unicef
Office of Research – Innocenti

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

Zara Rahman and Julia Keseru

UNICEF Office of Research - Innocenti Working Paper
WP 2021-08 | November 2021

unicef | for every child

UNICEF OFFICE OF RESEARCH – INNOCENTI

The Office of Research – Innocenti is UNICEF's dedicated research centre. It undertakes research on emerging or current issues to inform the strategic directions, policies and programmes of UNICEF and its partners; shape global debates on child rights and development; and inform the global research and policy agenda for all children, particularly the most vulnerable.

Publications produced by the Office are contributions to a global debate on children and may not necessarily reflect UNICEF policies or approaches. The views expressed are those of the authors.

The Office of Research – Innocenti receives financial support from the Government of Italy, while funding for specific projects is also provided by other governments, international institutions and private sources, including UNICEF National Committees.

For further information and to download or order this and other publications, please visit the website at www.unicef-irc.org.

INNOCENTI WORKING PAPERS

UNICEF Office of Research Working Papers are intended to disseminate initial research contributions within the programme of work, addressing social, economic and institutional aspects of the realization of the human rights of children.

The findings, interpretations and conclusions expressed in this paper are those of the authors and do not necessarily reflect the policies or views of UNICEF.

This paper has been peer reviewed both externally and within UNICEF. The text has not been edited to official publications standards and UNICEF accepts no responsibility for errors.

Correspondence should be addressed to:

UNICEF Office of Research – Innocenti
Via degli Alfani 58
50121 Florence, Italy
Tel: (+39) 055 20330
Fax: (+39) 055 2033 220
florence@unicef.org
www.unicef-irc.org
@UNICEFInnocenti
facebook.com/UnicefInnocenti

PREDICTIVE ANALYTICS FOR CHILDREN: AN ASSESSMENT OF ETHICAL CONSIDERATIONS, RISKS, AND BENEFITS

Zara Rahman[1] and Julia Keseru[1]

[1]The Engine Room

ACKNOWLEDGEMENTS

The development of this paper and the attendant brief 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) with the editorial support of Helen Kilby (The Engine Room).

This paper and the brief benefited immeasurably from the insights and expertise of the following reviewers: Gabrielle Berman*, Josiah Kaplan*, Eric Seigel***, Karen Carter**, Patrick Burton****, Kerry Albright*, and Tanja Serjesen**.

* UNICEF Office of Research
** UNICEF Data and Analytics
*** Predictive Analytics World
**** Centre for Justice and Crime Prevention

CONTENTS

GLOSSARY

Adolescent and youth: The UN Secretariat uses 'youth' and 'young people' interchangeably to denote ages 15–24 (United Nations Department of Economic and Social Affairs n.d.); WHO and UNICEF defines adolescents from 10–19 (WHO, 2014)

Algorithm: A procedure for solving a mathematical problem in a finite number of steps that frequently involves repetition of an operation (Merriam-Webster dictionary).

Big data: "'Big Data' refers to large amounts of different kinds of data produced from various types of sources, such as people, machines or sensors. This data could be climate information, satellite imagery, digital pictures and videos, transition records or GPS signals. Big Data may involve personal data too" (Directorate General for Justice and Consumers, European Commission, 2016).

Children: The United Nations Convention on the Rights of the Child (OHCHR, 1989) defines a child as "every human being below the age of 18 years unless, under the law applicable to the child, majority is attained earlier".

Demographically-identifiable information: Individual and/or aggregated data points allowing inferences to be drawn that enable the classification, identification, and/or tracking of both named and/or unnamed individuals, groups of individuals, and/or multiple groups of individuals according to ethnicity, economic class, religion, gender, age, health condition, location, occupation, and/or other demographically defining factors (Raymond, 2016).

Differential privacy: A technique which aims to maximize the accuracy of database queries or computations while minimizing the identifiability of individuals with records in the database, typically via obfuscation of query results; for example, by the addition of spurious information or 'noise' (Center for Open Data Enterprise, 2016; Zewain, 2021).

Filter bubble: Coined by internet activist Eli Pariser (2011), a 'filter bubble' is a form of personalization in the consumption of ideas and information – invisible to the lay user – that unwittingly isolates opposing viewpoints and is catered by algorithms generated through past searches.

Forecasting: A sub-discipline of prediction in which one makes predictions about the future on the basis of time-series data with a consideration for the temporal dimension (Döring, 2018).

Informed consent: The voluntary agreement of an individual, or his or her authorized representative who has the legal capacity to give consent and who exercises free power of choice, without undue inducement or any other form of constraint or coercion to participate in research. The individual must have sufficient knowledge and understanding of the nature of the proposed evidence generating activity, the anticipated risks and potential benefits, and the requirements or demands of the activity to be able to make an informed decision (Levine, 1998).

Low and Middle Income Countries (LMIC): The list of included countries follows the World Bank's classifications for the fiscal year 2021(World Bank, 2021).

Machine learning: A branch of artificial intelligence in which a computer generates rules and predictions based on raw data fed into it (Onuoha, 2018). Machine learning involves the computer continuously learning and in turn, adapting its prediction algorithms through the use of training and testing datasets.

Predictive analytics: A range of analytical and statistical techniques used for developing models that may be used to predict future events or behaviours (Techopedia, 2017) and a subset of machine learning. The techniques utilize historical data or past experience to predict future outcomes.

ACRONYMS

CERF	UN Central Emergency Response Fund
COE	Council of Europe
CRC	UN Committee on the Rights of the Child
ENOC	European Network of Ombudspersons for Children
GDPR	General Data Protection Regulation
ICO	Information Commissioner's Office
ML	Machine learning
PA	Predictive analytics
RD4C	Responsible Data for Children
UN	United Nations
UNCRC	United Nations Convention on the Rights of the Child
UNGA	United Nations General Assembly
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations International Children's Fund
UNOCHA	United Nations Office for the Coordination of Humanitarian Affairs

1. INTRODUCTION

This paper examines potential ethical issues, including benefits and risks, associated with predictive analytics as they pertain to children. It is designed to support readers in gaining an overview of the current state of the field, knowledge of real-world deployments of predictive analytics and ultimately, a deeper understanding of the opportunities and potential harms of deploying predictive analytics that directly or indirectly target children.

Predictive analytics (PA) is defined in this paper as the use of historical data or past experience to predict future outcomes (Ongsulee et al., 2018). For the purposes of this paper, we consider predictive analytics as a subset of machine learning; that is, predictive analytics is always powered by machine learning, but there are uses of machine learning that are not predictive analytics. PA offers the opportunity to take advantage of increasing amounts and different types of historical data, combined with increasing computational power.

PA has been proliferating over the last decade or so in a range of sectors, with direct or indirect impact on children, including health care, child protective systems, criminal justice, humanitarian planning, and consumer targeting. In the specific context of development and humanitarian sectors, PA systems have already been employed by development and humanitarian agencies in ways that affect children – for example, UNHCR's Project Jetson (UNHCR Innovation Service, 2019), the Danish Refugee Councils' Mixed Migration Foresight Project (Horwood and Powell, 2019), the World Food Programme's use of predictive analytics for food security analysis through a Hunger Map (Ong, 2020), and Save the Children's Migration and Displacement Initiative (MDI) (Save the Children International, 2021). If responsibly deployed and integrated into initiatives in a way that builds upon contextual knowledge and local expertise, PA models offer the possibility to strengthen proactive rather than reactive management of crises. Additionally, they can highlight trends and patterns from large amounts of data from different sources and enable more efficient resource allocation. At the same time, potential harms from predictive analytics are severe and potentially life-changing, particularly for children, families, and communities who may already be marginalized or disadvantaged.

This report posits that it is crucial to carefully consider the ethical implications of these projects specifically as they relate to children who, as a demographic, may have distinctly different needs, as reflected in the Convention on the Rights of the Child (UNCRC). Within the broader umbrella of predictive analytics, various approaches can either directly or indirectly impact children – and each of these approaches brings with it different assessments of risk and benefit. It is likely that some sort of predictive analytics will diffuse into children's lives in the near future, with direct and indirect impacts. However, there remains an open question as to what kinds of PA will be implemented, and how PA can be used efficiently and safely, especially with the recent adoption by the UN Committee on the Rights of the Child (CRC) of General Comment No. 25 on Children's Rights in Relation to the Digital Environment (OHCHR, 2021). This paper seeks to provide insights that will inform these decisions, contributing to a wider debate that has already begun in the humanitarian and development sector around the use of PA.

When discussing ethical concerns, there is a wide body of existing ethical frameworks and principles to draw upon, including the Responsible Data for Children (RD4C) principles (UNICEF and GovLab, 2019); the Sphere Standards (Sphere Association, 2018); and the principles outlined in 'The Case for Better Governance of Children's Data: A manifesto' (Byrne et al, 2021). This paper builds directly upon those frameworks, focusing in particular on the rights of the child as codified in the UNCRC. The Annex also includes practical tools to support decision-making.

The paper is structured as follows: First, we ground the paper in definitions of the technology at hand, the context amid which predictive analytics is being discussed, and the specificities of childhood and child development. We then explore existing use cases of predictive analytics that directly or indirectly impact children, before examining the benefits, risks, and mitigation strategies to support responsible deployment of predictive analytics. The Annex includes a practical checklist to support decision-making around the use of predictive analytics in relation to its application to child cohorts. This checklist can be used to ensure that adequate measures are taken before implementing a predictive analytics system for children.

2. BACKGROUND

This section provides context on what predictive analytics is, the different types of PA, and the various ethical and legal frameworks that need to be considered in their implementation. The section also provides justification and background as to why it is critical to consider the specific impacts of PA on children.

2.1 Defining predictive analytics

Predictive analytics (PA) can be defined as the iterative, machine-learning-driven analysis of historical data to predict future outcomes (Michard and Teboul, 2019). In the literature, and within public, international development, and humanitarian sectors, predictive analytics is used interchangeably with predictive modelling, automated decision-making (depending on how outputs are used), and algorithmic decision-making.

When reflecting on predictive analytics, explicit consideration needs to be given to the data upon which models used for predictive analytics are built. An increasing amount of data about children is available, produced by children themselves via their online activities; produced by others; and contained within data from populations and communities of which children are a part (Berman and Albright, 2017; Young, 2020). This raises the possibility of training machine learning algorithms on this data in order to support decision-makers to better predict potential future outcomes and plan for interventions that support children. However, specific risks arise in the context of predictive analytics as it pertains to children, as discussed in more detail below.

In this paper, we have defined predictive analytics as predictions generated via machine learning algorithms, noting that there exist different schools of thought as to how PA and machine learning overlap.Siegel (2016) states that all applications of predictive analytics are applications of machine learning, but that there are applications of machine learning that do not constitute predictive analytics. For the purposes of this paper, we will adopt Siegel's definition in order to reflect the applications of PA that are iterative (as per machine learning processes) while also highlighting the specific ethical concerns that arise from this type of process.

