Using a Machine Learning Tool to Support High-Stakes Decisions in Child Protection

Rhema Vaithianathan, Diana Benavides-Prado, Erin Dalton, Alexandra Chouldechova, Emily Putnam-Hornstein

■ *Machine learning decision support tools have become popular in a range of social domains including healthcare, criminal justice, and child welfare. But the design of these tools often fails to consider the potentially complex interactions that happen between the tools and humans. This lack of human-centered design is one reason that so few tools are actually deployed, and even if they are, struggle to achieve impact. In this article we present the example of the Allegheny Family Screening Tool, a machine learning model used since 2016 to support hotline screening of child maltreatment referrals. We describe aspects of human-centered design that contributed to the successful deployment of this tool, including agency leadership and ownership, transparency by design, ethical oversight, community engagement, and social license. Finally, we identify potential next-steps to encourage greater integration of human-centered design into the development and implementation of machine learning decision support tools.*

Although there has been considerable achievement in the use of artificial intelligence (AI) in the commercial sector, increasing the profits of existing businesses and disrupting industries, these techniques have not transferred easily to the social domain. Indeed, there are very few cases where even the most basic machine learning tools have made any difference to the many intractable social problems facing those same countries where these AI methods have been so enthusiastically embraced for profit. In areas such as homelessness, severe mental illness, substance-use, and child-maltreatment, many ideas have been proposed (Saxena et al. 2020) and theoretically shown to be useful, but few have been taken up, and upon evaluation, almost none of them have been found to have a positive impact.

Even when some extremely limited form of AI is proposed, the application has been so poorly executed that while intending to improve social disadvantage, they may have exacerbated existing harms. A case in point is the use of recidivism tools in criminal justice. These simple predictive analytic tools have been shown to be better than judges at deciding which prisoner should be allowed to be on parole, and if implemented could reduce the number of people in jail (Kleinberg et al. 2018). Yet, in what is now a seminal exposé by ProPublica, these tools were found to have disparate impacts on Black and White individuals (Chouldechova 2017). The subsequent furor undermined community trust in the use of machine learning not just in criminal justice, but in other social domains as well.

What the recidivism tool example and countless others like it show, is that there is a large gap between the theory and practice of AI for social good. The purpose of this article is to canvass why it is so hard to make an impact on the grand challenges in the social domain. Using the lessons from machine learning tools in Child Welfare, we outline some of the key features of a successful translation of theoretical possibilities of machine learning into social good. We use the Allegheny Family Screening Tool (AFST) as a case study because it is a rare exception, having been deployed and evaluated and shown to have an actual impact on key aspects of child protection decision-making.

The Use of Machine Learning in Child Abuse and Neglect

The rates of child abuse and neglect deaths in high income countries have shown no evidence of decline (Gilbert et al. 2009), and these deaths continue to be a major public health challenge. High profile deaths have excited considerable levels of public concern. Nonetheless, standard public health approaches have had poor success in reducing severe child abuse and neglect, posing the type of challenge that would seem to be an ideal candidate for the use of machine learning and AI tools.

Research papers on the theoretical possibilities of predicting child maltreatment using machine learning methods have been published in the last decade (Amrit et al. 2017; Schwartz et al. 2017; Vaithianathan et al. 2013). Yet, attempts to deploy machine learning tools in child protection had more often than not failed up to the time Allegheny County took up the challenge. For example, in 2012, the New Zealand Ministry of Social Development proposed to use a predictive risk model to proactively identify families who registered on welfare for risk of maltreatment. To test the validity of their model, they proposed that they would use this tool on all families and follow the high-risk families to see if, indeed, the predicted maltreatment occurred. This proposal naturally garnered a lot of public backlash, with critics arguing that it was unacceptable to deploy such a tool in an experimental setting without responding to the children who had received high scores. The attempt was put on hold and has not been resurrected (Radio New Zealand 2015). A similar public backlash ensued in Illinois, where the Eckerd Rapid Safety Feedback® tool — a machine learning tool trained to identify risk of maltreatment deaths — was deployed. In this case, the concerns that surfaced in media reports and published papers included the stark language used in communicating risk information (Jackson and Marx 2017) and that the contract terms prevented sharing details of how the tool works and the data it uses (Brauneis and Goodman 2018).

