Innocenti Research Brief

2021-06

unicef Office of Research – Innocenti

Predictive Analytics for Children: An assessment of ethical considerations, risks, and benefits

Zara Rahman[1] and Julia Keseru[1]

[1] The Engine Room

This brief examines the potential ethical issues, including benefits and risks, associated with predictive analytics as they pertain to children. It is based on a more in-depth working paper which provides further detail, guidance, and tools and accessible here.

A myriad of factors contribute to why a focus on the use of predictive analytics as it specifically pertains to children is necessary. These include the rapid datafication of children's lives; the complexities of gathering and responsibly using disaggregated data about children; the lack of child rights expertise among those responsible for analysis of data about children; and importantly, the biological and psychological differences between children and adults (which will be discussed in further detail below). To this end, this paper explores current and potential use cases of predictive analytics that may impact children, either directly or indirectly, and the ethical considerations arising from both the potential benefits and the risks these present.

DEFINITIONS

The following key definitions apply:

Predictive analytics (PA) can be defined as the iterative analysis of historical data to predict future outcomes, as powered by machine learning (Ongsulee et al., 2018). Within this paper we consider PA as a subset of machine learning and more specifically, define PA as being powered by machine learning. PA is typically used as a decision support mechanism, providing a source of information that can be considered in triangulation with other evidence and insights to, in the best case, strengthen decision-making.

Children: As defined by the United Nations Convention on the Rights of the Child (UNCR, 1989), children are referred to as any person below the age of 18 years, while adolescent and youth refers to those aged 10–19 years (WHO, 2014) and 15–24 years (United Nations Department of Economic and Social Affairs, n.d.) respectively. For the purposes of this paper, our working definition of "children" is inclusive of all three groups, considering those 24 years or younger.

WHAT CURRENT APPLICATIONS OF PREDICTIVE ANALYTICS ARE THERE?

The umbrella term of predictive analytics encompasses a variety of potential applications, use cases, and outcomes as they relate to children. Each of these applications brings with it a specific set of benefits and risks. For a detailed discussion of the different types of PA, please refer to the full paper. In short, predictive analytics applications differ in the following ways:

- The focus of the outcome: **population, phenomenon, and/or individual**
- The nature of the model, including:
  - The level of supervision and human intervention in the design of the model: **supervised, unsupervised, semi-supervised, reinforcement**
  - The level of interpretability of the algorithm used: **closed source (also known as black box), or open source**
- The way in which the outcome is used in decision-making: **supporting human decision-making or automating decisions**

WHY CONSIDER PREDICTIVE ANALYTICS AND CHILDREN?

PA models and, more broadly, data-driven systems in general, are "often designed with (consenting) adults in mind, without a focus on the unique needs and vulnerabilities of children" (UNICEF and GovLab, 2019).[1]

In addition, differences between children and adults manifest in a variety of ways that generate new benefits and risks of using PA:

- Differences in neurological and physical development mean that data must be analysed in a child/young-person-centric manner for insights to be accurate.
- Children's evolving capacity to exercise agency can and should impact how consent can be gathered and understood regarding the use of their data, and considered when assessing children's diverse understandings of data privacy.
- Children's lack of agency over their social network, decision-making and frequently, geography, affects the way in which they might be categorized or profiled within digital datasets (which then affects the predictions that arise from historical data).

WHAT ARE THE POTENTIAL BENEFITS OF USING PREDICTIVE ANALYTICS?

In sectors with both direct and indirect impact on children, PA offers the possibility to allocate resources in more effective and efficient ways; to reveal patterns and trends that are invisible to the human eye; to drive more complex and nuanced research; and to respond to changes and shocks proactively rather than reactively.

