%
Positive Social Support	5.23%	-
Unable to Self Protect	45.35%	2.30%
Unsafe Home Environment	2.73%	31.50%

Table 10: WARM to Narrative Coding Comparison. Variables found in both the WARM assessment and casenotes are used to examine how well WARM works in providing a framework for explaining the case.

investigate discrepancies in the computed risk. As previously noted, a risk rating of 0-2 results in a *low standard investigation*, and a risk rating of 3-5 results in a *high standard investigation*. Therefore, if the casenotes contained a risk rating of 0-2 and the WARM assessment for the same case contained a risk rating of 3-5 (or vice versa), a core discrepancy exists. Based on this, incongruities (i.e., *false negatives* and *false positives*) were calculated. False negatives were measured by dividing the number of cases for which casenotes explicitly described risk (rated 3-5) but were marked "No Risk" in the WARM assessment by the total number of cases for which risk was present in casenotes. On the other hand, false positives were calculated by adding all cases marked "Moderate", "Moderately High", and "High" risk in WARM (rated 3-5) where the casenotes explicitly mentioned no risk, and dividing that by the total number of cases for which the narrative coding determined there was no risk present. It is important to note that these percentages are not accuracy measures of WARM itself. The caseworker is the one to both determine the risk to the child according to WARM and also the one to write the casenotes. Instead, these percentages measure where WARM does not align with the content of casenotes. For example, a caseworker may determine that sexual abuse of a child did occur and indicated this on WARM resulting in a classification of high risk. However, they may also determine that due to the primary caregiver's protective actions, no ongoing risk (i.e., impending danger) exists within the family, resulting in a classification of no risk in the casenotes.

This further emphasizes the argument that critical information needed for decision-making exists in ***both*** the casenotes and quantitative WARM data. Casenotes offer more contextual details that highlight the complexities and uncertainties within a case and are necessary for explaining why the WARM assessment contains certain risk ratings as well as the protective factors within a family that may be mediating risk factors. In our recent work, we conducted computational inspection of child-welfare casenotes and highlighted how contextual factors such as patterns of invisible labor, the impact of systemic factors and constraints, as well as underlying power relationships, can be computationally derived from caseworkers' narratives [403].

5.3.2 Caseworkers' Impressions are Embedded in Risk Assessments (RQ2)

Table 3 depicts the most significant variables from WARM that predict the likelihood of maltreatment. These variables were selected via backward step-wise feature selection; the regression model created by forward step-wise feature selection had comparable results. Both models found *Cooperation with agency* and *Stress of caretaker* to be significant predictors of the overall risk of recurrence of neglect/abuse. Cooperation with the agency additionally has the highest parameter estimate and odds ratio. The above model had a correct negative classification of risk of 86.60% and a correct positive classification of risk of 80.00%. As the rating of risk for nurturance, stress of caretaker, age of child, and cooperation with agency increases, the probability

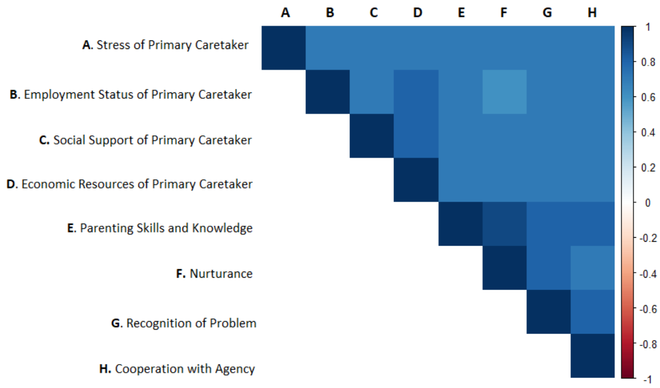


Figure 11: Correlations between subjective and related categories regarding caretaker's quality of care per the WARM assessment.

that the case is categorized as one with further overall risk to the child also increases. Notably, most of these significant variables are scored by caseworkers and are subjective based on the caseworkers' impression of the family.

As depicted in Figure 2, correlations among WARM risk factors were explored in order to assess relationships. Among those that were more expected, such as the high correlation between economic resources, employment status of the caretaker, and the caretaker's stress, there were some very high positive correlations of 0.5 and above between cooperation with the agency and other fairly subjective variables related to the quality of care provided by a caretaker. The caretaker's recognition of problem and cooperation with the agency were, among others, highly correlated with parenting skills, nurturance, and protection of a child.

This indicates a potential vector for bias as these predictors are singly scored by caseworkers and are influenced by caseworkers' own impressions of the family. As illustrated by prior studies, a caseworker's opinion of risk may account for most of the variance in whether children are removed from their homes, even when it comes to quantitative tools [161, 284, 380]. For example, a caseworker who is struggling to engage with a family may feel that the caretakers are not taking appropriate steps to protect their child by accepting the worker's assistance, or that the caretakers do not have the skills or disposition required to nurture their child due to their rejection of interventions. It is imperative to note here that predictors such as *Parent's Cooperation with Agency* measure the parent's response to the intervention rather than the effectiveness of the intervention itself with respect to child safety. In addition, there is a core discrepancy between how the risk assessment is designed to be used versus how the caseworkers are using it. The tool is designed to assess the likelihood of re-referral or recurrence of abuse (i.e., long-term trajectory of risk) [397], however, caseworkers are strictly focused on the substantiation of abuse/neglect at present (i.e., immediate risk posed).

Variable	Parameter Estimate	Standard Error	Confidence (95%)	Odds Ratio
Cooperation with agency^1	0.41961 ***	0.085	0.2543 to 0.588	1.5214
Stress of caretaker^1	0.34440 ***	0.0841	0.1802 to 0.5103	1.4111
Age of child	0.33551 ***	0.0849	0.1728 to 0.5061	1.3986
Nurturance^1	0.39401 **	0.1238	0.1525 to 0.6395	1.4829
Deviant arousal^1	-0.57698 **	0.2135	-1.0044 to -0.1671	0.5616
Frequency of abuse/neglect	0.21284 *	0.0845	0.047 to 0.3786	1.2372
Perpetrator access^2	0.16153 *	0.0665	0.0301 to 0.2913	1.1753
Attachment and bonding^2	0.34011 *	0.1401	0.0681 to 0.6179	1.4051
Child's role in family^2	-0.49729 *	0.207	-0.9093 to -0.097	0.6082
History of abuse or neglect as a child^2	0.27959 *	0.1268	0.0307 to 0.5289	1.3226
Deviant arousal^2	0.31524 *	0.1452	0.0262 to 0.5968	1.3706
Adequacy of medical/dental care	0.23922 *	0.1172	0.0128 to 0.4741	1.2703

Table 11: Significant Variables from Backward Feature Selection. Caretaker^1 is the primary caregiver (generally biological parent) and Caretaker^2 is the secondary caregiver in the household.

5.4 Discussion

Algorithms in child-welfare pose a unique set of challenges since interactions between different stakeholders (for e.g., parents and caseworkers) as well as interactions between systemic factors (for e.g., bureaucratic processes and policy-related factors) significantly impact which data is collected about families as well as how decisions are made. In this section, we first discuss the implications of inconsistencies between casenotes and WARM assessment followed by implications for child-welfare practice.

5.4.1 Inconsistencies between Casenotes and WARM Assessment

Our results indicate that there are significant inconsistencies between the caseworkers' casenotes and the WARM assessment scores. Risk assessments were designed to support decision-making and consistently record findings based on a risk framework. That is, they were supposed to be used in conjunction with caseworkers' contextual judgment about cases [413, 419]. However, structured data collected by these assessments over the past two decades is now used in algorithms that make high-stakes decisions about children and families [397, 137]. The lack of internal consistency between caseworker narratives and the WARM assessment discovered in this study is further evidence that algorithmic decision-making likely embeds incomplete assessment data. This further implies that there will be a high degree of uncertainty associated with any predicted outcome and human discretion is needed to fill in these gaps [354, 398]. In sum, there are inadequacies in both forms of assessments (caseworker's judgment and algorithmic) such that neither offers a holistic assessment of a family's circumstances. Here, the purpose of this study is to highlight these discrepancies to draw attention towards the **decision-making process** (and how to improve it) instead of the **decision outcome**. This requires recognition of the complexities within this socio-political domain, critical decision points, as well as the value-laden choices and heuristic decision-making that CWS staff must engage in. Ongoing engagement with domain experts (e.g., child-welfare program directors, social work academics, caseworkers, and impacted families) as partners is necessary, rather than a peripheral engagement where computing professionals singly focus on data-driven practices and exercise more agency. Recent work in SIGCHI has highlighted how engagement with public sector stakeholders can uncover the complex decision-making ecosystems [398, 20], needs of the stakeholders [101, 232], as well as utilize unstructured textual narratives to help contextualize decision-making processes [403]. In addition, the design of research questions, as well as how these questions are situated within the cultural and historical contexts of child-welfare, are of critical importance [377]. Any systems work that is conducted without equal partnerships with child-welfare domain experts will uncritically reproduce and embed similar points of failures (as illustrated by this study) into new sociotechnical systems.

5.4.2 Proxy Variables and Underlying Power Dynamics

Algorithms are meant to support standardized and evidence-based assessment of risk, thereby reducing bias. However, our results suggest *cooperation with agency* is a significant predictor of the risk of maltreatment per the WARM risk assessment. Prior work has established the importance of actively engaging families in interventions and its consequent impact on the recurrence of abuse [204, 451, 427]. When a tool like WARM quantifies a parent's cooperation with the agency, using it as a proxy for engagement and child safety, worker bias becomes deeply embedded in the algorithm without an adequate critique of the practices that improve (or deteriorate) a family's engagement. That is, risk assessment models often use a family's response to interventions rather than the effectiveness of the interventions themselves to assess the likelihood of recurrence of abuse. Bureaucratic processes, such as training, protocols, and organizational resources play a significant role in family engagement. For instance, the caseworkers' micro-skills, such as collaborative problem solving, focus on strengths, respect for diversity, listening, and reliability have all been presented as worker skills associated with improved engagement [376, 204, 427, 451]. Efforts have been made within child-welfare from both a policy and practice standpoint to transition towards a "Families as Partners" [375] model where parents are supposed to act as equal partners in the case planning process and have agency in decision-making. However, as our results indicate, *cooperation with agency* and *stress of caretaker* are embedded in risk assessment algorithms and scored by caseworkers with no input from parents. That is, a lack of proper examination of variables and underlying biases further shifts power away from parents. In sum, human discretionary work on the part of caseworkers and bureaucratic processes significantly impacts the context within which (and how) data is collected about families and needs to be critiqued as a critical part of algorithm design. Evidence-based practices in child-welfare service delivery, such as motivational interviewing, describe parent motivation as something that emerges as an outcome of skilled communication led by the caseworker, but motivation is often misplaced as a quality of a parent in risk assessment without an examination of the influence of the caseworker's actions in supporting meaningful engagement [218].

5.4.3 Risk Assessment Algorithms: Old Wine in New Bottles

Psychometric risk assessments have been used in CWS since the early 1980s, however, there are several imminent concerns regarding the reliability and predictive validity of variables and outcomes that have now been embedded into and inherited by risk assessment algorithms [188, 395]. Our results unpack some concerns about one such variable: *cooperation with the agency.* Magura et al. [310] developed the *Family Risk Scales* where the *Parent's Cooperation with Agency* scale originated. Although several predictors from the scale (e.g., parental recognition of problems, the capacity of parents to change, parental motivation, etc.) have been repurposed into risk assessment models, their association with the recurrence of abuse remains unknown [443]. Empirical knowledge related to child-welfare practice is fragmented, and social science theories are needed to fill in these gaps [188]. Computer scientists have continued to focus on the reliability of predictors while averting a closer inspection of predictive validity. A high inter-rater reliability does not necessarily mean that a causal relationship exists between a predictor and an outcome (i.e. - internal validity) [153]. As depicted in Table 3, we find that there is a disproportionate effect of specific variables on overall risk. Given that we know how some of those variables (e.g., Stress of Caretaker, Cooperation with Agency, etc.) are biased, contextual, and socially constructed, it is no surprise that these power dynamics become embedded in the decision-making outcomes of ascertaining risk. Moreover, a deeper understanding of the impact of interventions (e.g., parenting services), protective factors (e.g., parents' social support system), and the risk posed by the system itself is necessary for developing a more comprehensive understanding of risk [169, 188].

5.5 Conclusion

Unpacking a risk assessment algorithm is a good initial step towards understanding the *human discretionary work* and *bureaucratic processes* that influence the decision-making process and need to be further examined. Even though, such quantitative deconstruction makes visible some of the latent processes that impact the final decision outcome, it still only accounts for a small proportion of caseworkers' day-to-day practices. For instance, in a recent ethnographic study, Saxena et al. [398] found four different risk assessment algorithms that caseworkers were using on a daily basis. Moreover, distrust in an algorithm impacted how they interacted with systems. That is, quantitative deconstruction unveils how caseworker biases and underlying power dynamics can be concealed within structured assessments but a deeper ethnographic analysis of the impact of such tools on caseworkers is equally important. Based on the immediate findings, our first objective will be to assess how theoretical signals derived from case notes can be used in conjunction with quantitative risk scores to offer a more holistic perspective on the risks and needs of children and families.

In this study, we present the discrepancies that exist between caseworkers' narratives (i.e., what the caseworkers are witnessing on the grounds) versus the structured decision-making tool. Initially employed as a means to support decision-making, the balance has shifted significantly towards algorithmic risk assessment and away from caseworkers' judgment over the last decade. While neither method is perfect, the former presents a veneer of authority. Moreover, deconstruction of the algorithm reveals latent processes through which caseworkers' biases based on subjective impressions of the families can become embedded into such algorithmic tools. The contribution of this work is twofold - 1) it characterizes systemic deficiencies that impact how risk assessments are conducted and the kinds of data they seek to collect. Understanding such deficiencies allows us to reimagine the role and scope of algorithmic decision-making and its ability to support child-welfare practice, and 2) it explicates that a key role for data-driven practice might be to deconstruct, re-design, and evaluate data infrastructures in complex socio-political domains to uncover how such disparities impact street-level practice, their impact on communities, as well as implications for developing better sociotechnical systems.

CHAPTER 6: UNPACKING INVISIBLE WORK PRACTICES, CONSTRAINTS, AND LATENT POWER RELATIONSHIPS IN CHILD WELFARE THROUGH CASENOTE ANALYSIS

ABSTRACT: Caseworkers are trained to write detailed narratives about families in Child-Welfare (CW) which informs collaborative high-stakes decision-making. Unlike other administrative data, these narratives offer a more credible source of information with respect to workers' interactions with families as well as underscore the role of systemic factors in decision-making. SIGCHI researchers have emphasized the need to understand human discretion at the street-level to be able to design human-centered algorithms for the public sector. In this study, we conducted computational text analysis of casenotes at a child-welfare agency in the midwestern United States and highlight patterns of invisible street-level discretionary work and latent power structures that have direct implications for algorithm design. Casenotes offer a unique lens for policymakers and CW leadership towards understanding the experiences of on-the-ground caseworkers. As a result of this study, we highlight how street-level discretionary work needs to be supported by sociotechnical systems developed through worker-centered design. This study offers the first computational inspection of casenotes and introduces them as a critical data source for studying complex sociotechnical systems.

6.1 Introduction

Government agencies in the United States have sought to reduce costs and increase efficiencies in public policy and social services delivery by increasingly adopting information communication technologies (ICTs) [167, 465, 163] that aim to minimize repeated data collection and bureaucratic overhead, provide targeted client services, and improve decision-making processes [290]. These ICTs have helped public entities continually collect comprehensive cross-sector data including, structured data (e.g., quantitative assessments), unstructured data (e.g., case narratives), and metadata on different attributes of citizens' interactions with public services [326]. Academics, practitioners, and policymakers have used this data to develop algorithmic systems that purportedly lead to more consistent, objective, and defensible decision-making on critical matters related to human lives [163, 397, 76]. Various public sector services now use algorithms, such as in child-welfare [397], criminal justice [214], job placement [20], and public education [386], often in the form of risk assessments to preemptively recognize and mitigate "risk" posed to citizens and governmental apparatus [22].

The U.S. Child-Welfare System (CWS) faces significant challenges. CWS has limited resources, burdensome workloads, and high staff turnover [89, 397], and faces intense public scrutiny on harm caused to children who are removed from their parents [86] but also when child abuse tragedies occur [182]. These challenges have mounted pressure on CWS to employ algorithmic systems and prove that they follow consistent and objective decision-making processes. SIGCHI researchers have made significant contributions in developing algorithms that aid frontline caseworkers in deciding which calls (i.e., allegations of abuse) should be screened in for an investigation [137, 106]. SIGCHI researchers have also used crowdsourcing platforms such as Amazon Mechanical Turk (MTurk) to study people's perceptions of algorithmic decisions and their impact on human judgment [287, 211]. However, as highlighted by recent ethnographic work in CWS [396, 72], there are drawbacks in these studies that need redressing: **1)** algorithms built from quantitative administrative data in CWS only account for a narrow set of predictors, offering a deficit-based framing of families [397], and **2)** experiments conducted on crowdsourcing platforms do not account for organizational/legislative constraints or day-to-day bureaucratic protocols that impact decision-making for all cases [396]. In light of these concerns, SIGCHI researchers have suggested that collaboratively curated caseworker documentation (i.e., caseworkers' narratives) may offer a more holistic picture of street-level interactions and bureaucratic complexities [20, 396]. Unlike administrative quantitative data, caseworker narratives offer a more credible source of information by revealing workers' interactions with families, uncertainties in a case, and impact of bureaucratic constraints on decision-making. These narratives offer much of the desiderata necessary for computational narrative analysis [26]. Casenotes about families are highly contextual but also share core similarities because they describe similar pathways that most families follow in CWS [245]. For this study, we pose the following over-arching research questions –

- **RQ1:** *How can computational text analysis help uncover invisible patterns of street-level labor conducted by caseworkers?*
- **RQ2:** *How does computational text analysis highlight the systemic constraints placed on caseworkers' discretion?*
- **RQ3:** *How can computational text analysis help investigate latent power relationships in CWS?*

To answer these questions, we conducted computational text analysis of casenotes using topic modeling [470]. For **RQ1**, we analyzed dominant topics over time and uncover patterns of invisible labor conducted by caseworkers. For **RQ2**, we divided families into three groups based on their number of interactions with CWS and highlight that families in different groups have varying needs. For **RQ3**, we conducted computational power analysis of the casenotes to uncover latent power structures in CWS. This paper makes the below unique research contributions –

- We offer the first computational investigation of child-welfare casenotes and introduce them as an important and useful data source for studying complex sociotechnical systems.
- We highlight invisible patterns of street-level work that caseworkers do within the gaps of legislation (and beyond job duties). These patterns were not uncovered in prior ethnographic work at the same CW agency suggesting case narratives can provide rich contextual information.
- We show how caseworkers navigate different constraints (systemic, temporal, algorithmic, resource etc.) for different needs of families over the life of a case which uncovers nuances and implications for worker-centered technology design beyond algorithmic interventions.
- We found how power relationships for key personas in CW (i.e., CW staff, foster parents, birth parent, etc.) change for different family types, complicating the popular narrative of CW workers having the most power in CW cases.

We find support that computational text analysis of casenotes can be a powerful tool for developing holistic decision-support tools instead of the popular administrative data-centered risk assessment tools [106, 137] that have been found to be biased [396]. This answers calls in prior SIGCHI research about the possibility of using case narratives as an important research tool [397, 20]. We advocate combining computational analysis with qualitative explorations to critique sociotechnical systems. Multiple methodological lenses on the same phenomenon will likely provide holistic insights that any single approach may not [334, 47]. In the following sections, we first present the current public sector and computational text analysis research within SIGCHI. Next, we discuss our methodology for answering each of the research questions.

6.2 Related Work

We situate our research within the SIGCHI community and provide an overview of the work that has been done towards developing sociotechnical systems for the public sector followed by computational text analysis research conducted within SIGCHI.

6.2.1 Public Sector Research within SIGCHI

The SIGCHI community has been at the forefront of research on how sociotechnical systems are developed and employed within the public sector. The work has been wide-ranging, including studies that examine issues of civic engagement [150, 206], shaping emergent technologies for collaborative work [58, 325], designing systems centered on participation and empowerment of affected communities [72, 146], and expanding HCI methods for support labor [176]. Through the continued employment of digital technologies in the public sector, researchers have also studied how these systems have impacted the decision-making latitude of street-level bureaucrats[8] who traditionally exercised significant autonomy when implementing policies [81]. Recent studies have found that value conflicts arise when the logics embedded within the government's digital platforms do not align with street-level bureaucrats' discretion when they tried enacting the same shared values in practice [466, 146, 232]. SIGCHI researchers have also unpacked the forms, limits, and complexities of participatory design within the public sector that is now increasingly dictated by public-private partnerships [150, 299] where newer technologies are now being developed for the governance of smart cities [478, 416, 227].

The continued employment of digital technologies in the public sector has changed governance practices in two distinct ways. First, these systems have improved data sharing practices between different government sectors and purportedly allowed for minimal repeated information gathering, provided targeted services to clients, and allowed for *end-to-end* service delivery [167, 465, 163]. This has allowed government agencies to continually collect data about citizens through their daily operations [326], with the expectation that the data will be transformed into knowledge to inform future decisions that seek to efficiently allocate resources [232]. Here, **"data becomes the promise of future bureaucratic efficiencies"** [232]. Second, with a primary focus on efficiency and economy, scholars are questioning the core nature of public services as "caring platforms" designed for the public good as opposed to private corporate entities that focus more on optimizing profits [295]. That is, public services that exist to "care for" and "serve" citizens cannot and should not be optimized using performance metrics of the corporate world. SIGCHI scholars have thus begun studying data-driven practices that adopt *care* as a design lens to create systems that advocate for a caring democracy [325, 449, 206]. Despite two decades of adoption of digital technologies (often referred to as Digital Era Governance [81]) and promises of transformation, these tools have generally fused onto existing human discretionary practices rather than altering them at a deeper organizational level [464, 396, 232]. Digital technologies have raised the need to understand human discretionary work conducted by bureaucrats who must balance citizens' needs against the demands of policymakers as they acquire new skills and learn to make decisions through these systems [396].

As a result, recent HCI scholarly work has sought to unpack the nature of human discretionary work conducted at the street-level in public services [16, 354, 396, 232, 385, 109, 374]. Alkhatib and Bernstein introduced the theory of *street-level algorithms* to distinctly highlight the gaps in algorithmic decision-making that human discretion needed to address [16]. Unlike street-level bureaucrats who used discretion to reflexively make decisions about novel cases, street-level algo-

[8] A street-level bureaucrat is a professional service worker (e.g., social worker, police officer, teacher) who operates in the frontline of public service provision. They interact closely with clients and make decisions about them based on how they interpret policies relating to the situations at hand [297].

rithms produced illogical decisions that could only be redressed in the future through new data. Pääkkönen et al. further developed this theory to highlight that algorithm design must identify and cultivate important sources of uncertainty because it was at these locations that human discretionary work was most needed [354]. Recently, Saxena et al. [396] synthesized this prior work conducted in the public sector into a cohesive framework for algorithmic decision-making adapted for the public sector (ADMAPS) which accounts for and balances the complex interdependencies between human discretion, algorithmic decision-making, and bureaucratic processes. ADMAPS framework advocates for developing algorithms based on a holistic decision-making process, balancing complex dynamics within sociotechnical systems, and accounting for human discretion and bureaucratic processes [396]. Additionally, Ammitzbøll Flügge et al. [20] and Saxena et al. [396] highlighted the collaborative nature of caseworkers' decision-making processes and the impact of bureaucratic structures that algorithm design need to account for. In sum, HCI scholars have reached a general consensus that any algorithmic interventions in the public sector needed to understand the complexities of human discretion carried out at the street-level when implementing day-to-day bureaucratic processes and legislative policies.

6.2.2 Child Welfare Research within SIGCHI

Recent work within SIGCHI has focused on understanding how we can better support individuals and groups within CW. Gray at al. [210] have worked on designing technologies for foster youth by creating a new digital memory box for fostered and adopted children to create and store their childhood memories. Badillo-Urquiola et al. [38] have focused on addressing online safety within foster families by identifying the challenges foster parents face as they mediate teen technology use in the home. Recently, the community has expanded its efforts towards understanding algorithmic decision-making systems employed in CWS [396, 137, 101]. Algorithms are currently used to determine if a child should be removed from a parent's care [129], the level of care a child needs [331], and the type and intensity of services a family will receive [61]. While these decisions can be life-altering, a systematic review of CWS algorithms has shown that many failed to incorporate child-welfare literature or social science theories, instead primarily adopting a deficit-based framework that performed poorly against outliers and deviated from target outcomes [397]. SIGCHI researchers have also directly engaged with CWS stakeholders (i.e., families, frontline workers, and specialists) to understand community perspectives and the impact of algorithms on frontline workers' decisions. Brown et al. [72] conducted community-based co-design workshops with CWS stakeholders and found that they felt uncomfortable with algorithmic systems because decisions were centered in deficit-based frameworks that perpetuated biases and bolstered distrust. Complementing this work, De-Arteaga et al. [137] found that frontline workers sought supervisor approval to override an algorithmic decision when they considered it to be incorrect. Similarly, Cheng et al. [101] examined stakeholders' understanding of 'fairness' regarding machine learning systems in CWS and proposed a framework that allows stakeholders' notions of fairness to emerge organically by working directly with public sector agencies to develop systems that provide a higher comfort level to the community.

Recent ethnographic work in CWS also revealed caseworkers' frustrations with state-mandated algorithms as they did not account for an agency's resource constraints, legislative policies, or uncertainties inherently present in every case [396]. Saxena et al. [396] also found that all the caseworkers involved at the front-end of case planning collaboratively curated casenotes comprising details about interactions with families, uncertainties about the case, critical decisions, and sequence of events that offer a more holistic perspective of case circumstances. Prior work in CW has conducted qualitative exploration of case narratives for a small corpus of text to understand the experiences of both mothers and fathers [156, 319]. Our study sought to understand whether it is feasible to use computational text analysis of narratives to uncover critical details about CW cases such as patterns of human discretionary work conducted by caseworkers and the bureaucratic processes that constrain human discretion.

6.2.3 Computational Text Analysis Research within SIGCHI

The study of sociotechnical systems requires an understanding of how nuanced and contextualized activities of humans inform, shape, and are shaped by technical systems [10]. Studying sociotechnical systems often involves the analysis of text data to understand these types of interactions. While scholars often used qualitative methods to analyze such texts in the past, researchers such as Muller et al. [334] have found parallels between qualitative methods and machine learning (ML) techniques and explored the possibility of adopting computational text analysis for unstructured text-based datasets. Recently, computational text analysis methods, including ML methods, have become popular in studying sociotechnical systems in the SIGCHI community [27, 96, 217, 97]. This is because, as Molina and Garip [330] note, ML techniques can overcome the long-standing limitations of statistical modeling and provide contextual findings. Moreover, Nguyen et al. [339] state that applying computational text analysis on text which is inherently steeped in cultural and social factors can scale to large bodies of text, help discover insights that may only reveal themselves when text is aggregated, unpack subtle patterns, and detect sentiment.

While SIGCHI has widely adopted computational text analysis methods to study sociotechnical systems, few studies have examined complex sociotechnical systems in the public sector. Instead, much SIGCHI work has only indirectly touched upon areas of relevance in the public sector using computational text analysis. For example, Antoniak et al. [27] studied the experiences of pregnant women via Reddit posts, Chancellor et al. [96] predicted mental illness

Heading	Details
Family Interaction	Describe the frequency/location, quality of interaction, justification for the type and level of interaction (supervised/unsupervised), and conversations with the parent(s)/caregiver(s) regarding what needs to happen in order to move to a lesser restrictive setting of Family Time.
Concerns	Discuss any concern(s) surrounding family time, how they are being addressed, and enter information about future plans to resolve the concern(s)
Communication	Describe the parent's/caregiver's response or receptiveness to communicating with the child(ren)'s caregiver(s) and describe any schedule or method of communication.
Special Considerations	Include information on any special considerations for the child and parent(s) during family time (e.g., no contacts orders, parents confirming the visit, anyone who should not be at the visit).

Table 12: Agency guidelines on how to record visitations in casenotes.

severity from Instagram tags, and Guha et al. [217] examined the role of an individual's agency in social media non-use from web survey responses. Of these works, Antoniak et al. [27] revealed the versatility and applicability of using computational text analysis on unstructured narrative texts. The authors [27] show that topic modeling works well with stories that follow a formulaic sequence of events and can reveal latent power dynamics between personas and patterns of topic transitions. Recently, in the area of sociotechnical systems research in the public sector, Saxena et al. [397] conducted a systematic literature review of computational methods used in CWS and suggested employing computational text analysis techniques (e.g., topic modeling) to elicit context-specific information about CWS cases that current statistical and machine learning algorithms fail to draw out.

Our survey of prior literature shows that while much of SIGCHI research has indirectly examined sociotechnical systems in the public sector, there is a dearth of SIGCHI research that employs computational text analysis to examine these complex systems. And yet, outside of the SIGCHI community, scholars have actively examined the utility of applying computational text analysis methods (specifically topic modeling) to sociotechnical systems research in the public sector [240, 343, 200, 140] and have noted that topic modeling methods can aid qualitative methods by guiding the systematic discovery of information [240] and help reduce directionality bias that arises from manual interpretations of text [140]. Therefore, responding to these calls, we employed topic modeling techniques for analyzing child-welfare casenotes. Using topic modeling, we discovered invisible patterns of human discretionary work performed by caseworkers to gain a more holistic understanding of child-welfare work practices with direct implications for algorithmic decision-making and worker-centered systems design.

6.3 Research Context

We partnered with a child welfare agency that serves around 900 families and 1300 children in a metropolitan area in a U.S. Midwestern state. The state's Department of Children and Families

(DCF) has contracted this agency to provide child welfare and family services and must comply with all DCF standards, including the use of mandated decision-making algorithms. DCF's Initial Assessment (IA) workers investigate allegations of child maltreatment, and if abuse/neglect is substantiated, the case is referred to the agency to provide services. These services are negotiated between the parents' attorneys, district attorney's office, and the judge after caseworkers have conducted initial structured assessments and provided their recommendations to the court. As depicted in Table 13, this agency is comprised of several different child-welfare teams that work in collaboration based on the specific needs of families. From the onset of a case, a safety and permanency plan is developed which also establishes the frequency of interactions between caseworkers and birth parents, and consequently, the documentation of these interactions. The agency has established rigorous standards around writing casenotes and compiling case documentation since information needs to be shared among all involved parties (i.e., CW staff, parents' attorneys, district attorney's office, judge). Caseworkers are trained at the agency to write detailed, narrative-style casenotes to record information about families based on observations, pertinent details, and discussions with families. The agency's training guide on casenotes is informed by best practices in social work literature [173, 195]. For instance, Table 12 provides an example of how caseworkers must record supervised visits in casenotes. This collaboratively curated documentation by CW staff involved at the front-end of case planning also acts as a roadmap of decisions made (and the circumstances surrounding these decisions) if such decisions need to be critiqued and/or defended for any case. Narratives, unlike risk assessments, also capture the uncertainties inherent in any child-welfare case. Understanding these uncertainties (and their impact on caseworkers' decisions) becomes especially important for cases where a child-welfare tragedy may have occurred. Prior work [83] in CWS highlighted these uncertainties for a case where a child passed away -

> *"How can the uncertainties confounding workers be conveyed in such situations: the deep commitment of the mother to do well by her child, the remorse of the father and his agreement with a court order to stay away, the rallying around of family members and friends, the subsequent loss of the father's job, the worker's transfer to another caseload, the move of the family to another community, all occurring over time, amidst improvements in the child's care, and amongst all of the other factors taking place in the lives of the parents, workers, family members and others."*

Case management supervisors add another layer of accountability by ensuring that caseworkers are updating casenotes on a bi-weekly basis and providing detailed descriptions. The agency also has specific instructions in the "Case Note Content Guide" on how to record face-to-face interactions, phone calls, court hearings, and visitations. Many of these uncertainties and com-

Name	Details	Role
IIS	Intensive In-home Services	Provides in-home services to both birth and foster parents where the child has high medical needs
HART	Human Anti-Trafficking Response Team	Manages cases where the foster youth is at high-risk for human or sex trafficking
ICWA	Indian Child Welfare Act	Manages cases concerning children from native American tribes
YTA	Youth Transitioning to Adulthood	Work with foster youth who are about to age out of the foster care system and require independent living provisions
FPS	Family Preservation Services	Works with birth parents in their efforts to achieve reunification
FCA	Foster Care and Adoption	Works with foster parents for training and certification to manage children's needs, foster care licensing, and adoption
PC	Permanency Consultation	Works with case management through the legal process of achieving permanency (i.e., reunification, adoption, or guardianship)

Table 13: Different kinds of child-welfare teams at the agency

plexities are highlighted in casenotes, and we expected computational text analysis on these casenotes could reveal nuanced dynamics between caseworkers and families. CWS comprises of several different child-welfare teams (see Table 13) and works with families based on varying case circumstances. We specifically analyze casenotes written by the ***Family Preservation Services (FPS)*** team that works with birth parents in their efforts to achieve reunification with their children. However, every family is assigned a case management team (case manager and supervisor) that also works with the family and FPS and records their interactions in casenotes which are then compiled in case documentation and made available to all involved parties. We obtained Institutional Review Board (IRB) approval from our mid-sized private research university to use this child-welfare agency's casenotes for this research.

6.4 Methods

This section provides details about the casenotes dataset and the data cleaning process followed by our data analyses process. For this study, we employ methodology developed by Antoniak et al. [26] for computational narrative analysis using LDA topic modeling. The authors [26] showed that their methodology work well for corpus of text that follows a specific sequence of events with frequently occurring personas – characteristics that are observed in child welfare casenotes. The stories are highly individual but share core similarities in terms of personas, sequence of events, power hierarchies, and critical decision points.

6.4.1 Dataset

This dataset was acquired from Family Preservation Services (FPS); a specialized child-welfare team whose primary goal is to help birth parents achieve reunification with their children. FPS works closely with birth parents through parenting classes and other court-ordered services to ensure that a safe living environment can be achieved for children. FPS must provide substantial evidence to the DA's office and the judge in order to recommend reunification. They accomplish this by recording parents' progress in parenting classes and other services as well as risk factors within the households. Documenting casenotes is an important task for caseworkers because it guides the child welfare staff on the next steps, provides evidence that agency or caseworkers are making reasonable efforts to help children, and serves as a collaborative tool by demonstrating the collective efforts between families and caseworkers [173, 195]. FPS works closely with the case management team and other service providers and also has access to their casenotes which are compiled into case documentation. In this regard, casenotes are collaboratively written by CW staff. Our collaborators at the agency shared that CW staff spent about half their time working on documentation and updated casenotes on a bi-weekly basis (per on-boarding training) such that all parties have timely access to information. However, every casenote contains the date and time for all interactions, even if the case note is electronically updated at a later date.

Casenotes contain a rich source of information about a family's case and include details about caseworkers' interactions with and observations of parties involved in a case (e.g., birth parents, foster parents, relatives, and children). We obtained records of 9719 casenote entries (the 'dataset') for 310 families referred to the agency around May 1, 2019, and worked with Family Preservation until October 14, 2020, or were discharged sooner. Families that received services from the agency were assigned a family identification number, and caseworkers entered casenote details whenever a relevant interaction related to the family took place (e.g., phone call, home visit, parenting class, domestic violence class, court hearing, etc.). Specifically, the dataset contained detailed information on when an interaction related to the family occurred, the time and duration of the interaction, family member names related to the case, detailed narrative texts on what happened during the interactions, and the caseworker names.

Data Preparation, Cleaning and Anonymization

As we were interested in tracking the detailed sequence of interactions between families and CWS staff and inferring how interactions changed over time, we collated all narrative casenotes related to each family identification number in chronological order. Next, we extracted text columns and respective family identification numbers from the collated casenotes. All other columns were excluded from our analysis. We cleaned the collated casenotes by removing punctuation and stopwords from the text. We also anonymized all personal information to protect the privacy of the families. The anonymization process was conducted in the following two steps. First, we used the frequently occurring surnames dataset from the 2010 U.S. Census [78] and Social Security popular baby names dataset [13] to remove all first and last names from the casenotes. We, however, did not remove any first or last names that also functioned as common nouns, such as the last names List and Brown. Second, we replaced all numerical-related information in the text with the word *NUM*. Table 21 shows the summary corpus statistics after preprocessing the narrative text. Table 21 shows that of the 310 collated casenotes, 235 casenotes contain

Metric	Value
Number of casenotes with more than 1500 words	235
Average number of words per casenote	3,835
Number of words in longest casenote	38,748
Number of unique words	44,407

Table 14: Corpus Statistics

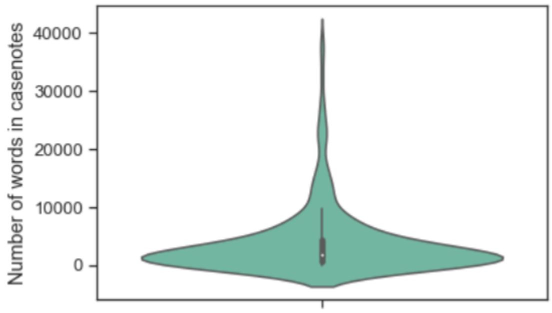


Figure 12: Word distribution in Casenotes.

text greater than 1500 words, and the maximum word length of a casenote is over 38,000 words. Figure 12, a violin plot depicts a skew in the word length distribution of casenotes after data curation where most casenotes are shorter in length.

6.4.2 Topic Modeling and Narrative Analysis Over Time (Life of Case)

Topic Modeling Solution

Topic modeling is one of the most widely used text mining methods in natural language processing (NLP) to infer latent themes from text documents and extract features from bag-of-words representations [471]. We decided to use LDA for our child-welfare casenotes because LDA can provide easily interpretable insights into densely structured texts which contain both formal and informal language such as ethnographic fields notes [351] and Reddit stories [27]. Following Antoniak et al. [26], we used Mallet's implementation of LDA topic model to train our topic model. As this implementation of LDA requires the number of topics as a hyperparameter, we took the next two steps to train the topic model. First, we determined the optimal number of topics by creating topic model solutions from 1 to 30 topics and calculated the coherence score and average topic overlap (using the Jaccard similarity statistic) when we assigned 15, 20, 25, and 30 keywords to each of the topics. We found that 14, 17, 22, and 29 topics maximize the divergence between the topic model's coherence score and average topic overlap. We then manually inspected the 14, 17, 22, and 29 topic model outputs to determine the optimal number of topics. After the interpretations were collaboratively discussed, we reached a consensus to use the 17 topic model solution depicted in Table 23.

Member Checks for Topic Model Qualitative Interpretation

Topic model outputs often identify thematic patterns in the text at lower abstraction levels than human interpretivist analyses and can benefit from grounded thematic methods to draw out themes in the text [47]. As such, three co-authors of this paper used an open-coding process on the original casenotes that have the highest probabilities assigned to each topic to capture patterns (themes) within the texts [67]. Each co-author individually identified dominant themes, labeled the topics, and then collaboratively discussed their interpretation and labels with co-authors. After this iterative process was complete, a consensus was reached between co-authors on the final trained topic model's themes. Having assigned themes to topics, we next conducted member checks by providing caseworkers with our interpretations of topics, top keywords, and examples of original casenotes with the highest probability (for each of the respective topics). Creswell and Miller [127] argue that member checking is crucial to establishing credibility to qualitative analyses as this technique brings study participants back to the data to judge how accurate and realistic researchers' interpretations are. Accordingly, we asked frontline caseworkers to determine the high-level themes based on their reading of the original casenotes and asked if they agreed with our interpretative themes. Caseworkers' feedback helped us further refine our interpretation and topic labels. After these iterative discussions were complete, we reached a consensus about the interpretations of the topics.