The AFST

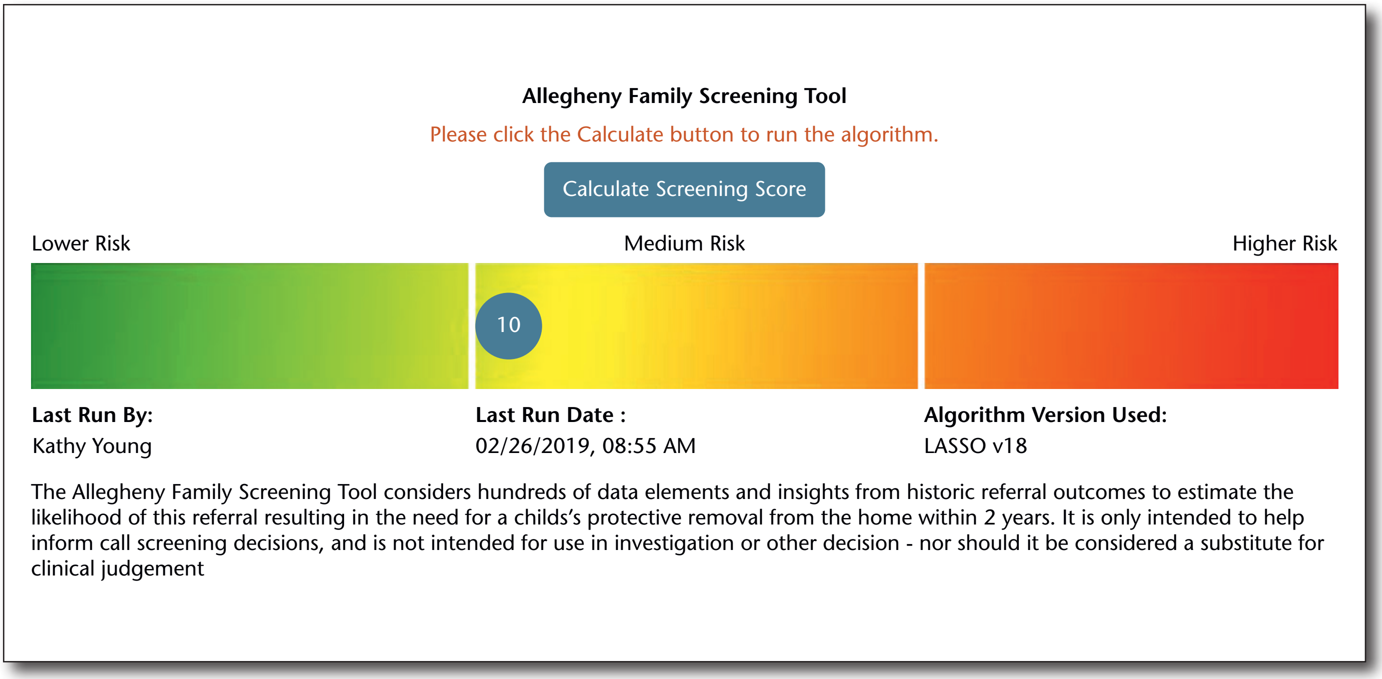
In 2016, Allegheny County, Pennsylvania, implemented a predictive risk model called the AFST, which is accessible to staff who triage calls received by the County regarding allegations of child abuse and neglect. The AFST helps workers decide whether to undertake further investigation of the calls or to screen them out — around half of all calls are screened out. These calls can come from mandated reporters such as teachers or physicians, who are required by law to report cases of suspected abuse or neglect, or from individuals in the community, such as family members and neighbors.

When a call is received by a staff member, they enter the details of the people (children, parents, alleged perpetrator, or others) involved in the call into the standard case management system. An AFST score is automatically generated and visualized from the predictive risk model (see figure 1).

The first version of the AFST was initially implemented in August 2016; the tool was fully deployed in November of that same year. The tool has now been in operation for over three years,[1] with a series of updates and modifications made during that time. The second version of the tool, deployed in December 2018, relies on a least absolute shrinkage and selection operator model to calculate a risk score reflecting the likelihood that the child (if screened in) would be removed from the home due to risk and safety concerns within the next two years. This score considers as input, information about the child for whom the allegation has been made and family members, including demographics, previous involvement in the Department of Human Services (DHS) child protection and jail systems, encounters for mental health and substance abuse disorders, and other data. Table 1 provides the performance of the model in terms of area under the receiver operator characteristic curve and true-positive rate on the test data set.

The impact of the tool on the screening decision can be easily seen in figure 2. Using the AFST, the researchers retrospectively risk-scored maltreatment referrals from the pre-implementation period (from January 2013 through to September 2016). The researchers also looked at screening decisions after AFST Version 2 was deployed in the field in December 2018. As is clear from figure 2, post-deployment decisions are more consistent with the tool. Before deployment, screening decisions bear no relationship to the AFST score.