Further benefits include:

- **More efficient resource allocation**: Population-based PA models used to support decision-making have the potential to yield more tailored interventions and resource allocations in ways that are informed by historical interventions, population movements and demographics, as well as by environmental changes.
- **Supporting better planning for environmental shocks and changes**: PA offers the possibility for humanitarian and development actors to pre-emptively mobilize resources before a natural disaster has taken place, based on risk potential rather than only acting once events have occurred.
- **Reducing the burden on overburdened systems**: PA could be used to predict future capacity constraints within institutions.
- **Generating new insights:** Data collected for machine learning (ML)-driven systems could strengthen humanitarian programming, development work, and other child focused interventions in addressing mismatches between resource allocation and actual needs. Disaggregated data on children could also allow for more focused advocacy interventions, especially for vulnerable children such as those with disabilities.
- **Identifying patterns across multiple data streams:** PA can find patterns and identify similarities within large amounts and diverse types of data.
- **Increasing accountability:** If designed appropriately, PA could be used to predict where institutions providing child-focused services are more likely to experience bottlenecks, or to monitor or forecast performance of relevant agencies in a way that supports them in mitigating predicted bottlenecks.
- **Targeted profiling for positive behaviour change:** PA-facilitated targeted profiling techniques can support behaviour change; for instance, in public health interventions to reduce risk-taking behaviour and promote positive health behaviours.

However, the evidence in relation to cost/benefit analysis and with respect to robust evaluations is not comprehensive. As a result, because these initiatives are implemented and established over time, a greater body of rigorous evidence will be required to substantiate both the rollout and upscaling of these programmes.

WHAT ARE THE POTENTIAL RISKS ARISING FROM USE OF PREDICTIVE ANALYTICS?

The potential risks of using PA involving children arise from the use of their data and the actual implementation and design of PA models.

Data-related risks

Child-specific risks that arise from the use of data in PA include the following:

- **Risks to children's autonomy in consent processes**: As children develop, their relationships and capacity to exercise agency and grant consent for use of their data evolves.
- **Lack of high quality, consistent and disaggregated data about children:** There is frequently a lack of high quality, large scale and disaggregated data about children, and this lack of data will negatively impact the accuracy of a PA model. However, when disaggregated data is available, it

also raises the risk of de-anonymization and identification of children.

- **Bias within children's datasets**: When data about children is available, it has been noted that this data can reflect systemic societal inequities (Munro, 2019),[2] which may be replicated if used to train PA models.
- **Security risks:** To increase the accuracy of predictive models used on children, they should be trained on data about children; however, this requires large amounts of children's data to be used and stored over time, which increases the risks of data breaches.
- **Risks to identity and agency of children:** As most data about children is gathered by professionals, this may mean there are limited opportunities for children (or their families/caregivers) to be aware of this data or to contest the data if they do not agree with it.

Implementation-related risks

Risks relating to the implementation of a predictive analytics stem may arise, in large part, from it being employed in such a way that there is a hyper-dependence or poor understanding of PA, meaning that staff (or end-users) do not have the skills or training to interpret the outputs with necessary critical assessment. These risks include:

- **PA affecting behaviours of those impacted by implementation**: For example, false positives leading to families and their communities losing trust in the intervening institution, leading to less responsiveness to service provision in the future.
- **PA undermining practitioner decision-making or agency**: Using PA may erode social workers' agency and critical thinking (Munro, 2019),[3] or they may feel influenced to adjust their own estimates to reflect PA outputs (Eubanks, 2018).[4]
- **Creating more work for over-burdened staff**: Overworked staff may be reluctant or simply not have time to learn how to properly integrate PA into their work.
- **Potential waste of resources:** In spite of the oft-touted benefit of PA to improve efficiency, there is a distinct lack of evidence to demonstrate this, suggesting it is more of an aspiration than a proven metric at this time.
- **Conflict of values between users and private sector providers of PA**: The private sector is frequently involved in the implementation of PA and their values and approaches may sit in contrast to those of, for example, humanitarian actors. This raises potential privacy risks involving future uses of the data or derivative models.
- **Privacy:** Many of these risks have serious implications for the way in which children, whose data is used in PA models or who are affected by PA outcomes, can exercise their right to privacy.
- **Risks to identity and personhood of children:** The preferences of children change over time and they must be free to explore and express themselves without fear of negative consequences. However, research on online profiling suggests that it may have a chilling effect on both online and offline behaviour (Marder et al., 2016).[5]
- **Risks to children's freedom of expression:** When PA is used to 'nudge' children's behaviour online, it may covertly contribute to commercial profiling of a child and affect their sense of agency and identity.
- **Risks to children's access to information**: When PA is used to profile children, recommendation algorithms that use their past online behaviour to recommend content can limit the information they are exposed to, thus affecting their right to access to information.