2.2 Context for predictive analytics

The contemporary landscapes and contexts where PA are being considered for use in public, humanitarian, and development sectors to support and protect children are frequently defined by:

- **Resource constraints:** Particularly amid the Covid-19 pandemic, there is both a significant financial shortfall in humanitarian and development agency budgets and a growing need for their work. Covid-19 has caused a 40 per cent increase in the number of people in need of humanitarian assistance since 2020 (UN, 2020). In addition, significant funding shortfalls have been noted regarding crises in Yemen and Ethiopia (Nasser, 2021; CGTN Africa, 2020). Globally, in terms of public policy, governments advancing an austerity agenda have created a "need for local authorities to look for ways to do more with less" (Redden et al., 2020). This climate, in which both authorities and humanitarian organizations are being forced to do more with less, has paved the way for technologies which promise increased resource efficiency.
- **Significant displacement**: 79.5 million people were forcibly displaced worldwide at the end of 2019, and an estimated 40 per cent of these were children below 18 years of age (UNHCR, 2019). This figure stands at an all-time high and is likely to have increased due to the Covid-19 pandemic.

- **Climate emergency**: The ongoing climate emergency means that in the near future natural hazards and previously unseen environmental shocks will increase in number and severity. A data-driven approach could help with early warnings and manage risks of climate change (Ford et al, 2016).
- **Increasing amounts of data available about children:** There is growing interest in children's data across sectors (Berman and Albright, 2017; Young 2020; Russell and Macgill, 2015), for example, via digital identification programs (Viola de Azevedo Cunha, 2017) children themselves are generating digital footprints from ever-younger ages (Livingstone and Stoilova, 2018), and the private sector is gathering more data on children's preferences and behaviour (UNICEF, 2018).
- **A regulatory environment trying to keep apace:** The General Data Protection Regulation (GDPR) was implemented in 2018 in the European Union (European Union: European Commission 2016), acting as a model for other countries to develop their own data protection legislation. It has also shaped national policy across Africa (AfDec, 2021) with 24 out of 53 African countries adopting laws or policies to protect personal data, with the number continuing to rise (Privacy International, 2020).
- **Increased focus on the role of algorithms and automated systems in shaping the public sphere**: Due to high-profile court cases and media investigations (Cadwalladr and Graham-Harrison, 2018), as well as major political events influenced by the technology industry (Zialcita, 2019), there has been increased public scrutiny, and legal challenges and awareness of the ways in which algorithms are shaping the public sphere, especially with regard to the processing of personal data – including the personal data of children (Council of Europe, 2020).

2.3 Why focus on children and machine learning?

It is important to recognize that in many instances children are, and need to be, understood as distinct from adults due to their physiological and psychological differences (Medical Research Council, 2004). As defined by the United Nations Convention on the Rights of the Child (UNCRC) (1989), children are referred to as those below the age of 18 while adolescent and youth respectively refers to those aged 10-19 (WHO, 2014) and 15-24 (United Nations Department of Economic and Social Affairs (UNDESA, n.d.). For the purposes of this paper, our working definition of 'children' is inclusive of all three definitions, considering those 24 years or younger.

The differences between children and adults are highlighted and recognized by researchers in a diverse number of fields, from medicine (CDC, 2019) to linguistics (Newport, 2020) and neuroscience (Furlong, 2009). Unlike adults, children's bodies are in an active state of change, moving through a series of what are referred to as 'developmental windows' (World Health Organization, 2007). This means that exposure to physical hazards and the psychological impacts of their socio-economic ecologies can and does have differential impacts on children affecting them over a greater duration of their life course in comparison to adults (Shea, 2008).

These differences are manifested in a variety of ways that generate new benefits and risks of using PA, as discussed in Sections 5 and 6. For example:

- **Differences in neurological and physical development** (World Health Organization, 2007 and 2018) means that data must be analyzed in a child-centric/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 when it comes to using their data and be considered when assessing children's diverse understanding of data privacy.
- **Children's frequent lack of agency** over their social network, physical environment, and in many cases, data provision, must be considered when PA is used, so as to respect their right to self-determination. It also acknowledges both the potential and actual lack of control over their own digital identities and the problematic ways in which they show up and do not show up in datasets. For example, children can often be invisible in big datasets through a lack of disaggregated data; at the same time, many children (such as migrant children) are part of populations that might intentionally seek to avoid increased data visibility because of concerns about privacy and vulnerability. Migrant children can further experience unreliability or duplication in data collection when separated or unaccompanied (IOM's Global Migration Data Analysis Centre [GMDAC], 2021).

Additionally, any potential negative impacts of PA may have consequences throughout children's life course. Particularly with PA that generates individual outcomes, the impacts can be significant and drastically affect a child's life trajectory – for example, their physical health, education pathways, home and family environment, and decisions regarding health interventions such as immunization programmes (Chandir et al, 2018). Because PA outputs pertain to events that have not yet happened, assessing accuracy can be difficult, but if these outputs are incorrect or inaccurate, subsequent interventions based on PA may curtail a number of fundamental child rights, as defined in the United Nations Convention on the Rights of the Child (UNCRC).

As the Responsible Data for Children (RD4C) report (UNICEF and GovLab, 2019) states, "data collection and data-based systems were often designed with (consenting) adults in mind, without a focus on the unique needs and vulnerabilities of children". A myriad of factors contributes to why focusing on the use of predictive analytics specifically pertaining to children is necessary, including 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).

This paper focuses primarily on applications of predictive analytics that involve machine learning (as opposed to predictive modelling based on traditional statistical approaches), for a number of reasons. First, the rapid spread of ML techniques has had a transformative impact on multiple sectors, with significant implications and direct and indirect impacts on the wellbeing, rights, and long-term opportunities of children (Leslie et al., 2020). Second, using traditional statistical methods to predict child-specific future outcomes already has a well-established literature (Munro, 2005), though it should be noted that the benefits and risks of traditional statistical models are often similar to ML-driven models. Finally, machine learning brings a new range of risks (Gillingham, 2016) and benefits which are particularly worthy of interrogation as they pertain to children, as explained below.

In relation to children, predictive analytics has been used for a broad range of purposes that affect children in both indirect and direct ways. This includes, but is not limited to, improving health diagnostics in pediatric and maternal care; preventing child maltreatment and abuse; enhancing educational programmes; forecasting emergencies; monitoring human rights violations; and predicting natural hazards.

While multiple potential and actual uses of PA can be, and are, used to support a child rights agenda, as discussed below, it is also particularly crucial to consider the unintended impacts of predictive analytics as they may affect children and young people in a variety of intersecting ways. These include potential and actual limitations in relation to agency and autonomy (Berman and Albright, 2017; UNICEF, 2020), implications for expression and identity development, potential short and longer-term consequences of inappropriate interventions and assessments, and (further) targeting and exclusion or marginalization based on biased datasets (UNICEF, 2021).

2.4 Different types of predictive analytics

The umbrella term of predictive analytics encompasses a variety of potential use cases as they relate to children. These varying use cases have different implications with regard to potential benefits and risks. In order to properly assess potential impacts on children, as mentioned earlier, we need to draw a distinction between different types of predictive analytics.

For the purposes of this paper, a basic taxonomy of predictive analytics is presented that is distinguished by:

- The focus of the outcome: **population, phenomenon and/or individual**
- The nature of the model, including:
    - 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**

The following section unpacks this taxonomy.

2.4.1 Focus of outcomes (population vs. individual)

Population-based outcomes: where PA is used to assess needs and derive predictions that affect specific populations (a distinct group of people), as opposed to specific individuals.

Examples include:

- Population-based needs assessment, such as mobility tracking of displaced populations or movement tracking of populations for disease containment.
- Natural or social phenomena predictions, such as predicting natural hazards, flagging potential conflict hotspots, disaster-risk reduction.

Individual-level outcomes: where outcomes of PA target specific individuals by translating and applying population-level trends and aggregations to individual cases.

Examples include:

- Inferential applications for individual assessment, such as risk assessment tools, to recommend child welfare interventions, or assessment of performance to suggest targeted learning programmes.
- Consumer, informational, or political targeting – drawing patterns between online and offline behaviour to suggest possibly undisclosed preferences, interests, and dispositions.

Crucially, the risk/benefit analysis for PA with population-based outcomes is considerably different to the risk/benefit analysis for PA with individual outcomes. As such, we draw a distinction between these in the later sections covering benefits and risks and suggest this to be an initial consideration in the risk assessment process to ensure the nuances of PA are properly understood.

2.4.2 The nature of supervision and human intervention in the design of the model

When considering the various ethical impacts of PA, it is important to note that a variety of different types of learning can be employed to create ML models, each of which involves distinct levels of human intervention or human input, and is suitable for different scenarios.

Learning types generally fall into four main categories: supervised, unsupervised, semi-supervised, and reinforcement learning.

- **Supervised learning** is used when the inputs and the desired output are already known. Supervised learning uses labelled data for the creation of a learning algorithm, where the dataset is used as the basis for predicting the classification of other unlabelled data (Talabis et al., 2014). Supervised learning algorithms are generally among the easiest to interpret and have the most human input. They are, however, limited by the labels within the dataset used, and are thus affected by the ways in which the dataset is structured and categorized.
- **Unsupervised learning** uses unlabelled data to uncover patterns not previously considered. For example, these algorithms are often used for large-scale image processing in which thousands of images are grouped together based on their similarity to each other. There is no human intervention in classifying the data or deciding what to look for; instead, the algorithm finds its own similarities between the images. This offers the possibility to uncover new patterns but is among the least interpretable, presenting the smallest opportunity for human intervention and review.
- **Semi-supervised learning** attempts to strike a balance between supervised and unsupervised learning algorithms, combining a smaller amount of labelled data with a larger amount of unlabelled data. These algorithms often work well in situations where there is a need to control the interpretability of the model, but a lack of time or resources to label the entire dataset (Fumo, 2017). This option reduces the need for labelled data, but also decreases interpretability in comparison to supervised learning.
- **Reinforcement learning** algorithms interact with their environment, providing a feedback loop between the learning system and the responses it receives. While these algorithms resemble supervised learning models in that they also receive feedback and correction, the feedback tends to be unclear and delayed (Brownlee, 2019). Since no guidance is provided on what the correct answer is (unlike supervised learning), the algorithm learns on its own, through a typically slower trial-and-error process.

In short, using **supervised learning** or **reinforcement learning** increases the possibility for human intervention, as the level of interpretability is higher than forother types. This provides a mitigation opportunity – in case of outputs being generated which might cause unintended harm to children – as humans involved in the process can have input into what the model learns from and provide feedback or adjust the model accordingly.

Using unsupervised or semi-supervised models means that incorrect outputs are harder to catch. Models with lower levels of interpretability also mean that it is much harder (if not impossible) for humans involved to adjust the model based on incorrect outputs, or to know exactly where in a model a mistake was made.

