Artificial Intelligence, Machine Learning and Algorithmic Decision-Making in Child Welfare - An Ongoing Scan of the Academic Literature

Curated by Melanie Sage and Laura Burney NIssen

November 2019

Increasingly, all aspects of our lives are mediated by technology-enhanced algorithmic decision making. Child welfare is no exception. In fact, we have worked for years in the child welfare sector to map and enhance the structure and explainability of child welfare decision-making. Still, we have much to learn related to how child welfare caseworkers make implicit and explicit decisions about which families are most at risk.

We have tried to address concerns about fairness through the use of structured risk assessments. Historically, these algorithmic decision making tools were paper-based processes completed by caseworkers, such as Risk-Assessment Scoring and Structured Decision Making forms that workers hand-calculated to produce risk scores. These were algorithms that assessed factors that contributed to greater risk or safety based on what we know about common protective and risk factors.

New methods of technology-enhanced algorithmic decision-making allow these scores to be computed automatically based on massive data sets. Mathematical models are typically built by looking at historical data and the variables associated with risk in past cases that resulted in negative child outcomes such as out-of-home placement, serious injury, or fatality. Which variables are used and how they are weighted vary by model, but typically include easily-collectible demographic data, history of child welfare involvement, and information about household members. Depending on what databases an agency has access to, these variables may include information related to the use of other public services, criminal history, and zip code, for example.

It is becoming common that child welfare agencies contract with outside vendors to purchase these models to use as an aid in human decision-making processes. A score may be calculated by the model, and then used to inform a decision about what kind of response the child welfare agency should make. It is very important that social service agencies who use or purchase risk-prediction models understand important issues about how they are made, including how they might promote bias. A risk assessment model should align with an agency's mission and practice model.

As social workers, our values support algorithmic decision making which is deployed in ways that are transparent, explainable, and accountable to stakeholders; that social workers and others who engage in the deployment of algorithmic decision making understand how the systems work and account for potential sources of bias; and that these processes should work

alongside human decision making, and should improve our ability to intervene and improve the safety and ongoing well-being of families.

This collection of resources offers a snapshot of uses, concerns, and information about algorithmic decision-making in child welfare to date. The bibliography offers an emerging picture of the strengths and concerns related to their use.

The purpose of this resource is to be an ongoing and evolving space to organize scholarship and resources for researchers, professors, scholars, community members, practitioners and students who may be interested in this topic.

Academic Papers

Amrit, C., Paauw, T., Aly, R., & Lavric, M. (2017). Identifying child abuse through text mining and machine learning. *Expert systems with applications*, *88*, 402-418.

In this paper, we describe how we used text mining and analysis to identify and predict cases of child abuse in a public health institution. Such institutions in the Netherlands try to identify and prevent different kinds of abuse. A significant part of the medical data that the institutions have on children is unstructured, found in the form of free text notes. We explore whether these consultation data contain meaningful patterns to determine abuse. Then we train machine learning models on cases of abuse as determined by over 500 child specialists from a municipality in The Netherlands. The resulting model achieves a high score in classifying cases of possible abuse. We methodologically evaluate and compare the performance of the classifiers. We then describe our implementation of the decision support API at a municipality in the Netherlands. (Author abstract.)

Brown, A., Chouldechova, A., Putnam-Hornstein, E., Tobin, A. & Vaithianathan, R. (2019). Toward algorithmic accountability in public service. *Glasgow, Scotland CHI Meeting.*

Algorithmic decision-making systems are increasingly being adopted by government public service agencies. Researchers, policy experts, and civil rights groups have all voiced concerns that such systems are being deployed without adequate consideration of potential harms, disparate impacts, and public accountability practices. Yet little is known about the concerns of those most likely to be affected by these systems. We report on workshops conducted to learn about the concerns of affected communities in the context of child welfare services. The workshops involved 83 study participants including families involved in the child welfare system, employees of child welfare agencies, and service providers. Our findings indicate that general distrust in the existing system contributes significantly to low comfort in algorithmic decision-making. We identify strategies for improving comfort through greater transparency and

improved communication strategies. We discuss the implications of our study for accountable algorithm design for child welfare applications. (Author abstract.)

Chouldechova, A., Putnam-Hornstein, E., Benavides-Prado, E., Fialko, O. * Vaithianathan, R. (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. *Proceedings of Machine Learning Research, 81*(1), 1-15.

