A Future for AI Governance Systems beyond Predictions

Devansh Saxena, Erina Moon, Shion Guha

Public sector agencies in several western liberal democracies have adopted algorithmic systems as a means to provide consistent and evidence-based decisions to citizens who utilize these services (e.g., Medicaid, unemployment, SNAP, SSI program) or come under the attention of these punitive systems (e.g., criminal justice, child welfare). AI systems hold the promise of transforming governmental interactions with people by supporting information processing and decision-making where digital platforms and data-driven methods simplify application processes and automate some aspects of decision-making.

It also purportedly allows agencies to “do more with less” by minimizing workers’ tasks through minimal repeated information gathering, identifying clients in the most risky circumstances, and reducing bureaucratic overhead such that clients receive services quickly and efficiently [6]. However, instead of transforming public sector operations at a deeper organizational level, several AI tools have fused onto existing practices where we witness complex street-level interactions between human decision-making, systemic constraints, policies/practices, and AI systems. Ethnographic work conducted by HCI and STS researchers in the public sector reveal much of these messy interactions where algorithms designed to augment human discretion are instead curtailing it and leading to poor and inconsistent decision-making on part of workers who are mandated to use them. Recognizing these complex interactions between human discretionary work, algorithmic decision-making, and bureaucratic processes and through our own ethnographic work, we developed a theoretical framework for algorithmic decision-making for the public sector that provides guidelines for designing algorithms that augment human discretionary work [3]. Central to this framework is the recognition that algorithms need to support workers’ decision-making processes instead of providing predicted outcomes that take discretionary work away from workers. As highlighted by recent work [1], predictive systems in high-stakes domains are *extractive by design* where they lead to systemic extraction of discretionary power such that probabilistic outcomes (based on quantifiable historical data) are increasingly being used to make definitive judgments and supplant workers’ contextual knowledge. However, in order to provide higher utility and improve decision-making processes, we need to learn from the discretionary choices that workers make as they employ their professional expertise and navigate complex sociotechnical systems. In the paragraphs below, we highlight some pertinent findings from our research conducted in the child-welfare system, how they relate to the themes of this workshop.

Prescribing away from Predictions

A decade of ethnographic and computational research conducted in the public sector shows that algorithms are exacerbating racial disparities and achieving worse outcomes for citizens. The fundamental issue here is that empirical knowledge in the public sector derived from historical administrative data is quite fragmented. This problem is further aggravated by the fact that several predictors and outcomes are poorly and inconsistently defined, are unreliable, and have not been validated [3]. Here, blackbox predictions frustrate workers who must continually provide additional labor on top of high caseloads to be able to 'work around' nonsensical predictions, or worse, allow the algorithm to supplant their judgment to limit their workload [2]. To address some of these concerns, the child-welfare agency that we collaborate with has moved away from predictive empirical models and designed a simple decision tool (7ei) using a trauma-informed care (TIC) framework [3]. The tool allows child-welfare teams to discuss, score, and track seven TIC variables over the life of every case. Here, the tool was designed with the intent to decompose the algorithm and turn it into an open-ended and transparent process centered in trauma-informed care where it tracked outcomes over time instead of predicting an outcome of interest. The tool fosters collaboration where the team is able to leverage strengths of both human intelligence (i.e., knowledge of TIC principles) and machine intelligence (i.e., trajectory of cases) to improve decision-making.

Designing for Collaborative Use and Facilitating Explanations

Decision-making in the public sector is a complex process where information needs to be shared among several parties (e.g., child-welfare staff, district attorney's office, parent's attorney, judges) and decisions are collaboratively made [3]. AI tools must be designed for collaborative use for them 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. Different parties can have conflicting interests and it is hard to agree upon a path forward without adequate explanations that support data-driven decisions. 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 [7]. The 7ei tool offers a case in point that was designed with these core values in mind, however, the agency had to make significant investments in terms of TIC trainings offered to staff, hiring experts, and creating the time and space (i.e., collaborative meetings) for the proper utilization of the tool [3].

Designing within Systemic Constraints

Domain-specific constraints relating to the organizational, bureaucratic, and legislative are intrinsically embedded in public sociotechnical systems and deeply affect how involved parties perceive and utilize public sector AI tools [2, 3]. These constraints can result in complex power dynamics between involved parties and influence the administrative data used to build public sector AI tools. When building AI tools for complex domains that involve multiple parties with differing motivations and responsibilities, we need to account for contextualized insights by affected parties. Our work investigating child welfare casenotes and risk assessment data used to build child welfare algorithmic tools finds that narrative documentation can illuminate nuanced and dynamic case details, which are often obfuscated in quantitative risk metrics [4, 5]. Narrative documentation can provide a useful critical lens into street-level perspectives of frontline workers as they describe the rationale for decisions made in the face of constraints and complex power asymmetries experienced by involved parties. While administrative data will flatten and reduce domain complexities into quantifiable metrics, casenote documentation can provide supporting and supplementary commentary on why data may be missing or incomplete due to privacy or organizational constraints [5].

REFERENCES

1. Sun-ha Hong. "Prediction as Extraction of Discretion." *2022 ACM Conference on Fairness, Accountability, and Transparency*. 2022.
2. Devansh Saxena, Karla Badillo-Urquiola, Pamela J. Wisniewski, and Shion Guha. 2020. A Human-Centered Review of Algorithms used within the U.S. Child Welfare System. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20).* Association for Computing Machinery, New York, NY, USA, 1–15.
3. Devansh Saxena, Karla Badillo-Urquiola, Pamela J. Wisniewski, and Shion Guha. 2021. A Framework of High-Stakes Algorithmic Decision-Making for the Public Sector Developed through a Case Study of Child-Welfare. *Proc. ACM Hum.-Comput. Interact. 5, CSCW2,* Article 348 (October 2021), 41 pages.
4. Devansh Saxena, Seh Young Moon, Dahlia Shehata, and Shion Guha. 2022. 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 (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 120, 1–22.
5. Devansh Saxena, Charles Repaci, Melanie D Sage, and Shion Guha. 2022. 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.*
6. Michael Veale and Irina Brass. "Administration by algorithm? Public management meets public sector machine learning." *Public management meets public sector machine learning* (2019).
7. 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 (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 41, 1–12. https://doi.org/10.1145/3290605.3300271