Recent developments have also uncovered new learning methods that can be used in additional scenarios. For example, the 'less than one' learning method allows the model to identify cases beyond the cases used for training. With this method, a computer can, for example, be trained to recognize the numbers one through nine, so that when it encounters a higher number (like 115), it can identify this number by what it already knows (in this case, it would predict that the number is 66 per cent of the digit '1' and 33 per cent of the digit '5'). As technology journalist Karen Hao (2020) notes, these improved algorithms "could make AI more accessible to companies and industries that have thus far been hampered by the field's data requirements. It could also improve data privacy, because less information would have to be extracted from individuals to train useful models."

2.4.3 Level of interpretability of the algorithm used: black box or open source

Machine learning models vary in terms of their interpretability (their intelligibility to human reasoning) and explainability (the conveyability of the logic behind its results) (Leslie et al., 2020). Algorithms that are low on interpretability and explainability are known as 'black box' models due to their opacity and the inability to see their internal workings, with only the inputs and outputs visible. On the other end of the interpretability and explainability spectrum are open source algorithms, sometimes known as 'white box' algorithms (USAID, 2018). Among the more interpretable models are linear regressions and simple decision trees; the harder to interpret models include neural networks and tree ensembles (USAID, 2018).

2.4.4 Role in decision-making systems

The way in which PA is integrated into existing decision-making systems also has a large impact on how opportunities can be taken or risks mitigated. Specifically, if the outputs of PA are automatically used without human review (which is often referred to as 'automated decision-making' technologies), a particular set of risks arises as errors and bias are more likely to slip through the cracks without being noticed. However, the most common form of predictive analytics currently within the sector has humans involved intentionally throughout the process, (known as decision support, or 'humans in the loop'), as well as humans considering the outputs of PA in triangulation with other data sources. Humanitarian experts interviewed for this report suggested that for PA to be effective within the sector, it should be a "decision support tool, not a decision-making tool" (Interviewee 1).

2.5 Existing ethical and legal frameworks

Numerous existing ethical and legal frameworks can and should be drawn upon to govern or suggest guidelines for the use of predictive analytics as it pertains to children. Among the foundational human rights documents that could be used as a basis for this type of analysis is the **United Nations Convention on the Rights of the Child** (UN General Assembly, 1989), which is grounded on principles that include non-discrimination, adherence to the best interests of the child, the right to life, survival and development, and the right to privacy. Its recent accompanying **General Comment No. 25** (OHCHR, 2021) establishes that the Convention in its entirety needs to be understood and implemented in relation to the digital environment. The document also explicitly calls on states to ensure transparency in any assessment and to explain what criteria have been applied with regard to automated decisions, to ensure that automated systems are not used in ways that limit the opportunities or development of children (Ibid).

The Responsible Data for Children principles and practices (RD4C, 2019) also provide important insights and guidelines for those committed to responsible and rights-based policy (UNICEF and GovLab, 2019). The RD4C guidelines are based on the rationale that "data about children requires an additional duty of care in comparison to data about adults, and responsible data approaches must accordingly adhere to higher standards and security measures" – all of which applies to the application of predictive analytics.

Within humanitarian contexts, **the Sphere Standards** articulate a further set of ethical principles relevant to deployment of predictive analytics (Sphere Association, 2018). The Sphere Standards outline four 'Protection Principles' for humanitarian action and actors. These principles are aimed at centering people's rights by avoiding exposing them to harm; ensuring people's access to assistance happens without discrimination; assisting people in recovery from threatened or actual violence; and supporting them in claiming their rights. In the case of children, this means ensuring that decisions based on predictive analytics do not reinforce policies that put them at risk but instead, offer safer alternatives for receiving support. The Inter-Agency Standing Committee (2021) also outlines, in their operational guidance report on Data Responsibility in Humanitarian Action, a set of actions crucial to ensuring the responsible management of data in humanitarian contexts and these can and should be considered in the use of humanitarian data for PA purposes.

The European Union's General Data Protection Regulation (GDPR), as one of the key international data protection frameworks, has a number of relevant regulations and provisions. While it mostly treats child data subjects the same as adults, there are a few key sections that specifically pertain to children, data, and automated processing (European Union: European Commission 2016). Recital 38 notes that children need "specific protection" since they are likely to have less understanding of data processing risks and related rights.

Regarding automated decision-making in particular, the **UK's Information Commissioner's Office** (ICO) advises that no decisions that have a legal or other significant impact on children should rely solely on automated decision-making unless certain thresholds are met. They specifically mandate the completion of a Data Protection Impact Assessment to first "establish whether automated processing will result in a high risk to their rights and freedoms" (Information Commissioner's Office, n.d.). This affects, in particular, the potential ways in which PA may be incorporated into existing decision-making systems. Furthermore, Recital 17 indicates that this use of children's data should be rare. The ICO determines compliance with the GDPR based in part on what is in the best interest of the child, following Article 12 of the UN Convention on the Rights of the Child, and recommends that a diverse range of children are consulted on the design of data collection processes.

Narrower in scope than the GDPR, **the United States' Children's Online Privacy Protection Act** (COPPA) of 1998 does not address processing per se but does reflect on the collection of children's personal data online. COPPA requires data collectors to obtain verifiable parental consent (aimed at ensuring the person giving consent is indeed the parent) for any child under the age of 13 in most cases. Rather than specifying an age, the GDPR relies on the child's competency "to understand the implications of the collection and processing of their personal data", which may be determined by member state legislation.

While some states in other regions are adapting their own data protection laws to be GDPR-compliant, eight African countries have formed the **African Declaration on Internet Rights and Freedoms** (AfDec) Coalition to standardize data protection laws and the digital economy across the region, much like the EU did with the GDPR (Nwafor, 2020; Velluet, 2021). These standards will be based on the Malabo Convention principles and African Union Personal Data Protection Guidelines (African Union, 2014), which mention children as an important stakeholder group. Details on how this legislation may affect the implementation of predictive analytics, and specifically PA for children, however, are as yet unclear.

The content and guidance in these frameworks will be reflected throughout this paper.

2.6 Key points: Background

- Predictive analytics (PA) is the use of historical data or past experience to predict future outcomes. For the purposes of this paper, we primarily consider predictive analytics as powered by machine learning.
- The broader context in which PA is being considered in the public, humanitarian, and development sectors is one of resource constraints; growing displacements; an ongoing climate emergency; increasing amounts of data available about children; increased scrutiny of emerging technologies; and a regulatory environment trying to keep up with these fast changes.
- Children are distinct from adults due to their physiological and psychological differences. In terms of the use of PA, these differences give rise to varying benefits and risks.
- PA can have individual-level outcomes, where specific children are targeted, or population-based outcomes, where the outcomes affect specific groups of people of which children form a part. Each of these has different risk/benefit analyses as they pertain to children.
- Other distinctions within PA include the nature and degree of supervision and human intervention; the level of interpretability of the algorithm used; and the way in which the outcome is used in decision-making.
- There are a number of existing ethical and legal frameworks to draw upon; in particular the UN Convention on the Rights of the Child; the Responsible Data for Children principles; and the EU's General Data Protection Regulation.

3. METHODOLOGY

An initial literature review was undertaken that focused on identifying current and proposed uses of predictive analytics impacting children directly and indirectly, and identifying benefits, risks, and possible mitigation strategies related to its use.

This review was conducted in three languages: English, French and Spanish, with English as the primary language. Our research included extrapolating from findings with other populations and sectors; reviewing relevant UNICEF documents; and conducting a literature review of journal articles, nonprofit studies, news reports, and grey literature such as blog posts.

Our primary databases were PubMed, Psycnet, Elsevier, Sage, Science Direct, Clio, and GoogleScholar, and our search terms included the following (in all three languages):

- predictive analytics, predictive analytics children, predictive analytics public policy, predictive analytics child welfare, predictive analytics humanitarian, predictive analytics ethics, predictive analytics juvenile justice, predictive analytics education/learning, predictive analytics health/nutrition, predictive analytics ethics, child definition, difference between children and adults, predictive models, machine learning models, forecasting, machine learning and children; neonatal care and predictive analytics, maternal care and predictive analytics, infant and predictive analytics, child malnutrition and predictive analytics, juvenile detention and predictive analytics, child homelessness and predictive analytics, child disability and predictive analytics, predictive analytics and criminal justice, predictive analytics and global south.
- machine learning and the same variations as above.

Case studies were selected from a longer list of real-world implementations of predictive analytics identified through outreach to The Engine Room's Responsible Data community and to various UNICEF and organization partners and allies, as well as through a call for cases in a project blog post on The Engine Room's website. Case studies were chosen for geographic and sector diversity, and variation in the type of implementation of predictive analytics.

Case studies were complemented by interviews conducted with 10 key experts and practitioners in the fields of humanitarian work; ethical technology; children's programming; data science; and data privacy located in Italy, Mexico, Paraguay, the United Kingdom and the United States. All interviews were conducted remotely from December 2020 through June 2021.

All interviewees were asked about their knowledge and experience with predictive analytics. This included perceived benefits, risks and mitigation measures, and participation of communities impacted by the use of PA in the redressal mechanisms.

Insights were drawn from coding and analysis of interviews and the literature review.

4. USE CASES: PREDICTIVE ANALYTICS AND CHILDREN

There has been an increase in sectors adopting ML in contexts where this may impact on children in both direct and indirect ways – from health care, education, and child protective services to humanitarian response planning. Access to large troves of data and complex data extraction techniques has presented a variety of new possibilities and challenges, both particularly pronounced when it comes to the rights and well-being of children. Crucial progress, driven by PA models and other non-predictive uptakes of ML techniques, is already happening in a number of areas affecting children, enabling more precise clinical decision-making in health care, promoting better humanitarian planning, and supporting targeted education programmes.

Nonetheless, the challenges and ethical concerns presented by these models are equally significant. The opacity of commercial and public sector algorithms (Berman et al., 2016; Executive Office of the President et al., 2016; Redden et al., 2020); the lack of sufficient accountability and oversight mechanisms (Andersen, 2019; Noble, 2018); a visible gap in "empirical research relating to the actual use of crisis data for preparedness activities" in humanitarian efforts (Hernandez and Roberts, 2020); the tendency of automated decision-making systems to reinforce deep-seated structural injustices (Eubanks 2017; Noble, 2018); and the spread of commercial and state surveillance enabled by emerging technologies (Zuboff, 2015; Tufekci, 2019), have all presented major challenges in areas where ML-driven models have proliferated. With respect to the uptake of PA in child-specific instances, some of the aforementioned concerns are even more pronounced, as described at greater length in Section 6.

In the following section, we provide a deeper analysis of how PA models work in a few child-specific areas, drawing upon a wide range of examples from health care, education, humanitarian planning, and child protective systems. Through these examples, we describe common trends prevalent across sectors and highlight key differences between the distinct uptakes of PA models in each area.