The AFST has been generally positively received. An independent impact evaluation concluded that the tool's implementation increased the accuracy of the decisions to further investigate and reduced some of the disparities. No evidence of unintended adverse consequences was found. The positive press included a report in *The New York Times Magazine* that featured



Allegheny Family Screening Tool

Please click the Calculate button to run the algorithm.

Last Run By: Kathy Young

Last Run Date : 02/26/2019, 08:55 AM

Algorithm Version Used: LASSO v18

The Allegheny Family Screening Tool considers hundreds of data elements and insights from historic referral outcomes to estimate the likelihood of this referral resulting in the need for a childs's protective removal from the home within 2 years. It is only intended to help inform call screening decisions, and is not intended for use in investigation or other decision - nor should it be considered a substitute for clinical judgement

Figure 1. Visual Display of AFST.

	Overall	Black	Non-Black
AUC	0.760 (0.748 to 0.771)	0.744 (0.728 to 0.759)	0.773 (0.756 to 0.791)
TPR at top 15% of risk	0.414	0.391	0.433

AUC, area under the receiver operator characteristic curve; TPR, true-positive rate.

Table 1. Performance of AFST.

the strengths of the project, as well as its transparency (Hurley 2018). A review of machine learning tools being used in the US government concluded that

> "Although this project [AFST] is not fully an open source project, it comes closer than any of the other five algorithms we studied" (Brauneis and Goodman 2018, p. 146).

An exception to the overall tenor of positive reporting was a book by Virginia Eubanks (2018). She was concerned that the AFST was tantamount to *poverty profiling,* that is, leading to increased removals of children from families just because they were poor. The County rebutted her assertions, and posted the point-by-point rebuttal on their website.[2]

Toward a Human-Centered Approach to AI

One of the main lessons learned from the deployment and adoption of the AFST is the complex interaction of humans and AI systems during decision-making processes. The crucial issues include: How do humans interpret the risk score? What types of human biases do these tools reduce, and what other biases in decision-making do these tools potentially *introduce*? How do humans anchor on their beliefs and experiences, and how do they potentially learn to anchor on automated tools?

Previous research has identified the need to address these concerns for a more collaborative approach during the design and deployment of algorithms that support decision-making. Areas such as interactive machine learning (Amershi et al. 2014), hybrid intelligence (Dellermann et al. 2019), and, more recently, human-centered machine learning (Riedl 2019), have proposed frameworks that consider humans in the process of designing algorithmic solutions.

Other paradigms such as machine teaching (Simard et al. 2017) placed humans at the core of the construction of machine learning systems. In this paradigm, humans transfer knowledge to machine learning