Design-related risks

Risks that relate to the design of PA models include:

- **Entrenching bias and discrimination**: There are multiple 'decision points' within the design and implementation of a machine learning model, at which human prejudices and bias can affect accuracy and efficacy (Glaberson, 2019).[6]
- **Networked effects of PA on children**: As individual-focused PA approaches often include demographic data to train the ML model, this means that a child's social network and demographic is likely to heavily shape the recommendations generated – which assumes that a child will behave in a certain way, based on the behaviour of others, and thus may deny their self-determination.
- **Low levels of accuracy:** The accuracy of currently deployed PA models raises concerns in almost all analysed areas, particularly when predictive models are focused on individuals as opposed to populations.
- **Inaccuracies arising from PA trained on adult data used for children:** Due to a lack of available data,

PA models are often trained on adult data, which can lead to problematic or inaccurate outputs as they pertain to children.

- **Environmental impacts of energy-intensive machine learning:** Machine learning and artificial intelligence has a significant cost on the environment due to the level of computational power required, as well as the costs of building the requisite hardware and infrastructure.
- **Lack of accountability, transparency, and explainability:** If PA systems with low levels of explainability are used to make decisions, it may be unclear at what point a decision is made, and – depending on the model used – why. This removes the opportunity for children and their caregivers to understand why a certain decision, that may affect their lives, is being made.

RECOMMENDATIONS: CONSIDERATIONS FOR IMPLEMENTATION OF PREDICTIVE ANALYTICS

When assessing potential use cases:

- Consider whether the project should take place or not and be open to the possibility that it should not.
- Identify and clearly articulate why the PA model is needed, and how the potential benefits relate to improving children's lives.
- Conduct child risk and child impact assessments of PA.
- Only use individual-focused PA with established requirements and risk mitigation measures.
- Establish transparency and accountability standards for private sector agreements.

When considering the data-related impacts of predictive analytics:

- Acknowledge tensions between minimizing data to protect children's rights and designing predictive analytics systems that reflect children's contexts.
- Practise data minimization when it comes to children's data.
- Where data from individual children is being used, adopt consent policies that recognize children's development.
- Ensure compliance with data protection regulations, but do not rely solely upon legislative regimes to ensure children's rights are protected.
- Avoid implementations that require the use of datasets with historically embedded inequities.

When designing and implementing predictive analytics, make sure to:

- Involve staff from diverse and multidisciplinary backgrounds from the very first stage, ensuring children's rights expertise and technical expertise is included.
- Consider options for reducing the environmental impact of any PA implementation.
- Provide training and skill development for staff involved in any PA implementation.
- Involve children and their communities in the design and validation of PA models.
- Use models that are interpretable and explainable to children and/or their caregivers or representatives.
- Establish redress mechanisms that are accessible and understood by children and/or their parents, caregivers, or guardians.
- Consider the potential risks of PA to children's agency and personhood.
- Ensure children's voices are valued alongside PA outputs by triangulating data.
- Provide clear guidance to staff on how to assess, interpret or use PA outputs as they affect children.
- Choose a PA model that fits the objectives, available data, and intended context.
- Incorporate community opinions and consistently iterate based on ongoing feedback.
- Continue to build a knowledge base on the potential impacts – negative and positive – of PA for children.