Every year there are more than 3.6 million referrals made to child protection agencies across the US. The practice of screening calls is left to each jurisdiction to follow local practices and policies, potentially leading to large variation in the way in which referrals are treated across the country. Whilst increasing access to linked administrative data is available, it is difficult for welfare workers to make systematic use of historical information about all the children and adults on a single referral call. Risk prediction models that use routinely collected administrative data can help call workers to better identify cases that are likely to result in adverse outcomes. However, the use of predictive analytics in the area of child welfare is contentious. There is a possibility that some communities— such as those in poverty or from particular racial and ethnic groups—will be disadvantaged by the reliance on government administrative data. On the other hand, these analytics tools can augment or replace human judgments, which themselves are biased and imperfect. In this paper we describe our work on developing, validating, fairness auditing, and deploying a risk prediction model in Allegheny County, PA, USA. We discuss the results of our analysis to-date, and also highlight key problems and data bias issues that present challenges for model evaluation and deployment. (Author abstract.)

Chung, H., Stewart, C.J., Rose, R.A. & D.F. Duncan. (2015). Using big data for evidence-based governance in child welfare. *Children and Youth Services Review, 58,* 127-136.

Numerous approaches are available for improving governance of the child welfare system, all of which require longitudinal data reporting on child welfare clients. A substantial amount of agency administrative information – big data – can be transformed into knowledge for policy and management actions through a rigorous information generation process. Important properties of the information generation process are that it must generate accurate, timely information while protecting the confidentiality of the clients. In addition, it must be extensible to serve an ever-changing policy and technology environment. Knowledge discovery and data mining (KDD), aka data science, is a method developed in the private sector to mine consumer data and can be used in public settings to support evidence based governance. KDD consists of a rigorous 5-step process that includes a Web based end-user interface. The relationship between KDD and governance is a continuous feedback cycle that enables ongoing development of new information and knowledge as stakeholders identify emerging needs. In this paper, we synthesis the different frameworks for utilizing big data for public governance, introduce the KDD process, describe the nature of big data in child welfare, and then present an updated KDD architecture that can support these frameworks to utilize big data for governance. We also demonstrate the role KDD plays in child welfare management through 2 case studies. We conclude with a

discussion on implications for agency–university partnerships and research-to-practice. (Author abstract.)

Church, C. E. & Fairchild, A.J. (2017). In search of a silver bullet: Child welfare's embrace of predictive analytics. *Juvenile and Family Court Journal, 68*(1).

Predictive analytics has shaken up a number of fields, including child welfare. Predictive analytics refers to the process of applying statistical algorithms to data to make informed guesses about future events. Although predictive analytics can help professionals make decisions more accurately, objectively, and quickly, there is a concern that some methods may result in discriminatory practices or consequences for vulnerable children and families. This paper examines a number of programmatic and ethical considerations for determining the appropriate role of predictive analytics in child welfare. (Author abstract.)

Cuccaro-Alamin, S., Foust, R., Vaithiananathan, R. & E. Putnam-Hornstein (2017). Risk assessment and decision making in child protective services: Predictive risk modeling in contect. *Children and Youth Services Review, 79*, 291-298.

In an era in which child protective service agencies face increased demands on their time and in an environment of stable or shrinking resources, great interest exists in improving risk assessment and decision support. In this article, we review the literature and provide a context for predictive risk modeling in the current risk assessment paradigm in child protective services. We describe how predictive analytics or predictive risk modeling using linked administrative data may provide a useful complement to current approaches. We argue that leveraging technology and using existing data to improve initial triage and assessment decisions will enable caseworkers to focus on what they do best: engaging families and providing needed services. (Author abstract.)

Daley, D., Bachmann, M., Bachmann, B. A., Pedigo, C., Bui, M. T., & Coffman, J. (2016). Risk terrain modeling predicts child maltreatment. *Child abuse & neglect*, *62*, 29-38.

As indicated by research on the long-term effects of adverse childhood experiences (ACEs), maltreatment has far-reaching consequences for affected children. Effective prevention measures have been elusive, partly due to difficulty in identifying vulnerable children before they are harmed. This study employs Risk Terrain Modeling (RTM), an analysis of the cumulative effect of environmental factors thought to be conducive for child maltreatment, to create a highly accurate prediction model for future substantiated child maltreatment cases in the City of Fort Worth, Texas. The model is superior to commonly used hotspot predictions and more beneficial in aiding prevention efforts in a number of ways: 1) it identifies the highest risk areas for future instances of child maltreatment with improved precision and accuracy; 2) it aids the prioritization of risk-mitigating efforts by informing about the relative importance of the most significant contributing risk factors; 3) since predictions are modeled as a function of easily obtainable data, practitioners do not have to undergo the difficult process of obtaining official child

maltreatment data to apply it; 4) the inclusion of a multitude of environmental risk factors creates a more robust model with higher predictive validity; and, 5) the model does not rely on a retrospective examination of past instances of child maltreatment, but adapts predictions to changing environmental conditions. The present study introduces and examines the predictive power of this new tool to aid prevention efforts seeking to improve the safety, health, and wellbeing of vulnerable children. (Author abstract.)