6.4.3 Group Analysis of Topic Popularity Over Time

Prior work in CWS [312, 89, 448] has found that caseworkers work with families for different lengths of time depending on the family's unique needs. In addition, CWS experiences a high turnover rate due to high caseloads. To mitigate this phenomenon, CW agencies often group cases in high, medium, and low needs groups based on case severity and assign them to caseworkers to ensure more equitable workloads [276]. Prior work has also highlighted that case complexity (e.g., type of maltreatment, age, number of children, need for financial assistance, drug abuse in the family) is directly associated with the time spent under the care of CWS [366, 89]. In line with these studies, we sought to examine if the length of casenotes can serve as a proxy for the family's needs and the severity of the case. To that end, we interrogated the distribution of number of interactions that families have with the child welfare agency. Figure 13 shows that most families interact with child welfare staff less than 10 times, and there are fewer families that interact with the agency as the number of interactions with the agency increases. Table 15 demonstrates that families in this dataset interacted with CW staff an average of 31 times. Based on the percentile distributions, we grouped families into roughly three equally sized buckets. We then conducted statistical and qualitative analysis into each group's casenotes to determine if the number of interactions with CW staff can serve as a proxy for a family's level

Descriptive Statistic	Value
N	9,616
Mean	31.1
Standard deviation	36.3
25 percentile	7.0
Median	19.0
75 percentile	40.0

Table 15: Descriptive statistics on the number of interactions between families and child-welfare staff

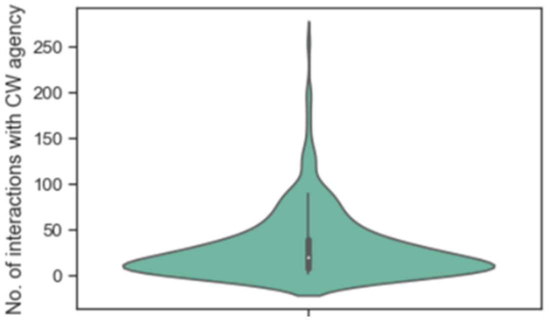


Figure 13: Families' interactions with the agency. The breadth of the violin represents the frequency of interactions.

of need. Finally, we applied the trained topic model from **??** to each group to track topic popularity over time. To accomplish this, we segmented each of the cleaned casenotes into ten equal sections and calculated how average topic probabilities differ for the groups. As casenotes follow a formulaic sequence of events, we were able to divide the texts into ten equal-length sections to create normalized sections (see Fig. 4-8). We define these normalized and chronologically arranged casenote sections as **"Life of a Case"** which further allowed us to study which topics emerged as significant at different temporal points in a case.

6.4.4 Power Analysis of Personas

Sentiment analysis

Child welfare cases involve many parties such as foster parents, family members, and CW staff who are bound by their own responsibilities, goals, and legal obligations. We were interested in examining the power dynamics between such parties by analyzing the day-to-day power relationships between them. However, we needed to first examine the sentiment of casenote sentences because the linguistic choices made by caseworkers could have important implications for how we examine the dynamic relationships between families and caseworkers.

Caseworkers are trained in writing detailed casenotes based on observations and facts and provide as much descriptive details as possible about their interactions with families [173]. As pre-

Sentiment	Number of sentences	Percentage
Positive	6,598	9.61%
Negative	2,415	3.52%
Neutral	59,619	86.87%

Table 16: Sentiment analysis of casenotes.

viously noted, this collaboratively curated documentation is imperative for creating a roadmap of critical decisions as well as the circumstances underscoring those decisions. We conducted sentiment analysis on all sentences in the casenotes using a sentiment analysis tool *Valence Aware Dictionary and sentiment Reasoner* (VADER) [235] to examine the writers' tone of these casenotes. As illustrated by Antoniak et al. [26], VADER was an appropriate tool to compute sentiment analysis since it was developed for social media text and textual data from other domains. Using only sentences with five or more words (to avoid mistakenly segmented sentences and sentences without meaningful information), we assigned a compound sentiment score (a normalized score ranging from -1, extreme negative to +1, extreme positive) to each sentence in the casenotes. As shown in Table 16, we noted that more than 86% of the sentences were neutral, and only 9.6% and 3.5% of the sentences were classified as positive and negative sentences, respectively. The predominantly neutral tone indicated that the casenotes were mostly descriptive in nature and provided for a suitable corpus of text for conducting power analysis and discovering underlying relationship patterns between key personas.

Personas of Interest

We were interested in examining how power relationships differed between personas in the groups defined in Section **??**. To do this, we first identified the personas of interest for the whole dataset by manually inspecting the casenotes. Table 17 illustrates the main personas that appear in all of the casenotes. After identifying personas of interest, we used a non-anonymized version of the casenotes to replace references made to the main personas (References column of Table 17) with normalized versions of the persona (Persona column of Table 17). For example, we assigned words like *grandmother*, *aunt*, *uncle* to Support System. Table 17 shows summary statistics on how often these personas appeared in the casenotes, including the total number of mentions of each persona, the number of casenotes that mention the personas, and the average number of times casenotes mention the personas. 'Legal parties', 'medical parties', and 'significant other' rarely appeared in casenotes. As we were interested in measuring the relative power scores between personas, we removed these three infrequently mentioned personas to prevent them from causing statistically spurious effects on power relationship analyses.

Persona	References	Total Mentions	Casenotes containing Mentions	Average Mentions per Casenote
Biological parent	Mother, Father, Parents, Mom, Dad, *Proper Name*	29,545	281	105.14
Child Welfare Staff (CWS)	FPS, OCM, CM, Case manager, Supervisor, FC, *Proper Name*	17,730	277	64.01
Child	Kid, Baby, Son, Daughter, *Proper Name*	17,935	262	68.45
Foster parent	Caregiver, FP, *Proper Name*	3,368	148	22.76
Support System	Grandparents, Aunt, Uncle, MGM, PGM, MGF, PGF, Friend, In-laws, Cousin, *Proper Name*	1,652	125	13.22
Medical Parties	Therapist, Dentist, Doctor, Nurse, *Proper Name*	334	96	3.48
Legal Parties	Lawyer, Judge, Law enforcement, Guardian ad-litem (GAL), Attorney, Assistant district attorney (ADA), District attorney (DA), Court, *Proper Name*	291	89	3.27
Significant Other	SO, Boyfriend, Girlfriend, Significant person, *Proper Name*	254	45	5.64

Table 17: Persona and their most frequent references in text. *References* column shows the common nouns that are frequently mentioned in the casenotes to represent each persona. Proper nouns (and related variations such as nicknames) are also extracted for the different personas.

Power computation

We adapted the works of Antoniak et al. [27] and Sap et al. [393] to compute power scores of and power relationships between personas of interest. Sap et al. [393] created a lexicon of power frames where an entity is assigned a positive power when the entity dominates or exerts a level of control over another entity. This definition of power was appropriate for our study as we anticipated that certain personas would exercise power over other personas in a similar manner. The aforementioned lexicon included 1737 verbs, of which each verb indicated directionality with respect to whom power is assigned. Table 18 shows examples of paraphrased sentences from our casenotes where verbs are assigned power. Next, we lemmatized the verbs in the lexicon and our casenotes, parsed the casenotes which contained the normalized personas from Section **??** using the spaCy parser, and computed power scores for personas of interest in each of the groups by extracting the subjects, verbs, and direct objects from each sentence. Finally, we incremented (or decremented) each persona's power score according to its position in the sentence and the verb power effect. In addition to the results of sentiment analysis, this power analysis method was appropriate here because the goals of all involved personas are aligned and centered in achieving reunification for children and birth parents.

6.5 Results

6.5.1 Topic model solution organized by dominant themes

We analyzed the results of a topic model solution trained on casenotes and determined 17 to be the optimal topic number based on topic comprehensiveness and interpretability. As illustrated in Table 23, we further grouped these 17 topics into 6 dominant themes to improve readability.

Example Sentences
Sarah [child] **demanded** some juice which made the mom upset.
This writer [child welfare staff] **communicated** with Pam [foster parent] via phone.
Ms. Jones [birth-mom] **refused** to speak with worker [CW staff] and continuously shrugged her shoulders when asked a question.

Table 18: Paraphrased exemplar sentences depicting power between personas. The child (Sarah) has a high power score; CW staff has equal power with foster parent (Pam); and the birth-mom (Ms. Jones) has a high power score and CW staff has a lower power score.

#	Theme	Topic	Unique keywords
1.	**Helping Families Secure Resources and Navigate Bureaucratic Processes**	**T2**: Helping families secure essential resources **T5**: Establishing roles and expectations for different parties **T7**: Coordinating virtual interactions during COVID **T12**: Helping families navigate court proceedings	*housing, appointment, employment, resources, services* *client, reported, shared, meeting, roles* *virtual, court, camera, communication, covid* *court, plan, safety, proceeding, reports*
2.	**Managing Medical Consent, Medication Administration, and Medical Appointments**	**T3**: Managing medical consent between caregivers and accompany clients to medical appointments **T11**: Helping establish medication schedules and manage logistics around medical appointments	*caregiver, discussed, consent, form, health* *medication, schedule, safety, therapy, appointment*
3.	**Coordinating Time, Travel, and Pickup Logistics for Visitations & Appointments**	**T1**: Managing conflicts between caregivers when scheduling visitations **T4**: Continued attempts to get in touch with birth parents **T6**: Managing logistics around visitations and appointments **T17**: Coordinating travel to and from school for foster children	*visitation, conflict, canceled, voicemail, email* *missed, reschedule, voicemail, phone, contact* *visit, arrived, room, residence, ride, issues* *school, attendance, missed, suspended, reports*
4.	**Conducting Structured Assessments to Determine Risks and Progress**	**T13**: Keeping track of parents' progress in court-ordered parenting classes **T14**: Conducting home visits, assessing safety concerns, and scoring assessments	*parenting, chapter, session, curriculum, completed* *observed, assessed, home, clean, beds*
5.	**Facilitating Interactions between Children and Parents during Supervised Visits**	**T8**: Observing and facilitating interactions with infants **T9**: Observing and facilitating visits between siblings and adolescents	*baby, visit, bottle, diaper, feeding* *children, play, room, toys, food*
6.	**Observing and Recording Concerns during Transportation**	**T10**: Observing and recording children's behavior during transportation **T15**: Observing and recording pre- and post-transportation concerns	*transported, slept, cried, picked, visit* *visit, concerns, weather, clothing, seat*

Table 19: 17 topic model solution organized by six dominant themes. Topics are labeled T1-T17.

The 17 topics are labeled T1-T17 and all names in exemplar sentences have been replaced with pseudonyms to protect the privacy of individuals.

Helping Families Secure Resources and Navigate Bureaucratic Processes

CW staff act as mediators between birth parents, relatives, and foster parents where they help establish roles and expectations for each party as well as bridge the administrative gap between community resource providers, clients, and the court system. CW staff work closely with birth parents and help them acquire essential resources that they require to meet their children's needs. They share information with parents about how and where to find resources as well as help them acquire these resources [259]. This often takes the form of helping parents find employment, transportation, and home essentials (e.g., food, clothing, toiletries) that would improve stability within the household and facilitate achieving reunification with children. Here, CW staff bridge the gap between community resource providers and clients (i.e., parents) in need. CW staff also work actively to alleviate ambiguity with respect to roles and expectations for each party from the onset of a case [50]. Prior work has established the need to improve communication and enhance teamwork in order to improve relationships in child-welfare practice [50, 194]. CW staff also work to ensure that birth parents and foster parents are in agreement with respect to parental roles and expectations. Specifically, CW staff explain to birth parents that foster parents are temporary caregivers who will care for the child's needs and give birth parents the time to make changes within their household so the child can safely return home.

CW staff play a critical role in helping families navigate court proceedings where they escort parents to court and advocate for them. CW staff share progress made by parents in parenting classes, court-ordered services, and parenting skills that they are exhibiting during supervised visits. As illustrated by the exemplar sentence below, CW staff help parents understand the court process and the changes they must make to receive a favorable decision in court. As illustrated by topic 7 (i.e., *virtual interactions during COVID*), CW staff also assumed newer responsibilities during the COVID pandemic in terms of facilitating virtual interactions between parents and children and also helping parents navigate through virtual court hearings. Reading through the case notes, we observed that CW staff also helped parents and caregivers troubleshoot technology issues and explained to them how to use Microsoft Teams or Skype for Business.

[T2 Probability: 0.65] "Case Manager would like the Family Preservation Specialist to visit with Ms. Davidson [birth-mom] at least 1x per week and assist with helping her secure resources, especially for the unborn baby."

[T5 Probability: 0.56] "Family Preservation Specialist [FPS] attended staffing with supervisor, Ongoing Case Manager, and Ongoing Case Manager supervisor to discuss referral, roles, and supportive services needed. FPS attended team meeting to introduce herself to Sarah [birth-mom] and explained her role in the process. FPS asked about what kinds of services Sarah [birth-mom] was in need of and she responded that housing is her main priority. In addition, Family Preservation Services will gather resources on rent assistance, emergency daycare, and baby supplies that Sarah [birth-mom] can then have at her disposal.

[T12 Probability: 0.59] "Family Preservation Services greeted the family and provided them with an introduction of their role and services that they will provide the family. Mr. B [birth-dad] shared that the baby may possibly be placed with his aunt and moving soon. Mr. B [birth-dad] stated that he has court tomorrow at 9am. Family Preservation Services asked the family if they would mind if she attended court with them. Family Preservation Services explained that she would be there for support and to answer any questions that they may possibly have. Mr. B [birth-dad] and Ms. M [birth-mom] agreed and stated that it would be perfect as court can sometimes become confusing."

Managing Medical Consent, Medication Administration, and Medical Appointments
Communication between involved parties (i.e., birth parents, foster parents, medical professionals, attorneys) about a child's medical needs and well-being is essential and is facilitated by CW staff. A foster child may be removed from a parent's care and placed in temporary protective custody with a foster parent; however, the birth parents still retain their parental rights and decision-making capacity regarding any health services extended to a foster child [74]. CW staff work with both the birth parents and foster parents to obtain and manage medical consent such that the foster child can receive medical care in the form of therapy, dental care, or other necessary services. CW staff also help supervise the day-to-day medical needs of foster children by establishing medication schedules as well as accompanying foster children to doctor's appointments. Here, CW staff's role as a mediator also helps alleviate any conflicts that may arise due to overlapping parenting roles.

[T3 Probability: 0.564] "Family Preservation Specialist met the caregiver, Yvette [foster parent], and Billy [child] at the doctor's office. Family Preservation Specialist observed Mrs. Olsen [birth-mom] holding Billy [child]. The doctor discussed how Billy [child] was doing and why there were being seen at the clinic. Family Preservation Specialist observed the doctor asking the caregiver questions. Family Preservation Specialist observed Billy [child] have his fists clenched while Mrs. Olsen [birth-mom] held him near the table. Family Preservation Specialist and the caregiver discussed meeting at her home after the appointment."

[T11 Probability: 0.68] "This worker [child welfare staff] attended the case transfer staffing with Ongoing Case Manager in the home of Ms. Brown [birth-mom]. Our group created a medication schedule with Family Preservation Services doing medication observation on Monday's at 7am and Thursday's at 11:00am. Paul [significant other] and Ms. Blar [relative] (maternal aunt to Billy [child]) will observe all other feedings and medication supervisions."

Coordinating Time, Travel, and Pickup Logistics for Visitations and Appointments
This theme is centered in the coordination and scheduling work that CW staff undertake in their

role as liaisons between birth parents, foster parents, and other professionals in child-welfare. CW staff is responsible for organizing and facilitating supervised visits between children and birth parents [429]. This involves scheduling the time and place of these visits with involved parties, managing scheduling conflicts, as well as transporting children (and parents, if necessary) to the location of supervised visits. These visits may occur at the child-welfare agency, a public space (e.g., public parks), or the parents' place of residence based on the presence and assessment of impending dangers in the household. CW staff also help parents get access to travel vouchers if they do not have the financial means or a vehicle for traveling long distances. While scheduling, conflicts within a family may also arise. For instance, birth parents might share a contentious relationship and may not want to work with each other. Here, CW staff must also work to ameliorate such concerns in order to promote congruence between involved parties [254]. This is necessary to ensure that progress is being made towards achieving permanency for the child. In addition, they also help coordinate travel to and from school for foster children who may not have access to a regular school bus route.

[T1 Probability: 0.620] "Family Services Counselor Nadine [child welfare staff] spoke to Mr. Smith [birth-dad] regarding Ms. Smith [birth-mom] and visitation with the children. Mr. Smith [birth-dad] stated that he will not allow Ms. Smith [birth-mom] in his home for visitation."

[T4 Probability: 0.949] "Family Preservation Services contacted Ms. Brow [birth-mom] to schedule a visit with her child and left a message. Family Preservation Services contacted Ms. Brow [birth-mom] and introduced themselves. Ms. Brow [birth-mom] stated that she needed to get off the phone and stated that she would call back. Family Preservation Services called Ms. Brow [birth-mom] to schedule a visit for the next week but Ms. Brow [birth-mom] did not answer the phone. Family Preservation Services left a message."

[T6 Probability: 0.598] Family Preservation Services arrived at Ms. Abel's [foster parent] residence to transport Bob [child] to a supervised visit. Ms. Abel [foster parent] did not report any issues with Bob [child] but she did need to assist with getting him into the vehicle. During the ride Bob [child] was crying but then fell asleep for most of the ride; he did not cause any issues or concerns.

Conducting Structured Assessments to Determine Risks and Progress

The child-welfare process is centered on assessing risk factors and helping parents develop protective capabilities to mediate these risks. CW staff, especially Family Preservation Services, works closely with parents through the parenting curriculum (keywords: parenting, chapter, session, curriculum) and other court-ordered services and score their progress on structured assessments. NFCAS (North Carolina Family Assessment Scale) and AAPI (Adult Adolescent Parenting Inventory) are examples of assessments especially used at the child-welfare agency by

Family Preservation [267, 285]. CW staff record a parent's level of engagement in these classes and whether they are exhibiting changes with respect to how they manage their child's behaviors. CW staff often refer to this as *perspective shift*, that is, whether the parent understands why their case was referred to CWS and if they are showing the willingness to make necessary changes in their lives. In addition, CW staff conduct home visits to assess safety concerns and any impending dangers within the household. This includes assessing the general cleanliness of the house, availability of food in the pantry and refrigerator, and both the children's and parents' hygiene. A sanitary and safe home help CW staff implement in-home services such that children do not have to be removed from their home and placed in foster care. Moreover, CW staff are also required to conduct and score quantitative assessments about risk factors associated with parents and children, safety within the household, parents' life experiences, and parenting skills. As illustrated by Saxena et al. [396], these quantitative structured assessments in CWS are now being used to develop algorithmic systems.

[T14 Probability: 0.82] "Family Preservation Services arrived at the home. Ms. Tazan [child welfare staff] was inside with the family. Family Preservation Services and Ms. Tazan did a walk through of the home. The home is not furnished and children don't have beds. Ms. Tazan had all the children clothing in black bags in the closet. Family Preservation Services did not observe any toys, books, etc. Family Preservation Services spoke to Ms. Tazan regarding the initial assessments that he needed to complete with her."

[T13 Probability: 0.61] "Family Preservation Services and Mr. Gibbs [bio-dad] watched the videos together and went through the power point presentation. It was apparent that Mr. Gibbs [bio-dad] had read the material as he was engaged in the discussion and talked about the examples in the book. First parenting assessment completed."

Facilitating Interactions Between Children and Parents During Supervised Visits

CW staff, especially Family Preservation Services, help facilitate interactions between children and birth parents and observe how these interactions are going during supervised visits every week [197]. Family Preservation Services use their expertise in parenting to work with the parents and help improve the quality of these interactions where the parents understand and attend to the needs of their children. Topic 8 (i.e., *interactions with infants*), however, emerged separately as compared to topic 9 (i.e., *interactions between siblings*) because an infant's interactions (e.g., eating well, sleeping, making eye contact, smiling, etc.) are essentially different from children's interactions (e.g., playing with siblings, playing with toys, running, etc.) and are noted distinctively by CW staff to assess well-being. For cases where multiple children are involved, CW staff also focus on ensuring that the parent(s) can manage their children's behaviors and establish some disciplinary boundaries. Family Preservation Services works with birth parents and advises them on how to manage interactions between siblings (e.g., fighting, yelling) and how to respond when being challenged by them [197]. Addressing these concerns helps ensure that time to reunification is reduced and the likelihood of case re-referral is lowered in the future.

[T8 Probability: 0.55] "Ms. Weldon [birth-mom] was excited to see the child as she kissed her and told her how much she missed the child. Ms. Weldon changed the child's clothes and did the child's hair while the child sat in her walker. Ms. Weldon continued to talk about her issues surrounding her case, Family Preservation Services had to remind Ms. Weldon to focus on her daughter instead of her situation she is in. Ms. Weldon praised the child for being able to wave and tried teaching the child how to clap her hands."

[T9 Probability: 0.84] "Ms. Tyndall [birth-mom] met Family Preservation Services outside to help bring in Ned [child], Phil [child], Pete [child], and Lawrence [child] into the family center. Upon entering the family room Ms. Tyndall who was holding Pete's hand and Lawrence in her arms told the boys that they have snacks in her bag for each of them. Ms. Tyndall sat on the floor and let Lawrence crawl and Pete explore in the visit room. Phil and Ned started playing with their little brothers and bringing them toys to play with."

Observing and Recording Concerns During Transportation

CW staff are trained to record any issues that may arise before, after, or during transportation [196]. Words in topic 10 (i.e., *children's behavior during transportation*) and topic 15 (i.e., *pre- and post-transportation concerns*) are associated with children's behavior and/or their interactions with Family Preservation Services while being transported for supervised visits. It helps CW staff assess how to best facilitate a supervised visit. For instance, if a child is anxious and agitated during the drive then CW staff might begin a supervised visit by engaging the child in activities that may help pacify them. This information is also shared and discussed with birth parents and foster parents to assess if there are any traumatic triggers that may be leading to emotional dysregulation. This also involves any concerns that might arise before or after the transportation. For instance, CW staff also ensure that children are dressed appropriately for the weather and look physically healthy.

[T10 Probability: 0.48] "This worker [CW staff] met the family at the Family Center. This worker transported Maya [child] and Jake [child] to their placement in [address]. Maya cried for roughly ten minutes for the car ride and then stopped and played with a stuffed animal. Coordinator Beth [CW staff] asked this worker to inform the caregiver that Maya had cried for roughly one hour during the visitation today. This worker did give this information to the caregiver upon arrival."

[T15 Probability: 0.409] "All three children were transported from maternal grandmother's home located at [address] and transported to McDonalds play land located at [address]. All three

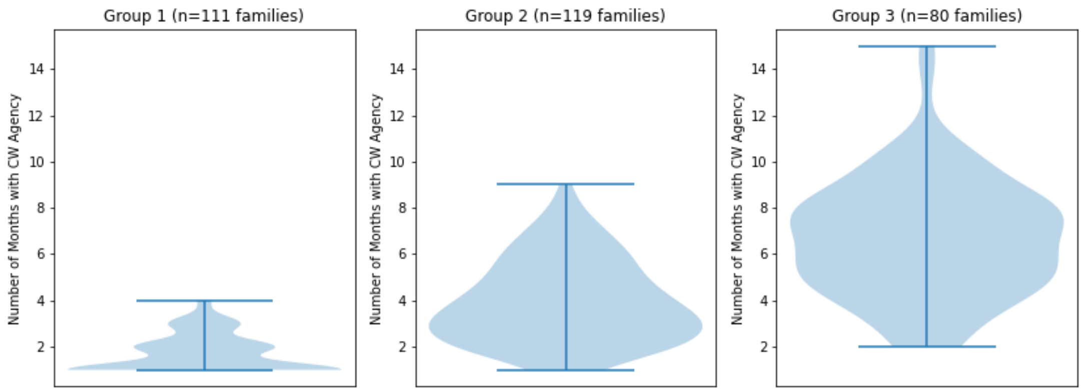


Figure 14: Distribution of the frequency of interactions with CW agency for groups G1 (low needs), G2 (medium needs), and G3 (high needs)

children were transported back to grandmother's. All three children were dressed appropriately for the weather and appeared free of injury, as they were able to walk, run and bend with ease"

6.5.2 Group analysis of topic popularity over time

We divided families into three groups based on their number of interactions with the child-welfare system. Figure 14 highlights the frequency of interactions that each group had with the agency and the number of months that families in each group worked with the agency. We notice a higher frequency of interactions at the onset of cases because CW staff must follow a 15-month timeline established by the Adoption and Safe Families Act (ASFA) where the State must proceed with the termination of parental rights if reunification has not been achieved in 15 months [12]. Therefore, CW staff work extensively with families from the onset of a case to gather relevant information and take necessary actions to expedite reunification. Below, we further discuss differences among the three groups based on the number of children, birth parents, and foster parents involved in each group. Descriptive characteristics about the three groups are available in Table 20.

Group 1 (G1) includes ***Low Needs Families*** that only had 1-10 interactions and generally involve cases of neglect (i.e., lack of childcare, lack of access to healthcare, lack of adequate food or clothing) where birth parents must make necessary changes within their household so as to provide a safe and nurturing environment for their children. As depicted in Table 20 (see Group 1), the majority of the children (n=68, 62%) were not removed from the care of birth parents, and instead, in-home services were provided to these families. Majority of these families were also single-parent households (n=78, 70%) and only involved one foster child (n=77, 69%).

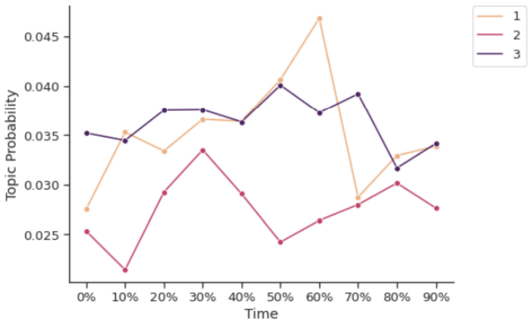
Group 2 (G2) includes ***Medium Needs Families*** that had 11-40 interactions with the child-welfare system. This group includes cases where most children were removed from the care of birth parents and placed with foster parents (or relatives) due to safety concerns within

Persona	Number of personas	Group 1 (1 - 10 interactions)	Group 2 (11 - 40 interactions)	Group 3 (40+ interactions)
Children	0	0	0	0
	1	77	35	20
	2	27	27	24
	3	3	24	10
	4	2	19	11
	5	2	10	10
	6	0	3	3
	7	0	1	2
Birth Parents	0	2	0	0
	1	78	70	48
	2	31	49	32
Foster Parents	0	68	32	19
	1	40	35	50
	2	3	42	6
	3	0	8	4
	4	0	1	1
	5	0	1	0

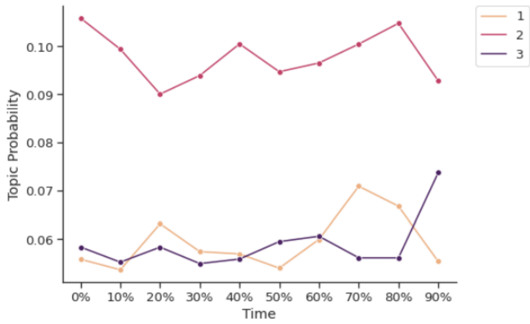
Table 20: Descriptive characteristics for the three groups based on the number of interactions with CWS. The table shows the total number of cases for each group $\{1, 2, 3\}$ having x members from the persona list y where $x \in \{[0,7],[0,2],[0,5]\} foreachy \in \{'Children', 'BirthParents', 'FosterParents'\} respectively. Zero$ value for foster parents means that the child was not removed and in-home services were provided to families.

the household. This is generally considered short-term foster care, where birth parents must complete parenting classes and court-ordered services (e.g., drug and alcohol services, domestic violence classes, etc.) and demonstrate stability within the household to achieve reunification with their children. Here, children are generally placed in short-term placements before long-term caregivers can be found. These cases generally involve multiple children placed with different foster parents since it is hard to find foster homes that can provide for all the children involved in a case. As depicted in Table 20, this group had 35 families with only one child, 27 families with two children, 24 families with three children, and so on. Group 2 also has 32 families where children were not removed, 35 families where children were placed with one foster parent, 42 families where two foster parents were involved, and 8 families where children were split between and placed with 3 foster parents.

Group 3 (G3) includes ***High Needs Families*** that had 40+ interactions with the child-welfare system and includes cases of more severe abuse and/or neglect. This group is generally considered long-term foster care, where children are placed with long-term caregivers who are trained and certified to care for high-needs children. Foster parents in this group may also be the next of kin since CW staff prioritize placing children with relatives. Prior work has established that children are more likely to achieve emotional and cognitive well-being when placed within the family [142, 246]. However, if children are placed in kinship care, the caregivers assume the role of foster parents (as active caretakers) and are no longer classified as a parent's support system (passive and occasional caretakers). CW staff work closely with birth parent(s) in parenting classes and other court-ordered services as well as help them find stable employment and other



(a) T2: Securing Resources



(b) T5: Establishing roles between parties

Figure 15: Time trends for topics focusing on helping families secure resources and role coordination

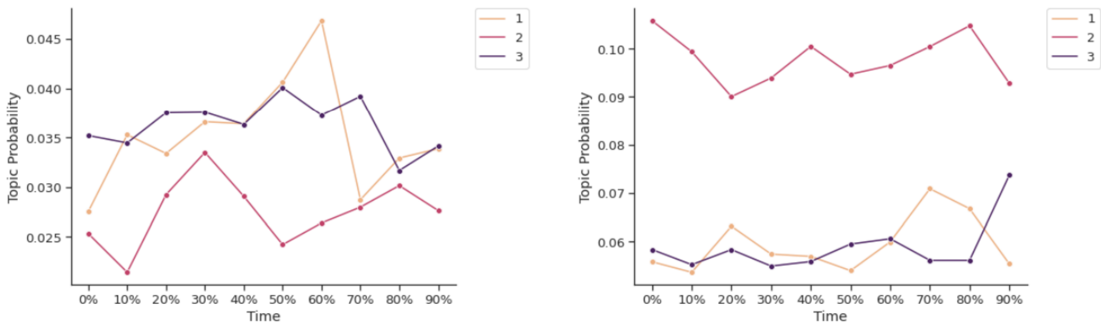
resources necessary to meet the needs of their children and eventually achieve reunification. As depicted in Table 20, this group consists of 19 families where children were not removed from the care of birth parents and in-home services were provided, 50 families where children were placed with one foster parent, 6 families where children were placed with two foster parents and so on. Moreover, similar to Group 2, families in this group consist of multiple children, which adds to the complexity of these cases.

Next, we discuss trends in topic popularity over time for the top four themes from Section 5.1 for each of the three groups. Following Antoniak et al. [26], we divided each of the casenotes into ten equal sections. As casenotes follow a formulaic sequence of events, we were able to divide the texts into ten chronologically arranged normalized sections (i.e., **Life of a Case**). This allowed us to track casenotes of varying lengths which begin and end at different times. Therefore, as depicted in Figures 4-8, 10% on the x-axis would point towards the events happening in 0-10% of the life of a case; 50% on the x-axis would point towards events happening in 40%-50% of the life of a case.

Helping Families Secure Resources and Navigate Bureaucratic Processes

Securing resources (topic 2) is a significant topic for both G1 and G3 families. For G1, we observe an upward trend through the life of a case as shown in Figure 15(a). CW staff work with birth parents from the onset of a case to acquire these resources to achieve a safe living environment. For G3, CW staff continually work with parents to ensure necessary changes are being made in the household from both an economic and behavioral perspective. However, this topic is less significant for G2 because the more dominant concerns are related to managing logistics (since G2 families involve multiple foster children). Specifically for G2 families (see Figure 15(b)), CW staff work on managing roles and expectations between birth parents and multiple foster parents as conflicts arise due to overlapping parental roles in managing the needs of multiple foster children.

In addition, court proceedings are a significant part of the child-welfare process, and this topic emerges as significant at key decision points of the life of a case. As depicted in Figure 16(a), for G1 and G2, we observe upticks in trends towards the beginning of the case as well as a rise in trends towards the end. This matches our expectations since critical court hearings occur at the onset and towards the closing of a case for these groups. For G3, we observe several upticks in trends (spread out evenly) since the more severe cases of neglect/abuse require more court appearances in terms of reunification hearings, transfer of guardianship, or termination of parental rights. As depicted in Figure 16(b), we anticipate that the COVID pandemic may have also influenced the trends for these groups. During the pandemic, resources were directed towards cases that most needed them. Court hearings, parenting classes, and services were rescheduled and/or postponed for several cases in G1 and G2. Our collaborators at the agency shared that virtual court hearings, virtual classes, and virtual visitations were still being conducted for high needs cases, i.e. – most families in group G3.



(a) T2: Securing Resources (b) T5: Establishing roles between parties

Figure 16: Time trends for topics focusing on virtual interactions and court proceedings.

Managing Medical Consent, Medication Administration, and Medical Appointments
CW staff help manage medical consent between caregivers (topic 3) and help manage medication schedules (topic 11) for foster children. Topic 3 consistently emerges for G1 because CW staff discuss medical consent with birth parents early on and take children to necessary medical appointments (e.g., neglected dental health). This topic is also more significant for G3 (as compared to G2) because these are cases where more significant abuse/neglect may have occurred, and consequently, children are enrolled in services (e.g., individual therapy) to address their needs and the underlying trauma. We also anticipated this topic to emerge as more significant for G2 since medical consent needs to be managed between birth parents and foster parents and can lead to conflict. However, as depicted in Figure 17(b), managing medication schedules takes precedence for G2 because CW staff must continually ensure foster parents (especially short-term caregivers in G2) understand the medical needs of children and are giving them their

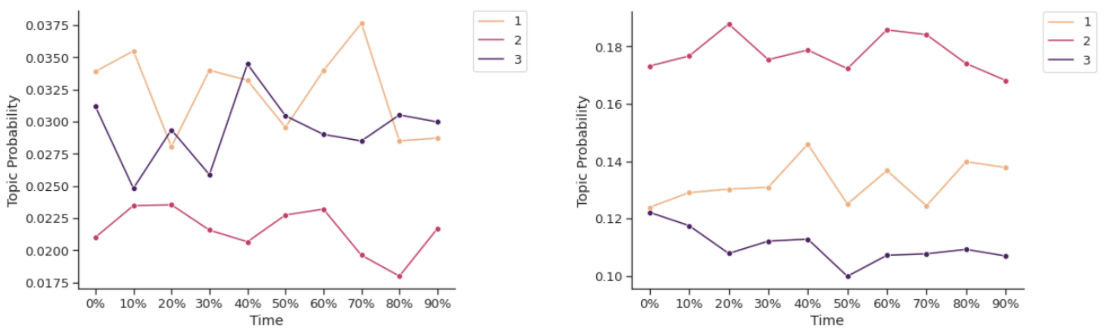
medications per set schedule. This topic is less significant for G1 because most children are placed with birth parents and less significant for G3 since long-term caregivers are trained and certified in caring for high-needs foster children.

Coordinating Time, Travel, and Pickup Logistics for Visitations and Appointments

Scheduling issues for supervised visits (topic 1) occur less frequently for groups G1 and G2; however, they are more common for G3 families. G3 includes cases of more severe neglect and/or abuse where intensive care is required in terms of medical appointments and supervised visits. For G3 cases, there may also be a no-contact order in place where parents can only see their children under proper supervision of family preservation caseworkers. However, as depicted by topic 6 (see Figure 18(b)), CW staff must also coordinate time, travel, and pickup logistics for court-ordered services, court hearings, visitations, and medical appointments. This topic emerges as significant for G2 at regular intervals since there may be multiple children involved in the case (and placed with different foster parents), and CW staff must coordinate these details among all parties. We observe two upticks in the trend for G1 and anticipate these to be medical appointments (general check-ups) conducted to assess children's well-being before case closure.

Conducting Structured Assessments to Determine Risks and Progress

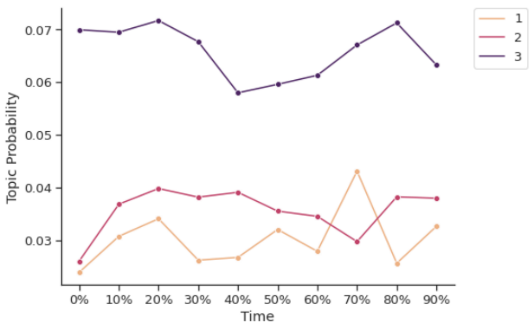
CW staff observe how parents respond to parenting classes and score their progress on quantitative structured assessments. This helps them assess the likelihood of the parents' employing these skills and strategies when addressing their children's needs and managing their behaviors. Topic 13 emerges consistently for both G1 and G3 (with upticks in trends spread out evenly) because parenting skills play an important role in achieving expedited reunification (as is the case with G1) but also in more severe cases of abuse/neglect as a means to assess if the parent is capable of meeting the needs of their children. We observe a similar trend for G2; however, the topic is less significant since more attention is being paid to managing logistics around multiple children, caregivers, and birth parents. CW staff also conduct home visits and score quantitative



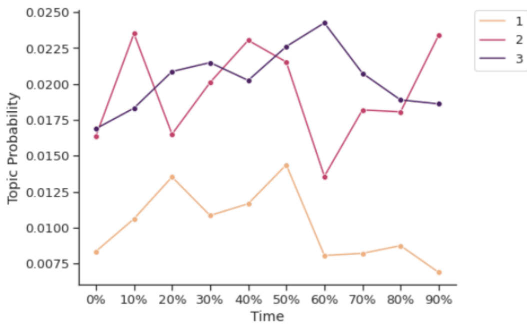
(a) T3: Managing Medical Consent

(b) T11: Managing Medication Schedules

Figure 17: Time trends for topics focusing on medical consent and schedules management.



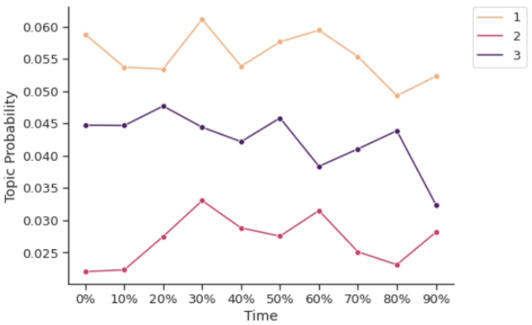
(a) T1: Scheduling Conflicts of Supervised Visitations



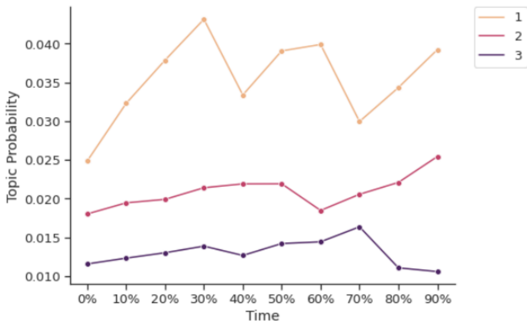
(b) T6: Managing time, travel, and pickup logistics

Figure 18: Time trends for topics focusing on scheduling conflicts and time management.

safety assessments from the onset of a case to assess if the home provides safe living conditions for children. Topic 14 emerges as being significant for G1, with several upticks in trends spread out evenly. For cases in G1, if the home provides a healthy and clean environment, then CW staff can provide in-home services to the families such that children's removal is not necessary. This topic did not emerge as significant for G2 and G3 (see Figure 19(b)) since the children are mostly placed with foster parents.