- **Predictive analytics with a direct impact on children:** Where children are the main target of a project, we refer to the impact of PA models as 'direct' (*see Section 4.1*).
- **Predictive analytics with an indirect impact on children:** Where children are not the primary focus of the project but are nonetheless impacted by a PA model, we refer to the impact as 'indirect' (*see Section 4.2*).

In order to draw out the most salient points, opportunities, risks and recommendations on how to effectively and ethically use PA models in child-specific initiatives, we also draw distinctions between population- and individual-focused use cases, as defined in Section 2.1. To briefly summarize, population-focused instances of PA target larger groups of people, in contrast to individual-focused PA models that apply population-level trends and aggregations to individual cases. As noted previously, lack of consensus relating to definitions has resulted in predictive models, forecasting, and PA being used interchangeably in the analyzed literature whenever possible, however this paper will attempt to distinguish between them.

4.1 Predictive analytics with a direct impact on children

PA models directed at children are most widespread in paediatric care; child-focused disease control; child protection services; juvenile justice systems; education initiatives; and consumer targeting. In the majority of analyzed sectors, real-world examples can be found for both individual-level uptakes of PA as well as population-focused predictive models that target children as a cohort, as elaborated in the examples below.

4.1.1 Health

PA is prevalent in the medical sector, in areas of environmental, physical, and cognitive health (Raj et al., 2020). In the past decade there has been exponential growth in the number and scope of clinical decision-aiding tools which have capitalized on both increased computational power and increased access to medically relevant data, including electronic health records; patient registries; health surveys; laboratory tests; and other data collected through intelligent medical devices.

PA has also become more widespread in paediatric care. Individual-focused risk assessment models have promised to improve the accuracy of child-focused prognostics, decrease infant mortality, and reduce both the length of hospitalization and the chances of hospital readmission for children. In medical research, predictive models have been tested to reduce neonatal mortality by using historical clinical data on pre-term infants (Singh et al., 2017), to identify paediatric patients at risk of developing blood clots (Walker et al., 2021); and to recognize the early signs of sepsis in children (Chandir et al., 2018; Spaeder et al., 2019).

With regard to the specific risks embedded in health-focused PA systems, Van Calster et al. (2019) point out that poorly calibrated algorithms pose a significant challenge in clinical practice, where inaccurate risk predictions can be particularly harmful. As an example, PA that predicts the chances of in-vitro fertilization (IVF) treatment leading to live birth can lead to a strong overestimation of the chances of a successful birth. This could give false hope to families already going through a stressful experience, whereas underestimation may lead to excessive and unnecessary treatments (Ibid). Van Calster et al. argue that better reporting on the “calibration performance” of clinical PA is essential, and that developers of health-focused risk models should always be required to test their algorithms in multiple contexts, and on sufficiently large samples. Parikh et al. (2019) also call for more regulatory guidance for the uptake of PA in clinical settings, noting that the predictive performance of health-focused PA models can vary significantly over time and across populations. Another key challenge for individual-focused PA models in healthcare is that they often build on sensitive patient data – particularly pressing in the case of children who, as described further in Section 6.2.1, have especially limited capacity to grant meaningful consent for data collection practices.

With regard to population-level examples, PA has been the most widespread in disease control, resource optimization, and epidemiological forecasting. PA models have been tested to support routine immunization programmes in Pakistan (see the case study below), and to predict and prevent parasite infections in West African children (Magalhães et al., 2011). It is important to note, however, that a greater body of evidence will be required to analyse the impact and limitations of existing predictive models, and that it is unclear whether population-level PA models yield better results than statistical forecasting.

Child-focused mental health services have also been experimenting with predictive models. Seattle Children, a regional mental health facility for children under 12, developed a PA model to monitor and forecast changes in the capacity of their under-resourced psychiatric unit, by collecting and analysing data on patient location, length of stay, patient census cohorts, and more (Children’s Hospital Association, 2020), although detailed information about the underlying PA model or the impact of the project is not currently available.

Case study: ***Routine immunization in Pakistan***

Using PA models to predict immunization adherence, Chandir et al. (2018) conducted a feasibility study to identify the children most likely to drop out of routine immunization programmes in Pakistan. They used data already routinely collected by health authorities (such as gender, languages spoken, town or city of residence, and more), with the aim of establishing a model that did not require additional data collection from already under-resourced public authorities.

Chandir et al. (ibid) used longitudinal immunization records and tested four different ML models (random forest; recursive partitioning; support vector machines [SVMs]; and C-forest) within their analysis. Assessing each of those models for categories such as accuracy, sensitivity, and predictive value, the authors concluded that all models had high levels of accuracy, and that PA could be deployed to help "accurately identify children who are at a higher risk for defaulting on follow-up immunization visits", and thus open a window for better interventions. According to the authors, uptake of such a model would enable targeted campaigns where families at greater risk of defaulting would receive additional information about the benefits of adhering to routine immunization.

In 2020, another feasibility study by Qazi et al. (2020) was conducted to address gaps and weaknesses identified in the previous model and to introduce an even more accurate prediction model. This data-driven framework achieved a "98 per cent accuracy for identifying children at different stages of risk of dropping out", according to the authors of this study.

While none of these models have been implemented by public agencies so far, the authors of both studies suggest that PA could help reinforce public health initiatives by strengthening targeted interventions.

On the individual level, this could mean improved chances for children to get crucial vaccines, and a reduction in the number of drop-outs from immunization programmes. A false positive would mean that a child, who would have otherwise not missed their routine vaccination, might receive extra information about its benefits – an event which would presumably not have negative impacts on the child or their family. The main concern in this scenario would be a waste of existing resources, especially since deploying complex PA models requires extensive expertise. Should false negatives occur, the consequences might be even more significant and such a scenario would mean that children in need of vaccine information would be overlooked by the campaign. On the population level, such models could support better resource allocation and more effective public health campaigns.

Similar interventions – where PA could potentially be useful in supporting decision-making around vaccine-related interventions – may become particularly important in the near future, given the disruption of routine immunization campaigns due to the COVID-19 pandemic (WHO 2020; Mansour et al., 2021).

4.1.2 Child protection

In a number of high-income countries like the UK and the US, child protection services are increasingly turning to PA models to improve their services and reduce their staff's workload. Pulling together administrative data from case reports, referrals, and individual assessment, ML-driven risk models have been implemented that aim to identify child maltreatment, abuse, and neglect to develop individual risk scores for children and families, and to provide better services for these children and families (Lanier et al., 2019; Leslie et al., 2020; Clayton and Sanders, 2020; Glaberson, 2019). As Glaberson (2019) notes, data used by PA models in the child welfare context often includes information "arising directly out of the child welfare agency's involvement with the family", but may also consist of other publicly available datasets, such as criminal justice records, medical data, birth records, and education records. Non-profits are also increasingly deploying PA to predict and prevent child maltreatment. For instance, Texas-based group Predict Align Prevent uses spatial risk modeling (Daley et al., 2016) to identify "high-risk places where child maltreatment and related fatality are likely to occur in the future" (Predict Align Prevent, n.d.).

ML-driven PA models often seek to do what statistical forecasting has done in the past, i.e., attempt to identify the predictive relationship between different variables, building on empirical research (Leslie et al., 2020). As a result, much of the specific risks embedded in more recent PA models are similar to the ones associated with traditional actuarial methods, described at greater length in Section 6. To briefly summarize, these include the lack of generalizability and applicability to different conditions and subpopulations, the over-representation of certain minority groups in the base data, or the challenges of setting appropriate thresholds for intervention (Ibid).

In the context of child protection, PA is not exclusive to child welfare systems. Similar to the way law enforcement agencies use PA for crime prevention (Lanier et al., 2019; Clayton et al., 2020), predictive models are being deployed to help prevent crimes against children; for example, there is exploratory research being done on predicting bullying and the spread of child sexual abuse material (Sanchez et al., 2019). While not always building on ML techniques, data-driven predictive models are also being tested to aid forensic work – as an example, to predict recidivism among online child sex offenders - through conducting linguistic analysis of available chat transcripts (Drouin, 2018).

With regard to population-level uptakes of PA in child protection, the authors of this paper have not found any noteworthy examples in the public domain. Relevant instances may include using PA to predict the likelihood of certain populations being more prone to child trafficking and other forms of child-specific organized crime, or to identify which child protective agencies will be more likely to use their resources effectively. However, given that these examples are hypothetical, some of these models may fall under the category of statistical forecasting and would not fit the description of PA as defined in this paper.

4.1.3 Juvenile justice

Thanks to the widespread use of PA within law enforcement (*see Section 4.2.2*), predictive models are increasingly common in the juvenile justice system. In the US, the Florida Department of Juvenile Justice was the first state agency to conduct PA-based risk assessments to reduce incarceration rates, using demographics data and arrest records of children with a history of trauma (Kelly, 2015; Hickey, 2015). Since then, several other US states have followed suit, with the aim of identifying at-risk youth and predicting future potential crime and recidivism among children (Taxman, 2016).

In Singapore, probation officers routinely collect data about youth offenders for case management purposes, to make informed recommendations on the sentences received by young offenders (Ting et al., 2018). Considering the time-consuming nature of processing and interpreting large troves of data, Ting et al. (Ibid) recommend that probation officers improve their efficiency by introducing ML algorithms into their processes. They also note, however, that automated risk assessment methods should never replace human judgement and that human involvement is still required in criminal decision-making processes.

There is extensive literature on the opportunities and challenges of using predictive risk modeling in law enforcement initiatives, described at greater length in Section 4.2. When it comes to juvenile justice, evidence-based techniques may hold great potential to increase the capacities of overburdened and under-resourced systems (Weber et al., 2018), but as Berman and Albright (2017) note, the use of predictive data in these contexts may be concerning. Similar to predictive policing programmes, where PA models often over-represent marginalized groups in their data sample, over-reliance on PA to identify at-risk youth can lead to "inaccurate conclusions, rankings and skewed decision making", particularly pertaining to marginalized communities (Ibid). For instance, in an analysis of how prediction software used by law enforcement in the US discriminates against black people, ProPublica (Angwin et al., 2016, cited in Church and Fairchild, 2017) reported that young black adults arrested and charged with petty theft were more likely to be flagged as at high risk for recidivism than white defendants with a longer history of crime.

4.1.4 Education

PA has been widespread in education settings as well, especially in high income countries with greater access to complex learning analytics. Universities and colleges in the US, for instance, have been engaging with PA to improve their institutional metrics; to manage enrolment; to help students with their learning progress; to identify students most in need; to develop adaptive learning courseware; and to advance education- and pedagogy-focused research (Ekowo and Palmer, 2016). In a policy paper that provides guidance on the uptake of learning analytics, UNESCO's Institute for Information Technologies in Education (2012) notes that when it comes to introducing complex analytics, a key motivating factor for education institutions has been the potential to enable timely interventions, as well as to free up human resources by replacing the work of a teacher or instructor.

