Designing for Human-Centered AI in the U.S. Child-Welfare System

Devansh Saxena
Dept. of Computer Science
Marquette University
Milwaukee, WI 53233, USA
devansh.saxena@marquette.edu

Shion Guha
Dept. of Computer Science
Marquette University
Milwaukee, WI 53233, USA
shion.guha@marquette.edu

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract
Child-Welfare System (CWS) in many states in the United States has come under escalating public and media scrutiny because of the potential damage done to children who are removed from the care of their parents. CWS has increasingly turned towards artificial intelligence (AI) as a way of standardizing decisions and demonstrating that these decisions are unbiased and evidence-based. Moreover, CWS in almost every state is underfunded and AI systems, from the perspective of policymakers, offer a potential means to reduce costs. Our research focuses on the collaborative work of child-welfare teams that participate in meetings mediated by policy, practice, and algorithms. In the following paragraphs, we first offer some background context in regards to the child-welfare system in the state of Wisconsin. Next, we establish a need for human-centered AI, and finally, we discuss how strategies proposed by human-centered algorithm design (HCAD) can help inform the development of algorithms in CWS.

Author Keywords
Child Welfare System, Algorithmic Decision-Making, Human-Centered Algorithm Design

CCS Concepts
•**Applied Computing** → **Computing in Government;**

Participant Quotes

Prescribing foster-parent rates: *"Foster parents need this money but we are punishing foster parents for being emotionally involved with children and helping them slowly recover. All of this because some policymaker thought it would be a good way to cut costs" -P1*

Impact of Trauma: *"It took my own adopted son 5 years to open up to me and tell me that his mother had set the apartment on fire to get away from paying rent. All this time he was trying to protect his mom even though she almost killed him and his brother." -P2*

Co-opting CANS: *"The subsidized rate is so low that there is no financial incentive in this for foster parents and the caseworkers are just trying to help foster parents by bending CANS any way they can" -P3*

Introduction

Contrary to popular belief, decisions in the child-welfare system are not made by individual caseworkers and consensus decisions are made by a multi-disciplinary team that brings in domain knowledge from family psychology, law, medicine, and social work [7]. This allows for an accountable and well-documented decision-making process directed by experts from different disciplines. Moreover, this team is trained in the medical practice of trauma-informed care [10] which places trauma front and center and allows the team to deliberate over a child's needs based on possible trauma symptoms resulting from trauma exposure. However, in recent years, algorithms have been introduced that this team must utilize as a means of standardizing decisions and purportedly add another layer of accountability. We are presenting a paper at CHI 2020 where we review the algorithms currently being used by CWS across the United States [8]. Below, we first discuss the changing ecology of work in CWS because of the introduction of AI systems. Next, we discuss exploratory findings from our ongoing ethnographic study that pertain to need for human-centered AI [7].

Ecological Impact of Algorithms in CWS

The caseworkers' job is to navigate through the system and find services for children and families as well as act as a mediator between the children's court, district attorney's office, and families. They are not trained to evaluate algorithmic decisions or think statistically, per se [5]. Algorithms are designed to assist CWS workers in the decision making process, however, CWS workers are being pressured to use the algorithmic judgment as a means of standardizing decisions [9]. Algorithms carry a perception of being value-free and more objective and subtly changes how CWS workers assess cases. For instance, Eubanks explains in her book *Automating Inequality* [4] how a risk assessment algorithm being used by CWS in Pittsburgh was inadvertently supplanting human decision-making. CWS workers began questioning their own assessment when faced with an algorithmic risk score even though the algorithm was routinely wrong about several individual cases. In another scenario, algorithms are being used as secondary decision aids, however, the caseworkers must translate information from both forms of assessments (clinical and algorithmic) leading to uncertainty and unreliable decision making [9]. Therefore, algorithms that are meant to aid decision-making often become the source of frustration and force caseworkers to abandon their contextual judgments [9]. These problems arise because of the opaque and isolated development of algorithms that does not account for the needs of the workers who are expected to use these systems.

Ethnographic Study in Wisconsin

We are conducting an ethnographic study in the state of Wisconsin to investigate the interaction of policy, practice and algorithms in CWS. Some of our initial findings from this study were presented at GROUP 2020 [7]. We are especially concerned with how an algorithm influences the work of child-welfare teams. The *Child and Adolescent Needs and Strength (CANS)* algorithm is used to assess the level of need of a foster child by determining the associated risk factors as well as well-being indicators (see [7] for details). Based on the level of need, it recommends a placement setting capable of meeting those needs. However, this algorithm often does not agree with theory-driven practices founded in trauma-informed care. Moreover, this algorithm has been reappropriated to perform a task that it was not design for which has led to unintended consequences and added to the frustrations of child-welfare teams.

Pitfalls in the Empirical Data

Empirical knowledge related to CWS is quite fragmented and social science theories must be used to fill in the gaps [5].