(a) T13: Progress in Parenting Classes



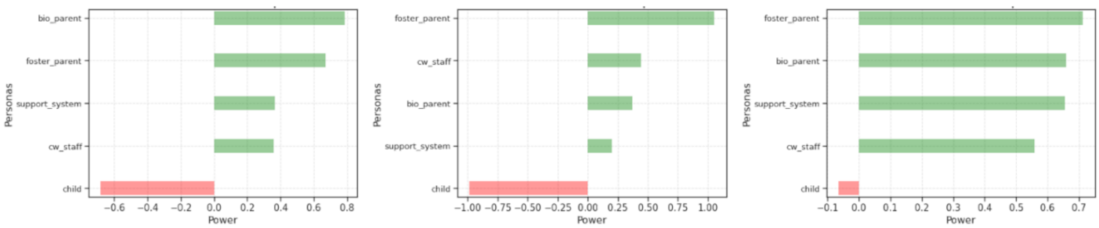
(b) T14: Safety Assessments during Home Visits

Figure 19: Time trends for topics focusing on parenting classes and safety assessments.

6.5.3 Power analysis of personas

Complex sociotechnical domains such as the child-welfare system consist of underlying power structures where some parties hold the majority of the power, exercise agency, and exert control over other parties. Power relationships with respect to CWS have been studied extensively in sociology literature [250, 234, 381, 77, 373], however, computational text analysis of caseworkers' narratives to uncover such underlying power structures is an understudied topic.

We conducted power analysis of casenotes and focused on five key personas which are actively involved at the front-end of child-welfare cases, namely, CW staff, birth parents, foster parents, birth parents' support system, and the foster child. Results of this analysis are depicted in Figure 20 which shows power scores for each persona across the 3 groups, and Figure 21 which



(a) Group 1 (b) Group 2 (c) Group 3

Figure 20: Power scores for each persona across the 3 groups.

demonstrates the estimated power of personas over other personas. Below, we first interpret our results for each of the three groups and then compare our findings across the three groups.

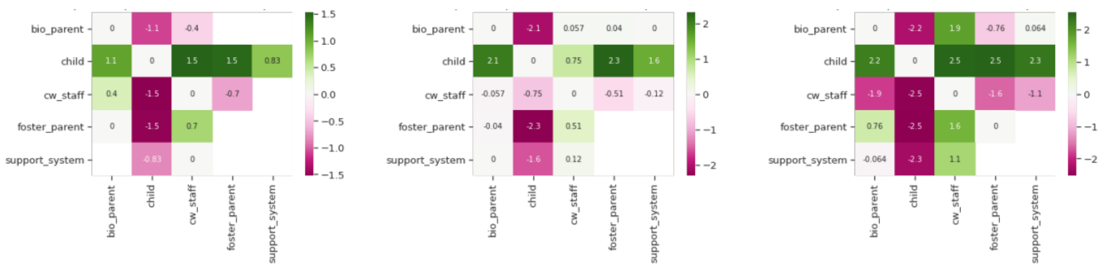
Group 1 (1-10 interactions with CWS): As illustrated in Figure 20 (a), birth parents have the most power for this group. As previously noted, cases in this group generally involve neglect (lack of childcare, lack of adequate food/clothing) and require birth parents to make adequate changes within their household to be able to provide a safe living environment for children. CW staff and the parents' support system are able to assist them but the parents must exercise their agency and demonstrate the necessary changes in their lives such that all agreed-upon court conditions are met and CW staff can recommend reunification at the court hearing. Foster parents exhibit the most power after the birth parents since they actively care for a child daily and share the child's needs and behaviors with CW staff (which informs case planning and service delivery). Finally, as expected, foster children exhibit the least power (negative score) among all the personas.

Group 2 (11-40 interactions with CWS): As illustrated in Figure 20 (b), foster parents exhibit the most power for this group. Foster parents are the primary caretakers for this group and are actively involved in case planning. Prior studies conducted with CWS in Wisconsin showed that foster parents exercise the most agency with respect to how the needs and risks associated with foster children are assessed [395] and how much they are compensated by the state [396]. CW staff exhibit the most power after foster parents since they manage all the logistics associated with foster placements, such as finding resources for children, managing medical consent and medication schedules, scheduling visits, etc. The primary goal of CW staff is to ensure that foster parents are fully supported, and placement is not disrupted. Moving between different foster homes adversely affects foster children who develop emotional and behavioral problems and are unable to form meaningful relationships [56]. Birth parents exhibit lower power scores as compared to foster parents and CW staff because they may feel disempowered by the process where their kids are removed and placed with multiple different foster parents. As previously noted, there may also be a lack of trust between birth parents and foster parents because of a lack of interpersonal relationships and ambiguity due to overlapping roles.

Group 3 (40+ interactions with CWS): As illustrated in Figure 20 (c), foster parents exhibit the most power for this group. Group 3 involves cases where severe abuse and/or neglect has occurred and requires trained and certified caregivers to meet these needs. There is a dearth of good foster homes in CWS where foster parents are trained in caring for high-needs kids [130, 113], and therefore, CW staff must prioritize maintaining and supporting these placements. As previously noted, foster parents in this group may also be next of kin. For either case, there is a stronger interpersonal relationship between the foster parents and birth parents, which would explain birth parents exhibiting the most power after foster parents. The higher magnitude of power scores across personas also provides some evidence of an integrated approach towards family reunification adopted by CW staff where all personas are involved in child care and provide caregiver support to each other. Birth parents in this group must also complete mandatory court-ordered parenting classes and services (domestic violence, AODA services, etc.), and consequently, reunification is contingent upon them fulfilling these requirements.

Comparing across Groups: CW staff act in a supporting role for groups G1 and G3 and exercise the least amount of agency (except for the child) compared to other personas. However, they take a lead role in G2 with respect to handling logistics and trying to address systemic barriers so that expedited reunification can be achieved for families. The agency has specialized meetings in place, called Permanency Consultations, designed to promote collaborative decision-making and expedite reunification [396]. As previously noted, if reunification does not occur within 15-months of a family being referred to CWS, the agency must begin exploring alternate placement options, that is, long-term foster care. CW staff's main objective is to prevent G2 cases from transitioning into G3 since long-term foster care leads to poor well-being outcomes for foster children. Moreover, finding good foster placements that can care for high-needs kids is hard because of a lack of good foster homes in the system. This is also why CW staff maintain a lower power profile with respect to foster parents. It is imperative that CW staff maintain good working relationships with both short-term and long-term foster parents so that there are homes to place children in need of care. Finally, CW staff (when acting in a supporting role for G1 and G3) also exercise less power as compared to birth parents' support system. They try to get the support system involved in the family's life such that birth parents have additional caregiver support and trusting relationships that they can rely on during times of crisis. This lowers the likelihood that the case would be re-referred to CWS due to instances of neglect (lack of childcare, lack of adequate food/clothing).

Comparing across Personas: Figure 21 depicts a heatmap of power relationships between pairs of personas. As highlighted in prior work [26], it is possible for a persona to have a lower (or higher) cumulative power score but a higher (or lower) power score when only their



(a) Group 1 (b) Group 2 (c) Group 3

Figure 21: Estimated power of personas (rows) on other personas (columns).

interactions with another persona are measured. Interestingly, foster children who exhibit the least cumulative power appear to exercise more power over all individual personas. This could provide evidence for why CW staff work closely with birth parents in parenting classes so that parents are able to manage the behaviors of their children and regain agency in setting healthy disciplinary boundaries. Similarly, a successful foster placement requires that foster parents are able to manage the behaviors and needs of children. Inability to manage these needs/behaviors leads to placement disruptions where foster parents feel disempowered and put in their notice to end a placement; a significant ongoing concern in CWS [89]. Surprisingly, CW staff appear to exercise the least amount of power (across all personas). Even for group G2, where CW staff assume a lead role, they appear to be sharing power across all individual personas.

6.6 Discussion

6.6.1 Unpacking Invisible Patterns of Street-Level Discretionary Work (RQ1)

Our topic modeling results highlight much of the hereto hidden, street-level discretionary work that caseworkers undertake while helping families (e.g., managing medication schedules, conducting quantitative assessments, establishing caregiving roles, navigating court proceedings etc.). These casenotes are collectively curated by CW staff involved at the front-end of case planning and offer a holistic picture for collaborative decision-making [173, 396]. What makes our results really important is that they revealed patterns of street-level work that were not even uncovered during an extensive ethnography at the same agency comprising of observations of collaborative meetings and interviews with caseworkers to understand their daily work practices and perspectives on algorithmic decision-making tools [396]. For instance, caseworkers help manage medical consent, medication administration (topics 3 and 11), as well as accompany clients to medical appointments and court hearings (topics 3, 11, 12). These topics were not highlighted during the ethnography even though they are collaboratively discussed in the casenotes. This suggests that qualitative deconstruction of work practices may not reveal all the nuances of invisible labor and in fact, demand complementary methodological lenses. By extension, we believe that this advocates for a need for both a qualitative and quantitative critique of sociotechnical systems. Critical computing has become popular at SIGCHI in recent years but remains bounded mostly by qualitative investigations [116, 265, 145]. Moreover, we also examined the most recent job descriptions of the child-welfare caseworker positions [484, 390] at two CW agencies in the region and found that these patterns of work were not formally outlined in them either. The job descriptions state that caseworkers must complete documentation for court work as mandated by state law, but as revealed by topic 12 (and exemplar sentences), caseworkers are accompanying parents to court in order to assist them through the court proceedings. On the other hand, both job descriptions reveal that caseworkers must "conduct and document safety assessments." However, as illustrated by Saxena et al. [396], data from these quantitative assessments are now being used to develop algorithmic risk assessments. As outlined by prior researchers [491, 212, 397, 396], these assessments and the administrative data used to build them are fundamentally biased. In contrast, our results point out that quantitative analysis of caseworker narratives can support strength-based, holistic assessments [450] without being bogged down in the quagmire of biased algorithmic risk assessments.

In sum, computational text analysis of casenotes helped uncover patterns of street-level discretion -ary work conducted by caseworkers that is otherwise hidden even from the findings of an ethnography or job descriptions of CW positions. This suggests two broader implications for SIGCHI research - a need for computational critique as well as a motivation to shift from biased risk assessments to more holistic strength-based assessments [38, 495, 450].

6.6.2 Understanding Constraints on Child-Welfare Practice (RQ2)

Our results also highlight how constraints affect the work (discretionary or otherwise) that caseworkers need to do in order to provide better outcomes for children. We find that all children in CWS are not treated the same as some have higher needs than others (hence, our groups - G1, G2, G3). This differential need is affected by constraints (e.g., resource, bureaucratic, temporal, algorithmic, or other) and has been noted in prior work [89, 397]. SIGCHI has become increasingly interested in the nature of work, especially when mediated/constrained by technology and algorithms [232, 20, 137]. As illustrated in our results in Section 5.2, different patterns (topics) of work are highlighted at different times through the life of a case and illustrate different interventions for different groups of families. For instance, as depicted in Figure 4(a), CW staff help secures essential resources for families. However, for G1 (less need), this generally takes the form of economic resources such as employment, food, clothing, and preventive services such as parenting classes. This requires CW staff to reach out to local parent support groups and family resource centers to connect clients to such services. Similarly, G2 (medium need) requires CW staff to find court-ordered services for their clients such as domestic violence classes, AODA (alcohol and other drug abuse) classes, therapy, etc. This requires CW staff to reach out to each of these service providers and find room for their clients. Much of this disparately available information can be curated into a system and made more accessible to CW staff. For instance, Yan et al. recently conducted an exploratory study to assess which systemic factors were associated with the services offered to clients. They offer direct implications for sociotechnical systems design in child welfare [488, 345].

Similarly, as Figure 6(b) illustrates, CW staff spend a significant amount of time through the life of a case for Group 2 in making sure that foster parents are actively following the medication schedules for foster children. As previously noted (see Table 9), Group 2 generally involved multiple children placed with multiple different foster parents. CW staff must call foster parents (and do this for all their G2 cases) and make sure that the schedules are being followed. On the other hand, families in G3 have significant needs and require more care. Here, CW staff develop individualized trauma-responsive services (e.g., cognitive behavioral therapy, cognitive processing therapy etc.) for clients through specialized consultation sessions with medical professionals [450]. As illustrated by Saxena et al. [396], the CWS agency designed a holistic, strength-based algorithm centered in trauma-informed care to improve collaborative decision-making for high-needs families.

Summarizing all of these, an important implication arises for algorithms in CWS. Much of the current work has focused on (a) developing more sophisticated machine learning based risk assessment algorithms to improve the status-quo [106, 137] or (b) understanding breakpoints, biases, and ways in which caseworkers make decisions from currently implemented algorithmic constraints [103, 105, 395, 396]. What is left unexplored at the current moment is whether (a) we need to be developing machine learning applications in CWS in the first place as well as (b) if simpler, non-algorithmic technological applications can help in removing some existing constraints that caseworkers work around (e.g., checking and notifying medication schedules). This aligns with recent work in worker-centered design in SIGCHI where caseworkers at a job placement center were averse to the introduction of a risk assessment algorithm (for profiling individuals) and instead, asked for sociotechnical systems that would help mitigate organizational constraints and help streamline bureaucratic processes [232]. Caseworkers also perceived algorithms to be useful if they could support caseworkers' practices in strengthening cases that required additional resources [20]. Similarly, caseworkers in child-welfare found utility in a simple decision-tools that help guide their decision-making processes through a trauma-informed care framework instead of predicting an outcome of interest [396].

6.6.3 Uncovering Latent Power Relationships in Child Welfare Systems (RQ3)

Limitations. We note important limitations of this computational power analysis approach that other researchers must consider before adopting this method. First, we acknowledge that this approach cannot uncover deep, structural power issues that are systemically embedded within CWS. We direct the readers to the works of J. Khadijah Abdurahman [5] and Victoria A. Copeland [121] who have deeply studied power hierarchies in CWS and illustrated that caseworkers both exercised power and experienced power asymmetries from supervisors, agency policies, and the court system. Second, since these casenotes are written by caseworkers, they do not capture families' firsthand accounts of their interactions with the system. We considered this methodology to be appropriate for this study for two reasons: 1) casenotes in our dataset are primarily written by the family preservation team whose goal is to achieve reunification for children and birth parents. That is, the team's objectives are aligned with those of the parents and centered in helping them prepare and achieve a favorable decision in court, and 2) casenotes are collaboratively written by case management and family preservation workers which adds a layer of accountability in regard to observations being recording in these casenotes. As illustrated by our recent study [406], this analysis would be inappropriate to study the casenotes of initial assessment/investigative caseworkers who exercised more power over families in regard to data being collected about parents and how critical decisions were made. However, such quantitative analyses help illustrate these complexities within child welfare where different teams assume different roles.

We draw from existing SIGCHI scholarship on sociotechnical power rooted in feminist HCI [43] and worker-centered design [176, 266] to unpack our findings. From this theoretical scaffolding, we further distinguish between two kinds of power - first, the structural power that is systemic within any complex sociotechnical system as well as the power that exists as a result of day-to-day work relationships. We acknowledge that computational power analysis cannot structural power issues but rather surface the power complexities that arise from daily work relationships. Further, we draw from Starhawk's [435] and Berger's [52] disentanglement of these relationships between power-over, power-with, power-to, and power-within relationships. Largely, we find that in addition to the expected power-over relationships that one might expect to find within CWS stakeholders, there also exist some other kinds of unexpected power relationships that complicate some popular media narratives on CWS [144, 342].

Our results find some evidence to support that CW staff exercised a more collaborative, power-with role (among adults) when they played a supporting role for groups G1 and G3 and only assumed more power-over relationships (in the case of group G2) when the primary goal was to expedite reunification such that cases did not transition into long-term foster care (i.e., group G3). This also provides some evidence for the efforts made within CWS from both a policy and practice standpoint to transition towards a "Families as Partners" model [375] where parents are supposed to act as equal partners in the case planning process and have agency in the decision-making process. As previously noted, critical decision-making power in regard to reunification and termination of parental rights sits with the legal parties (i.e., - district attorneys, judges) [90, 158] and often frustrates CW staff who are working with birth parents in their efforts to achieve reunification. These tensions between the court system and CW staff are well-documented in social work literature [90, 154, 158]. However, as previously noted, this is not to say that CWS is not riddled with deep systemic issues that disproportionately impact families of color [251, 369]. On the contrary, our hope with this analysis is to illustrate the daily, working power complexities within this domain of which CW staff is only a piece of the complex puzzle comprising several parties with conflicting interests. For instance, a case typically involves four attorneys - one for each parent, the agency, and the child(ren) where each of these attorneys advocates for the individual rights of their clients [461].

Different power relationships also help uncover the differences in different families (i.e., - the three groups) involved in child-welfare and highlight the need to support both the families and CW staff in different capacities. For instance, CW staff is involved in a supporting, power-to relationship in both G1 (less need) and G3 (most need) groups, where they help secure resources for families. However, for G1, this translates into finding material resources (adequate food/clothing, childcare). Whereas, for G3, CW staff must find ongoing professional services. On the other hand, G2 cases require that CW staff have a more power-over role in managing the needs of multiple foster placements. Moreover, different power relationships also directly impact how data is collected about children, how their needs are assessed, and have serious implications for algorithmic decision-making. For instance, our prior ethnographic study conducted at this agency [396] revealed that foster parents exercised significant control over how children's risks and needs were quantitatively scored, which impacted their compensation rates and the services offered to children [396]. This in turn leads to the manipulation of data and the algorithm such that foster parents received higher compensations. In prior work conducted in CWS [302], these power imbalances also generated perverse incentive structures for algorithmic decision-making based on mental health needs. Medical professionals exercise more power than other involved parties in regard to the quantitative scoring of the needs of children. Consequently, they are paid when needs are detected and interventions offered. That is, there were clear professional and financial incentives that encouraged the detection of needs and led to the manipulation of the algorithm [302]. On the other hand, CW staff were trained to conduct mental health assessments; however, the detection of needs invariably led to more work on their part because it required them to find and secure services for children. That is, the short-term incentive for CW staff was to not detect needs so as to limit the amount of work [302].

In sum, our analysis unpacks different kinds of work power relationships (e.g., power-over, power-to [435, 52] etc.) between CWS stakeholders depending on the context and aligns well with prior social work literature on power relationships in CWS [77]. These results imply that human-centered algorithm design in child welfare needs to understand and consider these power relationships to support the primary objective of providing positive outcomes for foster children.

6.7 Limitations

Our study only used casenotes from one CW agency in a US midwestern state, so our findings may not be generalizable to other states where different policies and regulations impact daily processes and decisions. Nevertheless, this study offers the methodology to perform computational narrative analysis in other CWS contexts and can help generate similar insights. Moreover, although all caseworkers are trained to record interactions and decisions in casenotes, their writing styles may vary. For instance, some caseworkers may not write details about characteristics captured in assessments (e.g., living conditions when scoring home-safety assessments). Moreover, it is imperative to note that casenotes may contain more contextual information, however, they are still based in workers' impression of family circumstances and could potentially introduce biases into decision-making [406].

6.8 Conclusion

This study offers the first computational inspection of casenotes and introduces them to the SIGCHI community as a critical data source for studying complex sociotechnical systems. We applied topic modeling with LDA on collaboratively curated case narratives by CW staff. The casenotes are highly contextual for every family yet carry similarities concerning the processes families follow in child-welfare, including critical decisions made and personas involved at the front end of case planning. Our results show that on-the-ground caseworkers engaged in several patterns of hidden labor that were not uncovered in prior ethnographic work or depicted in job descriptions. Analysis of different cases (based on the number of interactions) revealed that CW staff need to support families differently and further helped contextualize the meaning of topics. For instance, CW staff acquired different resources for G1 families (less need) vs. G3 families (high need). Finally, power analysis of casenotes revealed the power asymmetries within CWS that contest the dominant societal narrative that caseworkers exercise significant autonomy and are responsible for the removal of children. The power asymmetries have implications for algorithmic decision-making as these latent power structures directly impact algorithmic decisions and help us understand which personas exercised more agency in decision-making.

CHAPTER 7: RETHINKING "RISK" IN ALGORITHMIC SYSTEMS THROUGH A COMPUTATIONAL NARRATIVE ANALYSIS OF CASENOTES IN THE CHILD WELFARE SYSTEM

ABSTRACT: Risk assessment algorithms are being adopted by public sector agencies to make high-stakes decisions about human lives. Algorithms model "risk" based on individual client characteristics to identify clients most in need. However, this understanding of risk is primarily based on easily quantifiable risk factors that present an incomplete and biased perspective of clients. We conducted computational narrative analysis of child-welfare casenotes and draw attention to deeper systemic risk factors that are hard to quantify but directly impact families and street-level decision-making. We found that beyond individual risk factors, the system itself poses a significant amount of risk where parents are over-surveilled by caseworkers and lack agency in decision-making. We also problematize the notion of risk as a static construct by highlighting the temporality and mediating effects of different risk, protective, systemic, and procedural factors. Finally, we draw caution against using casenotes in NLP-based systems by unpacking their limitations and biases embedded within them.

7.1 Introduction

Public sector agencies such as the child-welfare, criminal justice, unemployment services, and public education have experienced a fundamental economic shift over the last two decades in regard to how governance practices are carried out and how clients are "assisted" by street-level civil servants. Economic principles centered in cost reduction, efficiency, and productivity are now being applied to public services where several sectors have experienced privatization with a core focus on optimization and austerity [163, 377]. "Risk" has been one of the core organizing principles in this economic shift in governance where administrative data accumulated by government agencies about citizens purportedly allows them to preemptively recognize clients in the riskiest circumstances [83, 22], i.e. – clients most in need of public assistance, clients most likely to harm others or engage in unlawful behavior, and clients who pose the most risk to governmental apparatus in terms of resources used. Government agencies are employing algorithmic systems to predict outcomes such as the risk of recidivism [212, 152], risk of child maltreatment [397, 99], risk of long-term unemployment [232, 20], risk of extended homelessness [163], among others. This preemptive recognition and mitigation of "risk" through predictive models is a defining characteristic of what scholars have called *digital era governance* [81] or *digital welfare states* [18]. However, there is a mismatch between how risk is quantified *empirically* [397] based on administrative data versus how it is understood *theoretically* [34] in these domains.

Empirical risk predictions hold the promise of providing consistent, cost-effective, and objective decisions, and bringing a new data-driven perspective to government agencies where data would bear the promise of future bureaucratic efficiencies [232]; however, audits of these systems have revealed that they instead achieve worse outcomes [396], embed human biases present in administrative data [152, 444], appear nonsensical to workers [256, 396], and exacerbate existing racial biases [99]. Consequently, researchers studying fairness, accountability, and transparency in algorithmic systems have developed technical definitions of "fairness" and "bias" and formulated them into systems design to achieve equitable outcomes. These approaches may lead to the development of systems that are mathematically fair, however, they still continue to focus on a narrow understanding of "risk" as derived from empirical administrative data while drawing attention (and resources) away from the complexities in the decision-making ecosystem and ecological nature of risk that families and caseworkers experience on the street-level [397]. To address these gaps in computational research, SIGCHI scholars have begun to examine how quantitative methods can be used to uncover complexities and latent patterns within sociotechnical systems [87, 207, 404]. In this study, we accept this call, and instead of quantifying risks using administrative data, we focus on uncovering human interactions between caseworkers, families, and other child-welfare stakeholders to understand the multiplicity and temporality of risk factors that arise in child-welfare cases through the lens of computational narrative analysis of casenotes, i.e. - rethinking "risks" as they occur on the street-level and recorded in caseworkers' narratives as opposed to what is quantified in the administrative data. We used the socioecological model of health and development that has been recently used to study risk, protective, systemic, and procedural factors associated with child maltreatment [34] as the theoretical lens for grounding our quantitative analysis. In this study, we pose the following research questions:

- **RQ1:** Which factors in the child-welfare ecosystem directly impact street-level decision-making and family well-being?
- **RQ2:** How do these critical factors interact and what is the impact of this interplay on decision-making?
- **RQ3:** How do these critical factors fluctuate and mediate each other throughout the life of child-welfare cases?

Abebe et al. [6] argue that computational research has meaningful roles to play in addressing social problems by highlighting deeper patterns of injustice and inequality. In this regard, they formulate roles that computing can play and define ***computing as rebuttal*** when it illuminates the boundaries of what is technically feasible and define ***computing as synecdoche*** when it makes long-standing social problems newly salient in the public eye. In this study, we assume these roles and make the following contributions:

- We use computational narrative analysis [404, 26] to uncover the different risk, protective, systemic, and procedural factors that impact decision-making and draw these connections to the theoretical understanding of risk in sociology and child-welfare literature [34].
- We showcase how these different factors interact with each other where systemic and procedural factors can amplify the risks that families experience in the child-welfare system.
- We highlight how these factors change over time and can compound uncertainty in decision-making due to a lack of clarity about the trajectory of cases. We further complicate the use of predictive risk models (PRMs) because no temporal point estimate of risk offers a complete picture of family well-being.
- We surface the limitations and biases embedded within child-welfare casenotes and draw caution against using these narratives for downstream tasks (e.g., predicting the risk of child maltreatment) in NLP-based systems. Alternately, an upstream approach, as adopted by this study can help uncover dynamic and transitory signals embedded within the sociotechnical practices of decision-making.

This study responds to calls within SIGCHI research to investigate complex sociotechnical systems from both a qualitative and quantitative lens to understand the opportunities and limitations of computational research towards highlighting social problems and addressing injustices [28, 330, 6].

7.2 Related Work

In this section, we first discuss recent research within SIGCHI conducted in the public sector followed by research conducted on computational text analysis of sociotechnical systems.

7.2.1 Public Sector Research within SIGCHI

The SIGCHI community has a long-standing history of conducting research in the public sector and designing sociotechnical systems that empower public sector workers [232, 20, 400, 257] and affected communities [433, 72, 432, 38]. Most relevant to this study, SIGCHI research spans across digital civics [150, 299], digital governance [81, 295], and algorithmic governance [256, 386] where HCI scholars have studied issues of citizen engagement in the public space [274, 150], citizen activism [317], empowerment of affected communities [72, 363], centering worker well-being in gig work [493, 453, 452, 394], and engaging in action research with community partners [120]. As government agencies experience renewed neoliberal market forces centered in austerity and privatization [290, 490], digital governance platforms are being developed through public-private partnerships [125] or by contracting tech startups [238, 371]. Here, HCI researchers have also questioned the forms and limitations of participatory design in the public sector that is increasingly experiencing the deployment of technologies developed through public-private partnerships for the administration of smart cities [299, 150]. In addition, HCI scholars have also brought into question the core function of government services that were designed to act as "caring platforms" by serving the public good but are now being operated based on business models of private corporations [295]. That is, public services designed to "serve" the people should not be optimized or reduced to the performance metrics of the business world. To oppose this, HCI scholars have also advocated for adopting "care" as a design framework for developing systems that upload values of a caring democracy [325, 449, 206]. Here, a critical aspect of civil servants' labor involves conducting care work in the context of risk. Gale et al. [184] describe this as "risk work" where civil servants are tasked with assessing and managing risks, minimizing risks in practice, and translating risk in different contexts. However, risk work (i.e., human discretionary work) in the public sector such as assessing the risk of child maltreatment [397], risk of recidivism [212], risk of long-term unemployment [20], risk of long-term homelessness [270], etc. is increasingly mediated through algorithmic systems.

Consequently, HCI researchers have also started studying how human discretionary work is changing in the public sector and adopted Lipsky's theory of street-level bureaucracy [297] to understand how street-level bureaucrats or civil servants (e.g., caseworkers, police officers, judges, educators) reflexively balance the needs of citizens against the demands of policymakers. With the adoption of digital technologies and several decisions about citizens being made from 'behind a screen', Bovens and Zouridis adopted Lipsky's theory to highlight how public services were transforming into screen-level bureaucracies [62]. Most recently, Alkhatib and Bernstein adapted Lipsky's theory into *street-level algorithms* [16] to further highlight the shift in governance as a result of algorithmic decision-making. This has further allowed researchers to investigate the intersection of human discretionary work conducted at the street-level and algorithmic decision-making [354, 396, 464, 232, 20, 385, 395]. Much of this scholarly work conducted in these domains has found that there is a mismatch between how risk is empirically quantified and predicted by algorithms versus how risk is theoretically understood, informs street-level practices, and impacts families in need of public services. This mismatch also leads to unreliable decision-making and frustrations on part of civil servants who are mandated to use algorithmic systems [256, 396]. Specific to the child-welfare system, algorithmic governance systems in the form of predictive risk models (PRMs) are being adopted as a means to proactively recognize cases where children are at high risk for maltreatment and offer targeted services to these families. However, recent studies have found that such systems exacerbate racial discrimination and inequalities and further undermine the rights of low-income communities [99, 355].

A nationwide survey on predictive analytics in child-welfare conducted by the American Civil Liberties Union (ACLU) in 2021 revealed that 26 states have considered employing predictive analytics in child-welfare [391]. Of these 26 states, 11 are currently using them [391]; however, audits of these systems reveal that they are achieving worse outcomes for families and exacerbating racial biases [171, 454, 344, 230]. Due to these concerns, Los Angeles County and Illinois have shut down their predictive analytics programs in the past [171, 454] with Oregon recently joining their ranks in June 2022 [344]. A recent study conducted by Cheng and Stapleton et al. [99] on the Allegheny Family Screening Tool (AFST) found that AFST-predicted decisions were racially biased, and workers reduced these biases by overriding erroneous decisions. AFST was designed to mitigate call screeners' biases and subjective decisions and augment decision-making by making it more objective through data. Ironically, AFST has introduced more complexities in decision-making and the call screeners are the ones mitigating algorithmic biases. Another recent study conducted on Eckerd Rapid Safety Feedback showed that the algorithm did not reduce incidences of subsequent child maltreatment [355]. A literature review of child-welfare algorithms in the United States also revealed other sources of biases embedded in the predictors, outcomes, and computational methods being used to develop these systems [397]. This study also highlights that the majority of algorithms used in CWS are predictive risk models. Finally, a study conducted by Kawakami et al. [256] on AFST showed that there were misalignments between AFST's predictive target and call screeners' decision-making objective where call screeners relied more on their contextual understanding of the family and risk factors to make decisions rather than empirical risk as predicted by AFST. That is, call screeners, focused more on contextual risk factors that families experienced on the street-level as opposed to risk quantified using administrative data.

Federal initiatives such as improved data infrastructures for CWS [225] have paved the way for tech startups to develop and pitch algorithmic systems to human services agencies across different states [239, 371, 447, 238]. However, there is a need to critically examine the current points of failure in the design of predictive risk models (PRMs). Critical to the conversation about PRMs is also the underlying principle of "risk" and how its understanding has shifted in response to the restructuring of public services to be economically efficient and productive [83, 377, 22]. Traditionally, child-welfare services have focused on risks and protective factors within families to be able to provide them with individualized care. However, with a shift towards an empirical understanding of risk and the introduction of PRMs, risk has now become a function of client characteristics as existing in prior cases and their impact on a predictive outcome (i.e., risk of maltreatment). That is, risk is estimated based on historical administrative data and is being used to identify the "deserving poor" who pose the most risk to governmental apparatus [163]. Redden et al. [377] refer to this as the embedded logic of actuarialism that also obfuscates and drives attention away from social and structural issues that bring poor and vulnerable communities under the attention of public services such as the child-welfare system [258].

7.2.2 Computational Text Analysis Research within SIGCHI

Recent works studying sociotechnical systems have employed computational text analysis techniques such as topic modeling, sentiment analysis, and part-of-speech tagging to understand sociotechnical systems [404, 27, 97]. Nguyen et al. [339] argue computational text analysis on texts involves unpacking textual information that is inherently socially and culturally situated where there exists no absolute ground truth. While this poses challenges, this method also offers opportunities to uncover dynamic and transitory phenomena present in sociotechnical systems [10]. Prior research has shown that computational text analysis can aid traditional qualitative methods by quickly scaling to large text corpora, aggregating text for analysis, and reducing directionality biases or qualitative oversimplifications [248, 140, 339]. Furthermore, recent research has found that machine learning techniques such as topic modeling carry similarities with qualitative methods such as grounded theory, and offer supporting and complementary insights into text [334]. For example, Baumer et al. [47] employed topic modeling and grounded theory on survey responses and found that the two methods yielded similar results, although the former uncovered patterns at lower abstraction levels.

Various domains, including public policy, child welfare, health, and communication, have applied computational text analysis on varying lengths of texts to investigate issues relevant to the public sector [387, 240, 404, 140, 27]. Notably, Saxena et al. [404], and Antoniak et al. [27] found that topic modeling and sentiment analysis on dense and unstructured narrative texts can provide insights not necessarily revealed via manual qualitative methods. Through topic modeling, they showed the technique could uncover invisible patterns of human activity, constraints that affect human decision-making within the domain, and latent day-to-day power dynamics between agents. In a similar vein of work, Abebe et al. [7] found that computational text analysis of texts can uncover holistic and contextualized details in pregnancy-related tweets and could predict maternal mortality rates at a higher accuracy rate than using socioeconomic and risk variables. Prior applications of topic modeling have also found evidence showing how the technique can support manual analyses of text. For example, Rodriguez and Storer [387] showed that by plotting a topic model correlation network for tweets related to domestic violence, topic modeling can provide a descriptive analysis of texts, which is comparable to first-round qualitative analysis. Isoaho et al. [240] also noted policy analysis journals extensively use topic modeling as a computational text analysis technique because it can aid manual analyses of texts. Most recently, Showkat and Baumer [420] engaged in speculative design workshops with journalists and legal experts and examined these domain experts' value expectations regarding automated NLP systems. Their study uncovered tensions around the technical implementation of such systems and implications for when 'not-to-design' them.

Public administrative work often involves collaborative decision-making where civil servants continuously negotiate with multiple stakeholders involved in cases and document case-related information drawn from multiple sources (e.g., meeting notes with other caseworkers, observations, and email or phone exchanges) [20, 404]. While this sector frequently uses predictive algorithms, civil servants have expressed doubt on the utility of such technologies [20, 100]. Instead, civil servants have expressed a desire for technology to support work processes and case management rather than profiling individuals [232]. Responding to these stakeholder needs; we applied computational narrative analysis to uncover critical aspects of child-welfare to better understand the domain. For this study, we applied CorEx [185], a semi-supervised topic model which can be used to uncover topics that are specifically associated with the above factors. Unlike unsupervised LDA topic models, which are prone to highlighting dominant themes in texts, CorEx incorporates user-provided domain knowledge in the form of anchor words that allow the topic model to uncover specific topics of interest associated with these anchor words [29, 383, 3].

7.3 Research Context

We partnered with a child-welfare agency in a metropolitan area in a Midwestern U.S. state that is part of the broader child-welfare system that was recently investigated by Saxena et al. (2021) [396]. This agency is contracted by the state's Department of Children and Families (DCF) and provides child-welfare services to families that are currently under investigation by DCF. Allegations of child maltreatment are investigated by DCF's Initial Assessment (IA) caseworkers and if maltreatment is substantiated and the case is opened for a CPS investigation, the family is then referred to this non-profit agency to provide child-welfare services. The agency must comply with all DCF standards and policies and meet its accountability requirements. During the initial court hearing, mandatory services and supervised visitation requirements are negotiated between each parent's attorney, the district attorney's office, and the judge. The agency provides case management services, parenting classes, permanency consultations, services to foster youth transitioning into adulthood, in-home services when children are not removed from the care of parents, foster care and adoption services, and family preservation services. It is important to note here that critical decision-making power in regard to reunification, termination of parental rights, transfer of guardianship, and adoption sits with the court system and caseworkers can only make recommendations to the district attorney's office. We obtained Institutional Review

Board (IRB) approval from our private research university to use casenotes for this research.

Critical to the understanding of child-welfare is also the **Adoption and Safe Families Act (1997).** This legislation introduced some of the most sweeping changes to the child-welfare system and shifted the focus primarily toward child safety concerns and away from the policy of reuniting children with parents regardless of prior neglect/abuse. It introduced federal funding to assist states with foster care, adoption, and guardianship assistance and expanded family preservation services. In addition, it also introduced a 15-month timeline where states must proceed with the termination of parental rights if the child has been in foster care for 15 out of the last 22 months [209]. This speedy termination of parental rights has received widespread criticism but still establishes the restrictive legislative framework within which caseworkers must conduct their work [216]. To ensure expedited permanency[9] for foster children, the agency employs concurrent planning such that two simultaneous plans begin when a child enters foster care – a plan for reunification with birth parents and a plan for adoption or transfer of guardianship if reunification is not possible (henceforth, *permanency plan*). The goal here is to ensure that children do not incessantly stay in foster care if reunification fails because extended stay and interaction with CWS lead to poor long-term outcomes for foster children where they are unable to form lasting relationships.

Saxena et al. (2022) [404] used unsupervised LDA topic models to study casenotes and uncovered patterns of invisible labor undertaken by caseworkers as well as showed how different systemic constraints impacted different families based on case complexity and their level of need. Their study conducted the first computational inspection of child-welfare casenotes and provided the computational basis for conducting similar studies in the public sector that seek to uncover latent contextual signals embedded in these sociotechnical systems. In this study, we go a step further and focus on the dynamic and transitory factors that impact caseworkers' decision-making and family well-being. We use semi-supervised topic models [186] that embed domain knowledge in the form of anchor words to specifically uncover different risk, protective, systemic, and procedural factors that impact street-level decision-making. By embedding domain knowledge based on Austin et al.'s framework [34] into the CorEx topic model, we are able to guide the model towards specific topics of interest and uncover the multiplicity and temporality of risk factors that are experienced by families and impact caseworkers' decision-making. We further problematize the notion of empirical risk by highlighting the various systemic and procedural factors that augment risks posed to families but can not be quantified.

[9]Permanency is defined as reunification with birth parents, adoption or legal guardianship.

7.4 Methods

In this section, we first introduce the casenotes dataset and the data cleaning process. Next, we discuss the data analysis and interpretation process. For this study, we adopted Correlation Explanation (CoRex), a semi-supervised topic modeling method developed by Gallagher et al. [186] that allows us to incorporate existing domain knowledge into the topic generation process via the use of anchor words. Unlike the generative topic modeling approach (i.e., LDA topic models) employed by Saxena et al. (2022) [404] which requires specifications for hyperparameters and detailed assumptions, this study uses semi-supervised CoRex topic models that do not assume an underlying generative model. CoRex allows us to embed domain knowledge through anchor words which further promote topic separability and representation. In addition, generative topic models may only portray dominant themes (or topics) in a corpus, however, CoRex, through the incorporation of meaningful domain words, allows us to surface topics that may otherwise be underrepresented in the corpus.

7.4.1 Dataset

We obtained casenotes written by the Family Preservation Services (FPS) team. FPS focuses on assisting birth parents achieve reunification with their children by providing crisis support, parenting classes, and helping improve family functioning [190]. This team works closely with families throughout the child-welfare process and interacts with them in-person on a regular basis. The success and effectiveness of FPS is assessed in terms of how many families are successfully reunified. Here, casenotes serve multiple purposes – 1) they provide a roadmap of all interactions and decisions made and are submitted to the DA's office if/when FPS recommends reunification for a family, 2) they highlight birth parent's progress in their efforts to achieve reunification, 3) they ensure accountability among all caseworkers (i.e., family preservation team and case management team) and consistent recording of interactions [173, 424]. Writing detailed casenotes is a central component of FPS caseworker duties who are mandated to follow documentation standards established at the agency [173]. We manually analyzed several sources of text data such as family assessments, safety plans, and discharge summaries, and settled on FPS casenotes for this study since they carried the most detailed and contextual information from ongoing face-to-face interactions with families as compared to the casenotes of investigative or initial assessment (IA) caseworkers that contained 'perceived' risks from initial interactions. That is, we conducted a significant amount of manual exploration to assess which data sources were useful and appropriate for analysis. We obtained records of 12,391 casenote entries (the 'dataset') for 462 families referred to the agency around May 1, 2019, and worked with Family Preservation until December 31, 2021, or were discharged sooner.