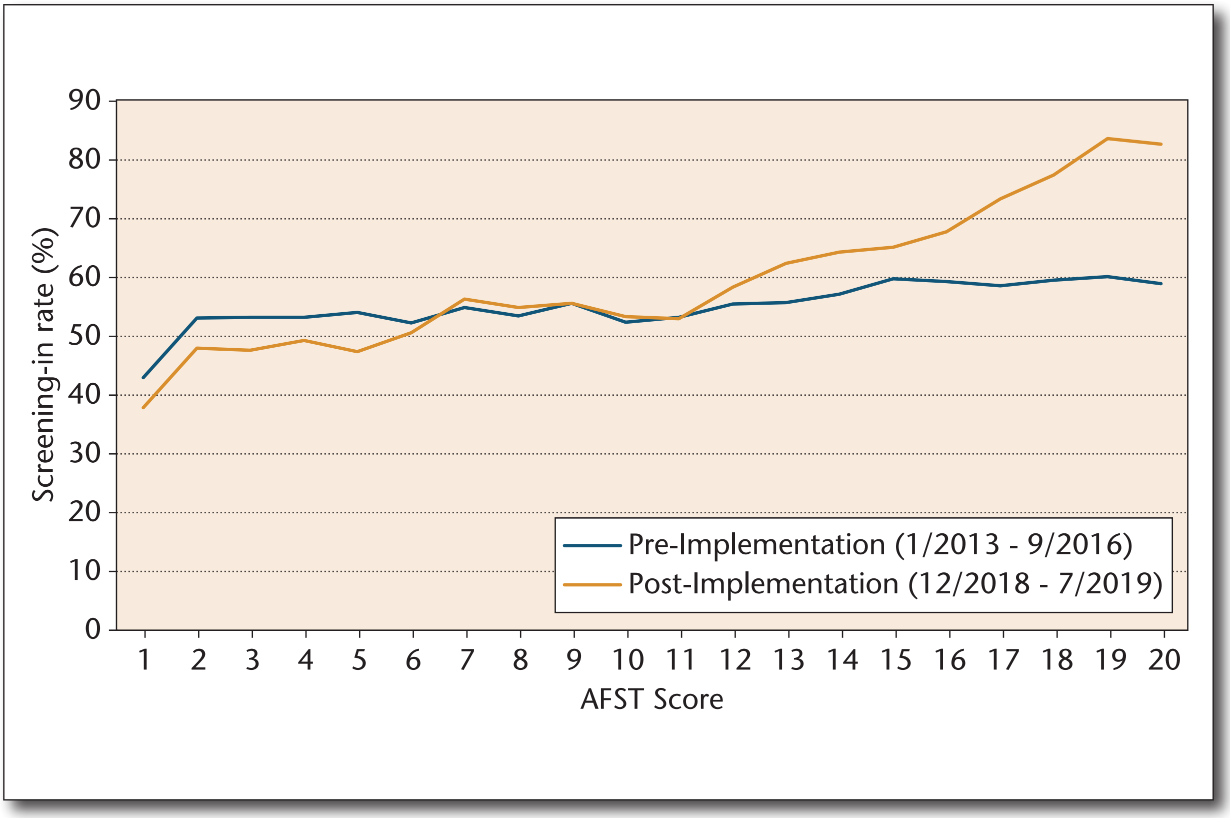


Figure 2. Screening-In Rates Pre- and Post-Implementation of the AFST.

systems in the form of features, comparisons between examples, and labels. Some research listed guidelines for the consideration of humans in the process of designing and deploying algorithms (Amershi et al. 2019).

Many of the papers on human-centered design and human-algorithm interaction model the algorithm as having dropped from the sky. Yet, before the decision-maker has access to the algorithm, many policy decisions have been made by leadership — in this case, the leadership of the Allegheny County DHS. In the social domain, these are crucial decisions that have a large impact on the success of the endeavor. We outline below the elements that led to successful adoption of the AFST in Allegheny County.

The outstanding question of just how to place humans at the center of designing and deploying machine learning-based algorithms must start with those humans that control the system as a whole.

Agency Ownership of All Intellectual Property

In essence, the AFST was acquired under a service contract rather than a software licensing model. The contract specified that the County would own all the code and have free use of the intellectual property developed through the project. This included the code to generate the features and weights. Equally importantly, the research data upon which the tool was built was also returned to the County with appropriate documentation, codebook, and variable definitions. A detailed methodology report was deposited with the County that would allow any outside researcher to replicate the tool and independently validate the results. These measures ensured that the County could easily share data and code with third-party researchers who were able to use the data for testing alternative modeling approaches for evaluation.

Agency Leadership and Engagement

Agency leadership and internal engagement was a crucial component of this project from start to finish. Allegheny DHS had secured the funds, defined the parameters of the project, and conducted an open call for proposals, choosing the research team from a pool of national and international applicants.

This was not a project where data scientists approached an agency looking for a problem to solve with machine learning. Rather, this was a project where the Agency had a strong desire to make better use of its integrated data systems to support decision-making and expressed an openness to machine learning tools providing one potential pathway.

Agency leadership also required that the research team conduct a large number (and broad range) of post-modeling analyses to examine how the proposed tool would or could change call-screening decisions. For example, post-modeling analysis showed that the Agency's then-practice was resulting in screening decisions that led to screen-outs of almost one-third of children who would have been scored at the highest risk by the AFST; additional data confirmed that these children were subsequently re-referred for maltreatment at very high rates. Meanwhile, the Agency's practice led to almost half the children with a low risk score being screened-in for investigation, with very few having any investigative findings that would have justified those decisions.

As a form of external validation, the research team also linked the maltreatment referral data to local pediatric hospitalization records (Vaithianathan et al. 2020) to show that the children who were classified as high risk by the AFST were also significantly more likely to be admitted to hospital for injuries (children who were classified in the highest five-percent of risk by the AFST were more likely to have a medical encounter for an injury than a child who scored in the lowest fifty-percent of risk). More detailed results of this external validation are explained later in this paper.

In summary, the Agency acted like an optimistic but critical purchaser of technology — requiring enough evidence that the tool would improve existing decisions.