Gillingham, P. (2016). Predictive risk modelling to prevent child maltreatment and other adverse outcomes for services users: Inside the 'black box' of machine learning. *British Journal of Social Work, 46,* 1044-1058.

Recent developments in digital technology have facilitated the recording and retrieval of administrative data from multiple sources about children and their families. Combined with new ways to mine such data using algorithms which can 'learn', it has been claimed that it is possible to develop tools that can predict which individual children within a population are most likely to be maltreated. The proposed benefit is that interventions can then be targeted to the most vulnerable children and their families to prevent maltreatment from occurring. As expertise in predictive modelling increases, the approach may also be applied in other areas of social work to predict and prevent adverse outcomes for vulnerable service users. In this article, a glimpse inside the 'black box' of predictive tools is provided to demonstrate how their development for use in social work may not be straightforward, given the nature of the data recorded about service users and service activity. The development of predictive risk modelling (PRM) in New Zealand is focused on as an example as it may be the first such tool to be applied as part of ongoing reforms to child protection services. (Author abstract.)

Glaberson, S.K. (2019). Coding over cracks: Predictive analytics in child welfare. *Fordham Urban Law Journal, 46*(2), 307-362.

Across the nation, child protective authorities are turning to machines to assist them in their work, developing predictive analytic tools to forecast risk to children and families. While there is clear evidence that current child welfare decision-making processes are flawed and in need of change, the advent of predictive analytics carries with it numerous risks to children and families that cannot be ignored. This Article explains the fundamentally human processes that go into the creation of predictive analytic tools and highlights some of the risks that these tools pose. It argues that the choices made in developing predictive tools implicate some of the most fundamental and as-yet unanswered questions in our child welfare system. As a result, the advent of predictive analytics in child welfare presents a moment for systemic reflection. Without careful attention to the issues that predictive analytics raise, communities risk simply coding over the cracks in the foundation of a flawed system, burying problems of bias, transparency, and accountability deeper, and imbuing the status quo with an undue patina of inevitability. Instead, communities should use this moment to demand more of their child welfare systems and see these tools as opportunities to build better, more humane systems that focus more on support and prevention and less on too-little, too-late crisis response. (Author abstract.)

Kedell, E. (2014). The ethics of predictive risk modelling in the Aotearoa/New Zealand child welfare context: Child abuse prevention or neo-liberal tool? *Critical Social Policy,* (Published online July 28, 2014).

The current White Paper on Vulnerable Children before the Aotearoa/New Zealand (A/NZ) parliament proposes changes that will significantly reconstruct the child welfare systems in this country, including the use of a predictive risk model (PRM). This article explores the ethics of this strategy in a child welfare context. Tensions exist, including significant ethical problems such as the use of information without consent, breaches of privacy and stigmatisation, without clear evidence of the benefits outweighing these costs. Broader implicit assumptions about the causes of child abuse and risk and their intersections with the wider discursive, political and systems design contexts are also discussed. Drawing on Houston et. al. (2010) this paper highlights the potential for a PRM to contribute to a neo-liberal agenda that individualises social problems, reifies risk and abuse, and narrowly prescribes service provision. However, with reference to child welfare and child protection orientations, the paper suggests ways the model could be used in a more ethical manner. (Author abstract.)

Lanier, P., Rodriguez, M., Verbiest, S., Bryant, K., Guan, T. & Zolotor, A. (2019). Preventing infant maltreatment with predictive analytics: Applying ethical principles to evidence-based child welfare policy. *Journal of Family Violence,* (Published online 7 JUne 2019).

Infant maltreatment is a devastating social and public health problem. Birth Match is an innovative policy solution to prevent infant maltreatment that leverages existing data systems to rapidly predict future risk through linkage of birth certificate and child welfare data then initiate a child protection response. Birth Match is one example of child welfare policy that capitalizes on recent advances in computing technology, predictive analytics, and algorithmic decision making. We apply frameworks from business and computer science as a case study in ethical decision-making in child welfare policy. Current Birth Match policy applications appear to lack key aspects of transparency and accountability identified in the frameworks. Although technology holds promise to help solve intractable social problems such as fatal infant maltreatment, the decision to deploy such policy innovations must consider ethical questions and tradeoffs. Technological advances hold great promise for prevention of fatal infant maltreatment, but numerous ethical considerations are lacking in current implementation and should be considered in future applications. (Author abstract.)