Metric	Value
Number of casenotes with more than 1500 words	134
Average number of words per casenote	1,461
Number of words in longest casenote	16,601
Number of unique words	20,751

Table 21: Corpus Statistics

7.4.2 Data Cleaning, Preparation, and Anonymization

The dataset contains casenote entries for families identifiable by their unique family identification numbers. To understand the dynamic relationship between CW staff and family members, we compiled the narrative casenotes for each of the 462 families by their family identifier (i.e., the 'family ID') in chronological order. We then cleaned and anonymized the casenotes by removing punctuation and names if they appeared in the 2010 U.S. Census ad Social Security names dataset [78, 13]. Numbers in the texts were also replaced with the label *NUM* to prevent numbers from raising confounding signals in our analysis. Lastly, consistent with Schofield et al. [408] who found removing stopwords led to superficial improvements in topic model solutions, we kept all short words and stop words in the texts as we found regardless of whether we removed these words, the topic models yielded no significant variations. Table 21 depicts the summary corpus statistics after we followed the above cleaning and preprocessing steps.

7.4.3 Data Analysis Approach

In this section, we discuss our data analysis approach for our three research questions. Saxena et al. (2022) [404] showed that Latent Dirichlet Allocation (LDA) can be an effective method to computationally study casenotes in the child-welfare system. However, LDA is an unsupervised generative probabilistic method [247], which does not have the option to incorporate domain knowledge in the modeling and topic generation process. As such, we build upon this prior study by adopting a semi-supervised Correlation Explanation (CorEx) topic modeling approach. This approach uses word-level domain knowledge by embedding anchor words. According to Gallagher et al. [186], anchored CorEx can offer the following advantages compared to LDA methods: 1) anchoring words allows for topic separability. The topic clusters generated by anchored CorEx have been found to be more homogeneous and contain adjusted mutual information, 2) Anchored CorEx can represent topics better. Anchoring domain knowledge to a single topic can help uncover representative topics, and 3) anchored CorEx allows the user to explore complex issues within a document by finding interesting and non-trivial aspects within the texts.

CorEx Topic Modeling (RQ1)

To answer RQ1 and inform the selection of anchoring words for analysis, we picked 10 families that had 10-15 interactions with the agency, another 10 families that had 30-35 interactions, and finally, 10 families with the most interactions and manually inspected their casenotes to understand the factors that impacted critical decisions and family well-being. A word map was made to facilitate our examination. Next, we used the socioecological model of health and development that has been recently used to study risk, protective, systemic, and procedural factors associated with child maltreatment [34] as our theoretical lens to be able to incorporate domain knowledge into our quantitative analysis (in the form of anchor words). Risk factors refer to parental experiences, behaviors, and characteristics that increase the likelihood of maltreatment (e.g., mental health, drug use, domestic violence) [34]. Protective factors are characteristics that mediate risk factors and reduce the likelihood of maltreatment (e.g., social support system, self-regulation, social skills). Systemic factors (or environmental/community factors) refer to socioeconomic factors such as employment, housing, health insurance, and transportation that impact low-income families. Finally, procedural factors (or societal factors) refer to the policies, protocols, and street-level regulations that underscore the entire child-welfare process and must be followed by families, caseworkers, and all other involved parties, i.e. - procedural factors establish the legislative framework within which all the decisions must be made.

Therefore, we specified anchoring words for four topics based on these critical factors from the socioecological model [34] and also our manual inspection of casenotes. The anchoring words for the four topics are shown in table 22. To select the optimal model for our data, we tried to maximize the Total Correlation (TC) value of the models. Total correlation measures the total dependence of topics on the document. The higher the TC value, the more effective the model is in describing the document. We also considered two other aspects, the number of topics and the anchor strength. The anchor strength controls how much weight CorEx puts toward maximizing the mutual information between the anchor words and their respective topics. Anchoring strength is positively correlated with TC. Gallagher et al. [186] suggest that setting anchor strength from 1.5-3 can nudge the topic model towards the anchor words and setting it to a value greater than 5 can strongly enforce the CorEx topic model to find topics associated with the anchor words. In our analysis, we found that TC in the CorEx model tends to increase as the number of topics increases. For interpretability, we limited the number of topics to under 20. In the model selection process, we ran all combinations with the topic number from 4 to 20 and the anchor strength from 1 to 6. The model with an anchor strength of 6 and topic number of 19 showed maximum TC. We, therefore, decided on these parameters for the final model for our analysis. Next, four co-authors of this paper individually interpreted and labeled topics based on top keywords and exemplar casenotes. Then, the authors discussed the interpretations and refined topic labels until all authors reached a consensus.

Topics	Anchoring Words
Risk Factors	'neglect', 'violent', 'anger', 'drug', 'criminal', 'behavior'
Protective factors	'encourage', 'receptive', 'protective', 'family', 'support', 'care'
Systemic factors	'rent', 'job', 'transport', 'insurance', 'medication', 'resource'
Procedural factors	'attorney', 'court', 'consent', 'appointment', 'evaluation', 'voicemail'

Table 22: Anchor Words associated with Risk, Protective, Systemic, and Procedural Factors

Qualitative Axial Coding (RQ2)

While interpreting our topics based on top keywords and exemplar casenotes, we learned that there was an overlap between several key factors (i.e., risk, protective, systemic, procedural) where a topic could belong to more than one category. For instance, lack of adequate housing is a systemic factor, however, it poses a direct risk to families. Therefore, we conducted qualitative axial coding [126] to understand how these different factors were related to one another. We placed risk factors at the center of this process (i.e., the core phenomena) and then assessed how other factors influenced the core phenomena, procedures formulated to influence the core phenomena, or general strategies carried out as a response. Figure 22 depicts these inter-relationships between the four categories.

Group Analysis of Topic Popularity Over Time (RQ3)

Prior social work studies have found that the duration of time that families spend in the child-welfare system is related to the complexity of their respective cases [366, 89]. Here, case complexity may depend on maltreatment type, financial need, substance abuse or health concerns, and the age or number of children. In light of heavy workloads carried by caseworkers and high turnover, agencies often group cases into high, medium, and low needs so caseworkers have more equal caseloads [276]. Saxena et al. [404] also found that time spent with CWS (i.e., number of interactions with CW services) can indicate case complexity. As such, we grouped families into three groups - Group 1 (low needs), Group 2 (medium needs), and Group 3 (high needs). Due to space considerations and to improve readability, we only focus on Group 3 in this study. Group 3 includes families with 40+ interactions with the agency. Next, we plotted topics from the trained CorEx topic model in Section 7.4.3 over time to understand which topics (i.e., risk, protective, systemic, and procedural factors) emerged as significant at different temporal points in a case. To accomplish this, we followed the methodology from Saxena et al. [404], where we concatenated casenotes for each family in each group and then chronologically arranged them. As these casenotes tracked the trajectory of CWS events, we then equally divided the casenotes into ten segments, so each segment had the same number of words [404]). Equal segmentation of casenotes thus allowed us to create normalized segments that can track the **"Life of a Case"** for different families involved with CW services at the agency for differing lengths of time.

7.5 Results

In this section, we discuss our results organized by our three research questions. For the sake of readability, we present our semi-supervised topic model solution organized by our set of anchor words, i.e. - risk, protective, systemic, and procedural factors. Topics are grouped in Table 1 based on anchor words and labeled T0-T18.

#	Theme	Topic	Unique keywords
1.	**Risk Factors that Impact Family Well-Being**	**T0**: Substance Misuse and Mental Health Issues **T11**: Risks arising from Inability to Manage Child Behaviors **T7**: Risks arising from Environmental Factors or Past Trauma **T18**: Risks arising from High Medical Needs of Children	*behavior, drug, anger, neglect, violent* *mother, feels, expressed, child, frustrated, understand* *choke, vented, trafficking, boyfriend, fight, steal* *observes, documentation, provides, informs, appointment, recover*
2.	**Protective Factors that Impact Family Well-Being**	**T1**: Building Protective Factors in a Child's Ecosystem **T4**: Recording Parents' Progress during Supervised Visits **T6**: Addressing Parenting Challenges through Parenting Classes **T8**: Employing Parenting Techniques through Parenting Curriculum	*family, care, support, encourage, preservation, receptive, growing* *kisses, burped, engaged, activity, redirected, attention* *communicated, related, clarified, reiterated, enrichment, negative* *parenting, curriculum, session, completed, chapter*
3.	**Systemic Factors that Impact Families and Decision-Making**	**T2**: Critical Economic Resources Needed for a Stable Household **T10**: Unforeseeable environmental or systemic factors that augment risk **T16**: Access to Household Necessities through Public Assistance and Community Providers	*rent, resource, insurance, pay, medications, landlord* *assisted, residence, supervised, settled, transition, issues* *communicated, collateral, home, pantry, bus*
4.	**Procedural Factors that Impact Decision-Making**	**T3**: Legal Processes Associated with Child-Welfare Cases **T5**: Caseworkers' efforts towards Finding Services for Clients **T9**: Risks Arising from Street-level Decisions and Time Constraints **T12**: Managing Logistics Associated with Supervised Visitations, Classes, and Appointments **T13**: Relationship between Caseworkers and Families **T14**: Barriers Associated with Following Permanency Plan **T15**: Conducting Home Visits and Safety Assessments **T17**: Recording and Sharing Details about Services, Classes, and Appointments	*appointment, court, consent, attorney, evaluation* *services, information, resources, health, mental, aoda* *services, help, mother, housing, feels, provided, therapy, health* *visits, case, shared, information, discuss, check, aware* *waited, voicemail, shared, frustrated, complaint, responded* *received, visits, missed, explained, called, reports, schedule* *observed, dressed, clean, free, marks, visible, injury* *room, arrived, played, visit, time, center*

Table 23: 19 topic semi-supervised model solution organized by four sets of anchor words. Topics are labeled T0-T18.

7.5.1 Critical Factors Arising in Child-Welfare Cases that Impact Decision-Making

In this section, we first discuss our results grouped by our sets of anchor words and explain the different risk, protective, systemic, and procedural factors that impact family well-being and the decision-making process. In exemplar casenotes below, FPS refers to the Family Preservation Specialist and CM refers to the Case Manager.

Risk Factors that Impact Family Well-Being

We first grouped our topics based on risk factors that arise in families and are noted by caseworkers in casenotes. Substance and/or alcohol misuse and mental health issues emerged as the most dominant risk factors (and also the most dominant topic) in the topic model solution. This finding aligns with prior literature that found that one-third to two-thirds of child abuse/neglect cases involve substance use disorder [192, 34]. Reading through casenotes for this topic, and as depicted by the exemplar sentence below, we also learned that substance use disorder (SUD) generally overlaps with some mental health issues. That is, in cases where SUD was a concern, mental health services were frequently discussed alongside AODA (Alcohol and Other Drug Abuse) services. This finding is also consistent with prior literature [193, 34].

> Topic 0 example: FPS and Mr. BN discussed why Mr. BN has not been in contact with FPS. Mr. BN discussed having alcohol poisoning and explained that he felt embarrassed and did not want to talk with FPS. Mr. BN continued saying that he did not want people thinking that he was not interested in getting his daughter back. Mr. BN informed FPS that he is in anger management and AODA classes and shared location of the AODA and anger management classes.

Risks arising from environmental factors or past trauma is the next dominant risk factor where domestic violence or intimate partner violence was consistently discussed in casenotes. Prior work has found that intimate partner violence, especially in the case of single parents can pose an ongoing risk to the family (i.e., the parent and child) [446, 34]. This is also challenging for child-welfare workers because they are unable to include a significant other in case planning because they are not a biological parent and are not legally bound to the case [123]. The exemplar sentence below depicts how intimate partner violence can create risky situations for the family. *Inability to manage child behaviors* emerged as the next significant risk factor. These cases generally involve minor cases of neglect that can be addressed with parents developing proper intervention and disciplining skills that reinforce positive behaviors in children. Caseworkers, especially Family Preservation Specialists (FPS), work with parents through parenting classes offered at the agency. Another risk factor arises as a result of a lack of trust and a poor relationship between parents and caseworkers. We witnessed several examples of this in casenotes where parents believed that the caseworker was unable to handle the case or needed assistance. In the top exemplar casenote for this topic, parents say that they will be filing a complaint against the worker. Prior work has found that a healthy working relationship between parents and caseworkers is essential for achieving positive outcomes for families and ensuring that cases are not re-referred in the future [89].

> Topic 7 example: [Significant other] kicked down the apartment door. Downstairs neighbors

> were upset and apartment landlord stated that she had called the police and they gave her a number for reference. Affordable Rental was called to fix the door. FPS contacted for affordable rental to see if they could move Ms. AP [parent] to another apartment. They stated that they had no apartment available and she would have to wait until September to see if something becomes available for rent. FPS then transported Ms. AP to MPD [police department]. MPD was unable to locate the purse. EM [child] was asleep when everything had happened.

Building Protective Factors within Families

The majority of cases of child maltreatment are cases of neglect that are referred to CWS due to deeper systemic issues such as lack of access to child care, lack of access to healthcare, and lack of affordable housing [34, 89]. Such issues can be addressed by ensuring that parent(s) have additional caregiver support. As depicted by Topic 1 and the exemplar casenote below, caseworkers work with parents to get other family members (e.g., relatives, grandparents) involved so that the parents have additional support, especially during stressful circumstances in their lives. However, as depicted by the exemplar casenote below, working with extended family also requires caseworkers to address any familial conflicts that arise to ensure all parties align with the permanency plan for reunification and that parents have ongoing support.

> Topic 1 example : FPS [Family Preservation Specialist] and FSS [Family Support Specialist] walked into the home and FPS introduced FSS to MGM [Maternal Grandmother], MGF[Maternal Grandfather], EM [child], CH [child], and MA [birth mother]. While waiting for BK [birth father] to arrive, FPS And FSS sat at the kitchen table while EM colored pictures and CH played with toys and walked to and from the table. MGF and MGM were in and out of the room. Conversations about the negative behaviors the kids are experiencing happened, including MGM and MGF talking about the kids' tantrums, and being violent towards each other and the adults in the home. FPS acknowledged that these behaviors are hard to deal with and can be caused for various reasons. MGM and MGF's displeasure at BK's continuing to have visits was also discussed and FPS stated that at this time the court order must be followed and FPS cannot cancel supervised visitations.

Topics 6 and 8 depict parents' progress in parenting classes where they are addressing their parenting challenges and employing parenting techniques learned from the parenting curriculum. As depicted in the exemplar casenote below, the family preservation team works with parents through intervention tactics and parenting techniques on how to manage children's behavior and employ positive enforcement techniques to build up children's self-confidence and promote healthy habits and behaviors. Parents' progress in these classes is observed and documented in the casenotes and summaries are submitted to the court as part of documentation upon completion of classes/services. We witnessed several instances in casenotes where child-welfare involvement began due to the child(ren) engaging in risky behaviors (e.g., running away) and the parents' inability to manage such behaviors. However, it was interesting to note that even for cases where the identified target problem was children's behavior or actions, parents were still referred to several services (e.g., AODA services, therapy) and not just parenting classes.

> Topic 8 example: FPS went over examples that MS [birth mother] could use. FPS suggested that MS use the activity 'Panda and the Frog' with KJ [child] at the next visitation. FPS then went over tantrums and MS stated that KJ is at that age. FPS went over the stages of tantrums and techniques to use. FPS then talked about courage and building your child up. FPS provided MS with techniques on building her child up and examples of building your child down. FPS asked MS to give encouragement to her children once a day

Systemic Factors that Impact Families

Systemic factors have been extensively discussed in prior social work literature [189, 34]. Environmental factors or community-level risk factors are other terms that are prominently used to describe characteristics that impact most families referred to CWS. These include access to affordable housing, employment, health services, public transportation, and public assistance, among others [189]. As depicted by Topic 2 and the accompanying casenote below, such systemic issues can periodically arise within families, impact parents' ability to provide a stable household, and need to be addressed promptly and intuitively by both the parents and caseworkers to maintain stability. Here, the CW staff keeps information on community resource providers, service providers, and community centers that are able to help parents during unforeseeable circumstances such as loss of employment and housing and provide financial assistance (e.g., food stamps, travel vouchers, household necessities) that would offer some temporary relief to parents. An exemplar casenote for topic 10 depicted below shows how the caseworker and parent work together towards addressing their current housing problem.

> Topic 10 example: FPS asked Mr. JP how his housing search was progressing. Mr. JP stated that the property management list that FPS provided was not as helpful as he hoped due to companies being out of state. FPS informed Mr. JP that FPS will provide him with more information when they meet next week. FPS also discussed with Mr. JP that [Community Provider] provides emergency financial assistance for housing once he secures a place. FPS asked Mr. JP if he had contacted City of [city name] Cribs for Kids program. Mr. JP stated that he had not contacted them but plans to. Mr. JP stated that he received a check in December and went on to state that court is requiring that he conduct x hours of volunteer work and x hours of application completion. FPS informed Mr. JP that FPS will contact him on Monday to schedule a meeting and provide additional housing information.

It is important to note the temporality of these systemic risk factors as they may arise and require the parents to seek temporary assistance through public programs, however, they are also collectively resolved by the parents and caseworkers and do not pose an ongoing risk to the family. The exemplar casenote below depicts how caseworkers and parents work together in resolving such risks within the restrictive framework of CWS.

> Topic 2 example: MS [parent] stated she lost her job and will be without a job by the end of the month. MS stated this job was through a agency and she will try to find another job by the end of the month. MS stated she has already contacted the agency and they are helping her look for another job. FPS suggested that MS continue to work and continue to find a job until then. FPS also offered help with finding a job. MS asked FPS if she could help her with her student loans that are currently in default in order to not have her taxes garnished. FPS told MS they could focus on applying for the income-based repayment plan and see if she is able to be on a low cost plan.

Procedural Factors that Impact Decision-Making

Procedural factors refer to the legislative framework (or legal 'procedures') that underscores the entire child-welfare process and must be followed by all involved parties. These processes also establish the constraints within which all decisions must be made. These include court proceedings, legal agreements, medical appointments and services, assessments and evaluations, and the signing of consent forms, among others. Topic 3 (exemplar casenote below), describes procedural factors associated with child-welfare cases. Topics 5 and 17 describe caseworkers' efforts in finding services for their clients and recording the details (i.e., time, location, frequency) of these services so they can be shared with other parties.

> Topic 3 example: FPS met with Ms. BR [parent] one-on-one at the agency in meeting room Innovation. During the meeting, Ms. BR discussed her CPS case and criminal court proceedings. Ms. BR was able to complete the following consent forms: Family Preservation consent forms, RISE youth consents, and medical consents. Ms. BR informed FPS that the meeting had to be short as she currently has to meet IA [Initial Assessment] worker and supervisor IASW [Initial Assessment Social Worker] at [City] CPS regarding her new CPS case. FPS arranged for a meeting next week [date] at [x]pm as Ms. BR informed FPS this was the only time she was available to meet.

Topic 15 describes caseworkers' observations during home visits, supervised visits, and completion of quantitative assessments. We witnessed several assessments in the form of home safety assessments, mental health assessments, and parenting assessments being continually used by caseworkers throughout the life of cases. Caseworkers must follow DCF policy and periodically complete these assessments because it allows the department to collect consistent information about all cases.

> Topic 15 example: BK [parent] is refusing to work with any child welfare agency and stated he will not engage in any kind of services. In fact, BK was very upset and stated that once paternity is established, he will be filing for custody and not work with CPS. FPS kindly wanted to do an assessment for safety with BK but he refused.

Reading through the casenotes, we learned that caseworkers made continued attempts (via phone calls, emails, and in-person visits) to get in touch with all involved parties (i.e., parents, foster parents, relatives, etc.) to plan and schedule these visits (i.e., topic 12). Even though it is quite a mundane task, it requires significant ongoing effort. Topic 14 describes efforts made towards following the permanency plan as established under court conditions. Caseworkers are intimately involved in the parents' lives where they continually gather details from services, medical appointments, classes, and home visits as a means to provide information on case progress. However, this in-depth involvement of government officials in the lives of vulnerable families has been described as over-surveillance and policing of families involved in CWS where parents are *recipients of support* but also *subjects of regulation* [384, 121, 5]. In addition, tensions can arise between caseworkers and parents because of caseworkers' paradoxical role, i.e. - policing vs. supporting families [384]. This is also coupled with caseworkers carrying high caseloads as well as high turnover in the caseworker position such that cases are continually transferred between caseworkers [418, 41]. As highlighted by the Topic 13 exemplar casenote below, such tensions can periodically arise when parents might feel that the caseworker is not doing enough to support them or the caseworker might believe that parents are not making enough progress towards the permanency plan. This casenote also provides a glimpse into how families are continually surveilled in their homes - the caseworker considers it necessary to record all interactions during a supervised visit and tells the family that they were not allowed to speak in their native language in his presence and that an interpreter would be needed.

> Topic 13 example: Caregiver made comments about believing FPS was not able to handle the family and that they felt FPS needed assistance. The caregiver offered to speak to FPS supervisor on behalf of FPS to get more assistance. FPS declined this offer stating that FPS would speak to their supervisor. Caregiver stated that she was giving FPS a heads up that the family was going to file a complaint against FPS. The caregiver also questioned FPS about the grandmother giving children prescription medications. FPS told the family they were not allowed to speak in Spanish. FPS clarified that it is not that the family is not allowed but that an interpreter would be needed as FPS is not fluent in Spanish.

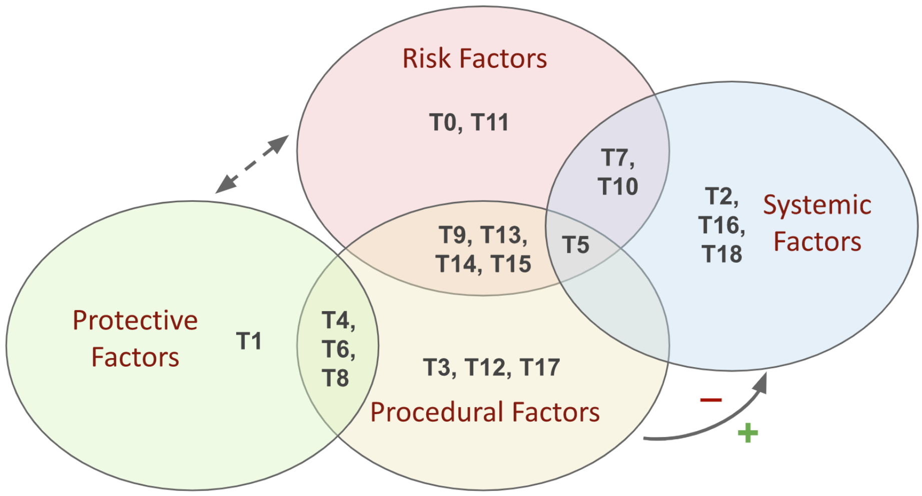


Figure 22: Axial Coding Paradigm: Relationship between Risk, Protective, Systemic, and Procedural Factors. Topics are labeled T0-T18 (see Table 3). The plus (+) and minus (-) signs between procedural and systemic factors mean that procedural factors can help mitigate systemic factors but can also amplify them. The dotted line between risk and protective factors means that there is a constant tension between these factors.

7.5.2 Interplay between Risk, Protective, Systemic, and Procedural Factors

We anticipated topics to emerge distinctly based on our anchoring of words associated with risk, protective, systemic, and procedural factors; however, we witnessed that several topics overlapped with other topics. For instance, we learned from reading the top exemplar casenotes that systemic factors can amplify risks within a family. In addition, procedural factors could help mitigate systemic factors and build protective factors, but also inadvertently amplify risk factors. Therefore, we conducted axial coding to understand how different factors were associated with one another and provide a visual representation in Figure 1. Below, we discuss the interplay between these factors.

Procedural factors can help mitigate systemic factors but can also amplify them

Caseworkers work closely with parents to address any systemic barriers (e.g., finding new employment, housing, etc.) that may inhibit case progress. They must intuitively come up with any solutions or even 'half-fixes' that may temporarily resolve a stressful circumstance for a parent. Here, the agency may be employing an evidence-based practice model, however, several arbitrary decisions are still made on the ground by caseworkers. In the exemplar casenote below, the child displays signs of underlying trauma and needs professional help. However, the caseworker draws an arbitrary conclusion and tells the parent to ensure that their child is not watching violent videos or playing violent games even though there is no evidence to suggest that this leads to aggressive behaviors [277]. It also puts the onus on the parent to 'do something' in order to address the child's immediate behavior. It is much later in this case that the caseworker acknowledges the need for a psychological evaluation and the child seeing a school psychologist.

Throughout our reading of casenotes, we witnessed several such instances where caseworkers engaged in defensive decision-making [336] where they formulated actions for parents to undertake just to be able to document that they were taking necessary steps and making decisions that purportedly addressed risks within the family.

> Example casenote: MS [parent] stated that the school called her again in regards to CJ [child] and his behavior at school. MS stated CJ was kicking and punched his teacher and assistant at school today. He has been more violent with other kids too as he purposely hit them in the head. CJ stated to his teachers he knows the head is the part that makes a person stay alive and that is the reason why he aims for people's head when he hits them. FPS asked MS to make sure CJ is not watching violent videos or playing violent games at night.

In addition, caseworkers may also engage in defensive decision-making when they anticipate risky situations or feel that they do not have enough expertise to effectively handle conflicts that might arise. In the exemplar casenote below, the caseworker feels uncomfortable managing jointly supervised visits for parents who have a history of domestic violence and bargains on an incomplete assessment to avoid these joint visits.

> Example casenote: TL [parent] stated that she spoke with her CM regarding HK [parent] attending joint visits. FPS explained that she would reach out to her CM regarding visits and CM explained that HK would need to complete the assessment prior to his enrollment in the program. TL appeared upset that BK wouldn't be able to sit in today's visit and stated that he would have to sit in the car until her visit is over. FPS provided CM with an update on the conversation with TL. CM stated that she spoke with TL and explained that joint visits with HK wouldn't be appropriate given the father's mental health condition and history of DV. CM stated that she didn't feel comfortable having joint visits at this time.

Similarly, in another case (see casenote below), the parent tells the caseworker about a rodent infestation in their place and the caseworker helps the parent by speaking with the landlord. However, this is also followed by the caseworker conducting a home safety assessment that permanently records these new risks on the parent's case documentation.

> Example casenote: MS [parent] discussed the issues that MS is having with her current landlord. MS stated that she thinks that there is a rodent infestation and that the landlord was not responding appropriately. MS stated that he was dragging his feet on an exterminator. FPS spoke with landlord to make sure he understands the urgency.

On the other hand, caseworkers within their capabilities, do help families address any arising risk or systemic factors (see example casenote below) associated with finding essential resources and getting access to public assistance. We also witnessed several instances where caseworkers helped parents create resumes for job applications, find new housing, apply for financial assistance, and get home essentials (e.g., beds for kids, clothing, toys, etc.). This underscores a need to understand the ***why*** and ***how*** street-level decisions are made by caseworkers within the restrictive legislative framework of CWS, as opposed to the broader focus on the service delivery model implemented at the agency.

> Example casenote: LC [parent] shared that KC[child] was being bullied at school and asked if FPS could help her look into another middle school in the area for KC to attend. LC stated its to the point of her daughter having anxiety when she is getting ready for school. LC thought about enrolling her for online classes but wants that to be the last resort. FPS told LC that she will look into the list of different schools around the area that KC could possibly attend.

Procedural Factors can Amplify Risk Factors

Caseworkers are central to the child-welfare process and act as mediators between birth parents, the court system, and service providers [389, 89]. That is, they bridge the administrative gap between legal processes established under court conditions (that the parents must conform to) and social work processes centered in helping families. These conflicting roles can create tensions between parents and caseworkers where caseworkers must help parents through services (e.g., therapy, parenting, domestic violence) but also police their actions to ensure that they are following court conditions for reunification [121, 384]. In the casenote below, the parents explain that they are more focused on finding stable housing and employment which is causing them to miss some supervised visits. Here, procedural factors add more stress to the lives of parents instead of helping them navigate child-welfare services. It is also important to note that the caseworker's primary concern here is receiving documentation about services so they are able to complete their procedural task. We witnessed similar tensions in other cases where the parent(s) shared that they were overwhelmed with several appointments for services and supervised visits throughout the week while trying to maintain full-time employment.

> Example casenote: FPS discussed with CM as to whether the documentation is sufficient and how visits look moving forward. Mr/Ms KD said that FPS could call the service and they are not lying. FPS explained that she has name and number of the service and they have to have documentation [of services]. This affects their [parents] supervised visits with their son. Mr/Ms KD said they are pretty sure that the judge is to want that they have somewhere sufficient to live and this is a necessary step for the kids. They understand that they are missing visits but they also have to focus on the bigger goal [stable housing]. FPS asked if both are able to meet with her on Friday so we can get consents signed in order to verify employment and discuss visits moving forward.

In addition, caseworkers are mandated to follow court orders regarding who attends supervised visits. During court hearings, each parent's attorney advocates for their client's parental rights and ability to visit their children. However, as depicted in Topic 1 casenote, inter-family conflicts can arise which pose an ongoing risk toward reunification efforts. Reading through the casenotes, we learned that there is a history of domestic violence in this case with one parent only peripherally involved in the children's life. Here, caseworkers must work with all involved parties even though uncertainty and conflicts persist within the family due to prior and ongoing risk factors. This further augments the overall uncertainty in decision-making because it is unclear whether familial conflicts would be resolved in the future so that both parents and relatives will provide caregiver support to each other.

Interactions between Risk, Systemic, and Procedural Factors

One critical aspect of CWS is to provide services (e.g., therapy, domestic violence, and Alcohol and Other Drug Abuse (AODA) classes) to parents to prevent future instances of child maltreatment. These services are agreed upon under court conditions, but it is up to caseworkers and parents to contact different service providers and find appointments. However, it can be challenging for both caseworkers and parents to find these services, especially accounting for parents' work schedules, supervised visitation appointments, and a lack of adequate service providers in the system. As depicted by the exemplar casenote for Topic 5 below, parents can find themselves waiting to hear back from service providers and may require assistance from caseworkers.

> Topic 5 casenote: Family Peace [service provider] will work with YW [parent] on issues related to domestic violence. YW stated that she is still waiting to hear back from them as well as the agency to start therapy. FPS encouraged YW to call both agencies and let them know that she has been waiting to start services with them. FPS asked if she would give FPS permission to discuss her case with them.

Another critical issue here is the efficacy and consistency of services that are offered to parents and children [321, 166, 132]. Prior work has highlighted concerns associated with over-medication of children, overuse of psychotherapy, and inappropriate use of psychological testing [321]. As depicted in the exemplar casenote below, the parent expresses her concerns regarding another psychological evaluation for her child but is unable to exercise agency. Several states also employ a standardized service model or a "cookie cutter" approach where judges order therapy, services, and evaluations for all clients regardless of case circumstances [321]. Psychological evaluations, especially, act as catalysts and are often used as a "staple tool" by judges for the provision of mental health services [166]. A program director at this agency confirmed that the Department of Children and Families (DCF) has used a cookie-cutter approach to services in the past where all parents had to complete the same set of services, i.e. - a standardized care approach was used instead of individualized care that recognizes target problems (e.g., mental health, drug use, domestic violence).

> Topic 5 casenote: FPS asked KL [parent] if she wanted to have another psychological evaluation done due to him [child] drawing disturbing pictures of hurting other people. KL stated she is afraid that they will only put him on more medication.

7.5.3 Temporal Dynamics between Factors through the Life of Cases

Following the methodology from Saxena et al. (2022) [404], we grouped families into three groups based on their number of interactions with the agency. Below, we only focus on the group with the most interactions with CWS (i.e., Group 3) because these are the more complicated child-welfare cases where interactions between risk, protective, systemic, and procedural factors are more evident. Prior work has also highlighted that case complexity (e.g., type of maltreatment, age, number of children, need for financial assistance, drug abuse in the family) is directly associated with the time spent under the care of CWS [366, 89].

Competing and Fluctuating Factors Lead to Uncertainty and Confounding Factors

As depicted in Figure 23, risk, protective, systemic, and procedural factors continually fluctuate and interplay with each other (i.e., changing topic probabilities over the life of cases). This may confound caseworkers' judgment and leads to uncertainty in decision-making because at any given time it is unclear what the trajectory of a case might look like. For instance, parents might be building protective factors where they now have additional caregiver support and learning parenting techniques that help them better manage child behaviors. However, systemic factors (e.g., loss of employment, housing, etc.) may also arise throughout the life of the case and pose risk to the permanency plan. As depicted in the previous section, procedural and systemic factors themselves may pose risks to families. Therefore, it is interesting to note that risk, systemic, and procedural factors oscillate together throughout the life of the case. This could be for two reasons - a discussion of needing services (i.e., *risk*) is generally coupled with finding services (i.e., *systemic*) and its association with the permanency plan (i.e., *procedural*), and 2) caseworkers discuss any arising systemic or procedural factors followed by their impact on the family. Caseworkers also discuss the development of protective factors in great detail as depicted by the green trend. This is primarily the case because the majority of these casenotes come from the Family Preservation Team which works closely with parents through parenting services. In sum, a post-hoc analysis of these trends of competing factors shows that uncertainty about the final outcome of cases (i.e., reunification or placement in foster care) persists even at case closure where several cases are re-referred to CWS in the future [272]. It also highlights that caseworkers continually face confounding factors (as a result of competing factors) in situ throughout the child-welfare process.

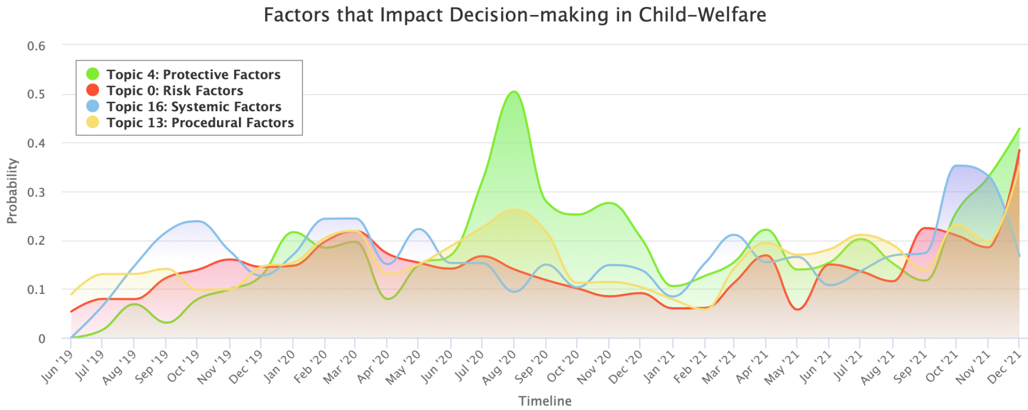


Figure 23: Relationship between Risk, Protective, Systemic, and Procedural Factors. Fluctuating and competing factors augment uncertainty and confound caseworkers' decisions such that uncertainty about long-term family well-being often persists even at case closure

Relationship between Different Protective Factors

Figure 24 depicts trends in protective factors. Family Preservation Services works with parents in parenting classes and supervised visits to build protective factors. Teaching parents proper intervention and disciplining techniques helps address risks arising from the inability to manage child behaviors. Topic 1 highlights protective factors in a child's ecosystem by assessing their interactions within their social support system (i.e., parents, relatives, grandparents). We notice Topic 8 (i.e., employing parenting techniques during visits) follows a much similar trend as Topic 1. This may be the case because both topics inherently assess healthy and positive interactions between adults and children. On the other hand, Topic 6 describes caseworkers' conversations with parents on how they could be addressing parenting challenges as they work through the parenting curriculum. We expected this trend to be higher at the onset of cases but gradually diminish as parents develop protective skills and are recorded as observations in casenotes during parenting classes (i.e., Topic 4). This highlights the need to understand the temporality of such protective factors that help children and parents achieve positive developmental outcomes over time. This is often described as "resilience" in social work literature [457, 294]. Resilience in children and parents is a result of interactions in their environment where caseworkers and other professionals can directly help shape this environment [286]. That is, an understanding of resilience can help assess which protective factors are pertinent for a family and would lead to better long-term outcomes.

Relationship between Systemic Factors

Figure 25 depicts trends in systemic factors. Topic 2 describes environmental and systemic factors that affect family well-being (e.g., employment, housing), and Topic 16 describes the essential household needs (e.g., food, clothing, utilities) as observed and recorded by caseworkers during home visits. In essence, both topics assess material resources necessary for maintaining

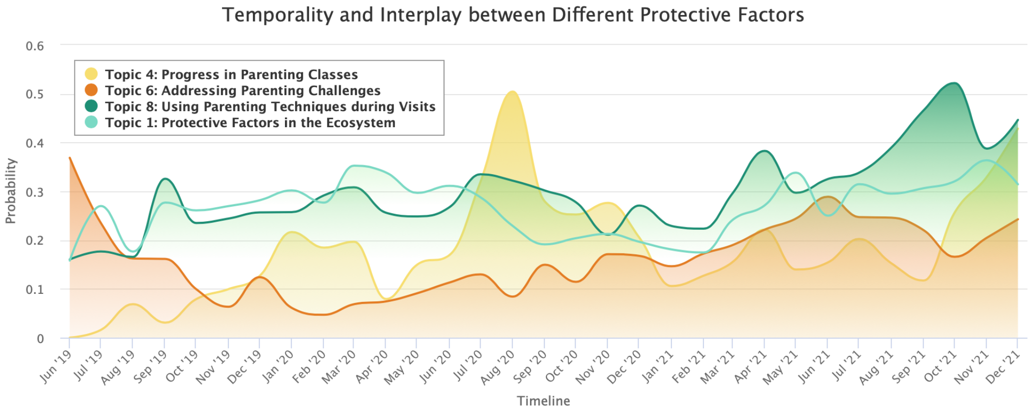


Figure 24: Relationship between different Protective Factors. Understanding the temporality and interplay between different protective factors can help assess long-term well-being outcomes for families.

a stable environment for children. This may explain why the trends for these two topics follow a much similar trajectory. On the other hand, Topic 7 describes emerging risk factors in a case due to unforeseeable circumstances such as intimate partner violence, medical needs arising from underlying trauma, and familial conflicts. We see a significant amount of fluctuation in this trend because unforeseeable systemic risks may arise but are also continually addressed through collaborative problem-solving between parents and caseworkers. Topic 18 discusses children's medical needs in terms of their medical appointments and medications. These needs, as well as systemic barriers associated with meeting these needs (i.e., finding proper services, consistent mental health assessments, and medical appointments), are consistently recorded by caseworkers in their casenotes because this information needs to be shared among several involved parties. It is imperative to note here that structural economic issues (e.g., stable employment, safe and affordable housing, affordable health care) underscore the majority of child-welfare cases and involve poor and low-income families [191]. These are the consistent risk factors (i.e., Topics 2 and 16) that impact most families. On the other hand, Topics 7 and 18 capture the transitory risk factors that the child-welfare staff is able to address with timely interventions. This underscores a need to understand both the socioeconomic risk factors that impact the majority of families as well as context-specific transitory risk factors specific to a family. Here, street-level interventions can help address some risks, however, systemic and policy-driven changes are equally necessary to improve social conditions that impact vulnerable and low-income communities.