Specific examples of individual-level PA models include early-stage alert systems that inform students of their chances of succeeding with their studies (Calvert, 2014); deep learning techniques that assess and predict individual student performance (Doleck et al., 2019; Uskov et al., 2019); automated grading mechanisms (*see Section 6.2.5 for a detailed case study*); adaptive learning programmes; and targeted student advising programmes that optimize learning routes (Ekowo and Palmer, 2016). For example, in the US, PA has been used to identify the risk levels of individual students dropping out of science, technology, engineering, and mathematics (STEM) classes (He et al., 2018). In Australia, universities reportedly use PA to predict a student's risk of attrition based on "first year student study behaviours", with the ultimate goal of identifying preventive measures and enhancing retention (Seidel and Kutieleh, 2017). The risks embedded in these systems are similar to the ones pertaining to child protection systems, i.e., predictive models relying too heavily on demographic data can easily entrench existing disparities. This is particularly relevant in the domain of student performance, where predictive models can have long-term consequences on the career choices and life opportunities of children and young adults (as described at greater length in Sections 6 and 7).

For educational institutions, population-level PA models also hold promise to optimize internal workflows, improve institutional metrics, and consequently, to attract more funding. US universities and colleges have been using historical data on enrolment practices to forecast the size of their incoming classes, to attract better applicants, and to unlock state funding that is tied to student performance (Ekowo and Palmer, 2016). Ultimately, population-focused PA models may also prove helpful to aid the design of national policies and initiatives that target children with traditionally lower access to education, or to enable a more effective allocation of relevant state-level funding.

UNESCO's Institute for Information Technologies in Education (2012) noted a caveat in relation to the uses of PA – the assessment of distinct pedagogical methods and learning routes is a "widely contested area", making the design of learning analytics a complex challenge. Consequently, they call for developers and users of learning analytics to recognize and mitigate the limitations of any model that builds on available learning data, especially if the goal is to account for all the skills and attributes that are needed for lifelong learning. Exploring the utility and applicability of deep learning in the education context, Doleck et al. (2019) also argue that "educational data mining and learning analytics researchers should aim for explanation over prediction".

4.1.5 Consumer targeting

Building on large sets of consumer data, online and offline targeting strategies have become increasingly sophisticated and consequently, the proliferation of, and growing dependence on, digital tools and services has resulted in an increased uptake of PA that directly targets children as customers (Lupton and Williamson, 2017). As Buckingham (2014) notes, companies and private sector advertisers are seeking to engage with children "more directly and at an ever-younger age", and children are increasingly contributing to online content (Berman and Albright, 2017). A new research report by 5Rights Foundation and Revealing Reality (2021b) shows children are directly targeted with sexual and suicide content within as little as 24 hours of creating an online social media account on platforms such as Facebook, Instagram, and Tiktok.

While YouTube on its own accounts for a significant majority of viewership with regard to children's programming (UNICEF Innovation and Human Rights Center, UC Berkeley, 2019), the most well-documented example of child-focused online targeting that builds on PA is YouTube Kids, a video app developed for children with curated selections of content, parental control features, and filtering options based on select age groups. YouTube Kids, like YouTube, uses machine learning-driven models to predict the individual preferences of its users and recommend content accordingly (often referred to as recommendation algorithms). As Mascheroni and Holloway (2019) note, the platform tends to corral children into a marketing space where they can receive targeted advertisements "disguised as programming" as well as "incentivizing sensational content"', something that children are especially susceptible to (UNICEF Innovation and Human Rights Center, UC Berkeley, 2019). Research by 5Rights Foundation (2021a), conducted to inform the introduction of 'age assurance' in product design and development in the UK, recommended disabling "intrusive or risky design features such as geolocation data tracking, private messaging, and targeted advertising".

However, beyond the unexpected exposure to disturbing content with 'shock value' (arising from 'related' videos and unregulated content upload), the risks can extend to potential exposure to child sexual abuse material, possible grooming and trafficking, and even indoctrination into an extreme ideology (Morris, 2016) – all instances of interactions that could potentially significantly and negatively shape children's experience and world view.

4.1.6 Humanitarian and development planning

The spread of PA has happened comparatively later in the humanitarian sector than in some of the other fields analysed in this paper and as a consequence, there appears to be little to no public evidence of programmes that are fully operational, nor has the impact of these programmes been widely evaluated so far. A majority of the known PA models deployed in humanitarian contexts target broader populations, not just children (as detailed in Section 4.2 below), but there have been examples of PA being used in humanitarian planning with a direct impact on children.

One example is the United Nations High Commissioner for Refugees (UNHCR) and the University of East Anglia PA model being developed to predict arrivals of unaccompanied and separated minors in South Sudan, the Democratic Republic of the Congo, and Somalia (Centre for Humanitarian Data, 2021). The proposed model is designed to predict the approximate number of "unaccompanied and separated children" (UASC) arriving on a monthly basis, based on historical registration data, and provides insights into displacement patterns in select countries. There are also examples emerging of initiatives targeting child cohorts that cut across the development and humanitarian space, such as Save the Children's Migration and Displacement Initiative (2021).

As far as the authors of this paper are aware, no examples of individual-level, humanitarian-focused, child-specific uptakes of PA have been documented in the public domain. Major implementation challenges remain with the potential use of PA in the humanitarian and development space due to lack of data and specifically, disaggregated data.

4.2 Predictive analytics with indirect impacts on children

Predictive models have been used in a variety of instances where children are not the primary focus of the project but are nonetheless affected by the use of the PA model. This includes humanitarian planning initiatives, social welfare initiatives, predictive policing, and maternal care interventions that target caregivers or families more broadly. What follows is a summary of key trends and examples for both individual- and population-level uptakes.

4.2.1. Humanitarian planning

As mentioned earlier, among humanitarian and development organizations PA is being developed as a tool to potentially support response programmes by anticipating shocks and forecasting the likely trajectory and features of emergencies, including pandemics, famines, disease outbreaks, natural disasters, and refugee movement (Hernandez and Roberts, 2020). A number of pilot initiatives are in place, with projects such as predicting and analysing issues related to food security and population malnutrition in order to better inform aid delivery, to tracking and predicting the movements of refugees. Proposed models are based on a range of data sources, including mobile phone records; social media data; satellite images; meteorological data; financial transactions; and government-held data (Ibid).

Noteworthy, documented PA initiatives include Project Jetson, a predictive analytics model developed by the UNHCR's Innovation Service (2019) that aims to provide predictions on the movement of displaced populations within and outside of Somalia, based on variables assessing violent conflict, climate, and weather anomalies, and market prices. Another is the World Food Programme's Food Prices Early Warning program, which tracks changes in food prices in low and middle income countries with the intention to proactively mitigate crisis response when local food prices hit abnormally high levels (Centre for Humanitarian Data, n.d.). The Foresight model developed by the Danish Refugee Council uses multiple data sources to predict forced displacement in Afghanistan, Myanmar, and the Sahel region (Ibid); Save the Children's Migration and Displacement Initiative uses machine learning algorithms to predict the duration, scale, and demographics of forced displacement (Kaplan and Morgan, 2018); and the Global Cholera Risk Model developed by the University of Florida and the University of Maryland, uses precipitation, temperature, and population data from different sources to predict the risk of cholera outbreaks (Centre for Humanitarian Data, n.d.).

Case study: *Predicting forced displacement in Somalia*

In 2017, UNHCR's Innovation Service launched Project Jetson, a predictive analytics initiative aimed at providing forecasts on the movement of displaced populations within and outside of Somalia, based on variables including violent conflict, climate and weather anomalies, and market prices (UNHCR Innovation Service, 2019).

The project originated out of a demand from UNHCR's in-country offices, which had dealt with a severe displacement crisis in 2011, during Somalia's worst famine in 25 years. Over 140,000 Somalis were forcibly displaced, and at a certain point 2,000 people per day were arriving at the UNHCR's facilities in Dollo Ado, in Ethiopia's southern border region. Facilities and teams were not equipped to deal with the situation, and crowded camps, as well as high levels of infant malnutrition, were some of the consequences (Ibid).

To find the variables that would be most relevant to a PA model, developers went to Ethiopia and conducted interviews with UNHCR staff in Dollo Ado and with displaced persons themselves. After desk and field research was done, 10 variables were identified as critical indicators of Somali displacement, including basic commodity market prices; rainfall; incidents of violent conflict; and historical population movement (UNHCR Innovation Service, 2019).

The model uses open data (UNHCR 2021) to predict forced displacement movement (i.e., population flow, influx-only) at least one month ahead, tracking arrivals of internally displaced persons and refugees in different regions of Somalia and in Dollo Ado (Centre for Humanitarian Data, n.d.). Implemented in 2017, by June 2018 predictions were accurate for 11 of the country's 18 regions, alerting teams with one month's notice as to when to expect an influx of displaced persons.

According to UNHCR, Jetson's outputs allow for better coordination among humanitarian actors as well as better resource allocation, which in turn improves readiness of service provision for refugees and internally displaced persons (UNHCR Innovation Service, 2019).

4.2.2 Criminal justice

Using large sets of data on past crimes, ML-driven risk models have been widespread in criminal justice systems, with indirect impact on children coming from families affected by, or involved in, criminal activity. Law enforcement agencies have been analysing historical data on the nature, time, and perpetrators of past offences, with the aim of preventing future crimes, predicting the likelihood of recidivism and reoffending, and identifying potential victims of crime (Perry et al., 2013).

Individual-focused examples include risk assessment software used by police, courtrooms, and correction facilities. The most well-known of these is the Correctional Offender Management Profiling for Alternative Sanctions system (COMPAS), a case management and decision-aiding tool deployed by several police departments in the US to assess the likelihood of defendants becoming recidivists. In Australia, the Suspect Targeting Management Plan (STMP) is a similar risk assessment tool deployed by New South Wales Police to prevent future crimes by focusing on repeat offenders (Sentas and Pandolfini, 2017).

As for population level instances, in the criminal justice context PA is most widely used to support neighborhood protection and monitoring programmes, and to forecast crime hotspots. These programmes build on historical crime data and include techniques such as 'near repeat modelling', denoting the likelihood of a neighborhood being criminally targeted repeatedly and 'risk terrain modelling', indicating locations with the propensity for criminal activities (Perry et al., 2013; Executive Office of the President et al, 2016).

The growing popularity of predictive policing is at least partially due to the presumed efficacy of data-driven programs designed to enhance 'human decision-making behaviour' (Church and Fairchild, 2017). However, researchers and journalists have warned of the dangers of over-reliance on algorithms that reproduce existing patterns of discrimination and bias, while human rights advocates have raised concerns around transparency, lack of awareness of community needs, and racial profiling (Angwin, 2016; Završnik, 2019; Meijer and Wessels, 2019; Benjamin, 2019).