Moreover, most predictors currently being used for risk assessment have not been properly validated, are unreliable, and overlook the broader ecology of risk resulting from systemic barriers [8]. Furthermore, the target outcome of *Child Maltreatment* is poorly defined [5]. Algorithms are trained on cases of substantiation, that is, cases where CWS concluded maltreatment to have occurred. This judgment in itself is very subjective because the CWS intervention criteria is dictated by the state's definition for child maltreatment, as well as the level of funding and caseloads at CWS [8].

Re-appropriation of the CANS algorithm

CANS algorithm was designed to assess the needs of a child and recommend a placement setting, however, it has also been re-appropriated to calculate the subsidized guardianship rate offered to foster parents by the state. Based on the algorithm's recommendation, the higher the needs of a foster child, the higher rate is offered to foster parents. However, *CANS* is recalculated every six months and as the child supposedly exhibits less trauma symptoms, their needs are lowered and so is the rate offered to foster parents. One child-welfare worker explained that by lowering the rate, foster parents were being punished for being emotionally involved and helping traumatized children (see P1 quote).

Theory-driven practice vs. CANS algorithm

CWS team members are now being trained in trauma informed care and one caseworker explained that trauma stays with a child for years and cannot be alleviated in a few months (see P2 quote). However, *CANS* does not account for trauma (or traumatic triggers) and expects the children to keep getting better every six months. This is directly at odds with theory-driven practice that the workers are trained in and adds to their frustration because it disincentivizes foster parents who are actively involved and help a child cope with trauma and progress emotionally.

Co-opting the CANS algorithm

Both the caseworkers and the foster parents have now begun to figure out how the algorithm calculates the subsidized guardianship rate so they have started to co-opt *CANS* to either account for trauma-informed care or simply anticipate child needs. This has, however, led to a tricky situation in regards to the data provenance of *CANS* algorithm. As P3 explains, there is no financial incentive in co-opting the system to receive a higher rate. It is the state's expectation that the children magically get better every few months so that they can lower the rate offered to foster parents to cut costs.

Human-Centered Algorithmic Design (HCAD)

Baumer proposed Human-Centered Algorithm Design [1], a conceptual framework that incorporates human and social interpretations into the design process of algorithms. It 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 human-centered theoretical design strategy to inform algorithm design as follows:

- **Meaning-making**: Designers need to study the sociocultural domain in which they intend to situate their work. It is imperative to understand how CWS workers make sense of algorithmic results. CWS workers are often not trained in statistics and misconstrue probabilities for perfect predictions [4], doubting their own contextual judgments.
- **Design**: Theoretical approaches aim to incorporate concepts and theories from social sciences into data science [1]. This is especially important in case of CWS where the empirical knowledge is fragmented and the predictors and/or outcomes are ill-defined. Algorithms need to be theoretically constructed by selecting predictors that are well-defined and properly validated to ensure a lower compounded error and more reliable predictions. From our

observation study we also discovered that CWS workers maintain a rich narrative on every case by writing detailed case-notes that are stored in a database. We hypothesize that valuable theoretical signals from these case notes can be considered within methodological approaches like topic modeling. In recent years, HCI researchers have used signals from such unstructured data as predictors within algorithms to study complex, sociotechnical systems [2, 6].

- **Evaluation**: The stakeholders' interpretations of results can help ensure that the algorithm has higher utility and integrates well with practice. Social interpretations also play an important role in understanding how the community perceives the role of AI systems. For instance, Brown et al. [3] discovered significant concerns about the deficit-based framing of risk assessment algorithms. These algorithms aggregate the negative aspects of a person's life while ignoring the positive aspects, thereby, driving negative perceptions towards a person and their community.

We see human-centered design as a paradigm that informs the methodological choices that we make with respect to a domain. For instance, models used in CWS need to be interpretable and explainable because multi-disciplinary CWS teams must discuss predictions and reach consensus decisions. Therefore, we need to use computational methods that offer higher transparency and are easier to explain. Moreover, we know that case-notes contain important contextual information that is more pertinent to a case than the quantitative data derived from psychometric assessments. Therefore, we could possibly extract rich qualitative signals from this unstructured text by using NLP techniques like topic modeling. Furthermore, it is imperative to engage in participatory design and involve the stakeholders because understanding the theoretical knowledge of a domain as well as the values and needs of stakeholders most affected by AI systems are the pre-requisites for building fair and responsible AI. These pre-requisites allow us to make careful methodological decisions that lead to development of AI systems that offer higher utility and help caseworkers instead of acting as yet another systemic barrier.

REFERENCES

[1] Eric PS Baumer. 2017. Toward Human-Centered Algorithm Design. Big Data & Society 4, 2 (2017). (2017).

[2] Eric PS Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K Gay. 2017. 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.

[3] 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*. 1–12.

[4] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

[5] Eileen Gambrill and Aron Shlonsky. 2000. Risk assessment in context. (2000).

[6] 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*. 3–8.

[7] 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.

[8] Devansh Saxena, Karla Badillo-Urquiola, Pamela J Wisniewski, and Shion Guha. 2020b. A Human-Centered Review of the Algorithms used within the U.S. Child Welfare System.". In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.

[9] 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.

[10] 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.