Community Engagement and Social License

The Agency already had a high-trust relationship with its community — having integrated data across systems over a period of twenty years. This social license made it easier for the data to be used for decision support tools and other analytics. However, the County didn't take this social license for granted. Right from the start, the community was informed and consulted about the intention to start using the integrated data for building predictive risk models. For example, one of the earliest meetings held (more than a year before deployment) was with families who were involved with the Child Welfare system. The research team and leadership explained the basics of a protype AFST and some of the value that the Agency expected to get out of using it. These sorts of community engagement gave the County and the research team some indication of the concerns of the community and what sort of guardrails had to be implemented to provide comfort.

Transparency by Design

The development of the AFST had high national interest within Child Welfare circles and beyond. At the time of development of the tool, a Presidential Commission on Eliminating Child Abuse and Neglect Fatalities was finalizing a report, arguing that the use of machine learning tools and analytics should be explored. At the same time, a number of critical voices argued that the use of this type of tool would increase disparities in Child Welfare.

As a result, the Agency and research team were routinely approached by journalists and researchers. By design, the research project was transparent — with a default setting (in attitude) that such external scrutiny, even if critical, would be welcomed. Indeed, there has been considerable press and researcher coverage of the AFST.[3,4]

When coverage was negative — as in a chapter of Virginia Eubank's book *Automating Inequality* that was devoted to a critique of the AFST — the research team and Agency produced analysis to show that some of these criticisms were not supported by the data. However, the default position remained that the research team and Agency would welcome critical voices and seek to promote greater understanding of the AFST and its use in cases where the criticism is due to a misunderstanding of system operations.

Appropriate Target to Train the Model

Predictive analytics in child protection raises a rich and complex array of ethical questions. One of the clear challenges in this area is, by its very nature, that these models can only be trained to predict events that are observed within administrative data systems. In child abuse and neglect, there is a substantial gap (up to ten-fold) between maltreatment substantiated by child-protection agencies and that reported by victims or parents (Gilbert et al. 2009). The only ground-truth measures of abuse and neglect that are recorded in administrative data are maltreatment-related fatalities — which are too rare to be a useful target for training a model. In the AFST, removal out of home is used as a proxy outcome for abuse and neglect.

Additional analysis is needed, however, to ensure that a tool that predicts systems outcomes such as removals or substantiation, is sensitive to maltreatment deaths. To do this the research team uses a strategy called *external validation* — by which the tool is calibrated against more universal measures of abuse and neglect. For example, the research team linked a cohort of children scored by the AFST with universal hospital encounter data and found that among children referred for maltreatment and who scored 20 in the AFST — highest risk — the rate of experiencing an any-cause injury encounter was 14.5 (95%CI, 13.1–15.9) per 100 compared to children who scored as low risk — score 1 to 10 in AFST — who had an any-cause injury encounter rate of 4.9 (95%CI, 4.7–5.2) per 100. Similar findings with data from a national study (Vaithianathan et al. 2018) showed that children identified as at-risk by a predictive risk model targeting child-protection agency contact were 9.0 times (95% CI, 3.9 – 20.7) as likely to experience a post-neonatal inflicted injury resulting in death than those with a low-risk score.

Ethical Oversight

There are also tricky privacy questions that predictive analytics brings. For example, should the score be shared with the family? If the score is high because Mom's new boyfriend has an extensive history of violence, is it ethical *not* to share the information with the Mom? If the tool is more accurate for one racial group than another — is this of concern?

An independent ethical review was commissioned from Tim Dare (professor of philosophy at The University of Auckland) and Eileen Gambrill (professor of social work at the University of California at Berkeley). The ethical report provided a set of recommendations and guidelines. The DHS responded to each of these recommendations — and both documents were posted on the DHS's website.

The value of the independent ethical review is to frame the problem and systematically uncover the pros and cons. Of course, such a document is not definitive; as with all decisions made in the project, the DHS could have chosen, and indeed did choose, to reject some recommendations.

Independent Evaluation

The impact of decision support tools, designed to assist rather than replace human decision-making, can only be assessed in concert with the humans whose decisions they support. While the raw predictive accuracy of the tool can be assessed on existing research data, the impact of the tool being added to the existing decision pathways must be assessed in an experimental or quasi-experimental design, where the counter-factual decisions without the tool can be compared with the decisions made with the tool.

The main concern to test is that there are no unintended adverse effects. This might arise, for example, if staff were unwilling to screen-in a low scored child even though they observed risk factors that were not incorporated into the AFST. This type of over-reliance on the algorithm cannot be evaluated except by looking specifically at measures of accuracy for atypical cases.