THE FUTURE OF SAFE, PREDICTIVE ANALYTICS FOR CHILDREN:

As interest in predictive analytics grows, and noting the need to mitigate against potential and significant risks to children, potential future directions of work include:

- Initiating pilot projects in safe, sandbox environments which "would institute restrictions and controls to mitigate possible negative consequences before any work is replicated" (RD4C, 2019) while allowing for innovation to take place in safe and controlled environments.

- Gathering evidence on how predictive analytics for children affects efficiency and resource allocation, noting the current lack of evidence.
- Assessing the impact of PA beyond accuracy or efficacy of outputs and instead, accounting for their impact as broader socio-technical systems with potential unintended impacts within the broader social and political system, as well as the environment, in which they are implemented.
- Taking a child-centred approach to innovation on PA; for example, identifying or creating algorithms and predictive models and associated policies that:
  - Have established redress mechanisms that are accessible to children;
  - Involve children in the development and/or design of the PA model itself;
  - Allow for children to understand how their lives may be impacted by PA.
- Developing training materials and capacity development to build the technical intuition of staff whose work may be impacted by PA.

Further information

To learn more about the ethical implications of predictive analytics on children, please see the following resources:

- **For an accessible checklist of key considerations please see Annex 2 of the Working Paper (WP 2021-08);**
- **For the full report, please see TK LINK;**
- **For an accessible checklist of key considerations please see TK LINK.**

ACKNOWLEDGEMENTS

The development of this brief, and the broader paper, was supported by the significant efforts of the project's research assistants, including Cathy Richards, Teresa Perosa, Helen Kilbey (The Engine Room), and Sudeshna Mukherjee (independent consultant) and with the editorial support of Helen Kilby (The Engine Room).

This brief and paper benefited immeasurably from the insights and expertise of the following reviewers: Gabrielle Berman, Josiah Kaplan and Kerry Albright, UNICEF Office of Research; Karen Carter and Tanja Serjesen, UNICEF Data & Analytics; Eric Siegel, Predictive Analytics World; Patrick Burton, Centre for Justice and Crime Prevention.

ENDNOTES

1 UNICEF and GovLab, Responsible Data for Children (RD4C), November 2019, rd4c.org, accessed February 2021.

2 Munro, E., 'Predictive Analytics in Child Protection', CHESS Working Paper No.3, Durham University, U.K., April 2019, www.researchgate.net/publication/332528200_Predictive_analytics_in_child_protection, accessed April 2021.

3 Munro, E., 'Predictive Analytics in Child Protection', CHESS Working Paper No.3, Durham University, U.K., April 2019, www.researchgate.net/publication/332528200_Predictive_analytics_in_child_protection, accessed April 2021.

4 Eubanks, V., *Automating Inequality: How high tech tools profile, police and punish the poor,* St. Martin's Press, New York, 2017.

5 Marder, B., et al., 'The Extended Chilling Effect of Facebook: The cold reality of ubiquitous social networking', *Computers in Human Behaviour,* vol. 60., July 2016, www.researchgate.net/publication/296637938_The_extended_%27chilling%27_effect_of_Facebook_The_cold_reality_of_ubiquitous_social_networking, accessed May 2021.

6 Glaberson, S. K., 'Coding Over the Cracks: Predictive analytics and child protection', *Fordham Urban Law Journal,* vol. 46, no. 2, 2019, ir.lawnet.fordham.edu/ulj/vol46/iss2/?utm_source=ir.lawnet.fordham.edu%2Fulj%2Fvol46%2Fiss2%2F3&utm_medium=PDF&utm_campaign=PDFCoverPages, accessed May 2021.

UNICEF OFFICE OF RESEARCH-INNOCENTI

Florence, Italy

www.unicef-irc.org

+39 055 20330

florence@unicef.org

@UNICEFInnocenti

facebook.com/UnicefInnocenti

Research for Children at Innocenti