An example of this was noted in Australia where research by the Youth Justice Coalition found that the Suspect Targeting Management Plan targeted a disproportionate number of Aboriginal and young people (as young as 10); perpetrated oppressive policing patterns detrimental to relationships between police and young people; and initiated unnecessary and expensive contact between young people and the criminal justice system, with no actual evidence in reduction of youth crime (Sentas and Pandolfini, 2017).

4.2.3 Maternal care

PA models that aim to improve maternal health often have significant impacts on children as well, particularly in instances of prenatal and postpartum interventions. As is the case with other health-focused uptakes of PA, there are a growing number of examples where predictive models have been used to prevent potential complications; for example, during pregnancy, child delivery, and postpartum care, and to consequently improve the health prospects and wellbeing of mothers and infants.

Individual-level examples include models that predict postpartum emergencies, a major cause for maternal mortality and readmission to hospital after delivery (Hoffman et al., 2021). In terms of population-level examples of use in maternal care, PA has been tested to identify underlying causes of maternal morbidity and mortality, as well as to uncover nuances of racial and ethnic disparities that put certain populations at higher risk of hospital readmission (Carroll, 2018). The majority of the specific benefits and risks of health-focused PA models, as described in 4.1.1, apply to maternal care interventions as well.

4.2.4 Social welfare

As with child protective services, social welfare services in high income countries have been deploying PA to better allocate services to their beneficiaries, including homeless families. Such services include adapting predictive models to develop a better understanding of what makes certain families more prone to losing their homes or re-entering shelters. To illustrate, using anonymized sociodemographic data for residents of New York City who have exited homeless shelters, Hong (2018) examined the factors associated with homeless families' stay patterns, including long-term stays and re-entries into the shelter system, in order to "enhance the likelihood of vulnerable populations getting the support and interventions they need." However, as the authors note, while such applications may strengthen research on homelessness, they are unable to help to radically reduce the number of entries or address the "overall magnitude of the homelessness problem" (Ibid). Similarly, the California Policy Lab and the University of Chicago Poverty Lab used county data to "predict homelessness among single adults receiving mainstream county services", noting that "predictive analytics can greatly improve our ability to identify single adults at risk of homelessness and more precisely target prevention programmes" (Von Wachter et al., 2019).

4.3 Key points: Use cases

- PA is deployed in a range of sectors with direct or indirect impact on children, with examples of both individual- and population-focused outputs. Existing examples with direct impact include prognostic models in paediatric care; child-focused disease control and immunization programmes; risk assessments in child protection systems; and targeted student learning programmes. Examples of indirect impact include humanitarian planning, maternal care, criminal justice, and social welfare initiatives.
- The predictive performance of PA models can vary significantly over time and across populations, posing the need for increased care in the use and application of PA.
- Predictive models in child protection, social welfare, criminal justice, and education settings hold the potential to expand the scope and impact of broader population-focused initiatives. However, the individual-level risk assessments that proliferate in these sectors raise significant concerns around efficacy, discrimination, and bias.
- Prediction algorithms are increasingly widespread in online targeting strategies, often directly pursuing children who are especially vulnerable to age-inappropriate content and extreme ideologies.

5. POTENTIAL BENEFITS OF PREDICTIVE ANALYTICS

In contexts and applications with direct or indirect impacts 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. This is particularly valuable given the current context. In many high and low income countries, social welfare, development and humanitarian sectors have seen increased demand without comparable funding increases, leaving agencies severely resource-constrained (OCHA, 2020; Leslie et al., 2021). As an example, according to OCHA (2020) the humanitarian sector received 46 per cent less funding than required to meet identified needs.

This section discusses this and other potential and observed benefits of PA, including:

- More efficient resource allocation;
- Supporting better planning;
- Reducing the burden on overburdened systems;
- Generating new insights;
- Identifying patterns across multiple streams;
- Increasing accountability; and,
- Targeted profiling for positive behaviour change.

For all of these potential and observed benefits, the following caveat applies: The currently-available evidence in relation to cost/benefit analysis and robust evaluations is not comprehensive and consequently, as 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.

5.1 More efficient resource allocation

Population-based PA models that are used to support decision-making have the potential to yield more tailored interventions and resource allocation informed by historical interventions, population movements, and demographics and environmental changes. Experts with direct experience of using PA in the humanitarian sector interviewed for this paper all noted that they see more efficient resource allocation as a major benefit of PA.

The changing nature of humanitarian emergencies means that policy makers, social workers, and aid workers need to decide whether to treat an emerging situation with short-term emergency relief, or plan for it as a longer-term, protracted emergency, particularly noting that two out of three acute crises eventually become protracted (Kaplan and Morgan, 2018). Save the Children notes that for them, "better data and predictive capability is one key to providing more effective support" in this area (Ibid). Interviewee 1, working at an international NGO, stated that "programming is still for short-term displacements", meaning that programming efforts are not focused on planning proactively for the changing nature of these emergencies, but rather are required to react to changing situations at short notice.

Yet the growing severity of the climate emergency means that the ability to triangulate complex climate data with other data streams could be particularly transformative in helping to understand and react to novel situations. The nature of these changing emergencies means that any technology that could support more effective use of resources holds great potential. As an example, the PA model developed by UNHCR and the University of East Anglia to more accurately estimate the number of unaccompanied children arriving in refugee camps may eventually result in better planning and resource allocation for nutrition and immunization programmes (Centre for Humanitarian Data, 2021).

5.2 Supporting better planning for environmental shocks and changes

Population-based PA models also offer the possibility for humanitarian and development actors to preemptively mobilize resources before a natural disaster has taken place – that is, acting on risk potential rather than acting only once events have taken place. This allows for more considered planning rather than acting under emergency conditions. The approach of acting based on risk is not new for humanitarian actors, but has seen increased confidence due to the introduction of PA, which has affected the scale at which funds are being mobilized as with, for example, the Anticipatory Humanitarian Action initiative, led by the UN's Central Emergency Response Fund (CERF) and OCHA.

Anticipatory pilot frameworks to release funding ahead of shocks and disasters are currently in place for five countries (Bangladesh, Chad, Ethiopia, Malawi, and Somalia), often supported by data generated by PA models. In July 2020, this allowed CERF to carry out its fastest-ever funding allocation of $5.2 million for anticipatory response ahead of seasonal floods in Bangladesh (UNCERF, 2020). Population-level prognostics, such as the drawing of predictive maps to analyse the role of water supply and sanitation in preventing helminth infections in West African children (Magalhães et al., 2011), offer promise for preventive disease control, with both direct and indirect impacts on children. However, it should be noted that predictive models in disease control do not automatically generate findings with higher accuracy than traditional statistical models and forecasting techniques.

If the outputs of PA models are used to support decision-making with, for example, preemptive plans that are ready to trigger if the predicted emergency does occur, this could be particularly transformative. In these cases, the 'worst case' scenario – if a model flags an emergency that ends up being less serious than originally predicted – would be that time may have been invested in making a plan, or resources may have been preemptively moved around ready for mobilization. The best case would be that an agency is able to respond faster to an emergency, and potentially, that predictive modelling of emergencies will allow for better resourcing and long-term planning, which is particularly crucial for sustaining responses that may last beyond the humanitarian response and move into the development stage.

5.3 Reducing the burden on overloaded systems

PA also holds the potential to decrease the burden on under-resourced and overloaded systems and agencies. PA models are routinely used to predict future capacity constraints in institutions – as observed in the demand for better estimations around the urgent care associated with the Covid-19 pandemic. When Chile witnessed a rapid increase in caseload and inability to accommodate all incoming cases, for example, ML-driven PA was employed. The solution included a combination of "autoregressive, machine learning, and epidemiological models" to forecast ICU utilization at the regional level (Goic 2021; Locey et al., 2020).

The implications could be significant for child welfare services and development and humanitarian agencies, especially in the context of low and decreasing resources and a fast-changing environment, as illustrated above. In fact, resource optimization has been a common motivating factor in many of the sectors where PA is already prevalent (as described in Section 4), with child protective services using predictive risk assessment models to reduce the workload of their overburdened staff; under-resourced education institutions turning to PA to replace the work of teachers and instructors; and child-focused mental health facilities testing PA to forecast future capacity constraints.

It is important to note that most PA models require extensive expertise in deciding what type of ML may work best in a given scenario or how to best train the algorithm. They also require long-term investment in all the iterations and updates that will be necessary to ensure that the model remains relevant and accurate over time and across populations. Resource allocation may therefore work best when expertise is already available – something that is often not the case for humanitarian agencies, development, and social welfare systems.

5.4 Generating new insights

Data gathered and used for ML-driven systems also holds the potential for being used to strengthen other facets of humanitarian programming, development work, and other child-focused interventions. Interviewee 1 stated that having more disaggregated data would be "transformational in terms of reach", in terms of addressing existing inefficiencies and mismatches between planning resource allocation and actually meeting needs. In addition, such disaggregated data has the potential to support more focused advocacy and strengthen funding – for example, better understanding the needs of children with disabilities.

Similarly, PA can have multiple purposes. When combined with other quantitative research methods and extensive subject matter expertise, predictive models hold the potential to drive complex analysis, especially when the goal is to understand the interaction of different factors in conjunction with each other (Ruigh et al., 2019). Relatedly, predictive models may offer the possibility to expand existing research methods and generate new themes for analysis, revealing novel factors that have not been accessible to researchers in the past. In a paper exploring the positive roles that computing may have on social issues, Abebe et al. (2020) note that innovative technological interventions can "offer us a tractable focus through which to notice anew, and bring renewed attention to, old problems" (Ibid). However, it is important to note that relevant uptakes of ML-driven methods aiming to generate new findings do not always fit the description of PA as defined by this paper.

5.5 Identifying trends and patterns across multiple data streams

PA can identify patterns in large amounts of data in a far more efficient way than humans can. In fact, ML-driven models excel at making sense of, and finding patterns in, large amounts and diverse types of data. It can process data quickly and find similarities that might otherwise be missed by the human eye. ML techniques are also able to adequately combine and analyse complex datasets from different sources, such as text, satellite imaging and geolocation – a process that would be difficult or near impossible to perform with traditional statistical methods.

For example, most of the use cases in humanitarian settings deploying PA models to predict the likely trajectory and features of emergencies – including UNHCR's Project Jetson, the Foresight model from the Danish Refugee Council, or Save the Children's Migration and Displacement Initiative (as described in Section 4) – utilize a wide range of large data sources, including mobile phone records; social media data; satellite images; meteorological data; financial transactions; public data; etc.

However, it is worth noting that the effectiveness of PA depends on the accuracy and quality of the diverse data streams that feed into it. The high false negative and false positive rates that stem from poorly designed and trained algorithms, inaccurate or incomplete data sources, or population disparities have reportedly resulted in less reliable risk models, unnecessary interventions, and inappropriate decisions in medical, educational, social welfare, and child protective settings. Furthermore, ML-driven models aimed at identifying patterns in large datasets do not always have a predictive focus, but rather aim at explaining existing problems instead of predicting future potential outcomes.