An impact evaluation using quasi-experimental methods was commissioned from Jeremy Goldhaber-Fiebert (associate professor of medicine at Stanford University).[5] The end-point he used was the number of children screened-in for investigation, where there was no evidence of any maltreatment identified and no re-referrals observed (that is, false-positives). He also measured the number of children screened-out without an investigation who were later re-referred — suggesting that these were incorrect screen-outs.

The evaluation found that the use of the tool increased the identification of children determined to need further child welfare intervention, and led to reduced disparities in case opening rates between Black and White children.

Next Steps

As the research team attempts to scale-out the AFST to other jurisdictions, it is facing multiple challenges that might be useful avenues for applied research.

On reflection, the emerging research program in algorithmic fairness, transparency, and explainability, which takes a technical approach to fairness (Corbett-Davis et al. 2017), although useful from an intellectual point, offers no direction on how to address community discomfort or achieve acceptance among the families and communities that might be subject to algorithmic tools. Demonstrating technical fairness does not contribute to an improved community acceptance or even leadership acceptance.

For one thing, most communities feel (and are correct in realizing) that these types of tools are only a (small) part of the way in which the system affects them. A related research team used a participatory design methodology to explore the question of community comfort in the use of predictive risk models. Unsurprisingly, the conclusion of this research was that as people trust (or don't trust) the Child Welfare system, that is how they would also trust the algorithm (Brown et al. 2019). A useful area of future research is to understand how these tools could bolster trust in the system overall.

While leadership was supportive of the tool, frontline screening staff were initially less inclined to respond to the tool, continuing to screen-in children with low risk scores. Often these children ended up with no abuse being found, and no cases being opened. There is a delicate balance between encouraging frontline workers to not become overly reliant on the algorithm, and ensuring that the full value of the algorithm is being exploited. There is a need to explore and incorporate more human-malleable forms of explanation that are more satisfying for the workers.

The AFST as currently deployed offers only the score — and no other explanation. More recently, human-centered machine learning (Riedl 2019) has proposed frameworks that consider humans in the process of designing algorithmic solutions, while machine teaching (Simard et al. 2017) placed humans at the core of the construction of machine learning systems. In this paradigm, humans transfer knowledge to machine learning systems in the form of features, comparisons between examples, and more. Humans can, for instance, provide information that is used as input for algorithms, such as labels and data (Cao and Ai 2015; Biswas and Jacobs 2013); can help to correct errors made by algorithms (Dusenberry et al. 2019; Bansal et al. 2019); and can help interpret results (Doshi-Velez and Kim 2017).

Conclusion

The implementation of the AFST was a far more complex exercise than adoption of machine learning tools in commercial environments. Child Welfare decisions are extremely high-stakes and made under the glare of public scrutiny, where mistakes can be fatal and heavily scrutinized in the public domain. Poorly funded and stretched Child Welfare systems are dealing with the fallout from decades of social crises, ranging from the incarceration of Black men in the 1990s to the present opioid crisis. These crises reverberate through the generations — ending up as child welfare cases to be dealt with by under-paid and over-loaded social workers.

The intention of the research team was to ask whether advances in machine learning of the kind that are being harnessed for commercial purposes, can be harnessed to solve some of the hardest problems facing society. What we discovered is that doing so is a socio-technological problem, much more so than it is a technical one.

Notes

[1] See www.alleghenycountyanalytics.us/index.php/2019/05/01/developing-predictive-risk-models-support-child-maltreatment-hotline-screening-decisions.

[2] www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx.

[3] See www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html

[4] www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx.

[5] www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Impact-Evaluation-Summary-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-5.pdf.

References

Amershi, S.; Cakmak, M.; Knox, W. B.; and Kulesza, T. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35(4): 105–20. doi.org/10.1609/aimag.v35i4.2513

Amershi, S., Weld, D.; Vorvoreanu, M.; Fourney, A.; Nushi, B.; Collisson, P.; Suh, J.; Iqbal, S.; Bennett, P. N.; Inkpen, K.; Teevan, J.; Kikin-Gil, R.; and Horvitz, E. J. 2019. Guidelines for Human–AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York: Association for Computing Machinery. doi.org/10.1145/3290605.3300233

Amrit, C.; Paauw, T.; Aly, R.; and Lavric, M. 2017. Identifying Child Abuse Through Text Mining and Machine Learning. *Expert Systems with Applications* 88(1): 402–18. doi.org/10.1016/j.eswa.2017.06.035

Bansal, G.; Nushi, B.; Kamar, E.; Lasecki, W. S.; Weld, D. S.; and Horvitz, E. 2019. Beyond accuracy: The Role of Mental Models in Human–AI Team Performance. In *Proceedings of the Seventh Association for the Advancement of Artificial Intelligence (AAAI) Conference on Human Computation and Crowdsourcing*. Palo Alto, CA: AAAI Press.