Relationship between Different Procedural Factors

Figure 26 depicts trends in procedural factors. Topic 13 describes strenuous relationships and interactions between caseworkers and parents. Child-welfare staff is legally mandated to follow a 15-month timeline which also establishes the permanency plan. That is, per the Adoption and Safe Families Act (ASFA), if parents do not fulfill all court conditions within 15 months, then

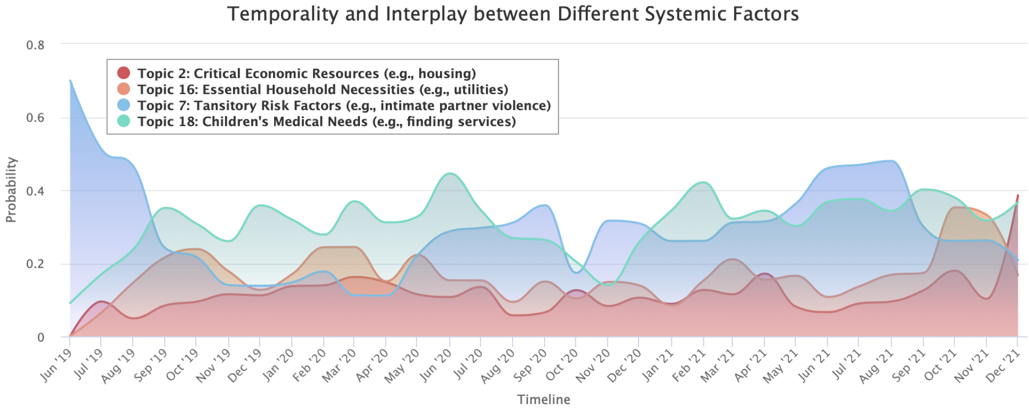


Figure 25: Relationship between different Systemic Factors. Both socioeconomic risk factors and transitory risk factors impact families. Caseworkers are able to address transitory risk with proper interventions but are unable to have a meaningful effect on socioeconomic risks.

parental rights must be terminated and child-welfare staff must find a more permanent placement for children in foster care. Here, caseworkers must work within this restrictive legislative framework to ensure that the permanency plan as established under court conditions is on track (i.e., Topic 14) where parents are completing court-ordered services, attending supervised visits, and working towards building a stable household, i.e., they must continually police the parents' actions to ensure progress towards permanency. Dorothy Roberts describes this as a dual and paradoxical role where caseworkers act as "investigators and helpers" and parents are both subjects of regulation and recipients of support [384]. These ongoing tensions between following the permanency plan and maintaining a working relationship with parents may explain why trends for these topics oscillate together. Moreover, Topic 12 describes scheduling and managing logistics around supervised visits and services. This trend is closely followed by Topic 9 which describes the risks emerging due to street-level decisions and time constraints. As noted in the previous section, parents in several cases shared that they were overwhelmed by the number of appointments and supervised visits while trying to maintain full-time employment and make necessary changes within their household. That is, procedural factors can themselves add risks to the stressful lives of parents who are fighting for reunification. Such risks arising due to the restrictive legislative framework of CWS cannot be quantified. It is also not possible to assess their long-term impact on families.

7.6 Discussion

Abebe et al. [6] highlight that much of the computational research that focuses on fairness, bias, and accountability in machine learning systems continues to formulate “fair” technical solutions while treating problems that underscore the sociotechnical environment as fixed and fail to address deeper systemic and structural injustices. Through this study, we bring attention back to the *sociotechnical* and highlight social problems in child-welfare and how these problems become embedded in algorithmic systems. Abebe et al. [6] also formulate four roles or ways in which computational research can help address social problems. This study assumes the dual roles of ***computing as rebuttal*** where we highlight the technical limitations and feasibility of predictive risk models (PRMs), and of ***computing as synecdoche*** by uncovering systemic complexities and social problems in child-welfare that directly impact families.

7.6.1 Rethinking "Risk" and the Underlying Data Collection Processes

Our results bring into question how “risk” is formalized in the child-welfare system by drawing attention to the broader ecosystem of decision-making processes where systemic and procedural barriers can also create and amplify new risks posed to families. Prior research on algorithmic systems used in CWS has found that the majority of these systems define risk as a function of child and parent-related risk factors (e.g., parent’s involvement in drug and alcohol services, criminal justice, housing authority, etc.) [397, 256], however, as our results show, the system itself can pose a significant amount of risk to families in regard to how protocols and practices (i.e., the legislative framework) are carried out on the street-level. This is further complicated by the fact that “risk of maltreatment” is poorly defined [397] (essentially comprising of three different outcomes – neglect, physical abuse, and sexual abuse), and this definition as well as criteria for investigating families can vary from one jurisdiction to another [187, 144]. These investigations may result in substantiation of child maltreatment, and consequently, the case is brought into the system.

Here, our results shed light on the data collection processes that ensue as parents are surveilled by caseworkers and mental health professionals. We learned that caseworkers used several different screening tools and risk assessments that quantitatively capture risk factors during home visits, risk factors associated with children’s mental health, parents’ progress in parenting classes, as well as parents’ engagement and progress towards the

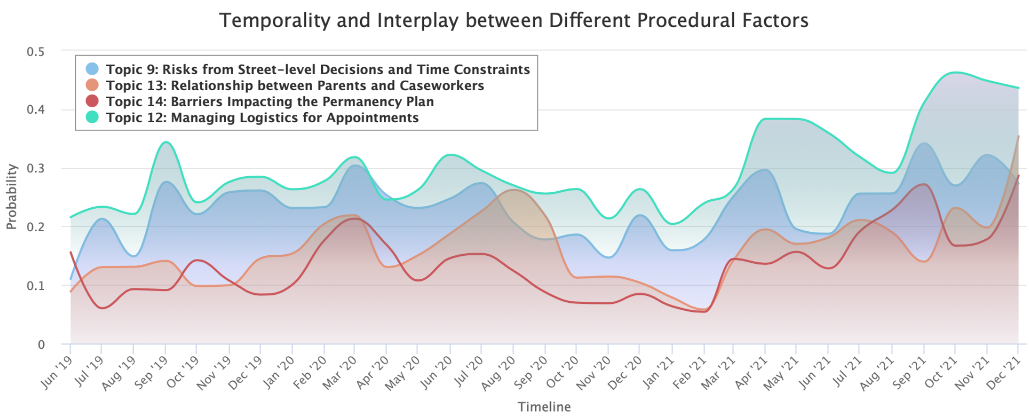


Figure 26: Relationship between Different Procedural Factors. Fluctuating procedural factors highlight tensions that arise between parents and caseworkers who must maintain a working relationship to make progress towards permanency.

permanency plan. The intent here is to collect as much information as possible and resolve any ambiguity resulting from missing information. However, this is problematic because CWS experiences a high turnover with a lack of well-trained caseworkers which leads to a lack of consistency in regard to how these assessments are completed [418, 41]. Here, **caseworkers rely more on their impressions of the family in completing these assessments rather than expertise developed over time** [121, 405].

In addition, algorithmic tools such as Allegheny Family Screening Tool (AFST) [99] and Eckerd Rapid Safety Feedback (ERSF) [355] use a family's prior involvement in public and medical services to assess the risk of maltreatment through proxy outcomes of *risk of re-referral* and *placement in foster care*. However, as our results show, **parents lack agency in the process and must consent to assessments and information disclosures. They are unable to turn down services or classes that they might consider unnecessary.** Parents may also face repercussions and subsequently experience over-surveillance if they refuse psychological evaluations, drug tests, and/or additional services [121, 384]. As highlighted by Saxena et al. [405], this refusal or disagreement with caseworkers might be captured under predictors such as "parents' cooperation with the agency" – a significant predictor of risk per the WARM risk assessment. Services for parents and children are court-ordered where several states (including the state where this study was conducted) have employed a standardized service model in the past where parents in all cases were referred to a fixed set of services (e.g., parenting classes, psychological evaluations, Alcohol and Other Drug Abuse (AODA) services, etc.) regardless of case circumstances [321, 166, 132]. That is, **parents were enrolled in services that they did not necessarily need, and consequently, more data was collected about them through multi-institution partnerships between child-welfare agencies, service providers, and the court system.** Therefore, it is problematic for algorithms such as AFST and ERSF to use this cross-departmental data collected through power asymmetries because it further puts these families at a significantly higher risk of being re-investigated since their prior involvement with CWS renders them to receive "high-risk" predictions for future child maltreatment events. Our findings here, act as *synecdoche* [6] by making visible child-welfare practices and power asymmetries through which vulnerable low-income families are continually targeted by the system.

On the other hand, let us assume that designers and technologists developing algorithmic systems are able to adequately model for organizational context in terms of protocols, practices, resource constraints, and policies as well as make founded assumptions for a specific social context; then by extension, the developed system is no longer portable to a different jurisdiction or social context because child-welfare practice can vary significantly from one state to another.

Selbst et al. [415] refer to this as the portability trap – "Failure to understand how repurposing algorithmic solutions designed for one social context may be misleading, inaccurate, or otherwise do harm when applied to a different context". Here, we want to draw caution against child-welfare agencies acquiring algorithmic systems from private companies developed in one jurisdiction but sold and employed in several other jurisdictions [239, 371, 447, 238, 33].

7.6.2 Confounding Factors, Uncertainties, and Implications for Predictive Risk Models

Prior literature in machine learning has discussed data and model uncertainties [88, 183, 263] and technical methods on how to mitigate these uncertainties that act as obstacles in the way of better predictive performance [311, 340, 262]. On the other hand, HCI scholars have recommended that we engage with uncertainties as opportunities for human-centered design rather than treat them as obstacles [430, 51, 354]. Pääkkönen et al. [354] note that "human discretionary power in algorithmic systems accumulates at locations where uncertainty about the operation of algorithms persists". They further note that the design of algorithmic systems could benefit from identifying and cultivating important sources of uncertainties because it is at these sources that human discretion was most needed.

Our results in Section 5.3 move beyond data and model uncertainties and show how uncertainties can arise throughout the child-welfare pathway as a result of fluctuating factors (i.e., risk, protective, systemic, and procedural factors) that continually interplay with each other and directly impact decision-making processes. Our results show that a parent may be developing protective factors through parenting services, however, transitory risk factors (e.g., loss of employment, housing, risks from intimate partner violence, etc.) may also periodically arise throughout the life of the case. In addition, systemic and procedural factors can themselves augment the risk posed to a family; however, at any given time there is a lack of clarity about their impact on the final outcome (i.e., reunification, adoption, or placement in foster care). **These competing and fluctuating factors confound caseworkers' decision-making and the situation is further aggravated by a lack of experienced caseworkers in the system [418, 41].**

This further brings into question our understanding of ecological risk and problematizes three core attributes regarding how risk is modeled in algorithms - 1) different risk factors present in a case are modeled as static variables, however, as our results show, transitory risk factors may arise but are also addressed collectively by caseworkers and parents. Here, **risk as a static construct is inherently biased because no temporal point estimate of risk taken at any given point in the child-welfare process offers a true picture of occurrences within the case**, 2) the baseline assumption underscoring predictive risk modeling is that risk within a family is likely to escalate if no interventions are made [397]. This leads to excessive CWS interventions and over-surveillance of vulnerable families [121, 5]. In addition, as our results show, the trends in risk factors oscillate throughout the life of cases where risk factors may arise but are also addressed. **Ignoring the temporality of different risk factors and treating them as static variables leads to elevated risk scores for families and excessive investigations**, and 3) prior work has established that empirical knowledge in child-welfare is quite limited and there is still a significant amount of debate regarding which risk factors (when taken together) lead to the accumulation of risk and which protective factors help mediate these risks [397, 188, 419]. As our results in Section 5.3 suggest, different factors mediate the effects of each other throughout the child-welfare process. **Without understanding and embedding empirical knowledge of interaction effects within predictive risk models such as AFST [99] and ERSF [355], risk predictions are likely to be elevated and biased.** As shown by a recent study conducted by Cheng and Stapleton et al. [99], call screeners helped reduce racial disparities in AFST-predicted decisions by using their contextual knowledge of cases to override erroneous decisions. That is, an algorithm designed to bring neutrality and objectivity to the decision-making process is itself producing racially biased predictions. Our findings here, act as *rebuttal* [6] by highlighting the limitations of predictive risk models and the core assumptions about risk factors that make predicted outcomes unfeasible.

In addition, our results also draw attention to how seemingly mundane tasks carried out by caseworkers such as continued attempts to contact birth parents, foster parents, and relatives to schedule supervised visits and services can pose risks to the 15-month timeline of the permanency plan [328] because it significantly impacts caseworkers' ability to work with parents and meet goals for completing set hours of visitations and services. This risk posed to families is hard to estimate but continually impacts street-level decision-making. It also highlights invisible patterns of labor that are only documented in casenotes and cannot be encapsulated by quantitative risk assessments.

7.6.3 Implications for Computational Narrative Analysis and NLP-based Systems

Selbst et al. [415] note that social context is often abstracted out so that machine learning tools can be applied to any given domain and evaluated based on predictive performance (i.e., *the algorithmic frame*). Here, fair machine learning researchers may further expand upon this approach to investigate ML system's inputs and outputs (i.e., *the data frame*), however, this is still an attempt to formulate mathematical notions of "fairness" and "bias" and continues to abstract out the broader social context within which the system is situated and interacts within organizational pressures, systemic constraints, and with a variety of stakeholders. Consequently, authors formulate the *sociotechnical frame* which recognizes that an ML model is simply a subsystem within a broader sociotechnical system where drawing stakeholders and institutions into the abstraction boundary allow us to investigate complex interactions. Through this study, we used computational narrative analysis to draw attention back to the sociotechnical frame and highlighted the complicated interactions between caseworkers and families, and brought into focus the critical structural issues within CWS.

Computational methods such as unsupervised and semi-superv -ised topic modeling [63, 186] facilitate a qualitative exploration of casenotes and allow us to understand street-level practices, systemic constraints, power asymmetries, as well as temporal dynamics between different factors. Prior work has hypothesized that using text-based narratives within risk assessment algorithms may offer more holistic and fair predictions by filling in the gaps of quantitative risk predictors [397]. However, we want to draw caution against this approach as tech companies are beginning to pitch NLP-based systems to human services agencies and are being acquired by several agencies across the United States [33]. Here, it is important to note whose values become embedded in these systems [55] and which (and whose) resources are directed towards these initiatives in an overburdened and underfunded system [203]. Through our reading of casenotes and data analysis, we recognized several limitations associated with caseworkers' narratives.

First, as previously noted in the Methods section, we manually analyzed several data sources to assess which casenotes contained detailed and credible information about interactions between parents and caseworkers. We settled upon casenotes written by the Family Preservation Team because they work closely with families throughout the process and understand the risks and needs associated with each family. On the other hand, **casenotes written by the initial assessment (IA) or investigative caseworkers carried information about *perceived risks* and the caseworker's impression of the family** because not enough information is available (and at times, contradicting facts are present) at the onset of a case. Second, **even though the agency has established rigorous standards on documentation, there is variability in the writing of casenotes** where some caseworkers captured details about children's demeanor during transportation, supervised visits, and medical appointments while other caseworkers only wrote pertinent details (e.g., child's response to parenting techniques, medication schedule created at a medical appointment, transportation logistics). Third, **inexperienced caseworkers are known to engage in defensive decision-making where they might omit their mistakes from casenotes** [336]. As highlighted in the second exemplar casenote in Section 5.2.1, the caseworker prioritized their own comfort over conducting joint supervised visits leading to the frustration of parents. This interaction was only uncovered because the Family Preservation Specialist wrote about it in their casenote entry. Here, collaboratively written documentation offers some accountability but we hypothesize that there may be several other such instances where caseworkers' actions went unchecked and undocumented. Fourth, qualitative exploration of casenotes driven by our computational approach allows us to understand the power asymmetries that both parents and caseworkers experience in the child-welfare process, however, **the contextual knowledge derived from casenotes can easily be stripped and instead exploited once quantified to be used in downstream tasks in NLP-based systems**. Fifth, as noted in the previous two sections, **a discussion of risk in casenotes does not necessarily mean that it is a persistent danger impacting family well-being**. It may simply be a noteworthy event that a caseworker recorded at that point in time (e.g., a child being bullied at school).

These five points are crucial for developing an understanding of this complex sociotechnical system and are especially important as it pertains to natural language processing. Recent studies in NLP have examined text datasets and seed lexicons and found that social hierarchies and racial, cultural, and cognitive biases can become embedded in and amplified by NLP systems and lead to significant disparities in downstream tasks [25, 442, 82, 57, 168]. Our findings here, act as a *rebuttal* by highlighting the limitations of casenotes as a data source for downstream NLP tasks and act as *synecdoche* by making visible the structural issues in child-welfare that become embedded in casenotes. Alternately, an upstream approach (i.e., - the corpus itself becomes an object of the study [24]) as adopted by this study can help uncover contextual street-level interactions, critical factors that are hard to quantify, and uncertainties and confounding factors that offer a more comprehensive view of the decision-making ecosystem. In addition, an upstream approach allowed us to uncover empirical evidence about how marginalized communities face systemic injustices in child-welfare.

7.7 Limitations and Future Work

This study conducts a computational narrative analysis of casenotes at one child-welfare agency in a midwestern state in the United States and uncovers several factors and street-level interactions that impact decision-making and family well-being. However, this study has some limitations that create opportunities for researchers to further expand upon this body of work. First, child-welfare practice can significantly vary from one state to another in terms of criteria for investigations and policies and protocols that all parties are mandated to follow. Similar analyses conducted in other jurisdictions would reveal hyperlocal and context-specific experiences of caseworkers and families in those regions. Second, this study only uncovers interactions and street-level decisions through the perspective of caseworkers and may omit or underplay the oppression, surveillance, and coercion experienced by many families and cannot reveal structural power dynamics that fundamentally underpin child-welfare interactions [121, 384]. These interactions are socially situated where parents are likely to experience the same events differently. Here, it is important to understand the perspective of affected communities (i.e., foster children, parents, and foster parents) about whom decisions are being made. For instance, a recent study conducted by Stapleton et al. [432] found that parents considered CWS to be punitive and unsupportive and instead wanted systems that would help them fight against CPS as well as evaluate CPS and the caseworkers themselves. Future research should continue to focus on uncovering street-level complexities within this complicated sociotechnical environment through the perspective of families and caseworkers. Finally, this study takes an upstream, corpus-focused approach (i.e., the corpus itself is the object of the study) where we sought to understand dynamic and transitory factors embedded in caseworkers' narratives that impact decision-making and family well-being. However, this requires a re-analysis and experimentation of different NLP techniques to focus on topic-specific corpora such as the corpus used in this study. That is, we recommend that researchers conducting similar analyses in various public sector domains experiment with and compare different NLP methods to assess which methods help uncover latent signals in the corpus as well as highlight limitations in the corpus itself.

7.8 Conclusion

We conducted a computational narrative analysis using Correlation Explanation (CorEx) [186], a semi-supervised topic modeling approach that allows us to incorporate domain knowledge in the form of anchor words. Using the socioecological model of health and development [34] as our theoretical lens, we incorporated domain knowledge about risks, protective, systemic, and procedural factors that impact decision-making and family well-being. We provide empirical evidence that the child-welfare system itself poses a significant amount of risk to the families that it is expected to serve. We show how parents are over-surveilled in the system, the lack of agency they experience in the child-welfare process, and problematize the data collection processes that take place as a result of this power asymmetry. We complicate the use of predictive risk models that treat risk factors as static constructs by highlighting the multiplicity and temporality of different risk factors that arise throughout the child-welfare pathway. Finally, we draw caution against using casenotes in NLP-based systems by highlighting the limitations and biases embedded within this data source.

CHAPTER 8: CONCLUSION

This dissertation is devoted to outlining responsible pathways forward for the design of human-centered algorithms in the public sector and contributes a holistic understanding of a complex sociotechnical system through deep ethnographic work, the design of a theoretical framework for algorithmic decision-making in the public sector, and computational narrative analysis of a critical data source. It showcases the practical tradeoffs that need to be balanced for algorithm design - 1) at the human discretion level, I highlight different insertion points and goals of algorithms for augmenting practitioners' decision-making processes, 2) at the bureaucratic level, I highlight the constraints within which all decisions (human or algorithmic) must be made and how organizational resources can be leveraged towards ensuring the proper integration and adoption of an algorithmic system, 3) at the algorithmic level, I showcase how algorithm design can account for the uncertainties inherent within cases and support decision-making processes instead of providing predicted outcomes. This dissertation work has provided actionable steps for human-centered algorithm design to child-welfare leadership that will further help ensure that the decisions made by child-welfare teams are centered in evidence-based practice and lead to positive outcomes for families.

8.1 Contributions to HCI and Critical Computing Research

Abebe et al. [6] highlight that much of the computational research that focuses on fairness, bias, and accountability on algorithmic systems continues to formulate "fair" technical solutions while failing to address deeper systemic and structural injustices. Through my dissertation work, I bring attention back to the *sociotechnical* and highlight social problems in child-welfare and how these problems become embedded in algorithmic systems. Through the studies discussed below, my dissertation assumes the dual roles of *computing as rebuttal* where I highlight the technical limitations and feasibility of risk assessment algorithms, and of *computing as synecdoche* by uncovering systemic complexities and social problems that directly impact families. This dissertation also makes contributions at the intersection of gaps highlighted by the literature review in Chapter 2, recommends solutions centered in strength- and asset-based approaches, and outlines responsible pathways forward for the design of human-centered algorithms in the public sector. This dissertation employs a human-centered data science research approach to the studies conducted and makes the following contributions -

8.1.1 Critical, Quantitative De-construction of Algorithms

To be able to assess the scope and utility of algorithmic decision-making in the public sector, it is imperative to unpack current algorithmic interventions and pay special attention to the points of failures such that these drawbacks are not uncritically reproduced in newer algorithmic systems. We conducted a comprehensive review of algorithms employed at child-welfare agencies across the United States and highlighted the deficiencies in current quantitative methods, predictors, and outcomes that lead to biased, inconsistent, and often unsensible decisions that frustrated caseworkers. This is of critical importance as federal initiatives have laid the groundwork for more algorithmic interventions in CWS [225, 425]. In a similar vein of work, we conducted a quantitative deconstruction of a prominent risk assessment algorithm and critically examined the data as well as the data collection processes [405]. We show how latent power structures are embedded into these decision tools from the very beginning (i.e., data collection process) where caseworkers exerted more power over families in how (and the kind of) data that is collected about them. Moreover, we show how several of these predictors are inherently biased, based on caseworkers' perceptions, and predominantly target low-income and historically marginalized groups. This analysis is of key importance because much of this historical data continues to be re-purposed in newer algorithmic systems without a critical examination of the predictors themselves. Through these studies, we show how algorithms can be deconstructed and examined, and connect significant predictors of maltreatment to social science literature in the domain to highlight how algorithmic tools can reinforce and further conceal existing systemic disparities within a sociotechnical system.

8.1.2 Theoretical Framework for Algorithmic Decision-Making in the Public Sector

The design of algorithmic systems in the public sector must account for complex interdependencies between human discretion, bureaucratic processes, and algorithmic decision-making. Drawing upon theories from Public Administration (PA), Science and Technology Studies (STS), and Human-Computer Interaction (HCI), we leveraged a sociotechnical perspective of algorithmic decision-making that captures these interactions. This theoretical framework showcases how practical trade-offs must be made to manage the cross-dependencies at both the macro- and micro-levels of the algorithmic model to offer practitioners autonomy and improve human discretionary work. The framework also draws attention to the high degree of uncertainty inherent in the administrative data which consequently means unreliable predictions. Therefore, the goal of algorithms in the public sector must be re-evaluated to support the decision-making processes of stakeholders instead of providing predicted outcomes. Specifically for child-welfare, this means 1) providing caseworkers with necessary information such that they can engage in heuristic decision-making, 2) providing explanations for recommended outcomes, and 3) demonstrating compliance with evidence-based practice. As depicted in Chapter 4, we also used this framework to conduct impact assessments on the nature of the practice, administration in public agencies, and street-level decision-making to critique value propositions of standardized, efficient, and cost-effective decision-making conducted through algorithms. We show that is necessary to decompose and rebuild the decision-making workflow such that algorithm-assisted decision-making is an essential part of the process and embedded at a decision point where it offers the most utility to stakeholders. In sum, it highlights that significant investments are required to ensure the proper adoption and integration of an algorithmic system into decision-making processes beyond the initial investment in the development and deployment phases.

8.1.3 Ethnographic Study to Understand Human-AI Interaction and Socio-Political Complexities that Impact Decision-Making

We conducted a two-year ethnography within the child-welfare system in Wisconsin to understand how caseworkers interacted with algorithms, their perspectives on these systems, as well as how decisions were made at the intersection of policies, child-welfare practice, and algorithms. This is important because caseworkers are not traditionally trained in "thinking statistically" about data, algorithms, and uncertainties but are legally mandated to input data, interact with algorithms, and make critical decisions about the most vulnerable children. This also raises new concerns about how the nature of professional practice is changing in the public sector where caseworkers are acquiring these new skills on the job. Continually engaging with caseworkers also allowed us to learn about the CWS ecosystem, critical decision points, institutional processes, and power imbalances. This further allowed us to learn about the discretionary choices that workers make as they fill in the gaps of algorithmic decision-making using their contextual knowledge of case circumstances, relevant protocols, and mandated policies. This helped us in examining how human-AI interaction plays out in practice and in developing design guidelines for building interactive AI systems that leverage aspects of both human intelligence and machine intelligence. Understanding these complexities in a sociotechnical system is imperative for mapping out the scope and utility of algorithmic systems as well as understanding the systemic constraints within which they must operate. Additionally, we examined the human factors of explainability and intelligibility and highlighted how different stakeholders (e.g., caseworkers, supervisors, clinical therapists, etc.) had different needs regarding these human factors and that a "one-size-fits-all" approach for algorithm design was not feasible. We also present a case study of an algorithmic tool that is collaboratively used by the child-welfare staff and instead of predicting outcomes, it tracks the trajectory of different outcomes over time. Being able to assess the trajectory of these outcomes over the life of cases offers higher utility to caseworkers and informs their decision-making processes because they are able to assess which cases require which levels of support, provide additional supervisory support for high-needs cases to inexperienced caseworkers, and ensure an equitable distribution of caseloads.

8.1.4 Computational Narrative Analysis to Derive Rich Contextual Signals

Government agencies collect and manage numerous amounts of data in the form of casenotes and plans, assessments, judicial reviews, home studies, and investigations through daily operations. Drawing upon CSCW theories of organizational memory, we can derive rich theoretical signals from this unstructured text to understand how human interactions and associated contextual factors translate into successful outcomes for citizens. This is especially true for child-welfare where casenotes are collaboratively written by child-welfare staff to create a roadmap of decisions made as well as circumstances surrounding those decisions. Casenotes also contain contextual details about a family's social support system, and children's well-being, as well as specific details about risk, protective, and systemic factors. We analyzed casenotes at the agency using computational narrative analysis techniques and highlighted patterns of invisible street-level discretionary work, systemic constraints, and power asymmetries. Interestingly, some of these patterns were not uncovered during the ethnography or outlined in job descriptions. These contextual factors have implications for algorithm design because they directly influence the data collection and decision-making processes. For instance, we learned that caseworkers support low, medium, and high-needs families in different capacities and that their labor in regard to these three groups can be supported by different technologies ranging from simple notification apps for checking medication schedules to algorithmic systems that can help caseworkers find services for their clients. We also use computational narrative analysis to highlight the uncertainties prevalent in cases that occur as a result of competing risk, protective, and systemic factors. These fluctuating uncertainties confound caseworkers' decision-making and the trajectory of cases. We further problematize the notion of risk as a static construct by highlighting the temporality and mediating effects of different risk and protective factors. Finally, we show how predictive risk models that are currently in use do more harm to families because they rely on data collection processes that lead to the over-surveillance of families and also harm social work practice because the empirical notion of risk does not align with the sociological understanding of risk which informs street-level decision-making.

8.1.5 Participatory Co-Design Process with Stakeholders

Building upon the foundation of the ethnography, theoretical framework, and computational narrative analysis, I contribute a holistic understanding of this complex ecosystem and showcase the tradeoffs that need to be managed for algorithm design: (1) at the human discretion level, I highlight different insertion points and goals of algorithms for augmenting practitioners' decision-making processes, (2) at the bureaucratic level, I highlight the constraints within which all decisions (human or algorithmic) must be made, (3) at the algorithmic level, I showcase how algorithm design can account for the uncertainties inherent within cases and support decision-making processes instead of providing predicted outcomes. My research has provided actionable steps for human-centered algorithm design to the agency leadership. This dissertation work highlights how the decision-making process through the 7ei algorithm can be further augmented by recommending trauma-responsive services based on case characteristics as well as further contextualizing data points using theoretical signals derived from casenote entries using NLP techniques. This dissertation work takes into account the decision-making ecosystem instead of simply focusing on a single decision point or task (e.g., assessing the risk of maltreatment at the hotline) and highlights how algorithmic decision-making can support workers' practices and augment the quality of human discretionary work. We provide guidelines for co-designing an NLP-based system that augments caseworkers' decision-making processes by surfacing critical factors derived from the casenotes of the family preservation team. The next step in this process is to recommend trauma-responsive services based on risk and protective factors derived from casenotes and provide explanations as to why such services would benefit the family. This will further help the agency leadership ensure that the decisions made by the child-welfare team are centered in evidence-based practice that leads to positive outcomes for families. The active participation of supervisors helped us co-design these requirements where we algorithmically support the caseworkers' cognitive environment without providing predicted decisions.

8.2 Implications for Human-AI Interaction

Public agencies offer an important setting for studying human-AI interaction for the following three reasons - 1) full automation in such settings is not desirable because of the ethical, legal, and safety concerns associated with high-stakes decision-making [280, 92]. This leads to observable and ongoing interactions between practitioners and algorithmic systems where practitioners learn to engage with such tools, assess their utility towards improving decision-making, as well as develop perspectives regarding which critical factors lead to the development of trust or distrust of such systems, 2) algorithms in the public sector are being used for several high-stakes decisions (e.g., risk of child maltreatment) which allows us to evaluate the key characteristics of algorithmic decision-making that offer utility to (or frustrate) practitioners at a given critical decision-point (e.g., assessing risk when an allegation is made at the hotline) and design systems that address these limitations and improve *interaction* and the practitioners' overall experience, and 3) decisions are collaboratively made which allows us to inspect how a team of caseworkers interacts and reasons with an algorithmic system and incorporates algorithmic decisions within their decision-making processes. Lai et al. [280] surveyed studies on Human-AI interaction and critiqued them based on decision tasks and the design space (i.e., the decision-making ecosystem) for AI assistance. Below, I follow these criteria and contextualize them with findings from this dissertation work and provide some implications for Human-AI interaction.

8.2.1 Defining and Evaluating Decision Tasks

Decision tasks or predicted outcomes (e.g., risk of recidivism, risk of maltreatment) in the public sector are poorly and inconsistently defined which poses a problem in assessing what constitutes a successful outcome or intervention. For instance, what is the likelihood of a person staying unemployed if no public assistance is offered to them? The answer is very likely dependent upon several circumstantial and contextual factors in the person's life that do not exist in the administrative data used to train machine learning algorithms. In this regard, we show in Chapter 7, how the empirical notion of the risk of maltreatment as predicted by risk assessment algorithms is inconsistent with the sociological understanding of risk that underscores child-welfare practice where caseworkers help mitigate several circumstantial and context-specific risk factors. Here, the underlying assumption is that human reasoning can benefit from counter-intuitive patterns derived from empirical models [280]. Such prediction tasks are also generally referred to as *AI for discovery* tasks [280]. For instance, ascertaining the risk of maltreatment as a function of prior involvement with public services (e.g., drug and alcohol services, housing authority, etc.) might help uncover patterns that aid decision-making and prevent future child maltreatment events. However, we found that this led to a cumulative distrust of such systems because the decision task was no longer aligned with the caseworkers' practice model and did not help inform their day-to-day decisions. Caseworkers shared that the knowledge of predicted long-term risk (i.e., the likelihood of being re-referred or placed in foster care within two years) did not help address risk factors that the families in the system experience (e.g., unstable housing or employment, lack of access to medical services) and did not help them acquire services for their clients. Call screeners in the Allegheny County case study also cast doubt on the Allegheny Family Screening Tool (AFST) by highlighting that the algorithm was not helping them find cases that might otherwise "slip through the cracks" but instead was highlighting the most obvious cases that seemed to "trip on every single crack they encountered" [256]. In addition, for empirical predictions to provide any utility, it is imperative that caseworkers are provided with explanations of predictions that especially highlight the counter-intuitive factors that do not traditionally align with their practice. Most of the *AI for discovery* tasks require domain expertise but domain experts are seldom included in the design of AI systems. The majority of studies conducted on human-AI interaction are conducted as isolated experimental trials of decision tasks on crowdsourcing platforms. Consequently, as highlighted in Chapters 3 and 4, we found that such predicted decision tasks developed without domain expertise did not generalize to decision tasks carried out by caseworkers within the bounds of policies and systemic resources. In addition, as is evident from the case study of the CANS algorithm presented in Chapter 3, seemingly cogent and well-aligned decision tasks can be at odds with one another in practice. In theory, it makes sense to predict the foster care placement setting and the subsidized guardianship rate based on the mental health assessment of foster children. That is, the higher the mental health needs of a foster child, the higher the level of care required and the higher the compensation offered to foster parents. However, in practice, we learned that due to a lack of foster homes in the system and low base compensation, the second-removed outcome (i.e., subsidized guardianship rate) became the primary outcome of interest and impacted how mental health assessments were conducted. It is imperative for caseworkers to maintain good working relationships with foster parents and financially support them so that caseworkers have homes to place children in. That is, a systemic issue is the critical driving factor that determines how decision tasks are carried out. Here, ethnographic engagement is necessary to fully understand the decision-making ecosystem and assess how human-AI interaction may occur within organizational pressures, time constraints, and limited resources. Next, I discuss the need to focus on the decision-making ecosystem (within which decision tasks are situated) and its implications for human-AI interaction.

8.2.2 The Need to Focus on the Decision-Making Ecosystem

There has been an overwhelming focus on high-stakes domains for studying human-AI interaction, however, as I note in the section above, the decision tasks are often evaluated in isolation and do not align with the decision-making reality of the domain. Moreover, AI systems often seek to replace or replicate human labor without a clear understanding of how they are expected to be integrated into existing workflows and inform decision-making [280]. Algorithmic decisions that are empirically driven often do not align with how practitioners engage in decision-making and as depicted in Chapters 3 and 4, they can find themselves translating information from two forms of assessments (algorithmic and clinical) which further leads to uncertainty and unreliable decision-making. As we show in Chapter 3, algorithms designed to support specific decision tasks (e.g., assess risk for sex trafficking, assess mental health for placement decisions and compensations) without a proper understanding of the decision-making process within which the decision task is embedded often erase vital contextual and systemic factors that are critical for informing decisions. For instance, caseworkers shared that the mental health assessment conducted via CANS did not help inform decisions because it only offered an isolated, live snapshot of exhibited behaviors and did not account for a child's ecosystem, the impact of trauma, or help acquire trauma-responsive interventions. Decision-making in high-stakes settings is a complex process that involves information sharing and discussions among several practitioners to reach consensus decisions. The first step here is to understand this decision-making ecosystem, critical decision points, and established workflows that inform day-to-day practice. An ongoing ethnographic engagement helped us understand which aspects of practice and decision-making processes can be supported technically and aided by AI systems as well as simple technologies. Examining the decision-making ecosystem is also imperative because researchers must understand the pragmatic constraints within which an AI system must operate. Creating the time and space for the collaborative use of the AI system is especially important in high-stress and fast-paced environments where practitioners carry high workloads and often do not have the time or cognitive bandwidth to engage with the system as initially intended. Consequently, regardless of their utility, AI systems hold the risk of becoming a safe default for practitioners and shifting accountability away from them and onto the system. The implementation and use of the CANS algorithms provide a case-in-point of this phenomenon. On the other hand, the 7ei algorithm was specifically designed with these constraints in mind where the agency leadership first created the time and space (i.e., specialized trauma-informed meetings) for its proper use as well as designed it for collaborative use. In addition, it is also necessary to recognize potential points of failure where the system might fail to offer utility. For instance, we highlight in Chapter 7, how uncertainty in cases can be visually depicted by deriving risk, protective, systemic, and procedural factors from casenotes and communicated to practitioners.

8.2.3 Bridging the HCI and AI Research Gap

Having understood the challenges associated with predicting decisions as well as the complexities and nuances of the decision-making ecosystem, we deliberately decided to design against the status quo of risk prediction and looked towards approaches where we could algorithmically support the caseworkers' decision-making processes. In high-stakes domains where improving human-AI interaction is paramount, AI research must be encapsulated within HCI methods because it helps us better understand stakeholders' interactions with the system and assess critical human factors such as trust, reliance, intelligibility, and explainability of the system. In addition, as noted above, AI systems designed for high-stakes domains often seek to replicate or replace human labor without an adequate understanding of the nature of practice and the reality of street-level decision-making. Here, HCI methods can help us understand organizational workflows, day-to-day protocols that workers are expected to follow, systemic factors (e.g., policies, resources, etc.), and the bottlenecks in decision-making. Consequently, this helps us outline the systemic mechanics as well as the insertion points, scope, and utility of algorithmic decision-making. That is, how do we design for and support the cognitive environment that would allow the practitioners to succeed? The *Seven Essential Ingredients* (7ei) algorithm discussed in Chapters 3 and 4, provides an interesting case study of a simple algorithmic tool specifically designed to overcome the obstacles described above. Having faced several frustrations with empirical models, the agency leadership engaged in theory-driven design and developed 7ei using the principles of trauma-informed care (TIC). This further allowed them to ensure that decisions made using

7ei are centered in the theory of practice. The agency also ensured that child-welfare staff were trained in TIC, created a collaborative setting to ensure the proper use of the tool, and decided to track multiple TIC outcomes over the life of cases instead of predicting them. Consequently, what we observed over an extended ethnography was that 7ei received collective buy-in and significantly better engagement from the caseworkers because the tool is centered in social work practice and facilitates collaborative decision-making.

As highlighted by Vaughan and Wallach [462], it is imperative to understand the needs of relevant stakeholders to be able to design for intelligibility techniques that meet these needs. Moreover, they show that simple models have proved to be just as accurate as complex neural networks in high-stakes domains. It is imperative to note here that 7ei was designed to cultivate the human factors of trust, reliance, and intelligibility. It facilitates the child-welfare staff's intelligibility needs by 1) providing relevant information, 2) centering this information in a theoretical framework (i.e., TIC), and 3) demonstrating compliance with evidence-based practice. The tool was also designed with the intent to decompose the algorithm and turn it into an open-ended and transparent process that promotes trust and reliance. From a human-AI interaction standpoint, it is imperative to note that designing for intelligibility needs preceded and impacted the caseworkers' engagement, trust, and reliance on the tool. Transparency and explainability are often endorsed as human factors that might facilitate engagement and improve decision-making (i.e., making the workings of the algorithm transparent and providing explanations for predictions). Here, the case study of the CANS algorithm provides a cautionary tale against prioritizing these human factors over the intelligibility needs of stakeholders. Even though CANS promotes transparency where all the predictors are accessible to caseworkers, it frustrates the child-welfare staff because it does not account for the principles of TIC and only offers an isolated view of the child. This further problematizes explainability because caseworkers were more invested in explanations of underlying trauma (based on their training in trauma-informed care). We also learned that different stakeholders had different needs in regard to explainability and a "one-size-fits-all" approach with respect to algorithm design may not be feasible here. For instance, the clinical therapist was more interested in explanations about the context surrounding worsening behaviors per CANS because worsening behaviors at the onset of therapy might signal that the child was finally working through and processing their underlying trauma. Consequently, we learned that transparency in regard to CANS promoted the manipulation of data and led to the diminishing of trust and reliance on the tool.