5.6 Predictive analytics as a tool for accountability

Finally, it is important to note that many of the identified use cases of PA in this paper are implemented and designed by those who hold the most structural power (e.g., law enforcement agencies, child protection services, education institutions), relying on the data collected about those who hold far less structural power (e.g., adolescents stopped by the police, families scrutinized by welfare institutions, health patients undergoing a treatment). Former UN Special Rapporteur Philip Alston (2019) notes that through the use of digital technologies such as predictive analytics in the public sector, "citizens become ever more visible to their governments, but not the other way around".

Civil rights practitioner interviewees observed that PA could be used to, for example, predict which service providers will be more likely to deliver better or worse results. Alston (2019) echoes this sentiment, recommending that governments move away from "obsessing about fraud, cost savings, sanctions, and market-driven definitions of efficiency" and instead focus, for instance, on using technology to "devise new ways of caring for those who have been left behind".

Building on the work of Krishnan (2021), who emphasizes the importance of de-colonial humanitarian digital governance, identifying ways to hold child-focused institutions accountable through ML-driven methods can help predict future instances of failure in the provision of relevant services, or to monitor and forecast the performance of relevant agencies. What this looks like in practice is yet to be defined but could include engaging with affected populations (in this case children and their families) to collect the data that matters most to them, under terms and conditions that they define, or co-designing PA implementations with affected children and families.

5.7 Targeted profiling for positive behaviour change in children

Targeted profiling techniques facilitated by PA can be used to support positive behavioural change. This is a fundamental practice in public health social marketing, where it is used for a wide range of positive social interventions around reducing risk-taking behaviours and promoting positive health behaviours (Jepson et al., 2010). A systematic review of behavioural interventions delivered to children (aged 5–11 years) revealed that they can be effective in prevention, treatment, and management of a range of health conditions, such as obesity, diabetes, chronic pain, asthma, and emotional problems (Brigden et al., 2019).

5.8 Making the most of available technologies to improve children's services and support

Given the current environment of austerity and limited resources available to fund children's services and support, the argument of integrating technology that may support more resource-efficient service provision, or better accountability outcomes for children, could be framed as an ethical imperative. However, there is a dearth of literature or existing work to back up this framing, particularly in the case of emerging technologies, where the potential benefits have yet to be proven and are largely speculative, or where they have been implemented in contexts very different to the intended end context.

5.9 Key points: Potential benefits of PA

- PA offers the possibility to allocate resources in more efficient ways; to design more tailored interventions; to allow for better preparedness for environmental shocks; to help reduce the burden on overburdened systems; to enable more accurate predictions about future capacity constraints; to generate new insights; to identify patterns in complex datasets; and to allow for stronger accountability mechanisms.
- The evidence in relation to cost/benefit analysis and with respect to robust evaluations is not comprehensive. A greater body of rigorous evidence will be required to substantiate both the rollout and upscaling of these programmes.
- PA models require specialist expertise in their design and frequent iterations to ensure the model remains relevant and accurate over time and across populations. More effective resource allocation may only work when such expertise is available for resource-constrained institutions.

6. POTENTIAL CHILD-SPECIFIC RISKS OF USING PA

To understand the risks involved in using predictive analytics on children, we draw upon evidence of risks arising from three aspects of PA: the actual implementation of PA; those that arise as a result of the data used to train a PA model; and those that arise from the design of a PA model.

6.1 Data-related risks

6.1.1 Children's autonomy in consent processes

As children develop, their relationships and levels of autonomy change. This can manifest in, for example, how capable an adolescent might be of granting meaningful, informed consent over how their data is used or shared, in comparison to how or whether a child may require a caregiver to make these decisions on their behalf. Under the GDPR, the lowest age at which a child can provide consent for the processing of their personal data is 13 years of age; for children under this age, consent from whoever holds parental responsibility for the child is necessary (European Union: European Commission 2016).

As a child's capacity to exercise agency and thus grant consent for the use of their data evolves, existing frameworks and processes for gathering consent are frequently insufficient. For example, consent is typically granted at a particular point in time rather than as an ongoing process, despite the fact that the data may be used for different purposes over time, which the child or their guardian may not have provided consent for if asked. This raises the risk that a child's evolving capacity, as identified in Article 5 of the CRC, will not be respected.

Additionally, there are often unknown future applications of data (Fossheim and Ingierd, 2015; Berman and Albright, 2017) that are impossible to know at the point of granting or requesting consent. As highlighted in the RD4C, "the consent of data subjects and their caregivers is important, but obtaining meaningful consent is a complex and, at times, impossible undertaking when dealing with children, especially in fragile humanitarian settings" (2019).

6.1.2 Inconsistencies and low data quality

Data about individuals is shaped by the context in which it is gathered, which has an impact on how and what a PA model seeking to generate individual-level outcomes might learn from it. For example, data gathered on individuals within child welfare or child health tends to be collected by professionals within the system, such as social workers or health professionals. "Inaccuracies, contested information, and consequentially systemically errors can enter into a dataset at multiple points along the extraction, collection, and consolidation workflow" (Leslie et al., 2020), and ensuring that a dataset is accurate, with relevant and recent information, is "a tall order to fill" (Ibid). There is frequently a significant lack of high quality, large scale, disaggregated data in the public, humanitarian (UNHCR, 2018) and development sectors, further exacerbated by a deficit in infrastructure (World Wide Web Foundation, 2017). This has implications for the accuracy of outcomes, as "models are only as good as the data on which they are trained, tested, and validated" (Leslie et al., 2020).

6.1.3 Bias within children's datasets

At the very first step of data collection, there can be systemic inequities present in the training data, which itself has been shaped by human decisions and prejudices (demonstrated in, for example, how data is collected, or not collected) (Barocas and Selbst, 2016). In the case of child welfare services, the use of risk scoring and risk assessments by staff is relatively well established as a tool, regardless of PA implementation. This has created historical data that could, theoretically, form part of training data for PA models. However, as Munro (2019) notes, "the incompleteness of many of the datasets that are being used for predictive analytics is also a concern since child protection datasets, in particular, are known to be incomplete in a non-random way" – indicating that they reflect systemic societal inequities, rather than simply being incomplete across all demographics or groups.

6.1.4 Availability of disaggregated data

Disaggregating children's data from non-disaggregated datasets raises a difficult dilemma with regard to privacy and de-anonymization. On the one hand, age-disaggregated data is needed to improve the accuracy of PA for children. On the other hand, age-disaggregated data increases the chances of de-anonymization, because the data concerns a smaller subset within the larger dataset. Interviewee 3 noted serious ethical concerns with pulling children's data out of a large dataset due to the increased ease of identifying children from this smaller child-specific subset. If these datasets are combined, the risk significantly increases as the combined dataset can increase the risk of revealing sensitive attributes about individuals (Metcalf and Crawford, 2016).

Similar concerns arise with data that is disaggregated along different lines – for example, data about children with disabilities, which is an even smaller subset, but which pertains to a group with particular needs that must be taken into account but who may be easier to identify once data is disaggregated.

6.1.5 Security risks

Part of the solution to addressing problems arising from PA models trained on adult data is to collect and gather large amounts of children's data, which must therefore also be stored over time and likely accessed by multiple different parties. USAID (2018) notes that "people may fear (often justifiably) that any collection or use of personal data could link to a government surveillance system". This increase in personal data collection increases the risk of data sharing and data breaches and thus the risk of violating children's rights to privacy, as enshrined in Article 16 of the Convention on the Rights of the Child.

Data breaches (an issue with any use of technology that gathers large amounts of personal data) of children's personal data can lead to targeted cyberattacks, or even child identity theft (Brody, 2011). Amid the current environment of increased scrutiny on how the data of individuals is being managed and shared between public authorities and the private sector, international development and humanitarian sector institutions risk reputational damage if PA is found to have compromised personal data, regardless of intention or attempted mitigation efforts. This damage could be even more severe if the data in question concerns children, or if children are directly negatively impacted, and has the potential to affect both fundraising efforts and trust (on the part of both the public and affected communities). Technology and society researcher Interviewee 1 noted that as details on the collection and processing of sensitive data for PA by public authorities in the UK have become public through the use of Freedom of Information requests, public trust has already been damaged.

6.1.6 Identity and agency of children

As the bulk of data held on children is gathered by professionals (rather than submitted directly by children and their families), this may mean there are limited opportunities for guardians and children to either be aware of what data is held about them and what this data says, or to contest the data held if they do not agree with it – thus suggesting a risk to children's agency in shaping information about them. Civil rights practitioner interviewees all acknowledged the "information asymmetry" that exists in relation to people in contact with child welfare agencies, i.e., those generating the data that could be used for PA are frequently not aware of whether or how their data is being used.

As Berman and Albright (2017) note, the formation of children's digital identities can be shaped by both parents; corporate third parties, such as social media companies; and by other parties who may contribute to a child's digital identity. Chaudron et al. (2018) highlight that most children under the age of two have a digital footprint through their parents, though this study focused on children in developed countries, highlighting that because children's digital identities and online profiles begin to be formed at such a young age, they effectively have zero agency over how data about them is put online. The long-term consequences of such identity formation are unclear, particularly given the opaque practices of social media companies – for example, understanding whether or not social media companies begin to establish individual digital profiles of children this early.

6.2 Potential risks relating to implementation of PA

Many of these risks stem from one overarching issue: hyper-dependence upon predictive analytics – meaning that staff, or end-users, interacting with the PA system do not have adequate skills to interpret the outputs with the necessary critical assessment. Given the critical variability in the capacity of implementing staff and organizations, the following risks must be considered.

6.2.1 PA affecting behaviours of those impacted by the implementation

Selbst et al. (2019) note the existence of a "ripple effect trap" where there is "failure to understand how the insertion of technology into an existing social system changes the behaviours and embedded values of the pre-existing system". Within the realm of PA and children, this failure could affect both children and their families, and also practitioners and decision-makers who are implementing or using the results of the PA system.

As interviewees noted, behavioural impacts could manifest, for example, in contexts where families are identified for a particular child protection intervention through individual-focused PA, and – in cases of false positives or where the output might be considered punitive (such as children being put into foster care) – subsequently losing trust in the intervening institution as a result. This could lead to less responsiveness to, or engagement with, these services in future. One can also reasonably hypothesize that this could have a network effect, where those within the close social network of affected families hear stories of inaccurate or inappropriate PA-driven interventions and change their perception of and behaviour, related to the PA-implementing institution. This concern is also consistent with false positives and errors that may arise from more traditional means of case management; however, the key differential here is that those using PA may place undue amounts of faith in the efficacy or effectiveness of results.