Biswas, A., and Jacobs, D. 2013. Active Image Clustering with Pairwise Constraints from Humans. *International Journal of Computer Vision* 108(1–2): 133–47. doi.org/10.1007/s11263-013-0680-6

Brauneis, R., & Goodman, E. P. (2018). Algorithmic transparency for the smart city. Yale JL & Tech., 20, 103.

Brown, A.; Chouldechova, A.; Putnam-Hornstein, E.; Tobin, A.; and Vaithianathan, R. 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*. New York: Association for Computing Machinery. doi.org/10.1145/3290605.3300271

Cao, C. J., and Ai, H. Z. 2015. Facial Similarity Learning with Humans in the Loop. *Journal of Computer Science and Technology* 30(3): 499–510. doi.org/10.1007/s11390-015-1540-3

Chouldechova, A. 2017. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data* 5(2): 153–63. doi.org/10.1089/big.2016.0047

Corbett-Davies, S.; Pierson, E.; Feller, A.; Goel, S.; and Huq, A. 2017. Algorithmic Decision Making and the Cost of Fairness. In *Proceedings of the 23rd Association for Computing Machinery (ACM) Special Interest Group on Knowledge Discovery and Data Mining SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM. doi.org/10.1145/3097983.3098095

Dellermann, D.; Ebel, P.; Sollner, M.; and Leimeister, J. M. 2019. Hybrid Intelligence. *Business & Information Systems Engineering* 61(March): 637–43. doi.org/10.1007/s12599-019-00595-2

Doshi-Velez, F., and Kim, B. 2017. Towards a Rigorous Science of Interpretable Machine Learning. arXiv preprint. arXiv:1702.08608[statML]. Ithaca, NY: Cornell University Library.

Dusenberry, M. W.; Tran, D.; Choi, E.; Kemp, J.; Nixon, J.; Jerfel, G.; Heller, K.; and Dai, A. M. 2019. Analyzing the Role of Model Uncertainty for Electronic Health Records. arXiv preprint. arXiv:1906.03842[csLG]. Ithaca, NY: Cornell University Library.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Gilbert, R.; Widom, C. S.; Browne, K.; Fergusson, D.; Webb, E.; and Janson, S. 2009. Burden and Consequences of Child Maltreatment in High-Income Countries. *Lancet* 373(9657): 68–81. doi.org/10.1016/S0140-6736(08)61706-7

Hurley, D. (2018). Can an algorithm tell when kids are in danger. New York Times, 2.

Jackson, D., and Marx, G. (2017). Data mining program designed to predict child abuse proves unreliable, DCFS says. Chicago Tribune.

Kleinberg, J.; Lakkaraju, H.; Leskovec, J.; Ludwig, J.; and Mullainathan, S. 2018. Human Decisions and Machine Predictions. *The Quarterly Journal of Economics* 133(1): 237–93.

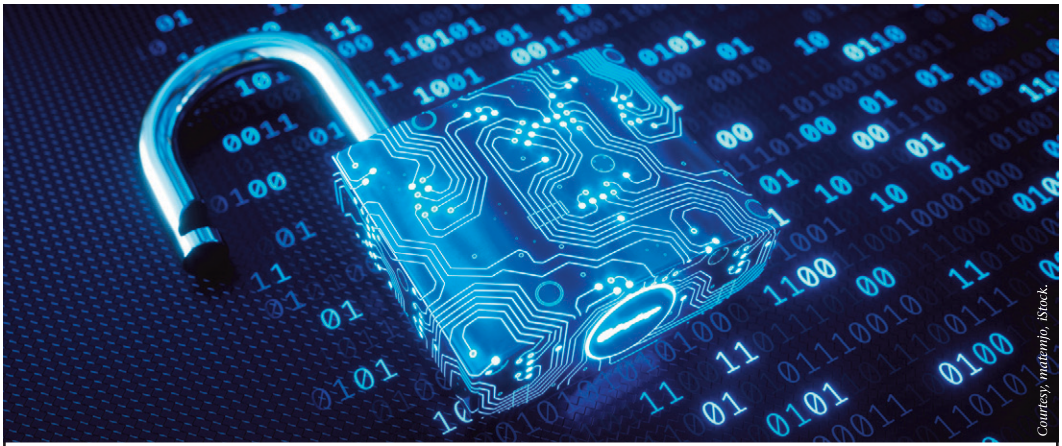
Radio New Zealand. 2015. MSD Urged to Adopt Predictive Tool to Identify at-Risk Children. Nine to Noon, 9:08 AM, 12 May. www.rnz.co.nz/national/programmes/ninetonoon/audio/201753988/msd-urged-to-adopt-predictive-tool-to-identify-at-risk-children.