8.3 Implications for HCDS Methodology

In this section, I discuss implications for the human-centered data science methodology where we used quantitative methods in conjunction with qualitative methods to approach critical factors that impact decisions from different angles and develop a deeper understanding of the decision-making ecosystem. In the subsections below, I discuss the nuances and patterns of practice as well as points of alignments and misalignments between qualitative and quantitative approaches and how combining the two helped us undertake a deeper critical analysis and uncover underlying systemic mechanics that could not have been accomplished by either approach alone.

8.3.1 Nuances of Practice

In Chapters 3 and 4, we show how algorithms are used at a child-welfare agency where they help caseworkers make day-to-day decisions about families. Instead of focusing on a single algorithmic tool, we outline how a suite of algorithms are embedded within bureaucratic processes at critical decision points that caseworkers engage with through the course of their practice. We conducted a two-year ethnography followed by an interview study to understand caseworkers' daily interactions, points of utility and failure of algorithmic tools in informing decisions, the impact of policy and systemic factors, and caseworkers' perspectives in regard to algorithmic decision-making. Here, qualitative methods helped us understand the nuances of practice and how cases are assessed by child-welfare staff, and which contextual factors added differentiating information for cases that may appear similar based on a broad set of predictors (i.e. - risk, protective, and systemic factors). For instance, we learned about power asymmetries between caseworkers and foster parents which determined how data was collected about children on the CANS assessment where foster parents exercised more power than caseworkers. This is the case because there is a dearth of good foster homes in the system and it is imperative for caseworkers to maintain good working relationships with foster parents (so that there is always a roster of foster homes where children can be placed) and support them to avoid closing of these homes. Therefore, two cases might have entered the system due to the same target problem but the foster children in these two cases can have significantly different mental health assessments per CANS. Here, the power asymmetry between foster parents and caseworkers is the critical contextual factor that determines how data is collected and how subsequent algorithmic decisions are made about foster children. A longitudinal ethnographic engagement also helped us learn about the systemic and policy-related constraints within which all decisions must be made. Consequently, we recognized that algorithms in the public sector continually interact with human discretionary work (which establishes the decision-making processes) and bureaucratic factors (which establish the constraints within which all decisions must be made). HCDS as an interdisciplinary field draws theories and techniques from the social sciences and other disciplines relevant to the problem at hand to better understand interactions within an ecosystem. Following this approach, we reviewed scholarly work in human-computer interaction, science and technology studies, and public administration and embedded that knowledge into a sociotechnical framework for algorithmic decision-making in the public sector. This framework demonstrates the practical trade-offs between the three dimensions (i.e. - human discretion, bureaucratic processes, and algorithmic decision-making) at both the macro- and micro-levels of an algorithmic model and if trade-offs cannot be adequately balanced then researchers must consider non-algorithmic approaches. Critically, for HCDS, we show how qualitative approaches in conjunction with domain knowledge from relevant disciplines can help us develop generalizable frameworks that capture the decision-making ecosystem.

An ongoing ethnographic engagement also helped us assess the scope, utility, and insertion points for algorithmic systems. For instance, we learned that the child-welfare staff engaged in collaborative meetings where they assessed each case from a trauma-informed perspective and used an algorithmic tool to conduct this work. This setting provides several key advantages - 1) it provides the time and space for the proper use of an algorithmic tool, 2) it allows caseworkers to engage in collaborative decision-making, and 3) the seasoned members of the staff are able to provide their expertise to inexperienced caseworkers; an important consideration since the caseworker position experiences high turnover. We also learned about the messy data collection processes in this domain where caseworkers undertake a significant amount of data labor where they are mandated to collect information about families ranging from parenting assessments, to progress in parenting classes and services, medical information, and mental health information that is necessary for developing and operating algorithms. In addition, caseworkers also performed a significant amount of *repair work* [242] to make the algorithms work for their clients and not just meet the demands of policymakers. Understanding these underlying and often hidden labor practices are essential for conducting human-centered data science and developing human-centered algorithms. Finally, we learned about a critical data source (namely, caseworkers' casenotes) that contains rich, contextual, and situated information about cases that are not captured by quantitative assessments but are vital for unpacking the decision-making ecosystem.

In sum, qualitative methods helped us understand the decision-making ecosystem, assess the design space for algorithms that would foster human-AI interaction, as well as understand why and how a simple algorithmic tool that tracks outcomes over time instead of predicting them offers higher utility to the child-welfare staff. This is especially important because most algorithmic introduced in public sector agencies, healthcare settings, etc. seek to replicate or replace human labor instead of augmenting it. We show that understanding the discretionary choices that workers make as they navigate bureaucratic processes, conflicting interests, and algorithmic decisions can help us design interactive AI systems that complement workers' expertise and augment the quality of human discretionary work. Critically, embedding ourselves in the domain for an extended period of time helped us acquire domain knowledge and develop a deeper understanding of insights derived from the quantitative analysis that we would have missed otherwise. In the next section, I highlight some of these quantitative insights that were further augmented due to our ongoing engagement with the domain and its practitioners.

8.3.2 Patterns of Practice

In Chapters 6 and 7, we conducted a computational narrative analysis of a critical data source, namely, caseworkers' narratives that account for the majority of the data stored in child-welfare digital archives but had not been computationally studied yet. These two studies offer the first computational inspection of child-welfare casenotes and introduce them as a significant data source to the HCI and AI research communities. These casenotes are well-positioned to be studied computationally because the narratives about families are highly individual (i.e., they contain contextual and situated details about the risk and protective factors, and the family's support system) but still share fundamental similarities in terms of personas, power hierarchies, and the sequence of events. That is, casenotes are constrained in terms of certain topics, personas, and events which allows us to examine the narrative structure of these documents and compare them across different types of cases.

In Chapter 6, we highlight the invisible patterns of labor that caseworkers engage in as they help families navigate the child-welfare system. Surprisingly, we uncovered two themes that we did not find during the ethnography. We learned that caseworkers helped manage medical consent between birth parents and foster parents, helped ensure medication schedules for children were being followed by caregivers, and accompanied clients to medical appointments. We also learned that caseworkers conducted several quantitative assessments throughout the life of cases to assess risks, protective factors, and progress in these cases. Interestingly, we did not find these patterns of labor during the ethnography because the field site for ethnographic work was the child-welfare agency, however, this street-level labor is undertaken at families' homes, the doctors' offices, and service providers. That is, this labor was not observable during the ethnography but we were able to derive these signals computationally from the casenotes. In addition, we reviewed job descriptions of caseworker positions and did not find these labor practices outlined in them either. This further highlights how computational methods can help us study patterns of labor, how practices are changing on the street level, as well as the new skills that caseworkers are acquiring that may otherwise be unobservable. Human-Centered Data Science as a discipline also recognizes that there are many ways of *seeing* the data and that different epistemological viewpoints may draw our attention to data-driven insights differently. Looking at the invisible labor practices described above from a worker-centered perspective shows us that caseworkers are undertaking a significant amount of added labor that is not part of their job descriptions in their efforts to help families. However, looking at these labor practices from a social justice perspective, we also begin to see the over-surveillance of vulnerable families that takes place in child welfare where these families are continually subjected to data collection processes that the caseworkers are mandated to follow.

We also conducted a timeline analysis of topic popularity for low, medium, and high-needs families where we plotted the distribution of different topics over the life of cases for each of these groups. We learned that case characteristics (e.g., type of maltreatment, number/age of children, need for financial assistance, etc.) dictate how much time each family spends in the child-welfare system, and consequently, affects their number of interactions and the number of casenote entries written about the family. Analyzing these casenotes also helped us understand that caseworkers support these three groups (i.e., low-, medium-, and high-needs families) in different capacities and their work is impacted by different systemic constraints. For instance, caseworkers help low-needs families acquire essential resources such as employment, food, clothing, and preventive services. However, for medium-needs families, this involves finding court-ordered services for their clients such as domestic violence classes, AODA (alcohol and other drug abuse) classes, therapy, etc. We further illustrate that this means that different systemic constraints arise for these different groups. Here, a deeper understanding of a family's needs and systemic constraints that may arise can help agency leadership divide the workload more equitably such that new caseworkers are not managing multiple high-needs cases which generally results in burnout and a high turnover.

From a practice standpoint, it is also imperative to understand the power relationships that different personas involved in child-welfare cases experience because 1) power asymmetries directly impact decision-making and as depicted in Chapter 6, also impact the data collection processes and subsequent algorithmic decisions, 2) power relationships can help us in identifying the right intervention strategies and decisions as caseworkers engage with families, and 3) it can also help us in identifying the right person who has the agency to make critical decisions. For instance, short-term foster parents exercise the most agency for medium-needs families, and here, a designated child-welfare team that works with them would help ensure that new caseworkers have enough support when developing working relationships with short-term foster parents. Similarly, it is imperative to ensure that long-term foster parents who are trained and certified to care for high-needs children are adequately supported by experienced staff.

In sum, from the studies presented in Chapters 5, 6, and 7, we highlighted that the analysis of caseworker narratives can help uncover patterns of invisible labor, the timeline of critical events for different types of cases, and power relationships that impact street-level practices. We also highlight how uncertainty can be visually communicated to agency leadership as a result of competing and fluctuating factors (i.e., risk, protective, systemic, and procedural factors).

Computationally deriving these contextual signals can help child-welfare staff make informed decisions about cases under their care. This is of critical importance because child-welfare staff who is carrying high caseloads do not have the resources to manually review these casenotes and make important time-sensitive decisions.

8.3.3 Nuances of Practice vs. Patterns of Practice

Combining insights from qualitative and quantitative methods allowed us to understand patterns of labor practices, systemic mechanics, and the decision-making ecosystem at scale as well as deeply understand the nuances of street-level practices at depth [28]. Qualitative methods are oftentimes proposed as a means to further contextual and fill in the gaps not addressed by quantitative methods. Here, qualitative methods were indeed indispensable and as noted above, allowed us to fully gauge the scope and utility of algorithmic decision-making and understand how to design AI systems that augment human discretionary work instead of seeking to replicate or replace it. However, as noted in the section above, quantitative methods can help uncover surprising insights that are even visible to researchers conducting an extended ethnography. Therefore, beyond just complementing each other, the two approaches can generate two distinct sets of insights that may contextual and ratify each other but may also problematize findings from the other. Below I discuss these alignments, misalignments, and their implications.

Alignments. We learned about the power asymmetry between foster parents and caseworkers from the ethnography which was further ratified by the findings of the computational power analysis (CPA) presented in Chapter 6. However, CPA added more depth to this analysis by helping uncover power relationships between several different personas (i.e., caseworkers, foster parents, birth parents, support system, and children) that we had not initially considered but appeared regularly in casenotes. The results from CPA allowed us to discuss these findings with program directors and gather more contextual and hyperlocal insights regarding power relationships. That is, CPA positioned us to ask questions we did not know to ask initially and allowed us to gather more qualitative insights. Here, the relationship between qualitative and quantitative methods is better understood as a cyclic process where findings from one approach can facilitate and signal a need for a deeper investigation into the findings from the other approach. It is also interesting to note that the caseworkers' exercising less agency than foster parents in their day-to-day relationships impacted the linguistic choices that they made as they wrote the casenotes and were clearly picked by CPA. However, a computational method can only help derive insights that consistently exist in the data to be able to consider it a reliable underlying pattern. From the ethnography, we learned that legal and medical parties exercised significant agency in decision-making where the child-welfare staff is only able to make recommendations to them. However, these personas also appeared rather sparsely in the casenotes because the caseworkers interact significantly less with them as compared to foster parents, birth parents, children, and the family's support system. Similarly, from the ethnography, we understood that systemic and procedural factors play a significant role in decision-making, however, computational analysis in Chapter 7 added more nuance and depth to this analysis by outlining how these factors interplayed with each other and fluctuated together throughout the life of a case. This further allowed us to provide empirical evidence about uncertainty in decision-making that arises as a result of these competing factors.

Misalignments. Comparing across the qualitative and quantitative findings, we also found several serious discrepancies where we had initially expected the two sources to complement each other and contain consistent information (i.e., qualitative findings further contextualize quantitative findings). However, as depicted in Chapter 5, we found that several differences existed between the risk factors mentioned in the narrative coding of casenotes versus risk factors scored on the WARM risk assessment. For instance, a caseworker might determine that sexual abuse of a child occurred in a case and recorded it on the WARM assessment. However, due to the caregivers' protective actions (e.g., filing a police report against the perpetrator and getting a restraining order), no impending risk to the child exists and is recorded as such in the casenotes. As discussed in Chapter 5, here the discrepancy is significant enough that it would result in a high standard investigation with significant CPS involvement if the family is re-referred in the future. This prodded us to inspect these incongruities between the two data sources and draw attention to the data collection processes and faulty underlying assumptions that become embedded in risk assessments. For instance, we learned that initial assessment (IA) caseworkers relied more on their impressions of the family when conducting risk assessments because not much information is available at the onset of a case. They score variables such as "parents' cooperation with the agency" that inadvertently captures a parent's response to an intervention rather than the effectiveness of the intervention itself. Connecting these findings to social work literature, we learned that this problem is further compounded by a high turnover problem in the caseworker position. Inexperienced caseworkers are known to engage in self-defensive decision-making where they over-estimate risks and record them as such in risk assessments. Subsequently, this data becomes embedded in risk assessment algorithms that predict high-risk scores for families that have had prior involvement with CPS. Similarly, we learned that computational narrative analysis found signals that further contextualize and validated findings from the ethnography but we also found signals that problematize some qualitative findings. We attended numerous meetings where the child-welfare staff made collaborative decisions about cases from a trauma-informed perspective; centering them in evidence-based practice that promotes family well-being. However, the computational narrative analysis uncovered that several critical street-level decisions are still individually and arbitrarily made by caseworkers when interacting with families. We also learned that with high caseloads, most caseworkers engaged in defensive decision-making where they were more focused on completing their procedural tasks than helping families (e.g., caseworkers caring more about documentation about services, content forms, completion of assessments than helping parents navigate the system). That is, we learned that decisions were collaboratively reached in the trauma-informed meetings but did not necessarily mean that caseworkers followed through on these decisions when interacting with families. This key difference highlights why we need to pay close attention to the street-level practices of caseworkers (highlighted by computational narrative analysis) and not just the practice model implemented by the agency (highlighted by the ethnography). Finally, we learned during the ethnography that the child-welfare staff writes detailed narratives about cases, however, as depicted in Chapters 6 and 7, we were further able to contextualize this finding computationally by interrogating these casenotes. This allowed us to learn that unlike the casenotes written by IA caseworkers that contained information about perceived risks and were not credible, the casenotes written by the family preservation team contained detailed and credible information about cases since this team works closely with parents throughout the child-welfare process and understands their needs.

In sum, qualitative methods are often proposed as filling in the gaps of computational research and further contextualizing them. However, we learned that it is also possible for computational methods to contextualize, ratify, and/or problematize qualitative findings. This further highlights the need to employ both methodologies for developing a deeper understanding of complex sociotechnical systems.

8.4 Implications for AI Governance

Child-welfare (CW) agencies in the United States increasingly rely on federal and philanthropic funding to support their operations [425] with more grant funding opportunities made available to agencies that invest in innovative evidence-based programs. As highlighted by Simon et al. [425], 231 federal grants (total sum of $515,092,496) were awarded to agencies implementing **evidence-based programs** and 96 grants (total sum of $224,809,279) specifically awarded for **systems reform and integration**. In addition, the most funded projects by philanthropic organizations fall under that umbrella of **Data/mapping data sources** that seek to improve the collection of data and development of data systems related to child well-being and public child welfare systems [425]. Even though it is unclear how much of the federal grant funding seeking to support evidence-based programs were awarded to analytics projects, a recurring theme across the literature has been the need to improve child welfare data systems to facilitate the breadth and depth of data collection processes, improve case management, and data tracking [425, 226, 439, 225]. Recently, more federal funding has been made available to CW agencies that implement the new Comprehensive Child Welfare Information System (CCWIS) data system that further improves data collection mechanisms, integrates data from previously siloed departments, and supports bi-directional data exchange between courts, education systems, and Medicaid [225].

Even though it is inherently good practice for CW agencies to develop improved case management systems, we want to draw caution against these agencies uncritically acquiring CCWIS-based algorithmic systems developed by tech startups [239, 371, 447, 238, 33]. In Chapter 7, we provide evidence regarding how such off-the-shelf systems developed in one jurisdiction can do harm to families in another jurisdiction because the system does not account for the varying policies, practices, and systemic constraints across these different regions. CW agencies have continued to rely on counsel from the federal government in the form of initiatives, regulations, and evidence-based approaches yet federal directives have continually focused on the need for CW agencies to adopt data-driven practices without providing adequate guidelines that focus on the *why* and *how* to integrate these systems and train the workforce on their use. Consequently, CW agencies in several states have rushed to adopt "something" in order to prove that they are employing scientific and evidence-based practices — but without ensuring that child welfare stakeholders have a strong understanding of how the model works, how to assure fidelity, and how to assess the model for issues of ethics and equity. Here, there is a dire need for federal guidelines on AI regulation and governance especially as more states begin adopting algorithmic decision-making in child welfare. A recent survey conducted by the ACLU highlighted that 11 states are currently using predictive analytics in their CW agencies while 26 states are considering the use of these systems [391].

Moreover, as we show in Chapters 3 and 4, there are serious data provenance and ethical concerns regarding the use of algorithmic decision-making where caseworkers must not only undertake the data work necessary to run AI systems but also undertake added labor in the form of *repair work* [242] to make these systems work for their clients. Algorithms, in their current form, have also added more barriers to consistent and evidence-based decision-making and diminished the quality of social work practice. Additionally, tech startups are only focused on the costs associated with the initial development and deployment of an AI system. However, already underfunded CW agencies are assuming the costs of the added labor required to adopt and integrate the system into their decision-making processes. On the other hand, the 7ei algorithm discussed in Chapters 3 and 4 offers an interesting case study regarding the ethical and responsible use of a system that has received collective buy-in from the caseworkers and led to better outcomes for families under their care. However, agency leadership had to invest in a significant amount of resources in terms of trainings, specialized consultations, hiring experts, and creating the time and space (in terms of collaborative 7ei meetings) to ensure proper utilization of the algorithmic tool. That is, there is a significant amount of human labor that went into the integration of 7ei into daily work routines. It is imperative to note that these investments must be made in order to rebuild and improve decision-making processes that utilize algorithmic systems. Here, it is imperative to assess the sustainability of AI projects in the public sector and consider the costs associated with training workers, developing integration processes, and evaluating these systems.

Per the Family First Prevention Services Act (FFPSA) passed in 2018, child welfare is supposed to transition towards a "families as partners" model where the primary goal is to achieve reunification for children with their birth parents and where parents play an equal role in decision-making. However, as we highlight in Chapter 5, power asymmetries become embedded within risk assessments where caseworkers exert more power over birth parents, engage in defensive decision-making where they overestimate risks and make arbitrary decisions that harm families due to high caseloads and inability to devote the necessary time to each case. Consequently, algorithmic systems are creating a layer of obfuscation where such power asymmetries can become embedded within them. Algorithmic decision-making, as proposed by tech startups, is at odds with the goals of the Family First Prevention Services Act.

8.5 Future Work

In this section, I describe some future research trajectories that will further build upon the findings and lessons learned from this dissertation work.

8.5.1 Developing Responsible Deliberation Processes to Support AI Innovation.

We learned from the CANS algorithm case study that theoretical design and involving domain experts in the process can fall short if there is a gap in knowledge regarding the capabilities and limitations of algorithmic systems. The re-appropriation of the CANS assessment into the CANS algorithm stripped away the temporality of variables; a key consideration for domain experts when conducting the individualized assessment. The assessment was also designed using communimetrics theory to facilitate sharing of information among parties and not to measure or predict any outcomes [301]. Therefore, as a precursor to the algorithm design process, it is first necessary to deliberate on the capabilities and limitations of AI systems in non-technical terms with domain experts. To accomplish this in the future, researchers should consider developing toolkits (consisting of a blueprint of AI capabilities, their limitations, and practical, real-world use cases) that facilitate stakeholder deliberation and ideation of technologies that offer higher utility. We will conduct focus groups and/or workshops with the practitioners and anticipate this to be an iterative process. The goal here will be to create a standardized interactive process that employs the toolkit and is repeatable across other domains such as healthcare, and human resources, among others.

8.5.2 Human-AI Complementarity and Collaboration

Between the gaps of all data-driven decisions, practitioners make several implicit and explicit discretionary choices that arise from their contextual knowledge of the domain as well as professional expertise developed over time. In future studies, we will seek to further explore the implications of these discretionary choices and how to translate them into the design of interactive AI systems that augment the quality of human discretionary work. That is, how does the practitioners' tacit and explicit knowledge inform their choices and what are their implications for human-centered design? Incorporating these insights into systems design will ensure that the system is human-centered and situated in the nature of practice. This will further ensure that machine intelligence complements human intelligence and augments human decision-making. In addition, it is imperative to recognize that most critical decisions in high-stakes domains are collaboratively made and AI systems must not only complement the nature of practice but also make room for consensus decisions. Here, the next step will be to have discussions with practitioners about where and how the system integrates into their decision-making processes and what are the pragmatic constraints within which it must operate. That is, developing processes for the collaborative use of an AI system and recognizing the points of failure. Creating the time and space for the proper use of the AI system is especially important in high-stress and fast-paced environments where practitioners carry high workloads and often do not have the time or cognitive bandwidth to engage with the system as initially intended. Consequently, regardless of their utility, AI systems hold the risk of becoming a safe default for practitioners and shifting accountability away from them and onto the system. In addition, it is also necessary to recognize potential points of failure where the system might fail to offer utility. For instance, what does uncertainty in decision-making look like, and in which cases is it most likely to persist and accumulate? How can the system communicate this uncertainty (and possible sources) with practitioners to facilitate collaborative discussions?

8.5.3 Incorporating Newer Methodologies into Human-Centered Data Science

Human-Centered Data Science is a relatively new interdisciplinary field and it is necessary to introduce new methodological lenses to this field that facilitate a deeper and critical study of complex sociotechnical systems. For instance, researchers recently released a Python library for causal inference that supports the modeling and testing of causal assumptions [417]. There is special interest in the child-welfare research community comprising political scientists, social scientists, and computer scientists to study the causal pathways that children and families take through the child-welfare system and the impact of policies and systemic factors on these pathways. In addition, combining quantitative insights derived from casual modeling with qualitative member checks as conducted in Chapters 6 and 7 can reveal hyperlocal insights that further contextual the impact of policies on vulnerable communities. For instance, it would be interesting to study whether the Family First Prevention Services Act (FFPSA) passed in 2018 is beginning to improve reunification outcomes for families five years later. Similarly, Cambo and Gergle [87] recently introduced the concepts of model positionality and computational reflexivity that allow data scientists to deeply investigate datasets and communicate the social and cultural context surrounding a model's development and use. This is of critical importance in the public sector where the datasets are inherently a product of human data collection processes and embed human biases within the datasets. Similarly, Gordon et al. introduced jury learning [207] to showcase how practitioners can make explicit value judgments when seeking to resolve disagreements in contested tasks (e.g., content moderation) and disagreement deconvolution [208] that transforms machine learning metrics such as precision or ROC AUC to account for the underlying distribution of labels and better reflect the decision-making reality of often contested human-facing tasks, and consequently, facilitate a more human-centered evaluation of human-facing systems.

8.6 Concluding Remarks

This dissertation assumes the dual roles of *computing as rebuttal* and *computing as synecdoche* and beyond the specific contributions highlighted above, makes the following three broader contributions. First, it showcases how algorithmic decision-making in the public sector needs to be understood as a three-way interaction between human discretion, bureaucratic processes, and algorithms and contributes this knowledge to public interest technology literature. Second, it offers the first computational inspection of child-welfare casenotes and provides a critical case study to human-computer interaction (HCI) and critical computing literature of *designing against the status quo*, when *not-to-design*, and designing to algorithmically support the practitioners' cognitive environment without providing predicted decisions. Finally, it contributes a critical case study to the Human-Centered Data Science (HCDS) literature and showcases how HCDS methodologies can be used to uncover critical systemic interactions, invisible patterns of labor, power asymmetries, as well as experiences of affected communities that have direct implications for the design of human-centered algorithms in the public sector.

References

[1] 2013. *How the Child Welfare System Works.* Technical Report. Children's Bureau: Child Welfare Information Gateway.

[2] 2019. Child Welfare Goals, Legislation, and Monitoring. (2019). Retrieved April 2, 2019 from https://www.nccdglobal.org/what-we-do/children-s-research-center

[3] 2020. Charting the Landscape of Online Cryptocurrency Manipulation. *IEEE Access* 8 (2020), 113230–113245. DOI:http://dx.doi.org/10.1109/ACCESS.2020.3003370

[4] Senate of the United States 115th Congress. 2018. S. 3039 - Using Data To Help Protect Children and Families Act. (2018).

[5] J Khadijah Abdurahman. 2021. Calculating the Souls of Black Folk: Predictive Analytics in the New York City Administration for Children's Services. *Columbia Journal of Race and Law* 11, 4 (2021), 75–110.

[6] Rediet Abebe, Solon Barocas, Jon Kleinberg, Karen Levy, Manish Raghavan, and David G Robinson. 2020a. Roles for computing in social change. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency.* 252–260.

[7] Rediet Abebe, Salvatore Giorgi, Anna Tedijanto, Anneke Buffone, and H. Andrew Andrew Schwartz. 2020b. Quantifying Community Characteristics of Maternal Mortality Using Social Media *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 2976–2983. DOI:http://dx.doi.org/10.1145/3366423.3380066

[8] Abdurazzag A Aburas, Mohammad Hassan, Hilary Lin, and Shreshtha Batshu. 2018. Child Maltreatment Forecast Using Bigdata Intelligent Approaches. In *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS).* IEEE, 302–308.

[9] Mark S Ackerman. 2000a. The intellectual challenge of CSCW: The gap between social requirements and technical feasibility. *Human–Computer Interaction* 15, 2-3 (2000), 179–203.

[10] Mark S. Ackerman. 2000b. The Intellectual Challenge of CSCW: The Gap between Social Requirements and Technical Feasibility. *Hum.-Comput. Interact.* 15, 2 (Sept. 2000), 179–203.

[11] Mark S Ackerman and Christine Halverson. 2004. Organizational memory as objects, processes, and trajectories: An examination of organizational memory in use. *Computer Supported Cooperative Work (CSCW)* 13, 2 (2004), 155–189.

[12] Libby S Adler. 2001. The meanings of permanence: A critical analysis of the Adoption and Safe Families Act of 1997. *Harv. J. on Legis.* 38 (2001), 1.

[13] Social Security Administration. 2021. State-specific data. (2021). https://www.ssa.gov/OACT/babynames/limits.html

[14] Stefanía Ægisdóttir, Michael J White, Paul M Spengler, Alan S Maugherman, Linda A Anderson, Robert S Cook, Cassandra N Nichols, Georgios K Lampropoulos, Blain S Walker, Genna Cohen, and others. 2006. The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist* 34, 3 (2006), 341–382.

[15] Alekh Agarwal, Alina Beygelzimer, Miroslav Dudík, John Langford, and Hanna Wallach. 2018. A reductions approach to fair classification. In *International Conference on Machine Learning.* PMLR, 60–69.

[16] Ali Alkhatib and Michael Bernstein. 2019. Street-Level Algorithms: A Theory at the Gaps Between Policy and Decisions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* ACM, 530.

[17] Doris Allhutter, Florian Cech, Fabian Fischer, Gabriel Grill, and Astrid Mager. 2020. Algorithmic profiling of job seekers in Austria: how austerity politics are made effective. *Front. Big Data 3: 5. doi: 10.3389/fdata* (2020).

[18] Philip Alston. 2019. Report of the Special Rapporteur on Extreme Poverty and Human Rights, Seventy-Fourth Session. (2019).

[19] Samaneh Aminikhanghahi and Diane J Cook. 2017. A survey of methods for time series change point detection. *Knowledge and information systems* 51, 2 (2017), 339–367.

[20] Asbjørn Ammitzbøll Flügge, Thomas Hildebrandt, and Naja Holten Møller. 2021. Street-Level Algorithms and AI in Bureaucratic Decision-Making: A Caseworker Perspective. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–23.

[21] Kim Normann Andersen, Jungwoo Lee, and Helle Zinner Henriksen. 2020. Digital sclerosis? Wind of change for government and the employees. *Digital Government: Research and Practice* 1, 1 (2020), 1–14.

[22] Mark Andrejevic. 2019. Automating surveillance. *Surveillance & Society* 17, 1/2 (2019), 7–13.

[23] Rachmadita Andreswari, Irfan Darmawan, and Warih Puspitasari. 2018. A Preliminary Study on Detection System for Assessing Children and Foster Parents Suitability. In *2018 6th International Conference on Information and Communication Technology (ICoICT)*. IEEE, 376–379.

[24] Maria Antoniak. 2022. Modeling Personal Experiences Shared in Online Communities. (2022).

[25] Maria Antoniak and David Mimno. 2021. Bad Seeds: Evaluating Lexical Methods for Bias Measurement. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online, 1889–1904. DOI:http://dx.doi.org/10.18653/v1/2021.acl-long.148

[26] Maria Antoniak, David Mimno, and Karen Levy. 2019a. Narrative paths and negotiation of power in birth stories. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–27.

[27] Maria Antoniak, David Mimno, and Karen Levy. 2019b. Narrative Paths and Negotiation of Power in Birth Stories. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 27.

[28] Cecilia Aragon, Shion Guha, Marina Kogan, Michael Muller, and Gina Neff. 2022. *Human-Centered Data Science: An Introduction*. MIT Press.

[29] Veronika Arefieva, Roman Egger, and Joanne Yu. 2021. A machine learning approach to cluster destination image on Instagram. *Tourism Management* 85 (2021), 104318. DOI:http://dx.doi.org/https://doi.org/10.1016/j.tourman.2021.104318

[30] Florian Artinger, Malte Petersen, Gerd Gigerenzer, and Jürgen Weibler. 2015. Heuristics as adaptive decision strategies in management. *Journal of Organizational Behavior* 36, S1 (2015), S33–S52.

[31] Jos ARM Arts, Wim H Gijselaers, and Henny Boshuizen. 2000. Expertise Development in Managerial Sciences: The Use of Knowledge Types in Problem-Solving. (2000).

[32] Mariam Asad and Christopher A Le Dantec. 2015. Illegitimate civic participation: supporting community activists on the ground. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1694–1703.

[33] Augintel. 2022. SOCIAL IMPACT MEETS AI: We exist to give clinicians a deeper understanding of the people they care for. Link: https://www.augintel.us/about. (2022).

[34] Anna E Austin, Alexandria M Lesak, and Meghan E Shanahan. 2020. Risk and protective factors for child maltreatment: a review. *Current epidemiology reports* 7, 4 (2020), 334–342.

[35] Karla Badillo-Urquiola, Jaclyn Abraham, Arup Kumar Ghosh, and Pamela Wisniewski. 2018. A Stakeholders' Analysis of the Systems that Support Foster Care. In *Proceedings of the 2018 ACM Conference on Supporting Groupwork*. ACM, 158–161.

[36] Karla Badillo-Urquiola, Arup Kumar Ghosh, and Pamela Wisniewski. 2017a. Understanding the unique online challenges faced by teens in the foster care system. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 139–142.

[37] Karla Badillo-Urquiola, Scott Harpin, and Pamela Wisniewski. 2017b. Abandoned but Not Forgotten: Providing Access While Protecting Foster Youth from Online Risks. In *Proceedings of the 2017 Conference on Interaction Design and Children*. ACM, 17–26.

[38] Karla Badillo-Urquiola, Xinru Page, and Pamela Wisniewski. 2019. Risk vs. Restriction: The Digital Divide between Providing a Sense of Normalcy and Keeping Foster Teens Safe Online. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM.

[39] Christopher Baird, Dennis Wagner, Theresa Healy, and Kristen Johnson. 1999. Risk assessment in child protective services: Consensus and actuarial model reliability. *Child Welfare* 78, 6 (1999), 723.

[40] Zuriana Abu Bakar, Rosmayati Mohemad, Akbar Ahmad, and Mustafa Mat Deris. 2006. A comparative study for outlier detection techniques in data mining. In *2006 IEEE conference on cybernetics and intelligent systems*. IEEE, 1–6.

[41] Michàlle E Mor Barak, Amy Levin, Jan A Nissly, and Christianne J Lane. 2006. Why do they leave? Modeling child welfare workers' turnover intentions. *Children and Youth Services Review* 28, 5 (2006), 548–577.

[42] Shaowen Bardzell. 2014. Utopias of participation: design, criticality, and emancipation. In *Proceedings of the 13th Participatory Design Conference: Short Papers, Industry Cases, Workshop Descriptions, Doctoral Consortium papers, and Keynote abstracts-Volume 2*. ACM, 189–190.

[43] Shaowen Bardzell and Jeffrey Bardzell. 2011. Towards a feminist HCI methodology: social science, feminism, and HCI. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 675–684.

[44] Solon Barocas, Asia J Biega, Benjamin Fish, Jędrzej Niklas, and Luke Stark. 2020. When not to design, build, or deploy. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 695–695.

[45] Solon Barocas and Andrew D Selbst. 2016. Big data's disparate impact. *Calif. L. Rev.* 104 (2016), 671.

[46] Eric PS Baumer. 2017. Toward human-centered algorithm design. *Big Data & Society* 4, 2 (2017), 2053951717718854.

[47] Eric PS Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K Gay. 2017a. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology* 68, 6 (2017), 1397–1410.

[48] Eric PS Baumer and M Six Silberman. 2011. When the implication is not to design (technology). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2271–2274.

[49] Eric PS Baumer, Xiaotong Xu, Christine Chu, Shion Guha, and Geri K Gay. 2017b. When Subjects Interpret the Data: Social Media Non-use as a Case for Adapting the Delphi Method to CSCW. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1527–1543.

[50] S Bekaert, E Paavilainen, H Scheke, A Baldacchino, E Jouet, L Zablocka-Zytka, B Bachi, F Bartoli, G Carra, RM Cioni, and others. 2021. Family members' perspectives of child protection services, a metasynthesis of the literature. *Children and Youth Services Review* (2021), 106094.

[51] Jesse Josua Benjamin, Arne Berger, Nick Merrill, and James Pierce. 2021. Machine Learning Uncertainty as a Design Material: A Post-Phenomenological Inquiry. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[52] Bruce K Berger. 2005. Power over, power with, and power to relations: Critical reflections on public relations, the dominant coalition, and activism. *Journal of Public Relations Research* 17, 1 (2005), 5–28.

[53] Lawrence M Berger, Sarah K Bruch, Elizabeth I Johnson, Sigrid James, and David Rubin. 2009. Estimating the "impact" of out-of-home placement on child well-being: Approaching the problem of selection bias. *Child development* 80, 6 (2009), 1856–1876.

[54] Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. 2018. Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research* (2018), 0049124118782533.

[55] Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022. The values encoded in machine learning research. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 173–184.

[56] Joan M Blakey, Sonya J Leathers, Michelle Lawler, Tyreasa Washington, Chiralaine Natschke, Tonya Strand, and Quenette Walton. 2012. A review of how states are addressing placement stability. *Children and Youth Services Review* 34, 2 (2012), 369–378.

[57] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (Technology) is Power: A Critical Survey of "Bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 5454–5476. DOI:http://dx.doi.org/10.18653/v1/2020.acl-main.485

[58] Susanne Bødker, Christian Dindler, and Ole Sejer Iversen. 2017. Tying knots: Participatory infrastructuring at work. *Computer Supported Cooperative Work (CSCW)* 26, 1-2 (2017), 245–273.

[59] Alan Borning and Michael Muller. 2012. Next steps for value sensitive design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1125–1134.

[60] Herman L Boschken. 1994. Organizational performance and multiple constituencies. *Public Administration Review* 54, 3 (1994), 308–312.

[61] Emily Adlin Bosk. 2018. What counts? quantification, worker judgment, and divergence in child welfare decision making. *Human Service Organizations: Management, Leadership & Governance* 42, 2 (2018), 205–224.

[62] Mark Bovens and Stavros Zouridis. 2002. From street-level to system-level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control. *Public administration review* 62, 2 (2002), 174–184.

[63] Jordan Boyd-Graber, Yuening Hu, David Mimno, and others. 2017. Applications of topic models. *Foundations and Trends® in Information Retrieval* 11, 2-3 (2017), 143–296.

[64] Engin Bozdag. 2013. Bias in algorithmic filtering and personalization. *Ethics and information technology* 15, 3 (2013), 209–227.

[65] Ann E Brand and Paul M Brinich. 1999. Behavior problems and mental health contacts in adopted, foster, and nonadopted children. *The journal of child psychology and psychiatry and allied disciplines* 40, 8 (1999), 1221–1229.

[66] Virginia Braun and Victoria Clarke. 2006a. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[67] Virginia Braun and Victoria Clarke. 2006b. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. DOI:http://dx.doi.org/10.1191/1478088706qp063oa

[68] Leo Breiman and others. 2001. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science* 16, 3 (2001), 199–231.

[69] Meredith Brindley, James P Heyes, and Darrell Booker. 2018. Can Machine Learning Create an Advocate for Foster Youth? *Journal of Technology in Human Services* 36, 1 (2018), 31–36.

[70] Evelyn Z Brodkin. 2008. Accountability in street-level organizations. *Intl Journal of Public Administration* 31, 3 (2008), 317–336.

[71] Urie Bronfenbrenner. 1975. Reality and research in the ecology of human development. *Proceedings of the American Philosophical Society* 119, 6 (1975), 439–469.

[72] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-making in Child Welfare Services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 41.

[73] Jason D Brown, Natalie George, Jennifer Sintzel, and David St Arnault. 2009. Benefits of cultural matching in foster care. *Children and Youth Services Review* 31, 9 (2009), 1019–1024.

[74] Per Ms Brown and Per Jeremy. 2021. Caring for Children in Foster Care: The Challenges and How to Overcome Them. *Child Abuse & Neglect Conference: Prevention, Assessment, and Treatment* (2021).

[75] Justin B Bullock. 2019. Artificial intelligence, discretion, and bureaucracy. *The American Review of Public Administration* 49, 7 (2019), 751–761.

[76] Justin B Bullock, Jesper Rosenberg Hansen, and David J Houston. 2018. Sector Differences in Employee's Perceived Importance of Income and Job Security: Can These be Found Across the Contexts of Countries, Cultures, and Occupations? *International Public Management Journal* 21, 2 (2018), 243–271.

[77] Kimberly Bundy-Fazioli, Katharine Briar-Lawson, and Eric R Hardiman. 2009. A qualitative examination of power between child welfare workers and parents. *British Journal of Social Work* 39, 8 (2009), 1447–1464.

[78] United States Census Bureau. 2016. File B: Surnames occurring 100 or more times. (2016). https://www.census.gov/topics/population/genealogy/data/2010_surnames.html

[79] Jenna Burrell and Marion Fourcade. 2021. The society of algorithms. *Annual Review of Sociology* 47 (2021), 213–237.

[80] Peter André Busch. 2020. Crafting or mass-producing decisions: Technology as professional or managerial imperative in public policy implementation. *Information Polity* 25, 1 (2020), 111–128.