Pryce, J., Yelick, A., Zhang, Y. & Fields, K. (2018). Using artificial intelligence, machine learning and predictive analytics in decision making. Tallahassee, FL: Florida Institute for Child Welfare at Florida State University.

This is a “101” style guide to basic AI, machine learning and predictive analytics concepts for child welfare workers.

Thurston, H. & Miyamoto, S. (2018). The use of model based partitioning as an analytic tool in child welfare. *Child Abuse and Neglect, 79,* 293-301.

Child welfare agencies are tasked with investigating allegations of child maltreatment and intervening when necessary. Researchers are turning to the field of predictive analytics to optimize data analysis and data-driven decision making. To demonstrate the utility of statistical algorithms that preceded the current predictive analytics, we used Model Based (MOB) recursive partitioning, a variant of regression analysis known as decision trees, on a dataset of cases and controls with a binary outcome of serious maltreatment (defined as hospitalization or death). We ran two models, one which split a robust set of variables significantly correlated with the outcome on the partitioning of a proxy variable for environmental poverty, and one which ran the same variable set partitioned on a variable representing confirmed prior maltreatment. Both models found that what most differentiated children was spending greater than 2% of the timeframe of interest in foster care, and that for some children, lack of Medicaid eligibility almost doubled or tripled the odds of serious maltreatment. We find that decision trees such as MOB can augment risk assessment tools and other data analyses, informing data-driven program and policy decision making. We caution that decision trees, as with any other predictive tool, must be evaluated for inherent biases that may be contained in the proxy variables and the results interpreted carefully. Predictive analytics, as a class, should be used to augment, but not replace, critical thinking in child welfare decision making. (Author abstract.)

Schwartz, I.M., York, P., Nowalkowski-Sims, E. & Ramos-Hernandez (2017). Predictive and prescriptive analytics, machine learning and child welfare risk assessment: The Broward County Experience. *Children and Youth Services Review, 81,* 309-320.

This paper presents the findings from a study designed to explore whether predictive analytics and machine learning could improve the accuracy and utility of the child welfare risk assessment instrument used in Broward County (Ft. Lauderdale, Florida). The findings from this study indicate that, indeed, predictive analytics and machine learning would significantly improve the accuracy and utility of the child welfare risk assessment instrument being used. If the predictive analytic and machine learning algorithms developed in this study would be deployed, there would be improved accuracy in identifying low, moderate and high risk cases, better matching between the needs of children and families and available services and improved child and family outcomes. This paper also identifies further areas of research and study. (Author abstract.)

White Papers and Reports

Alleghenhy County Reports (May 2019) - Developing Predictive Risk Models to Support Child Maltreatment Hotline Screening Decisions
https://www.alleghenycountyanalytics.us/index.php/2019/05/01/developing-predictive-risk-models-support-child-maltreatment-hotline-screening-decisions/

U.S. Dept. of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation - numerous papers on predictive analytics for child welfare.
https://aspe.hhs.gov/predictive-analytics-child-welfare

Chapin Hall & Chadwick Center Policy Brief (September 2018). Making the most of predictive analytics: Responsible and innovative uses in child welfare policy and practice.
https://www.chapinhall.org/wp-content/uploads/Making-the-Most-of-Predictive-Analytics.pdf

Casey Family Programs (April 2018). Considerations for implementing predictive analytics in child welfare.
https://caseyfamilypro-wpengine.netdna-ssl.com/media/Considerations-for-Applying-Predictive-Analytics-in-Child-Welfare.pdf

Primers and Strong Overview Resources on Artificial Intelligence

Benjamin, R. (2019). Assessing risk, automating racism. ***Science, 366*****(6464), 421-422.**

Desai, D.R. & J. A. Kroll (2017). Trust but verify: A guide to algorithms and the law. ***Harvard Journal of Law & Technology, Forthcoming Georgia Tech Scheller College of Business Research Paper No. 17-19.***

Gillingham, P. & Graham, T. (2016). Big data in social welfare: The development of a critical perspective on social work's latest "electronic turn." ***Austrailian Social Work,*** (Published online 16 March 2016).