Riedl, M. O. 2019. Human-Centered Artificial Intelligence and Machine Learning. *Human Behavior and Emerging Technologies* 1(1): 33–6. doi.org/10.1002/hbe2.117

Saxena, D.; Badillo-Urquiola, K.; Wisniewski, P. J.; and Guha, S. 2020. 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. New York: Association for Computing Machinery.

Schwartz, I. M.; York, P.; Nowakowski-Sims, E.; and Ramos-Hernandez, A. 2017. Predictive and Prescriptive Analytics, Machine Learning and Child Welfare Risk Assessment: The Broward County Experience. *Children and Youth Services Review* 81(October): 309–20. doi.org/10.1016/j.childyouth.2017.08.020

Simard, P. Y.; Amershi, S.; Chickering, D. M.; Pelton, A. E.; Ghorashi, S.; Meek, C.; Ramos, G.; Suh, J.; Verwey, J.; Wang, M.; and Wernsing, J. 2017. Machine Teaching: A New Paradigm for Building Machine Learning Systems. arXiv preprint. arXiv:1707.06742[csLG]. Ithaca, NY: Cornell University Library.



Courtesy, matemjo, iStock.

Vaithianathan, R.; Maloney, T.; Putnam-Hornstein, E.; and Jiang, N. 2013. Children in the Public Benefit System at Risk of Maltreatment: Identification via Predictive Modeling. *American Journal of Preventive Medicine* 45(3): 354–9. doi.org/10.1016/j.amepre.2013.04.022

Vaithianathan, R., Putnam-Hornstein, E., Chouldechova, A., Benavides-Prado, D., & Berger, R. (2020). Hospital injury encounters of children identified by a predictive risk model for screening child maltreatment referrals: evidence from the Allegheny Family Screening Tool. JAMA pediatrics, 174(11), e202770–e202770.

Vaithianathan, R.; Rouland, B.; and Putnam-Hornstein, E. 2018. Injury and Mortality Among Children Identified as at High Risk of Maltreatment. *Pediatrics* 141(2): e20172882. doi.org/10.1542/peds.2017-2882

Rhema Vaithianathan is director of the Centre for Social Data Analytics, Auckland University of Technology, New Zealand and The University of Queensland, Australia. She is also a professor of economics at the Auckland University of Technology and a professor of social data analytics at The University of Queensland. Predictive risk modeling is Rhema's main research interest, with a strong focus on the methodologies, implementation, and implications of predictive risk modeling in human services settings. She is leading the development, implementation, and quality assurance of a range of predictive risk modeling tools in the USA, for domains including child welfare and homelessness.

Diana Benavides-Prado is a senior research fellow in data science at the Centre for Social Data Analytics at Auckland University of Technology, New Zealand. Her research interests span transfer learning, continual learning, and human–algorithm collaboration in decision-making contexts. She investigates fundamental aspects of these paradigms and leads the construction of decision-making support tools aligned with current advancements in these areas.

Erin Dalton is deputy director of the Office of Analytics, Technology and Planning at the Allegheny County Department of Human Services. Dalton is responsible for directing the research and evaluation, evidence-based planning, and information technology activities of the Department. She has held policy positions with the Allegheny County Executive's Office and the US Department of Justice and was an adjunct staff member at the RAND Corporation. She is a board member of Neighborhood Allies and was a mayoral appointee to the Pittsburgh Civilian Police Review Board and a county executive appointee to the Allegheny County Juvenile Detention Board of Advisers.

Alexandra Chouldechova is the Estella Loomis McCandless assistant professor of statistics and public policy at Carnegie Mellon University's Heinz College of Information Systems and Public Policy. Her research investigates questions of algorithmic fairness and accountability in data-driven decision-making systems, with a domain focus on criminal justice and human services.

Emily Putnam-Hornstein is the John A. Tate distinguished professor for children in need and the director of policy practice at the School of Social Work at the University of North Carolina at Chapel Hill. She also maintains appointments as a distinguished scholar at the University of Southern California, where she co-directs the Children's Data Network and as a research specialist with the California Child Welfare Indicators Project at the University of California Berkeley. Her current research focuses on the application of epidemiological methods to improve the surveillance of non-fatal and fatal child abuse and neglect.