[81] Peter André Busch and Helle Zinner Henriksen. 2018. Digital discretion: A systematic literature review of ICT and street-level discretion. *Information Polity* 23, 1 (2018), 3–28.

[82] Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356, 6334 (2017), 183–186. DOI:http://dx.doi.org/10.1126/science.aal4230

[83] Marilyn Callahan and Karen Swift. 2018. The paradox of risk assessment, child safety and empowerment in child welfare. In *Revitalising communities in a globalising world*. Routledge, 67–77.

[84] Michael J Camasso and Radha Jagannathan. 1995. Prediction accuracy of the Washington and Illinois risk assessment instruments: An application of receiver operating characteristic curve analysis. *Social Work Research* 19, 3 (1995), 174–183.

[85] Michael J Camasso and Radha Jagannathan. 2000. Modeling the reliability and predictive validity of risk assessment in child protective services. *Children and Youth Services Review* 22, 11-12 (2000), 873–896.

[86] Michael J Camasso and Radha Jagannathan. 2013. Decision making in child protective services: A risky business? *Risk analysis* 33, 9 (2013), 1636–1649.

[87] Scott Allen Cambo and Darren Gergle. 2022. Model Positionality and Computational Reflexivity: Promoting Reflexivity in Data Science. In *CHI Conference on Human Factors in Computing Systems*. 1–19.

[88] Andrea Campagner, Federico Cabitza, and Davide Ciucci. 2020. Three-way decision for handling uncertainty in machine learning: A narrative review. In *International Joint Conference on Rough Sets*. Springer, 137–152.

[89] Sarah Carnochan, Megan Moore, and Michael J Austin. 2013. Achieving placement stability. *Journal of Evidence-Based Social Work* 10, 3 (2013), 235–253.

[90] Sarah Carnochan, Sarah Taylor, Anne Abramson-Madden, Meekyung Han, Sonja Rashid, Jennifer Maney, Sarah Teuwen, and Michael J Austin. 2006. Child welfare and the courts: An exploratory study of the relationship between two complex systems. *Journal of Public Child Welfare* 1, 1 (2006), 117–136.

[91] Eleanor Carter, Clare FitzGerald, Ruth Dixon, Christina Economy, Tanyah Hameed, and Mara Airoldi. 2018. Building the tools for public services to secure better outcomes: Collaboration, Prevention, Innovation. *Government Outcomes Lab, University of Oxford, Blavatnik School of Government* (2018).

[92] Anto Čartolovni, Ana Tomičić, and Elvira Lazić Mosler. 2022. Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review. *International Journal of Medical Informatics* 161 (2022), 104738.

[93] Spencer P Chainey. 2013. Examining the influence of cell size and bandwidth size on kernel density estimation crime hotspot maps for predicting spatial patterns of crime. *Bulletin of the Geographical Society of Liege* 60 (2013), 7–19.

[94] Paul Chalekian. 2013. POSDCORB: Core patterns of administration. In *Proceedings of the 20th Conference on Pattern Languages of Programs*. The Hillside Group, 17.

[95] Stevie Chancellor, Zhiyuan Lin, Erica L Goodman, Stephanie Zerwas, and Munmun De Choudhury. 2016a. Quantifying and predicting mental illness severity in online pro-eating disorder communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 1171–1184.

[96] Stevie Chancellor, Zhiyuan Lin, Erica L Goodman, Stephanie Zerwas, and Munmun De Choudhury. 2016b. Quantifying and predicting mental illness severity in online pro-eating disorder communities. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 1171–1184.

[97] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016c. #thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing (CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1201–1213.

[98] Janet X Chen, Allison McDonald, Yixin Zou, Emily Tseng, Kevin Roundy, Acar Tamersoy, Florian Schaub, Thomas Ristenpart, and Nicola Dell. 2022. Trauma-Informed Computing: Towards Safer Technology Experiences for All. *Under review* (2022).

[99] Hao-Fei Cheng, Logan Stapleton, Anna Kawakami, Venkatesh Sivaraman, Yanghuidi Cheng, Diana Qing, Adam Perer, Kenneth Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022a. How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions. In *CHI Conference on Human Factors in Computing Systems*. 1–22.

[100] Hao-Fei Cheng, Logan Stapleton, Anna Kawakami, Venkatesh Sivaraman, Yanghuidi Cheng, Diana Qing, Adam Perer, Kenneth Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022b. How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 162, 22 pages. DOI:http://dx.doi.org/10.1145/3491102.3501831

[101] Hao-Fei Cheng, Logan Stapleton, Ruiqi Wang, Paige Bullock, Alexandra Chouldechova, Zhiwei Steven Steven Wu, and Haiyi Zhu. 2021. Soliciting Stakeholders' Fairness Notions in Child Maltreatment Predictive Systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[102] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'Connell, Terrance Gray, F Maxwell Harper, and Haiyi Zhu. 2019. Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–12.

[103] Ka Ho Brian Chor. 2013. Overview of out-of-home placements and placement decision-making in child welfare. *Journal of Public Child Welfare* 7, 3 (2013), 298–328.

[104] Ka Ho Brian Chor, Gary M McClelland, Dana A Weiner, Neil Jordan, and John S Lyons. 2012. Predicting outcomes of children in residential treatment: A comparison of a decision support algorithm and a multidisciplinary team decision model. *Children and Youth Services Review* 34, 12 (2012), 2345–2352.

[105] Ka Ho Brian Chor, Gary M McClelland, Dana A Weiner, Neil Jordan, and John S Lyons. 2015. Out-of-home placement decision-making and outcomes in child welfare: A longitudinal study. *Administration and Policy in Mental Health and Mental Health Services Research* 42, 1 (2015), 70–86.

[106] Alexandra Chouldechova, Diana Benavides-Prado, Oleksandr Fialko, and Rhema Vaithianathan. 2018. A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Conference on Fairness, Accountability and Transparency*. PMLR, 134–148.

[107] Angèle Christin, Alex Rosenblat, and Danah Boyd. 2015. Courts and predictive algorithms. *Data & CivilRight* (2015).

[108] Christopher E Church and Amanda J Fairchild. 2017. In Search of a Silver Bullet: Child Welfare's Embrace of Predictive Analytics. *Juvenile and Family Court Journal* 68, 1 (2017), 67–81.

[109] Kevin Clancy, Joseph Chudzik, Aleksandra J Snowden, and Shion Guha. 2022. Reconciling data-driven crime analysis with human-centered algorithms. *Cities* 124 (2022), 103604.

[110] Madeleine Clare Elish and Elizabeth Anne Watkins. 2020. Repairing Innovation: A Study of Integrating AI in Clinical Care. *Data and Society* (2020).

[111] Shelby L Clark, Ashley N Palmer, Becci A Akin, Stacy Dunkerley, and Jody Brook. 2020. Investigating the relationship between trauma symptoms and placement instability. *Child Abuse & Neglect* 108 (2020), 104660.

[112] Victoria Clarke, Virginia Braun, and Nikki Hayfield. 2015. Thematic analysis. *Qualitative psychology: A practical guide to research methods* (2015), 222–248.

[113] Elysia V Clemens, Kristin Klopfenstein, Matt Tis, and Trent L Lalonde. 2017. Educational stability policy and the interplay between child welfare placements and school moves. *Children and Youth Services Review* 83 (2017), 209–217.

[114] James Clifford and George E Marcus. 1986. *Writing culture: The poetics and politics of ethnography.* Univ of California Press.

[115] Patricia Cohen, Stephen G West, and Leona S Aiken. 2014. *Applied multiple regression/correlation analysis for the behavioral sciences.* Psychology Press.

[116] Rob Comber, Shaowen Bardzell, Jeffrey Bardzell, Mike Hazas, and Michael Muller. 2020. Announcing a new CHI subcommittee: Critical and Sustainable Computing. *ACM Interactions* (July 2020). https://interactions.acm.org/blog/view/announcing-a-new-chi-subcommittee-critical-and-sustainable-computing

[117] US Congress. 2008. Fostering connections to success and increasing adoptions act of 2008. (2008).

[118] Roxanne Connelly, Christopher J Playford, Vernon Gayle, and Chris Dibben. 2016. The role of administrative data in the big data revolution in social science research. *Social science research* 59 (2016), 1–12.

[119] Lindsay D Cooper. 2005. Implications of media scrutiny for a child protection agency. *J. Soc. & Soc. Welfare* 32 (2005), 107.

[120] Ned Cooper, Tiffanie Horne, Gillian R Hayes, Courtney Heldreth, Michal Lahav, Jess Holbrook, and Lauren Wilcox. 2022. A Systematic Review and Thematic Analysis of Community-Collaborative Approaches to Computing Research. In *CHI Conference on Human Factors in Computing Systems.* 1–18.

[121] Victoria A Copeland. 2021. “It’s the Only System We’ve Got”: Exploring Emergency Response Decision-Making in Child Welfare. *Columbia Journal of Race and Law* 11, 3 (2021), 43–74.

[122] Katharan D Cordell, Lonnie R Snowden, and Laura Hosier. 2016. Patterns and priorities of service need identified through the Child and Adolescent Needs and Strengths (CANS) assessment. *Children and Youth Services Review* 60 (2016), 129–135.

[123] Common Core. 2018. Knowledge & Skill Reinforcement Lab: Case Planning Service Delivery Block. (2018).

[124] Michael Corrigan. 2019. Building A Comprehensive Child Welfare Information System. (Jan 2019). https://chronicleofsocialchange.org/child-welfare-2/building-comprehensive-child-welfare-information-system/33426

[125] Allegheny County. 2017. Developing predictive risk models to support child maltreatment hotline screening decisions. *Pittsburgh, PA, Allegheny County Department of Human Services, available online at: https://www. alleghenycountyanalytics. us/index. php/2017/04/17/developing-predictive-risk-models-support-child-maltreatment-hotline-screening-decisions/(accessed November 22, 2018)* (2017).

[126] John W Cresswell and others. 2012. Educational research: Planning, conducting, and evaluating quantitative and qualitative research. *Lincoln: Pearson* (2012).

[127] John W. Creswell and Dana L. Miller. 2000. Determining Validity in Qualitative Inquiry. *Theory Into Practice* 39, 3 (2000), 124–130.

[128] Theodore P Cross, EUN Koh, Nancy Rolock, and Jennifer Eblen-Manning. 2013. Why do children experience multiple placement changes in foster care? Content analysis on reasons for instability. *Journal of Public Child Welfare* 7, 1 (2013), 39–58.

[129] Stephanie Cuccaro-Alamin, Regan Foust, Rhema Vaithianathan, and Emily Putnam-Hornstein. 2017. Risk assessment and decision making in child protective services: Predictive risk modeling in context. *Children and Youth Services Review* 79 (2017), 291–298.

[130] Hannah Curtis. 2020. *Dimensions of Discomfort: Examining Child Welfare Professionals' Approach Toward Gender Diverse Foster Youth.* University of Washington.

[131] Amy D'andrade, Michael J Austin, and Amy Benton. 2008. Risk and safety assessment in child welfare: Instrument comparisons. *Journal of evidence-based social work* 5, 1-2 (2008), 31–56.

[132] Amy C D'Andrade and Ruth M Chambers. 2012. Parental problems, case plan requirements, and service targeting in child welfare reunification. *Children and Youth Services Review* 34, 10 (2012), 2131–2138.

[133] David Danks and Alex John London. 2017. Algorithmic Bias in Autonomous Systems.. In *IJCAI*. 4691–4697.

[134] EW Danktert and Kristen Johnson. 2013. Risk assessment validation: A prospective study. *Los Angeles: California Department of Social Services, Children and Family Services Division* (2013).

[135] Mark Davidson. 2020. Extreme municipal fiscal stress and austerity? A case study of fiscal reform after Chapter 9 bankruptcy. *Environment and Planning C: Politics and Space* 38, 3 (2020), 522–538.

[136] Elizabeth Davoren. 1975. Foster placement of abused children. *Children today* 4, 3 (1975), 41.

[137] Maria De-Arteaga, Riccardo Fogliato, and Alexandra Chouldechova. 2020. A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[138] Munmun De Choudhury and Emre Kiciman. 2018. Integrating Artificial and Human Intelligence in Complex, Sensitive Problem Domains: Experiences from Mental Health. *AI Magazine* 39, 3 (2018), 69–80.

[139] James W Dean Jr and Mark P Sharfman. 1993. Procedural rationality in the strategic decision-making process. *Journal of management Studies* 30, 4 (1993), 587–610.

[140] Ramit Debnath, Sarah Darby, Ronita Bardhan, Kamiar Mohaddes, and Minna Sunikka-Blank. 2020. Grounded reality meets machine learning: A deep-narrative analysis framework for energy policy research. *Energy Research & Social Science* 69 (2020), 101704.

[141] Ganesh Chandra Deka. 2014. Big data predictive and prescriptive analytics. In *Handbook of research on cloud infrastructures for Big Data analytics*. IGI Global, 370–391.

[142] Ramona W Denby, Mark F Testa, Keith A Alford, Chad L Cross, and Jesse A Brinson. 2017. Protective factors as mediators and moderators of risk effects on perceptions of child well-being in kinship care. *Child Welfare* 95, 4 (2017), 111–136.

[143] Benoît Desmarchelier, Faridah Djellal, and Faïz Gallouj. 2019. Innovation in public services in the light of public administration paradigms and service innovation perspectives. *European Review of Service Economics and Management* 8 (2019), 91–120.

[144] Alan J Dettlaff, Stephanie L Rivaux, Donald J Baumann, John D Fluke, Joan R Rycraft, and Joyce James. 2011. Disentangling substantiation: The influence of race, income, and risk on the substantiation decision in child welfare. *Children and Youth Services Review* 33, 9 (2011), 1630–1637.

[145] Lynn Dombrowski, Adriana Alvarado Garcia, and Jessica Despard. 2017. Low-wage precarious workers' sociotechnical practices working towards addressing wage theft. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 4585–4598.

[146] Lynn Dombrowski, Gillian R Hayes, Melissa Mazmanian, and Amy Voida. 2014. E-government intermediaries and the challenges of access and trust. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 2 (2014), 1–22.

[147] David Donoho. 2017. 50 years of data science. *Journal of Computational and Graphical Statistics* 26, 4 (2017), 745–766.

[148] Howard J Doueck, Diana J English, Diane DePanfilis, and Gerald T Moote. 1993. Decision-making in child protective services: a comparison of selected risk-assessment systems. *Child Welfare* 72, 5 (1993), 441–452.

[149] Paul Dourish and Edgar Gómez Cruz. 2018. Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society* 5, 2 (2018), 2053951718784083.

[150] Andy Dow, Rob Comber, and John Vines. 2018. Between grassroots and the hierarchy: Lessons learned from the design of a public services directory. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[151] Grant Drawve, Joel Caplan, and Michael Ostermann. 2019. Utilizing a risk of crime measure for recidivism research: Moving environmental corrections forward. *Crime & Delinquency* 65, 5 (2019), 606–629.

[152] Julia Dressel and Hany Farid. 2018. The accuracy, fairness, and limits of predicting recidivism. *Science advances* 4, 1 (2018), eaao5580.

[153] Ellen A Drost. 2011. Validity and reliability in social science research. *Education Research and perspectives* 38, 1 (2011), 105–123.

[154] Joe Duffy and Mary Elizabeth Collins. 2010. Macro impacts on caseworker decision-making in child welfare: A cross-national comparison. *European Journal of Social Work* 13, 1 (2010), 35–54.

[155] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*. 214–226.

[156] Andrea Lane Eastman, Lisa Schelbe, and Jacquelyn McCroskey. 2019. A content analysis of case records: Two-generations of child protective services involvement. *Children and Youth Services Review* 99 (2019), 308–318.

[157] Frank Edwards and Christopher Wildeman. 2018. Characteristics of the front-line child welfare workforce. *Children and Youth Services Review* 89 (2018), 13–26.

[158] Raquel T Ellis. 2010. Child welfare workers' perceptions of juvenile court influence on child welfare practices. *Journal of public child welfare* 4, 2 (2010), 158–173.

[159] Phoebe C Ellsworth. 2005. Legal Reasoning (The Cambridge Handbook of Thinking and Reasoning, KJ Holyoak and RG Morrison Jr.(eds),(New York. (2005).

[160] Diana J English, Sherry C Brummel, J Christopher Graham, and Laura K Coghlan. 2002. Factors that influence the decision not to substantiate a CPS referral phase II: Mail and telephone surveys of child protective services social workers. *Management Services Division, Office of Children's Administration Research* (2002).

[161] Guy Enosh and Tali Bayer-Topilsky. 2015. Reasoning and bias: Heuristics in safety assessment and placement decisions for children at risk. *The British Journal of Social Work* 45, 6 (2015), 1771–1787.

[162] K Anders Ericsson and Jacqui Smith. 1991. *Toward a general theory of expertise: Prospects and limits.* Cambridge University Press.

[163] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

[164] Ali Farazmand. 2009. *Bureaucracy and administration.* CRC Press.

[165] Ali Farazmand. 2017. Governance reforms: The good, the bad, and the ugly; and the sound: Examining the past and exploring the future of public organizations. *Public Organization Review* 17, 4 (2017), 595–617.

[166] Nicole Fedoravicius, J Curtis McMillen, Jill E Rowe, Njeri Kagotho, and Norma C Ware. 2008. Funneling child welfare consumers into and through the mental health system: Assessment, referral, and quality issues. *Social Service Review* 82, 2 (2008), 273–290.

[167] Sergio Fernandez and Hal G Rainey. 2017. Managing successful organizational change in the public sector. In *Debating public administration*. Routledge, 7–26.

[168] Anjalie Field, Su Lin Blodgett, Zeerak Waseem, and Yulia Tsvetkov. 2021. A Survey of Race, Racism, and Anti-Racism in NLP. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online, 1905–1925. DOI:http://dx.doi.org/10.18653/v1/2021.acl-long.149

[169] David Finkelhor. 2018. Screening for Adverse Childhood Experiences (ACEs): Cautions and Suggestions. *Child abuse & neglect* 85 (2018), 174–179.

[170] John D Fluke, Martin Chabot, Barbara Fallon, Bruce MacLaurin, and Cindy Blackstock. 2010. Placement decisions and disparities among aboriginal groups: An application of the decision making ecology through multi-level analysis. *Child Abuse & Neglect* 34, 1 (2010), 57–69.

[171] NCCPR: National Coalition for Child Protection Reform. 2017. Los Angeles County quietly drops its first child welfare predictive analytics experiment. Link: https://www.nccprblog.org/2017/05/los-angeles-county-quietly-drops-its.html. (2017).

[172] Administration for Children and Families. *Comprehensive Child Welfare Information System* (106 ed.). Vol. 81. Federal Register: The Daily Journal of the United States.

[173] Capacity Building Center for States. 2018. Child Protective Services: A Guide for Caseworkers. (2018). https://www.childwelfare.gov/pubPDFs/cps2018.pdf

[174] Patrick J Fowler, Katherine E Marcal, Saras Chung, Derek S Brown, Melissa Jonson-Reid, and Peter S Hovmand. 2019. Scaling up housing services within the child welfare system: policy insights from simulation modeling. *Child maltreatment* (2019), 1077559519846431.

[175] Sarah Fox, Jill Dimond, Lilly Irani, Tad Hirsch, Michael Muller, and Shaowen Bardzell. 2017. Social Justice and Design: Power and oppression in collaborative systems. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 117–122.

[176] Sarah E Fox, Vera Khovanskaya, Clara Crivellaro, Niloufar Salehi, Lynn Dombrowski, Chinmay Kulkarni, Lilly Irani, and Jodi Forlizzi. 2020. Worker-Centered Design: Expanding HCI Methods for Supporting Labor. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[177] Patricia Fredericksen and Rosanne London. 2000. Disconnect in the hollow state: The pivotal role of organizational capacity in community-based development organizations. *Public Administration Review* 60, 3 (2000), 230–239.

[178] H George Frederickson, Kevin B Smith, Christopher W Larimer, and Michael J Licari. 2015. *The public administration theory primer*. Westview Press.

[179] Batya Friedman, Peter Kahn, and Alan Borning. 2002. Value sensitive design: Theory and methods. *University of Washington technical report* 2-12 (2002).

[180] David J Fritzsche. 1991. A model of decision-making incorporating ethical values. *Journal of Business Ethics* 10, 11 (1991), 841–852.

[181] R Michael Furr. 2017. *Psychometrics: an introduction*. Sage Publications.

[182] Mahita Gajanan. 2020. The Heartbreaking Story Behind Netflix's Documentary Series The Trials of Gabriel Fernandez. *Time.com* (Mar 2020). https://time.com/5790549/gabriel-fernandez-netflix-documentary/

[183] Yarin Gal and Zoubin Ghahramani. 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*. PMLR, 1050–1059.

[184] Nicola K Gale, Gareth M Thomas, Rachel Thwaites, Sheila Greenfield, and Patrick Brown. 2016. Towards a sociology of risk work: A narrative review and synthesis. *Sociology Compass* 10, 11 (2016), 1046–1071.

[185] Ryan J. Gallagher, Kyle Reing, David Kale, and Greg Ver Steeg. 2017a. Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge. *Transactions of the Association for Computational Linguistics* 5 (12 2017), 529–542. DOI:http://dx.doi.org/10.1162/tacl_a_00078

[186] Ryan J Gallagher, Kyle Reing, David Kale, and Greg Ver Steeg. 2017b. Anchored correlation explanation: Topic modeling with minimal domain knowledge. *Transactions of the Association for Computational Linguistics* 5 (2017), 529–542.

[187] Eileen Gambrill and Aron Shlonsky. 2000. Risk assessment in context. (2000).

[188] Eileen Gambrill and Aron Shlonsky. 2001. The need for comprehensive risk management systems in child welfare. *Children and Youth Services Review* 23, 1 (2001), 79–107.

[189] Child Welfare Information Gateway. 2022a. Community and Environmental Factors That Contribute to Child Abuse and Neglect.
Link: https://www.childwelfare.gov/topics/can/factors/environmental/. (2022).

[190] Child Welfare Information Gateway. 2022b. Family Preservation Services. (2022). https://www.childwelfare.gov/topics/supporting/preservation

[191] Child Welfare Information Gateway. 2022c. Poverty and Economic Conditions.
https://www.childwelfare.gov/topics/can/factors/contribute/environmental. (2022).

[192] Child Welfare Information Gateway. 2022d. Risk Factors That Contribute to Child Abuse and Neglect.
Link: https://www.childwelfare.gov/topics/can/factors/. (2022).

[193] Child Welfare Information Gateway. 2022e. Substance Use by Parents and Caregivers.
Link: www.childwelfare.gov/topics/can/factors/parentcaregiver/substance. (2022).

[194] Jennifer M Geiger, Megan Hayes Piel, and Francie J Julien-Chinn. 2017. Improving relationships in child welfare practice: Perspectives of foster care providers. *Child and Adolescent Social Work Journal* 34, 1 (2017), 23–33.

[195] Jennifer M Geiger and Lisa Schelbe. 2021a. Assessment in Child Welfare Practice. In *The Handbook on Child Welfare Practice*. Springer, 195–217.

[196] Jennifer M Geiger and Lisa Schelbe. 2021b. Foster Care Placement. In *The Handbook on Child Welfare Practice*. Springer, 219–248.

[197] Rebecca Giallo, Holly Rominov, Catherine Fisher, Kirsty Evans, and Ali Fogarty. 2020. Preservation and reunification for families of young children: case file review of a home-visiting program. *Journal of reproductive and infant psychology* (2020), 1–13.

[198] Sarah Giest and Nadine Raaphorst. 2018. Unraveling the hindering factors of digital public service delivery at street-level: the case of electronic health records. *Policy Design and Practice* 1, 2 (2018), 141–154.

[199] Gerd Gigerenzer and Wolfgang Gaissmaier. 2011. Heuristic decision making. *Annual review of psychology* 62 (2011), 451–482.

[200] Fabrizio Gilardi, Charles R. Shipan, and Bruno Wüest. 2021. Policy Diffusion: The Issue-Definition Stage. *American Journal of Political Science* 65, 1 (2021), 21–35.

[201] Philip Gillingham. 2018. Decision-making about the adoption of information technology in social welfare agencies: Some key considerations. *European Journal of Social Work* 21, 4 (2018), 521–529.

[202] Philip Gillingham. 2019. Decision Support Systems, Social Justice and Algorithmic Accountability in Social Work: A New Challenge. *Practice* 31, 4 (2019), 277–290.

[203] Joan Gilroy. 2018. Critical issues in child welfare: Perspectives from the field. In *Community Approaches to Child Welfare*. Routledge, 25–44.

[204] James Gladstone, Gary Dumbrill, Bruce Leslie, Andrew Koster, Michelle Young, and Afisi Ismaila. 2014. Understanding worker–parent engagement in child protection casework. *Children and Youth Services Review* 44 (2014), 56–64.

[205] James P Gleeson. 1987. Implementing Structured Decision-Making Procedures at Child Welfare Intake. *Child Welfare* 66, 2 (1987).

[206] Connie Golsteijn, Sarah Gallacher, Licia Capra, and Yvonne Rogers. 2016. Sens-Us: Designing innovative civic technology for the public good. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 39–49.

[207] Mitchell L Gordon, Michelle S Lam, Joon Sung Park, Kayur Patel, Jeff Hancock, Tatsunori Hashimoto, and Michael S Bernstein. 2022. Jury learning: Integrating dissenting voices into machine learning models. In *CHI Conference on Human Factors in Computing Systems*. 1–19.

[208] Mitchell L Gordon, Kaitlyn Zhou, Kayur Patel, Tatsunori Hashimoto, and Michael S Bernstein. 2021. The disagreement deconvolution: Bringing machine learning performance metrics in line with reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[209] DeLeith Duke Gossett. 2017. The Client: How States Are Profiting from the Child's Right to Protection. *U. Mem. L. Rev.* 48 (2017), 753.

[210] Stuart Gray, Kirsten Cater, Chloe Meineck, Rachel Hahn, Debbie Watson, and Tom Metcalfe. 2019. trove: A digitally enhanced memory box for looked after and adopted children. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. ACM, 458–463.

[211] Ben Green and Yiling Chen. 2020. Algorithmic risk assessments can alter human decision-making processes in high-stakes government contexts. *arXiv preprint arXiv:2012.05370* (2020).

[212] Travis Greene, Galit Shmueli, Jan Fell, Ching-Fu Lin, Mark L Shope, and Han-Wei Liu. 2020. The Hidden Inconsistencies Introduced by Predictive Algorithms in Judicial Decision Making. *arXiv preprint arXiv:2012.00289* (2020).

[213] Robin Gregory, Lee Failing, Michael Harstone, Graham Long, Tim McDaniels, and Dan Ohlson. 2012. *Structured decision making: a practical guide to environmental management choices*. John Wiley & Sons.

[214] Nina Grgić-Hlača, Christoph Engel, and Krishna P Gummadi. 2019. Human decision making with machine assistance: An experiment on bailing and jailing. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–25.

[215] William M Grove, David H Zald, Boyd S Lebow, Beth E Snitz, and Chad Nelson. 2000. Clinical versus mechanical prediction: a meta-analysis. *Psychological assessment* 12, 1 (2000), 19.

[216] Martin Guggenheim. 2021. How Racial Politics Led Directly to the Enactment of the Adoption and Safe Families Act of 1997—the Worst Law Affecting Families Ever Enacted by Congress. *Columbia Journal of Race and Law* 11, 3 (2021).

[217] Shion Guha, Eric Baumer, and Geri Gay. 2018. Regrets, I've Had a Few: When Regretful Experiences Do (and Don't) Compel Users to Leave Facebook. In *Proceedings of the 2018 ACM Conference on Supporting Groupwork*. 166–177. DOI:http://dx.doi.org/10.1145/3148330.3148338

[218] Martin T Hall, Jeanelle Sears, and Matthew T Walton. 2020. Motivational interviewing in child welfare services: a systematic review. *Child maltreatment* 25, 3 (2020), 263–276.

[219] Kenneth R Hammond. 1996. *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford University Press on Demand.

[220] Melissa Hanson. 2021. After release of scathing report on 14-year-old David Almond's death, Massachusetts DCF says worker fired and changes are underway. *MASSLIVE* (Mar 2021). https://tinyurl.com/5e7h2xs7

[221] MD Romael Haque, Katherine Weathington, Joseph Chudzik, and Shion Guha. 2020. Understanding Law Enforcement and Common Peoples' Perspectives on Designing Explainable Crime Mapping Algorithms. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*. 269–273.

[222] MD Romael Haque, Katherine Weathington, and Shion Guha. 2019. Exploring the Impact of (Not) Changing Default Settings in Algorithmic Crime Mapping-A Case Study of Milwaukee, Wisconsin. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 206–210.

[223] Moritz Hardt, Eric Price, and Nathan Srebro. 2016. Equality of opportunity in supervised learning. *arXiv preprint arXiv:1610.02413* (2016).

[224] Ellie Harmon, Matthias Korn, Ann Light, and Amy Voida. 2016. Designing against the status quo. In *Proceedings of the 2016 ACM Conference Companion Publication on Designing Interactive Systems*. ACM, 65–68.

[225] Teresa M Harrison, Donna Canestraro, Theresa Pardo, Martha Avila-Marilla, Nicolas Soto, Megan Sutherland, Brian Burke, and Mila Gasco. 2018. A tale of two information systems: transitioning to a data-centric information system for child welfare. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*. 1–2.

[226] Ron Haskins. 2020. Child welfare financing: What do we fund, how, and what could be improved? *The ANNALS of the American Academy of Political and Social Science* 692, 1 (2020), 50–67.

[227] Sara Heitlinger, Nick Bryan-Kinns, and Rob Comber. 2019. The right to the sustainable smart city. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[228] Alison Hendricks, Lisa Conradi, and Charles Wilson. 2011. Creating trauma-informed child welfare systems using a community assessment process. *Child Welfare* 90, 6 (2011), 187.

[229] Robin K Hill. 2016. What an algorithm is. *Philosophy & Technology* 29, 1 (2016), 35–59.

[230] Sally Ho and Garance Burke. Associated Press. April 2022. An algorithm that screens for child neglect raises concerns. Link: https://apnews.com/article/child-welfare-algorithm-investigation-9497ee937e0053ad4144a86c68241ef1. (April 2022).

[231] Joana Hois, Dimitra Theofanou-Fuelbier, and Alischa Janine Junk. 2019. How to Achieve Explainability and Transparency in Human AI Interaction. In *International Conference on Human-Computer Interaction*. Springer, 177–183.

[232] Naja Holten Møller, Irina Shklovski, and Thomas T Hildebrandt. 2020. Shifting concepts of value: Designing algorithmic decision-support systems for public services. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–12.

[233] Sun-ha Hong. 2022. Prediction as Extraction of Discretion. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 925–934.

[234] Inger Sofie Dahlø Husby, Tor Slettebø, and Randi Juul. 2018. Partnerships with children in child welfare: The importance of trust and pedagogical support. *Child & Family Social Work* 23, 3 (2018), 443–450.

[235] C. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*.

[236] Kori Inkpen, Stevie Chancellor, Munmun De Choudhury, Michael Veale, and Eric PS Baumer. 2019. Where is the Human?: Bridging the Gap Between AI and HCI. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, W09.

[237] Luma Institute. 2012. *Innovating for people: Handbook of human-centered design methods*. LUMA Institute, LLC.

[238] SAS Institute. 2022a. Analytics for Child Well-Being. Link: https://www.sas.com/en_us/software/analytics-for-child-well-being.html. (2022).

[239] SAS Institute. 2022b. ECKERD RAPID SAFETY FEEDBACK. Link: https://eckerd.org/family-children-services/ersf/. (2022).

[240] Karoliina Isoaho, Daria Gritsenko, and Eetu Mäkelä. 2021. Topic Modeling and Text Analysis for Qualitative Policy Research. *Policy Studies Journal* 49, 1 (2021), 300–324.

[241] David Jackson and Gary Marx. 2017. Data mining program designed to predict child abuse proves unreliable, DCFS says. (dec 2017). https://www.chicagotribune.com/investigations/ct-dcfs-eckerd-met-20171206-story.html

[242] Steven J Jackson. 2013. Rethinking repair: breakdown, maintenance and repair in media and technology studies today. *Media meets technology, MIT Press, forthcoming* 68 (2013).

[243] Steven J Jackson, Tarleton Gillespie, and Sandy Payette. 2014. The policy knot: re-integrating policy, practice and design in cscw studies of social computing. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 588–602.

[244] Sigrid James. 2004. Why do foster care placements disrupt? An investigation of reasons for placement change in foster care. *Social service review* 78, 4 (2004), 601–627.

[245] Merav Jedwab, Anusha Chatterjee, and Terry V Shaw. 2018. Caseworkers' insights and experiences with successful reunification. *Children and Youth Services Review* 86 (2018), 56–63.

[246] Merav Jedwab, Yanfeng Xu, and Terry V Shaw. 2020. Kinship care first? Factors associated with placement moves in out-of-home care. *Children and Youth Services Review* 115 (2020), 105104.

[247] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications* 78, 11 (2019), 15169–15211.

[248] Matthew L. Jockers and David Mimno. 2013. Significant themes in 19th-century literature. *Poetics* 41, 6 (2013), 750–769. DOI:http://dx.doi.org/https://doi.org/10.1016/j.poetic.2013.08.005 Topic Models and the Cultural Sciences.

[249] K Johnson and D Wagner. 2003. California Structured Decision Making. Risk Assessment Revalidation: A Prospective Study. Children's Research Center. SAMSHA's National Registry of Evidence-based Programs and Practices (NREPP). (2003).

[250] Laura Johnson, Beth Sapiro, Catherine Buttner, and Judy L Postmus. 2018. Ambiguous agency as a diagnostic of power: Efforts of child welfare providers to promote responsible agency among youth involved in sex trades. *Journal of Aggression, Maltreatment & Trauma* 27, 6 (2018), 577–597.

[251] Lisa M Johnson, Becky F Antle, and Anita P Barbee. 2009. Addressing disproportionality and disparity in child welfare: Evaluation of an anti-racism training for community service providers. *Children and Youth Services Review* 31, 6 (2009), 688–696.

[252] Will Johnson. 2004. Effectiveness of California's child welfare structured decision making (SDM) model: a prospective study of the validity of the California Family Risk Assessment. *Madison (Wisconsin, USA): Children's Research Center* (2004).

[253] Jongbin Jung, Connor Concannon, Ravi Shroff, Sharad Goel, and Daniel G Goldstein. 2017. Simple rules for complex decisions. *arXiv preprint arXiv:1702.04690* (2017).

[254] Randi Juul and Inger Sofie Dahlø Husby. 2020. Collaboration and conversations with children in Child Welfare Services—Parents' viewpoint. *Child & Family Social Work* 25 (2020), 9–17.

[255] Elizabeth K Hopper, Ellen L Bassuk, and Jeffrey Olivet. 2010. Shelter from the storm: Trauma-informed care in homelessness services settings. *The Open Health Services and Policy Journal* 3, 1 (2010).

[256] Anna Kawakami, Venkatesh Sivaraman, Hao-Fei Cheng, Logan Stapleton, Yanghuidi Cheng, Diana Qing, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022a. Improving Human-AI Partnerships in Child Welfare: Understanding Worker Practices, Challenges, and Desires for Algorithmic Decision Support. In *CHI Conference on Human Factors in Computing Systems*. 1–18.

[257] Anna Kawakami, Venkatesh Sivaraman, Logan Stapleton, Hao-Fei Cheng, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022b. "Why Do I Care What's Similar?" Probing Challenges in AI-Assisted Child Welfare Decision-Making through Worker-AI Interface Design Concepts. In *Designing Interactive Systems Conference*. 454–470.

[258] Emily Keddell. 2015. The ethics of predictive risk modelling in the Aotearoa/New Zealand child welfare context: Child abuse prevention or neo-liberal tool? *Critical Social Policy* 35, 1 (2015), 69–88.

[259] Carol Kellison. 2019. *" How Can You Send Me Back to the Same Place?" A Qualitative Examination of the Professional Perspectives on the Implementation of Intensive Family Reunification Services into the Missouri Model of Juvenile Justice*. Ph.D. Dissertation. Southeast Missouri State University.

[260] Katherine C Kellogg, Melissa A Valentine, and Angele Christin. 2020. Algorithms at work: The new contested terrain of control. *Academy of Management Annals* 14, 1 (2020), 366–410.

[261] Jakko Kemper and Daan Kolkman. 2019. Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society* 22, 14 (2019), 2081–2096.

[262] Alex Kendall and Yarin Gal. 2017. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems* 30 (2017).

[263] Alex Kendall, Yarin Gal, and Roberto Cipolla. 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7482–7491.

[264] Geraldine Kennett and others. 2013. The impact of training practices on individual, organisation, and industry skill development. *Australian Bulletin of Labour* 39, 1 (2013), 112.

[265] Vera Khovanskaya, Phoebe Sengers, and Lynn Dombrowski. 2020. Bottom-Up Organizing with Tools from On High: Understanding the Data Practices of Labor Organizers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[266] Vera D Khovanskaya. 2021. *The Tools of Management: Data Practices for Worker Advocacy*. Ph.D. Dissertation. Cornell University.

[267] Raymond S Kirk, Mimi M Kim, and Diane P Griffith. 2005. Advances in the reliability and validity of the North Carolina Family Assessment Scale. *Journal of Human Behavior in the Social Environment* 11, 3-4 (2005), 157–176.

[268] Keith Kirkpatrick. 2016. Battling algorithmic bias: how do we ensure algorithms treat us fairly? *Commun. ACM* 59, 10 (2016), 16–17.

[269] Rob Kitchin and Tracey Lauriault. 2014. Towards critical data studies: Charting and unpacking data assemblages and their work. (2014).

[270] Chamari I. Kithulgoda, Rhema Vaithianathan, and Dennis P. Culhane. 2022. Predictive Risk Modeling to Identify Homeless Clients at Risk for Prioritizing Services using Routinely Collected Data. *Journal of Technology in Human Services* 40, 2 (2022), 134–156. DOI:http://dx.doi.org/10.1080/15228835.2022.2042461

[271] René F Kizilcec. 2016. How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2390–2395.

[272] Sacha Klein and Darcey H Merritt. 2014. Neighborhood racial & ethnic diversity as a predictor of child welfare system involvement. *Children and Youth Services Review* 41 (2014), 95–105.

[273] Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. Prediction policy problems. *American Economic Review* 105, 5 (2015), 491–95.

[274] Lisa Koeman, Vaiva Kalnikaitė, and Yvonne Rogers. 2015. " Everyone Is Talking about It!" A Distributed Approach to Urban Voting Technology and Visualisations. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 3127–3136.

[275] Eun Koh and Mark F Testa. 2008. Propensity score matching of children in kinship and nonkinship foster care: Do permanency outcomes differ? *Social Work Research* 32, 2 (2008), 105–116.

[276] Brianne H Kothari, Kelly D Chandler, Andrew Waugh, Kara K McElvaine, Jamie Jaramillo, and Shannon Lipscomb. 2021. Retention of child welfare caseworkers: The role of case severity and workplace resources. *Children and Youth Services Review* 126 (2021), 106039.

[277] Simone Kühn, Dimitrij Tycho Kugler, Katharina Schmalen, Markus Weichenberger, Charlotte Witt, and Jürgen Gallinat. 2019. Does playing violent video games cause aggression? A longitudinal intervention study. *Molecular psychiatry* 24, 8 (2019), 1220–1234.

[278] Todd Kulesza, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng-Keen Wong. 2013. Too much, too little, or just right? Ways explanations impact end users' mental models. In *2013 IEEE Symposium on visual languages and human centric computing*. IEEE, 3–10.

[279] J Nathan Kutz. 2013. *Data-driven modeling & scientific computation: methods for complex systems & big data*. Oxford University Press.