Lee, N.T., Resnik, P. & Barton, G. (2019). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. ***Brookings Institute Report.*** ***Available here.***

Lee, N.T. (2018). Detecting racial bias in algorithms and machine learning. ***Journal of Information, Communications and Ethics in Society, 16*****(3), 252-260.**

The online economy has not resolved the issue of racial bias in its applications. While algorithms are procedures that facilitate automated decision-making, or a sequence of unambiguous instructions, bias is a byproduct of these computations, bringing harm to historically disadvantaged populations. This paper argues that algorithmic biases explicitly and implicitly harm racial groups and lead to forms of discrimination. Relying upon sociological and technical research, the paper offers commentary on the need for more workplace diversity within high-tech industries and public policies that can detect or reduce the likelihood of racial bias in algorithmic design and execution. The paper shares examples in the US where algorithmic biases have been reported and the strategies for explaining and addressing them. The findings of the paper suggest that explicit racial bias in algorithms can be mitigated by existing laws, including those governing housing, employment, and the extension of credit. Implicit, or unconscious, biases are harder to redress without more diverse workplaces and public policies that have an approach to bias detection and mitigation. The major implication of this research is that further research needs to be done. Increasing the scholarly research in this area will be a major contribution in understanding how emerging technologies are creating disparate and unfair treatment for certain populations. The practical implications of the work point to areas within industries and the government that can tackle the question of algorithmic bias, fairness and accountability, especially African Americans. The social implications are that emerging technologies are not devoid of societal influences that constantly define positions of power, values, and norms. The paper joins a scarcity of existing research, especially in the area that intersects race and algorithmic development.

Mehr, H. (2017). Artificial intelligence for citizen services and government. Boston, MA: Harvard Kennedy School Ash Center for Technology and Democracy.

Naccarato, T. (2010). Child welfare informatics: A proposed subspecialty for social work. *Children and Youth Services Review, 32,* 1729-1734.

Informatics is a term that has been used and applied to data collection, analysis, and information and communication technologies across many disciplines including public health, nursing, medicine, and, more recently, to social work. To date, no collective discussion involving policy makers, practitioners, and researchers in the social work field defining child welfare informatics and its implications to the discipline, including curriculum development has occurred. This paper offers a perspective to begin the dialogue of child welfare informatics and presents a working definition and role specification for those working as child welfare informaticians. Finally, recommendations are made on how to evolve child welfare informatics. These recommendations include highlighting the importance of informatics as a subspecialty in social work, its prospectus for child welfare policy reform, and implications for interdisciplinary, social work curriculum development. (Author abstract.)

Noriega, A., Garcia-Bulle, B., Pentland, A. & L. Tejernia (2018). Algorithmic fairness in targeting social welfare programs at scale. *Bloomberg Data for Good Exchange Conference, September 2018.*

Targeted social programs, such as conditional cash transfers (CCTs), are a major vehicle for poverty alleviation throughout the developing world. Only in Mexico and Brazil, these reach nearly 80 million people (25% of population), distributing +8 billion USD yearly. We study the potential efficiency and fairness gains of targeting CCTs by means of artificial intelligence algorithms. In particular, we analyze the targeting decision rules and underlying poverty prediction models used by national-wide CCTs in three middle income countries (Mexico, Ecuador, and Costa Rica). Our contribution is three-fold: 1) We show that, absent explicit measures aimed at limiting algorithmic bias, targeting rules can systematically disadvantage population subgroups, such as incurring exclusion errors 2.3 times higher on poor urban households compared to their rural counterparts, or exclusion errors 2.2 times higher on poor elderly households compared with poor traditional nuclear families. 2) We constrain the targeting algorithms towards achieving fairness, and show that, for example, mitigating urban/rural unfairness in Ecuador can imply substantial costs in overall accuracy, yet, we also show that in the case of Mexico mitigating unfairness across four different types of family structures can be achieved at no significant accuracy costs. 3) Finally, we provide an interactive decision-support platform that allows even non-expert stakeholders to explore the space of possible AI-based decision rules, visualize their implications in terms of efficiency, fairness, and their trade-offs; and ultimately choose designs that best fit their preferences and context. (Author abstract.)

Additional Links from the Web

Algorithmic fairness: A code-based primer for public-sector data scientists (2019)

Algorithmic solutions to bias: A technical guide (2019)

Applying AI for social good (2018)

Assessing risk/Automating racism (2019)

Discriminating systems: Gender, race and and power in AI (2019)

Joy Buolamwini TED talks - “How I’m fighting bias in algorithms (2016)

Kriti Sharma TED talks - “How to keep human bias out of AI” (2019)

Social work tech notes: Social work and future technology/what can be automated, will be (2018)

The Guardian view of AI in social work: Algorithms don’t have all the answers (2018)

This is how AI bias really happens - and why it's so hard to fix (2019)

Untold history of AI: Algorithmic bias was born in the 1980's (2019)

What is this "AI for social good?" (2019)