[280] Vivian Lai, Chacha Chen, Q Vera Liao, Alison Smith-Renner, and Chenhao Tan. 2021. Towards a science of human-ai decision making: a survey of empirical studies. *arXiv preprint arXiv:2112.11471* (2021).

[281] Anja Lambrecht and Catherine Tucker. 2019. Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science* (2019).

[282] Jan-Erik Lane. 2000. *New public management*. Taylor & Francis US.

[283] Mark D Lardner. 2015. Are restrictiveness of care decisions based on youth level of need? A multilevel model analysis of placement levels using the child and adolescent needs and strengths assessment. *Residential Treatment for Children & Youth* 32, 3 (2015), 195–207.

[284] Camilla Lauritzen, Svein A Vis, and Sturla Fossum. 2018. Factors that determine decision making in child protection investigations: A review of the literature. *Child & Family Social Work* 23, 4 (2018), 743–756.

[285] Michael A Lawson, Tania Alameda-Lawson, and Edward Byrnes. 2017. Analyzing the validity of the Adult-Adolescent parenting inventory for low-income populations. *Research on Social Work Practice* 27, 4 (2017), 441–455.

[286] Bonnie Leadbeater, Dan Dodgen, and Andrea Solarz. 2005. The resilience revolution. In *Resilience in children, families, and communities*. Springer, 47–61.

[287] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (2018), 2053951718756684.

[288] Min Kyung Lee and Su Baykal. 2017. Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1035–1048.

[289] Soo-Young Lee and Andrew B Whitford. 2013. Assessing the effects of organizational resources on public agency performance: Evidence from the US federal government. *Journal of Public Administration Research and Theory* 23, 3 (2013), 687–712.

[290] Per LÊgreid. 2017. *Transcending new public management: the transformation of public sector reforms*. Routledge.

[291] Agnieszka Leszczynski. 2016. Speculative futures: Cities, data, and governance beyond smart urbanism. *Environment and Planning A: Economy and Space* 48, 9 (2016), 1691–1708.

[292] Jill Levenson. 2017. Trauma-informed social work practice. *Social Work* 62, 2 (2017), 105–113.

[293] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[294] Cynthia A Lietz and Margaret Strength. 2011. Stories of successful reunification: A narrative study of family resilience in child welfare. *Families in society* 92, 2 (2011), 203–210.

[295] Ann Light and Anna Seravalli. 2019. The breakdown of the municipality as caring platform: lessons for co-design and co-learning in the age of platform capitalism. *CoDesign* 15, 3 (2019), 192–211.

[296] Ida Lindgren, Christian Østergaard Madsen, Sara Hofmann, and Ulf Melin. 2019. Close encounters of the digital kind: A research agenda for the digitalization of public services. *Government Information Quarterly* 36, 3 (2019), 427–436.

[297] Michael Lipsky. 2010. *Street-level bureaucracy: Dilemmas of the individual in public service*. Russell Sage Foundation.

[298] Linda Lobao, Mia Gray, Kevin Cox, and Michael Kitson. 2018. The shrinking state? Understanding the assault on the public sector. (2018).

[299] Thomas Lodato and Carl DiSalvo. 2018. Institutional constraints: the forms and limits of participatory design in the public realm. In *Proceedings of the 15th Participatory Design Conference: Full Papers-Volume 1*. 1–12.

[300] Kim Loyens and Jeroen Maesschalck. 2010. Toward a theoretical framework for ethical decision making of street-level bureaucracy: Existing models reconsidered. *Administration & Society* 42, 1 (2010), 66–100.

[301] John S Lyons. 2009. *Communimetrics: A communication theory of measurement in human service settings.* Springer Science & Business Media.

[302] John S Lyons. 2014. Use of the Child and Adolescent Needs and Strengths (CANS) in Child Welfare in the United States. (2014).

[303] John S Lyons, Dana Aron Weiner, and Melanie Buddin Lyons. 2004. Measurement as communication in outcomes management: The child and adolescent needs and strengths (CANS). *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment. Volume 2: Instruments for Children and Adolescents* (2004).

[304] John S Lyons, Dana Aron Weiner, and Melanie Buddin Lyons. 2014. Measurement as communication in outcomes management: The child and adolescent needs and strengths (CANS). In *The use of psychological testing for treatment planning and outcomes assessment.* Routledge, 461–476.

[305] Peter Lyons, Howard J Doueck, and John S Wodarski. 1996. Risk assessment for child protective services: A review of the empirical literature on instrument performance. *Social Work Research* 20, 3 (1996), 143–155.

[306] Thomas M. Vogl. 2020. Artificial intelligence and organizational memory in government: the experience of record duplication in the child welfare sector in Canada. In *The 21st Annual International Conference on Digital Government Research.* 223–231.

[307] Mariusz Maciejewski. 2017. To do more, better, faster and more cheaply: Using big data in public administration. *International Review of Administrative Sciences* 83, 1_suppl (2017), 120–135.

[308] Michael J MacKenzie, Jonathan B Kotch, and Li-Ching Lee. 2011. Toward a cumulative ecological risk model for the etiology of child maltreatment. *Children and youth services review* 33, 9 (2011), 1638–1647.

[309] Diana MacLean, Sonal Gupta, Anna Lembke, Christopher Manning, and Jeffrey Heer. 2015. Forum77: An analysis of an online health forum dedicated to addiction recovery. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing.* ACM, 1511–1526.

[310] Stephen Magura, Beth Silverman Moses, and Mary Ann Jones. 1987. *Assessing risk and measuring change in families: The Family Risk Scales.* Vol. 51. Child Welfare League of Amer.

[311] Andrey Malinin and Mark Gales. 2018. Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems* 31 (2018).

[312] Anthony N Maluccio, Frank Ainsworth, and June Thoburn. 2000. *Child welfare outcome research in the United States, the United Kingdom, and Australia.* Child Welfare League of America.

[313] Azhar Manzoor. 2014. A look at efficiency in public administration: Past and future. *Sage Open* 4, 4 (2014), 2158244014564936.

[314] James G March. 1991. Exploration and exploitation in organizational learning. *Organization science* 2, 1 (1991), 71–87.

[315] Helen Margetts. 2012. *Information technology in government: Britain and America.* Routledge.

[316] David B Marshall and Diana J English. 2000. Neural network modeling of risk assessment in child protective services. *Psychological Methods* 5, 1 (2000), 102.

[317] Elaine Massung, David Coyle, Kirsten F Cater, Marc Jay, and Chris Preist. 2013. Using crowdsourcing to support pro-environmental community activism. In *Proceedings of the SIGCHI Conference on human factors in Computing systems.* 371–380.

[318] Steven Maynard-Moody and Michael Musheno. 2003. Cops, teachers, counselors: Narratives of street-level judgment. *Ann Arbor, MI: University of Michigan Press* (2003).

[319] Nicola McConnell, Matt Barnard, and Julie Taylor. 2017. Caring Dads Safer Children: Families' perspectives on an intervention for maltreating fathers. *Psychology of Violence* 7, 3 (2017), 406.

[320] Thomas P McDonald, John Poertner, and Gardenia Harris. 2002. Predicting placement in foster care: A comparison of logistic regression and neural network analysis. *Journal of social service research* 28, 2 (2002), 1–20.

[321] Curtis McMillen, Jill Rowe, and Nicole Fedoravicuis. 2006. Views of the Quality of Mental Health Services Received by Child Welfare Consumers. (2006).

[322] Paul E Meehl. 1954. Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. (1954).

[323] Albert Meijer. 2018. Datapolis: A public governance perspective on "smart cities". *Perspectives on Public Management and Governance* 1, 3 (2018), 195–206.

[324] Gary B Melton. 2005. Mandated reporting: A policy without reason. *Child abuse & neglect* 29, 1 (2005), 9–18.

[325] Amanda Meng, Carl DiSalvo, and Ellen Zegura. 2019. Collaborative data work towards a caring democracy. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

[326] Ines Mergel, R Karl Rethemeyer, and Kimberley Isett. 2016. Big data in public affairs. *Public Administration Review* 76, 6 (2016), 928–937.

[327] Michael E Milakovich and George J Gordon. 2013. *Public administration in America*. Cengage Learning.

[328] Lynne Miller. 2009. "You Have to Get It Together" ASFA's Impact on Parents and Families. *Intentions and Results* (2009), 36.

[329] Hannah Miller Hillberg, Zachary Levonian, Daniel Kluver, Loren Terveen, and Brent Hecht. 2018. What I See is What You Don't Get: The Effects of (Not) Seeing Emoji Rendering Differences across Platforms. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 124.

[330] Mario Molina and Filiz Garip. 2019. Machine Learning for Sociology. *Annual Review of Sociology* 45, 1 (2019), 27–45.

[331] Terry D Moore, Thomas P McDonald, and Kari Cronbaugh-Auld. 2016. Assessing risk of placement instability to aid foster care placement decision making. *Journal of Public Child Welfare* 10, 2 (2016), 117–131.

[332] Shabnam Mousavi and Gerd Gigerenzer. 2014. Risk, uncertainty, and heuristics. *Journal of Business Research* 67, 8 (2014), 1671–1678.

[333] Henrik Mucha, Sebastian Robert, Rüdiger Breitschwerdt, and Michael Fellmann. Towards Participatory Design Spaces for Explainable AI Interfaces in Expert Domains. (????).

[334] Michael Muller, Shion Guha, Eric PS Baumer, David Mimno, and N Sadat Shami. 2016. Machine learning and grounded theory method: Convergence, divergence, and combination. In *Proceedings of the 19th International Conference on Supporting Group Work*. ACM, 3–8.

[335] Michael J Muller. 2009. Participatory design: the third space in HCI. In *Human-computer interaction*. CRC press, 181–202.

[336] Eileen Munro. 2019. Decision-making under uncertainty in child protection: Creating a just and learning culture. *Child & Family Social Work* 24, 1 (2019), 123–130.

[337] Judge Michael Nash. 2017. Examination of using Structured Decision Making and Predictive Analytics in assessing Safety and Risk in Child Welfare. *County of Los Angeles Office of Child Protection* (May 2017).

[338] Barbara Needell, M Alan Brookhart, and Seon Lee. 2003. Black children and foster care placement in California. *Children and Youth Services Review* 25, 5-6 (2003), 393–408.

[339] Dong Nguyen, Maria Liakata, Simon DeDeo, Jacob Eisenstein, David Mimno, Rebekah Tromble, and Jane Winters. 2020. How We Do Things With Words: Analyzing Text as Social and Cultural Data. *Frontiers in Artificial Intelligence* 3 (2020), 62.

[340] Chao Ning and Fengqi You. 2019. Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Computers & Chemical Engineering* 125 (2019), 434–448.

[341] Kathleen G Noonan, Charles F Sabel, and William H Simon. 2009. Legal accountability in the service-based welfare state: Lessons from child welfare reform. *Law & Social Inquiry* 34, 3 (2009), 523–568.

[342] Amy E Nourie. 2021. Child Welfare Abolition: Critical Theories, Human Rights, and Heteronormativity. *Journal of Human Rights and Social Work* (2021), 1–10.

[343] Matthew C. Nowlin. 2016. Modeling Issue Definitions Using Quantitative Text Analysis. *Policy Studies Journal* 44, 3 (2016), 309–331.

[344] NPR. June 2022. Oregon is dropping an artificial intelligence tool used in child welfare system. Link: https://www.npr.org/2022/06/02/1102661376/oregon-drops-artificial-intelligence-child-abuse-cases. (June 2022).

[345] Uniersity of Buffalo News. 2020. UB receives $800,000 NSF/Amazon grant to improve AI fairness in foster care. (2020). http://www.buffalo.edu/ubnow/stories/2020/02/grant-ai-foster-care.html

[346] Wisconsin Department of Children and Families. 2016. Jeanine B. Settlement Agreement Report of the Division of Milwaukee Child Protective Services. (2016).

[347] US Department of Health and Human Services. 2017. Child Maltreatment 2017. *Children's Bureau (Ed.)* (2017).

[348] US Department of Health, Human Services, and others. 2017. The AFCARS report: Preliminary FY 2016 estimates as of Oct 2017. *Children's Bureau (Ed.)* 21 (2017), 6.

[349] US Department of Health, Human Services, and others. 2020. The AFCARS Report. Preliminary FY 2019 Estimates as of June 23, 2020 (27). *Retrieved December* 23 (2020), 2019.

[350] Cathy O'neil. 2017. *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown.

[351] Yotam Ophir, Dror Walter, and Eleanor R Marchant. 2020. A Collaborative Way of Knowing: Bridging Computational Communication Research and Grounded Theory Ethnography. *Journal of Communication* 70, 3 (2020), 447–472.

[352] Stephen P Osborne and Louise Brown. 2013. *Handbook of innovation in public services.* Edward Elgar Publishing.

[353] R Lyman Ott and Micheal T Longnecker. 2015. *An introduction to statistical methods and data analysis.* Nelson Education.

[354] Juho Pääkkönen, Matti Nelimarkka, Jesse Haapoja, and Airi Lampinen. 2020. Bureaucracy as a Lens for Analyzing and Designing Algorithmic Systems. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–14.

[355] Elizabeth M Parker, Jason R Williams, Peter J Pecora, and Daniel Despard. 2022. Examining the effects of the Eckerd rapid safety feedback process on the occurrence of repeat maltreatment among children involved in the child welfare system. *Child Abuse & Neglect* 133 (2022), 105856.

[356] Samir Passi and Steven Jackson. 2017. Data vision: Learning to see through algorithmic abstraction. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2436–2447.

[357] Samir Passi and Steven J Jackson. 2018. Trust in data science: collaboration, translation, and accountability in corporate data science projects. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–28.

[358] Samir Passi and Phoebe Sengers. 2020. Making data science systems work. *Big Data & Society* 7, 2 (2020), 2053951720939605.

[359] John W Payne, James R Bettman, and Eric J Johnson. 1993. The use of multiple strategies in judgment and choice. (1993).

[360] Rik Peeters. 2020. The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity* Preprint (2020), 1–16.

[361] Bruce D Perry. 2009. Examining child maltreatment through a neurodevelopmental lens: Clinical applications of the neurosequential model of therapeutics. *Journal of Loss and Trauma* 14, 4 (2009), 240–255.

[362] B Guy Peters. 2014. Accountability in public administration. *The Oxford handbook of public accountability* (2014), 211–225.

[363] Jennifer Pierre, Roderic Crooks, Morgan Currie, Britt Paris, and Irene Pasquetto. 2021. Getting Ourselves Together: Data-centered participatory design research & epistemic burden. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.

[364] Hitendra Pillay and Andrea R McCrindle. 2005. Distributed and relative nature of professional expertise. *Studies in Continuing Education* 27, 1 (2005), 67–88.

[365] Sarah Pink, Kerstin Leder Mackley, Val Mitchell, Marcus Hanratty, Carolina Escobar-Tello, Tracy Bhamra, and Roxana Morosanu. 2013. Applying the lens of sensory ethnography to sustainable HCI. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 4 (2013), 25.

[366] Keri LM Pinna, Lynn K Lewis, Canan Karatekin, Angela Lamb-Onyiga, Ashley Hirilall, and Sabrina D Jones. 2015. Evidence-based parenting programs for maltreating parents: Views of child protective services caseworkers. *Journal of public child welfare* 9, 4 (2015), 362–381.

[367] Anthony T Pinter, Pamela J Wisniewski, Heng Xu, Mary Beth Rosson, and Jack M Caroll. 2017. Adolescent online safety: Moving beyond formative evaluations to designing solutions for the future. In *Proceedings of the 2017 Conference on Interaction Design and Children*. ACM, 352–357.

[368] Nancy D Polikoff and Jane M Spinak. 2021. Strengthened Bonds: Abolishing the Child Welfare System and Re-Envisioning Child Well-Being. *Colum. J. Race & L.* 11 (2021), 427.

[369] Gordon Pon, Kevin Gosine, and Doret Phillips. 2011. Immediate response: Addressing anti-Native and anti-Black racism in child welfare. *International Journal of Child, Youth and Family Studies* 2, 3/4 (2011), 385–409.

[370] Anja Pors and Jannick Schou. 2020. Street-level morality at the digital frontlines: An ethnographic study of moral mediation in welfare work. *Administrative Theory & Praxis* (2020), 1–18.

[371] Predict-Align-Prevent. 2022. Stopping Child Maltreatment Before it Happens. Link: https://papreports.org/little-rock-ar/. (2022).

[372] Emily Putnam-Hornstein, Barbara Needell, Bryn King, and Michelle Johnson-Motoyama. 2013. Racial and ethnic disparities: A population-based examination of risk factors for involvement with child protective services. *Child Abuse & Neglect* 37, 1 (2013), 33–46.

[373] Rajendra Rambajue and Christopher O'Connor. 2021. Intersectional individualization: toward a theoretical framework for youth transitioning out of the child welfare system. *Journal of Public Child Welfare* (2021), 1–21.

[374] Trine Rask Nielsen and Naja Holten Møller. 2022. Data as a Lens for Understanding what Constitutes Credibility in Asylum Decision-making. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–23.

[375] Diane Boyd Rauber. 2009. From the courthouse to the statehouse: Parents as partners in child welfare. *Child Law Practice* 28, 10 (2009), 149–156.

[376] Mary A Rawlings and Emily R Blackmer. 2019. Assessing engagement skills in public child welfare using OSCE: A pilot study. *Journal of Public Child Welfare* 13, 4 (2019), 441–461.

[377] Joanna Redden, Lina Dencik, and Harry Warne. 2020. Datafied child welfare services: unpacking politics, economics and power. *Policy Studies* (2020), 1–20.

[378] Richard E Redding, Carrie Fried, and Preston A Britner. 2000. Predictors of placement outcomes in treatment foster care: Implications for foster parent selection and service delivery. *Journal of child and family studies* 9, 4 (2000), 425–447.

[379] Kellie Reed-Ashcraft, Raymond S Kirk, and Mark W Fraser. 2001. The reliability and validity of the North Carolina family assessment scale. *Research on Social Work Practice* 11, 4 (2001), 503–520.

[380] Cheryl Regehr, Marion Bogo, Aron Shlonsky, and Vicki LeBlanc. 2010. Confidence and professional judgment in assessing children's risk of abuse. *Research on social work practice* 20, 6 (2010), 621–628.

[381] Jennifer A Reich. 2012. *Fixing families: Parents, power, and the child welfare system.* Routledge.

[382] Jeanne S Ringel, Dana Schultz, Joshua Mendelsohn, Stephanie Brooks Holliday, Katharine Sieck, Ifeanyi Edochie, and Lauren Davis. 2018. Improving child welfare outcomes: balancing investments in prevention and treatment. *Rand health quarterly* 7, 4 (2018).

[383] Rubina F Rizvi, Yefeng Wang, Thao Nguyen, Jake Vasilakes, Jiang Bian, Zhe He, and Rui Zhang. 2019. Analyzing Social Media Data to Understand Consumer Information Needs on Dietary Supplements. *Stud Health Technol Inform* 264 (Aug. 2019), 323–327.

[384] Dorothy E Roberts. 2007. Child welfare's paradox. *Wm. & Mary L. Rev.* 49 (2007), 881.

[385] Samantha Robertson, Tonya Nguyen, and Niloufar Salehi. 2021. Modeling Assumptions Clash with the Real World: Transparency, Equity, and Community Challenges for Student Assignment Algorithms. *arXiv preprint arXiv:2101.10367* (2021).

[386] Samantha Robertson and Niloufar Salehi. 2020. What If I Don't Like Any Of The Choices? The Limits of Preference Elicitation for Participatory Algorithm Design. *arXiv preprint arXiv:2007.06718* (2020).

[387] Maria Y. Rodriguez and Heather Storer. 2020. A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data*. *Journal of Technology in Human Services* 38, 1 (2020), 54–86.

[388] Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1, 5 (2019), 206–215.

[389] Joseph P Ryan, Philip Garnier, Michael Zyphur, and Fuhua Zhai. 2006. Investigating the effects of caseworker characteristics in child welfare. *Children and Youth Services Review* 28, 9 (2006), 993–1006.

[390] SaintA. 2021. Child Welfare Case Manager. (2021). https://www.milwaukeejobs.com/job/detail/35323239/Child-Welfare-Case-Manager-1000-Sign-on-Bonus

[391] Anjana Samant, Aaron Horowitz, Kath Xu, and Sophie Beiers. 2021. Family surveillance by algorithm: The rapidly spreading tools few have heard of. American Civil Liberties Union (ACLU)(2021). (2021).

[392] Jodi R Sandfort. 2000. Moving beyond discretion and outcomes: Examining public management from the front lines of the welfare system. *Journal of Public Administration Research and Theory* 10, 4 (2000), 729–756.

[393] Maarten Sap, Marcella Cindy Prasettio, Ari Holtzman, Hannah Rashkin, and Yejin Choi. 2017. Connotation Frames of Power and Agency in Modern Films. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2329–2334.

[394] S Savage, C Toxtli, and E Betanzos. 2021. Research Methods to Study & Empower Crowd Workers. *Oxford* (2021).

[395] Devansh Saxena, Karla Badillo-Urquiola, Pamela Wisniewski, and Shion Guha. 2020a. Child Welfare System: Interaction of Policy, Practice and Algorithms. In *Companion of the 2020 ACM International Conference on Supporting Group Work*. 119–122.

[396] Devansh Saxena, Karla Badillo-Urquiola, Pamela Wisniewski, and Shion Guha. 2021a. A Framework of High-Stakes Algorithmic Decision-Making for the Public Sector Developed through a Case Study of Child-Welfare. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021).

[397] Devansh Saxena, Karla Badillo-Urquiola, Pamela J Wisniewski, and Shion Guha. 2020b. A Human-Centered Review of Algorithms used within the US Child Welfare System. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[398] Devansh Saxena, Karla Badillo-Urquiola, Pamela J Wisniewski, and Shion Guha. 2021b. A framework of high-stakes algorithmic decision-making for the public sector developed through a case study of child-welfare. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–41.

[399] Devansh Saxena, Erhardt Graeff, Shion Guha, EunJeong Cheon, Pedro Reynolds-Cuéllar, Dawn Walker, Christoph Becker, and Kenneth R Fleischmann. 2020. Collective Organizing and Social Responsibility at CSCW. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*. 503–509.

[400] Devansh Saxena and Shion Guha. 2020. Conducting Participatory Design to Improve Algorithms in Public Services: Lessons and Challenges. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*. 383–388.

[401] Devansh Saxena and Shion Guha. 2023. Algorithmic Impacts in Child-Welfare: Uncertainties in Practice, Administration, and Street-level Decision-Making. *ACM Journal on Responsible Computing* (2023).

[402] Devansh Saxena, Erina Seh-Young Moon, Aryan Chaurasia, Yixin Guan, and Shion Guha. 2023. Rethinking" Risk" in Algorithmic Systems Through A Computational Narrative Analysis of Casenotes in Child-Welfare. (2023), 1–22.

[403] Devansh Saxena, Erina Seh-young Moon, Dahlia Shehata, and Shion Guha. 2022a. Unpacking Invisible Work Practices, Constraints, and Latent Power Relationships in Child Welfare through Casenote Analysis. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*.

[404] Devansh Saxena, Seh Young Moon, Dahlia Shehata, and Shion Guha. 2022b. Unpacking Invisible Work Practices, Constraints, and Latent Power Relationships in Child Welfare through Casenote Analysis. In *CHI Conference on Human Factors in Computing Systems*. 1–22.

[405] Devansh Saxena, Charlie Repaci, Melanie Sage, and Shion Guha. 2022c. How to Train a (Bad) Algorithmic Caseworker: A Quantitative Deconstruction of Risk Assessments in Child Welfare. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–7.

[406] Devansh Saxena, Charles Repaci, Melanie D Sage, and Shion Guha. 2022d. How to Train a (Bad) Algorithmic Caseworker: A Quantitative Deconstruction of Risk Assessments in Child Welfare. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.

[407] Philipp Schmidt and Felix Biessmann. 2020. Calibrating Human-AI Collaboration: Impact of Risk, Ambiguity and Transparency on Algorithmic Bias. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 431–449.

[408] Alexandra Schofield, Måns Magnusson, and David Mimno. 2017. Pulling out the stops: Rethinking stopword removal for topic models. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, short papers*. 432–436.

[409] John R Schuerman and Lynn Harold Vogel. 1986. Computer support of placement planning: the use of expert systems in child welfare. *Child welfare* 65, 6 (1986), 531–543.

[410] A James Schwab and Susan S Wilson. 1989. The continuum of care system: Decision support for practitioners. *Computers in Human Services* 4, 1-2 (1989), 123–140.

[411] A James Schwab Jr, Michael E Bruce, and Ruth G McRoy. 1984. Matching children with placements. *Children and youth services review* 6, 2 (1984), 125–133.

[412] Craig Schwalbe. 2004. Re-visioning risk assessment for human service decision making. *Children and Youth Services Review* 26, 6 (2004), 561–576.

[413] Craig S Schwalbe. 2008. Strengthening the integration of actuarial risk assessment with clinical judgment in an evidence based practice framework. *Children and Youth Services Review* 30, 12 (2008), 1458–1464.

[414] Nick Seaver. 2017. Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4, 2 (2017), 2053951717738104.

[415] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*. 59–68.

[416] Nicole Shadowen, Thomas Lodato, and Daria Loi. 2020. Participatory governance in smart cities: Future scenarios and opportunities. In *International Conference on Human-Computer Interaction*. Springer, 443–463.

[417] Amit Sharma, Vasilis Syrgkanis, Cheng Zhang, and Emre Kıcıman. 2021. Dowhy: Addressing challenges in expressing and validating causal assumptions. *arXiv preprint arXiv:2108.13518* (2021).

[418] Miseung Shim. 2010. Factors influencing child welfare employee's turnover: Focusing on organizational culture and climate. *Children and Youth Services Review* 32, 6 (2010), 847–856.

[419] Aron Shlonsky and Dennis Wagner. 2005. The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management. *Children and youth services review* 27, 4 (2005), 409–427.

[420] Dilruba Showkat and Eric PS Baumer. 2022. "It's Like the Value System in the Loop": Domain Experts' Values Expectations for NLP Automation. (2022).

[421] Ravi Shroff. 2017. Predictive Analytics for City Agencies: Lessons from Children's Services. *Big data* 5, 3 (2017), 189–196.

[422] Fiore Sicoly. 1989a. Computer-aided decisions in human services: Expert systems and multivariate models. *Computers in Human Behavior* 5, 1 (1989), 47–60.

[423] Fiore Sicoly. 1989b. Prediction and decision making in child welfare. *Computers in Human Services* 5, 3-4 (1989), 43–56.

[424] Nancy L. Sidell. 2015. *Social Work Documentation.* NASW Press, Washington, DC, USA.

[425] June Simon, Jeeyeon Hong, Peter J Pecora, Geoffrey Cloepfil, Troy Gillespie, Elizabeth Parker, Reed Klein, Thalia Garcia, and Mariah Hardy. 2023. An examination of recent federal and philanthropic funding for child welfare research in the United States. *Children and Youth Services Review* 148 (2023), 106779.

[426] Douglas G Simpson, Peter B Imrey, Olga Geling, and Susan Butkus. 2000. Statistical estimation of child abuse rates from administrative databases. *Children and Youth Services Review* 22, 11-12 (2000), 951–971.

[427] Brenda D Smith. 2008. Child welfare service plan compliance: Perceptions of parents and caseworkers. *Families in Society* 89, 4 (2008), 521–532.

[428] Elizabeth A Smith. 2001. The role of tacit and explicit knowledge in the workplace. *Journal of knowledge Management* (2001).

[429] Gabriel Tobin Smith, Valerie B Shapiro, Rachel Wagner Sperry, and Paul A LeBuffe. 2014. A strengths-based approach to supervised visitation in child welfare. *Child Care in Practice* 20, 1 (2014), 98–119.

[430] Robert Soden, Laura Devendorf, Richmond Wong, Yoko Akama, Ann Light, and others. 2022. Modes of Uncertainty in HCI. *Foundations and Trends® in Human–Computer Interaction* 15, 4 (2022), 317–426.

[431] Michael R Sosin. 2010. Discretion in human service organizations. *Human services as complex organizations* (2010), 381–403.

[432] Logan Stapleton, Min Hun Lee, Diana Qing, Marya Wright, Alexandra Chouldechova, Ken Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022a. Imagining new futures beyond predictive systems in child welfare: A qualitative study with impacted stakeholders. In *2022 ACM Conference on Fairness, Accountability, and Transparency.* 1162–1177.

[433] Logan Stapleton, Devansh Saxena, Anna Kawakami, Tonya Nguyen, Asbjørn Ammitzbøll Flügge, Motahhare Eslami, Naja Holten Møller, Min Kyung Lee, Shion Guha, Kenneth Holstein, and others. 2022b. Who Has an Interest in "Public Interest Technology"?: Critical Questions for Working with Local Governments & Impacted Communities. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing.* 282–286.

[434] Susan Leigh Star and Anselm Strauss. 1999. Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer supported cooperative work (CSCW)* 8, 1-2 (1999), 9–30.

[435] Starhawk. 1987. *Truth or dare: Encounters with power, authority, and mystery.* Harper & Row.

[436] James P Stevens. 1984. Outliers and influential data points in regression analysis. *Psychological Bulletin* 95, 2 (1984), 334.

[437] Megan Stevenson. 2018. Assessing risk assessment in action. *Minn. L. Rev.* 103 (2018), 303.

[438] Kele Stewart. 2022. Re-envisioning Child Well-being: Dismantling the Inequitable Intersections Among Child Welfare, Juvenile Justice, and Education. *Columbia Journal of Race and Law* 12, 1 (2022).

[439] Emilie Stoltzfus. 2014. Child welfare: An overview of federal programs and their current funding. (2014).

[440] Norman M Stone and Susan F Stone. 1983. The prediction of successful foster placement. *Social Casework* 64, 1 (1983), 11–17.

[441] Angelika Strohmayer, Rob Comber, and Madeline Balaam. 2015. Exploring learning ecologies among people experiencing homelessness. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2275–2284.

[442] Tony Sun, Andrew Gaut, Shirlyn Tang, Yuxin Huang, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang, and William Yang Wang. 2019. Mitigating Gender Bias in Natural Language Processing: Literature Review. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 1630–1640. DOI: http://dx.doi.org/10.18653/v1/P19-1159

[443] Paul A Sunseri. 2020. Hidden figures: Is improving family functioning a key to better treatment outcomes for seriously mentally ill children? *Residential Treatment for Children & Youth* 37, 1 (2020), 46–64.

[444] Sarah Tan, Julius Adebayo, Kori Inkpen, and Ece Kamar. 2018. Investigating human+ machine complementarity for recidivism predictions. *arXiv preprint arXiv:1808.09123* (2018).

[445] Brian J Taylor and Karen McQuillan. 2014. Perspectives of foster parents and social workers on foster placement disruption. *Child Care in Practice* 20, 2 (2014), 232–249.

[446] CA Taylor, NB Guterman, SJ Lee, and PJ Rathouz. 2009. Intimate partner violence, maternal stress, nativity, and risk for maternal maltreatment of young children. *American Journal of Public Health* 99, 1 (2009), 175–83.

[447] MindShare Technology. 2022. Improving Outcomes using data you already have. Link: https://mindshare-technology.com/analytics/. (2022).

[448] Clare Tilbury. 2005. Counting family support. *Child & Family Social Work* 10, 2 (2005), 149–157.

[449] Austin L Toombs, Andy Dow, John Vines, Colin M Gray, Barbara Dennis, Rachel Clarke, and Ann Light. 2018. Designing for Everyday Care in Communities. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*. 391–394.

[450] James Topitzes, Timothy Grove, Erika E Meyer, Stacey M Pangratz, and Caitlin M Sprague. 2019. Trauma-responsive child welfare services: A mixed methods study assessing safety, stability, and permanency. *Journal of Child Custody* 16, 3 (2019), 291–312.

[451] Karmen Toros, Diana Maria DiNitto, and Anne Tiko. 2018. Family engagement in the child welfare system: A scoping review. *Children and Youth Services Review* 88 (2018), 598–607.

[452] Carlos Toxtli, Angela Richmond-Fuller, and Saiph Savage. 2020. Reputation agent: Prompting fair reviews in gig markets. In *Proceedings of The Web Conference 2020*. 1228–1240.

[453] Carlos Toxtli, Siddharth Suri, and Saiph Savage. 2021. Quantifying the invisible labor in crowd work. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–26.

[454] Chicago Tribune. 2017. Data mining program designed to predict child abuse proves unreliable, DCFS says. Link: https://www.chicagotribune.com/investigations/ct-dcfs-eckerd-met-20171206-story.html. (2017).

[455] Rebecca Tushnet. 2018. The Difference Engine: Perpetuating Poverty through Algorithms. *Jotwell: J. Things We Like* (2018).

[456] Rebecca Uliasz. 2020. Seeing like an algorithm: operative images and emergent subjects. *AI & SOCIETY* (2020), 1–9.

[457] Michael Ungar. 2007. Contextual and cultural aspects of resilience in child welfare settings. *Putting a human face on child welfare: Voices from the Prairies* (2007), 1–23.

[458] Rhema Vaithianathan, Emily Putnam-Hornstein, Nan Jiang, Parma Nand, and Tim Maloney. 2017. Developing predictive models to support child maltreatment hotline screening decisions: Allegheny County methodology and implementation. *Center for Social data Analytics* (2017).

[459] Sarah Valentine. 2019. Impoverished algorithms: Misguided governments, flawed technologies, and social control. *Fordham Urb. LJ* 46 (2019), 364.

[460] Marcel F van Assen. 2021. Training, employee involvement and continuous improvement–the moderating effect of a common improvement method. *Production Planning & Control* 32, 2 (2021), 132–144.

[461] Frank E Vandervort, Robbin Pott Gonzalez, and Katlheen Coulborn Faller. 2008. Legal ethics and high child welfare worker turnover: An unexplored connection. *Children and Youth Services Review* 30, 5 (2008), 546–563.

[462] Jennifer Wortman Vaughan and Hanna Wallach. 2020. A human-centered agenda for intelligible machine learning. *Machines We Trust: Getting Along with Artificial Intelligence* (2020).

[463] Michael Veale and Irina Brass. 2019. Administration by algorithm? Public management meets public sector machine learning. *Public management meets public sector machine learning* (2019).

[464] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–14.

[465] Eran Vigoda-Gadot and Itai Beeri. 2011. Change-oriented organizational citizenship behavior in public administration: The power of leadership and the cost of organizational politics. *Journal of Public Administration Research and Theory* 22, 3 (2011), 573–596.

[466] Amy Voida, Lynn Dombrowski, Gillian R Hayes, and Melissa Mazmanian. 2014. Shared values/conflicting logics: working around e-government systems. In *Proceedings of the sigchi conference on human factors in computing systems*. 3583–3592.

[467] Paul Voigt and Axel Von dem Bussche. 2017. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing* (2017).

[468] Claudia Wagner, Markus Strohmaier, Alexandra Olteanu, Emre Kıcıman, Noshir Contractor, and Tina Eliassi-Rad. 2021. Measuring algorithmically infused societies. *Nature* 595, 7866 (2021), 197–204.

[469] D Wagner, K Johnson, and W Johnson. 1998. Using actuarial risk assessment to target service nterventions in pilot California Counties. *13th National Roundtable on CPS Risk Assessment, San Francisco, CA* (1998).

[470] Hanna M Wallach. 2006a. Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd international conference on Machine learning*. ACM, 977–984.

[471] Hanna M. Wallach. 2006b. Topic Modeling: Beyond Bag-of-Words. In *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*. Association for Computing Machinery, New York, NY, USA, 977–984. DOI:http://dx.doi.org/10.1145/1143844.1143967

[472] James A Walsh and Roberta A Walsh. 1990. Studies of the maintenance of subsidized foster placements in the Casey Family Program. *Child Welfare* 69, 2 (1990), 99–114.

[473] Anne L Washington. 2018. How to argue with an algorithm: Lessons from the COMPAS-ProPublica debate. *Colo. Tech. LJ* 17 (2018), 131.

[474] Max Weber. 1978. *Economy and society: An outline of interpretive sociology*. Vol. 2. University of California press.

[475] Daniel Webster, Richard P Barth, and Barbara Needell. 2000. Placement stability for children in out-of-home care: A longitudinal analysis. *CHILD WELFARE-NEW YORK-* 79, 5 (2000), 614–632.

[476] Jane Webster and Richard T Watson. 2002. Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly* (2002), xiii–xxiii.

[477] Sandra K Welch. 2008. *A metasynthesis of the transition from novice to expert: Can instructional interventions shorten the process?* Ph.D. Dissertation. Capella University.

[478] Cedric Deslandes Whitney, Teresa Naval, Elizabeth Quepons, Simrandeep Singh, Steven R Rick, and Lilly Irani. 2021. HCI Tactics for Politics from Below: Meeting the Challenges of Smart Cities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[479] James K Whittaker. 2017. *The child welfare challenge: Policy, practice, and research*. Routledge.

[480] Maranke Wieringa. 2020. What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 1–18.

[481] Melody Williams. 2014. Child welfare in Milwaukee, Wisconsin. (2014).

[482] Ben Williamson. 2016. Digital education governance: data visualization, predictive analytics, and 'real-time'policy instruments. *Journal of Education Policy* 31, 2 (2016), 123–141.

[483] C Wilson. 2014. Integrating safety, permanency, and well-being: A view from the field. *Integrating safety, permanency and well-being series* (2014).

[484] Children's Wisconsin. 2021. Family Case Manager - Child Welfare. (2021). https://indeedhi.re/3yun3aj

[485] Jacob O Wobbrock and Julie A Kientz. 2016. Research contributions in human-computer interaction. *interactions* 23, 3 (2016), 38–44.

[486] Jill Palzkill Woelfer and David G Hendry. 2010. Homeless young people's experiences with information systems: life and work in a community technology center. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1291–1300.

[487] Susan P Wyche, Paul M Aoki, and Rebecca E Grinter. 2008. Re-placing faith: reconsidering the secular-religious use divide in the United States and Kenya. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 11–20.

[488] Jason Yan, Melanie Sage, Seventy F Hall, Yuhao Du, and Kenneth Joseph. 2021. A Computational Social Science Approach to Understanding Predictors of Chafee Service Receipt. *arXiv preprint arXiv:2111.14901* (2021).

[489] Karen Yeung. 2018. Algorithmic regulation: a critical interrogation. *Regulation & Governance* 12, 4 (2018), 505–523.

[490] Matthew M Young, Justin B Bullock, and Jesse D Lecy. 2019. Artificial discretion as a tool of governance: a framework for understanding the impact of artificial intelligence on public administration. *Perspectives on Public Management and Governance* 2, 4 (2019), 301–313.

[491] Aleš Završnik. 2019. Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology* (2019), 1477370819876762.

[492] Kristen M Zgoba and Nishi P Dayal. 2015. Recidivism. *The Encyclopedia of Crime and Punishment* (2015), 1–5.

[493] Angie Zhang, Alexander Boltz, Chun Wei Wang, and Min Kyung Lee. 2022. Algorithmic management reimagined for workers and by workers: Centering worker well-being in gig work. In *CHI Conference on Human Factors in Computing Systems*. 1–20.

[494] Haiyi Zhu, Bowen Yu, Aaron Halfaker, and Loren Terveen. 2018. Value-sensitive algorithm design: Method, case study, and lessons. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–23.

[495] Marc A Zimmerman. 2013. Resiliency theory: A strengths-based approach to research and practice for adolescent health. (2013).

[496] Susan J Zuravin. 1999. Child neglect. *Neglected children: Research, practice, and policy* (1999), 24–46.