This is a particularly high risk for child welfare services – if families are targeted or red-flagged as, for example, having an increased risk of child maltreatment before anything has even happened. If this approach were to be implemented it is reasonable to expect that families and potentially entire communities would lose trust in these services, particularly if the reason for being red-flagged could not be clearly and concisely explained to them. A consequence of this may be that they then refuse to engage with key services, thus potentially exacerbating inequities and denying children access to as many potential futures as possible, in direct contravention to UNCRC. Berman and Albright (2017) have highlighted the risks of largely unpredictable applications and usage of data, and lack of data governance mechanisms resulting in "children [being] likely to suffer the consequences hardest and longest" (ibid). While currently, predictive analytics is primarily used as a decision support tool – that is, with a human in the loop – there are significant real-world examples of negative impacts from predictive analytics and individual-based scoring where humans are involved but rely too heavily upon the PA outputs to determine their course of action (Eubanks, 2017). This is explored in more detail below.

6.2.2 PA undermining practitioner decision-making and agency

Evidence shows that the use of PA to support human decision-making may have unintended negative consequences for how social workers and other child-focused practitioners conduct their work and make decisions. Munro (2019) notes that the use of a risk assessment tool such as a predictive algorithm might erode social workers' agency, critical thinking, and "professional judgment skills, including the ability to define key concepts such as 'risk' or 'abuse' and to recognize that they are socially constructed and contested entities". Eubanks' (2017) research on PA implemented in child welfare services in Allegheny County shows that social workers often feel influenced to adjust their own estimates of risk to reflect the assessments made by the predictive models being used (see the case study below for more detail on this project). Similarly, Green and Chen (2020) highlight that algorithmic risk assessments "can systematically alter [human] decision-making processes" and note that their research conducted in government contexts highlights that "human prediction accuracy with algorithms does not necessarily improve human decisions".

This is particularly concerning in cases where PA (particularly individual-based PA) may be implemented in such a way that neither practitioners nor those affected by PA understand the limitations of the PA system, or where hyper-dependence on a PA system is established – that is, without adequate human oversight or ability to 'overrule' incorrect outputs.

Research by Zerilli et al. (2019) shows that the more sophisticated and reliable a tool is, the more difficult it becomes for the human responsible for supervising it to act against it. Jacobsen (2015) documents such a case in Pakistan, finding that humanitarian workers believed and acted upon the output data from biometric identification technology, even when the data contrasted directly with what people were saying to them face-to-face. In terms of using PA for children this is a particular concern, as children and their guardians may not have the capacity to directly demand accountability, nor themselves identify when 'incorrect' decisions have been made that affect their lives.

6.2.3 Staff capacity and willingness to learn how to effectively integrate PA

When it comes to the assertion that PA could potentially reduce the burden on overburdened systems, literature highlights the possibility that overworked staff may be reluctant, or simply may not have time, to learn how to properly integrate PA into their work. For the effective use of ML in children's social care, Leslie et al. 2020) recommend integration of data collection and analysis in professional development and training, with an emphasis on accuracy and impartiality.

Additionally, PA that generates a high number of false positives (e.g., falsely recommending low-risk situations for intervention) could "place erroneous burdens" on already overworked staff (Clayton and Sanders, 2020). Echoing this point, tech and society researcher Interviewee 1 noted that in the course of their research they had heard informally, in multiple countries, that child welfare social workers had "quietly decided" not to use the PA systems available because they did not find them valuable for their work.

In addition, the ways in which practitioners integrate the outputs of PA will vary depending on factors such as the time they have available; workload pressures; the guidance they have received; decision fatigue; and their range of beliefs or behaviours (Leslie et al., 2020). Another risk is that staff may have excessive trust in the technology, alongside general difficulty in interpreting predictions, which can result in inaccurate decisions being made based on the PA outputs (USAID, 2021).

Case study: *Allegheny County Family Screening Tool (AFST)*

In 2016, Allegheny County in the US state of Pennsylvania implemented its Family Screening Tool (AFST), a predictive risk modelling system that aimed to improve child welfare call screening decisions. Specifically, the tool was designed to support the county's child abuse hotline screening staff in determining which reports of maltreatment pertaining to children who were at high risk of future abuse and neglect and their involvement with child protective services, and/or critical incidents (i.e., near-fatalities or fatalities) (Dare and Gambrill, 2017). In this sense, the tool was designed to support hotline staff in deciding whether or not a call warranted a visit and "whether there is a justification for screening the child in and carrying out an investigation" (Ibid.).

The model was developed using administrative data already routinely collected by different agencies and service providers between April 2010 and April 2014, including data from "drug and alcohol services ... mental health services, the county housing authority, the county jail, the state's Department of Public Welfare, Medicaid, and the Pittsburgh public schools" (Eubanks, 2017). For each call made to protective services ('referrals'), developers were able to construct data on the family's history, such as the number of referrals within intervals of 90, 180, 365, and up to 548 days. Using this history, the model would predict the likelihood of events in the next two years.

AFST is a supervised learning model based on non-linear regression methods. After review, the methodology and modelling technique was adjusted to improve accuracy (Vaithianathan et al., 2019). The predictor variables considered in the model as well as model rationale, ethics review, and impact evaluation results have been made readily available to the public, though weights given to the predictor variables are available only upon request (Goldhaber-Fiebert, 2019).

AFST's own impact evaluation concluded that the tool's introduction led to moderate improvements in the accuracy of families being rightfully selected for further investigation ('screened in'), accompanied by a small decrease in accuracy in families not selected ('screened out') (Ibid.).

Beyond technical accuracy, however, questions have been raised concerning the model's rationale as well as the data upon which it was built. Though the internal evaluation found "no large or consistent differences across race/ethnic or age-specific subgroups in these outcomes" (Goldhaber-Fiebert et al., 2019), research by Eubanks (2017) on the project showed that the data processed in the model overrepresented the poor, putting poor and working-class families at a higher risk of being 'screened in' for investigation by the system.

The above case study demonstrates the importance of external evaluations that examine not just the technical accuracy of a tool, but also consider potential unintended social impacts, particularly along the lines of existing social inequities.

6.2.4 Potential waste of resources

Though improved efficiency is an often touted benefit of predictive analytics, e.g., in juvenile justice where PA is deployed to advance the work of law enforcement, there is currently a distinct lack of evidence, suggesting that this is currently more of an aspiration than a proven metric. In cataloguing PA models designed with humanitarian action in mind, Hernandez and Roberts (2020) argue that while close to half of the analyzed initiatives claimed that PA would "improve efficiency by saving time or money", they were unable to validate such assertions, given the lack of documented evidence. Such limited evidence sits in sharp contrast to the significant claims relating to PA, particularly notable in the humanitarian sector, which Interviewee 2 described as "the wild west of data science".

If the decision to invest resources is taken without appropriate evidence or robust assessments, this raises the potential for a number of interlinked risks; notably, wasted resources and potentially, a lack of trust and reputation for the involved organization or agency. The question of resources is also crucial to consider in light of the many mitigation measures necessary to ensure that PA has positive benefits for children. In turn, these mitigation measures (as outlined in Section 7) may require significant resources which again would need to be factored into assessments of efficiency and resource gains.

6.2.5 Conflicting priorities between users and private sector actors and suppliers of technologies and PA services

Due to capacity issues, as well as the specialist expertise required to implement complex technical projects, the private sector is frequently involved in implementing predictive analytics projects within the public, development, and humanitarian sectors. This can raise questions about data sharing, access, consent, and the value derived from derivative models. An example of the latter is the use of PA models trained on data from high income countries that are then used in low income contexts. This may compromise the validity and accuracy of the model given that in many instances, these would not reflect the complexity and socio, cultural, legal, and/or ethical norms of the new environments.

Additionally, the terms under which public sector and humanitarian and development agencies enter into partnerships with the private sector are typically not shared, which means that children and their caregivers are left without access to knowledge or agency in relation to how their data is being used, stored, and retained, and with whom it is shared (Berman and Albright, 2017). Lack of this fundamental information can infringe on children and their guardians' rights to form views on how the use of predictive analytics affects them, meaning that demonstrating respect for their views – a right enshrined in Article 12 of the CRC – may be difficult, if not impossible, to follow.

Notably, associations in both public and private sectors frequently have large networks of third-party organizations with whom they share data. In these instances, public, development, or humanitarian agencies utilizing their models may need to accept or explicitly consent to this data being broadly shared with these third-party organizations as part of the terms of service.

6.2.6 Privacy

Many of the issues identified in this section have serious implications for the ways in which children whose data is used in PA models – or children who are affected by PA outcomes – can exercise their right to privacy. With regard to individual-based PA, Leslie et al. (2020) identify that the generation of false positives may "interfere with the right to privacy" of families affected. Livingstone and Stoilova (2018) note that "privacy is vital for child development" and identify three types of contexts where privacy is important – interpersonal privacy, institutional privacy, and commercial privacy. Each of these privacy risks are raised in the use of PA, and they sit at the intersection of the issues explored in greater detail in this section – such as data ownership and data sharing, identity and autonomy, and consent. More broadly, these risks affect what Wang (2020) calls "personhood" – our "agency to determine one's own life decisions and outcomes". This framing of personhood – instead of just privacy or data rights

– may offer a useful frame of analysis to ensure the intersecting issues of privacy, ownership, agency, autonomy, and identity, which are so crucial to child development, are all factored in.

6.2.7 Limiting children's voices

An increased reliance upon PA based on datasets for decision-making could lead to a reduction in data collection through direct engagement with children, leading to the potential "silencing [of] the voice of the child" (Berman and Albright, 2017). This is particularly concerning if children are not given the chance to contribute directly to data held about them, or if they remain unaware of how data about them (generated directly or indirectly) is affecting their lives. As covered previously, Article 12 of the CRC establishes children's right to have their views heard in matters that affect them (CRC, 1989). In this sense, over-reliance on these systems as opposed to direct engagement might interfere with the realization of this right.

6.2.8 Preferences and personhood

The preferences of children and youth change over time in fluid ways, but the nature of PA that uses historical data (i.e., data on children's past behaviour and past preferences) in order to predict their future behaviour sits in contrast to this fluidity. As UNICEF (2018) states, adolescence is "a critical period for individual identity development", where youth should be free to explore and express themselves without fear of negative consequences.

The UNCRC enshrines the right to freedom of expression for children, but research on online profiling – defined as "any technique that automatically processes data related to individuals in order to develop predictive knowledge for the purpose of constructing profiles which form a basis for future decision-making" (Büchi et al., 2019), and as such potentially forming part of PA as understood in this paper – suggests it may have a chilling effect on both online and offline behaviour (Marder et al., 2016). However, any given child's capacity to understand, consent to, or manage data and privacy, and their level of agency, are influenced at a much more granular level by factors such as the family or individual child attributes like knowledge, personal resilience, opportunities, and resources. These individual and familial factors may be more pronounced in low resource settings with ample evidence of lack of "awareness … time and resources, or the understanding to protect and empower their children online" (Livingstone et al., 2016).

6.2.9 Nudging and profiling

Under the umbrella of targeted profiling, when predictive analytics is used to 'nudge' children's behaviour online, it may covertly contribute to commercial profiling of a child and can affect their sense of agency and identity. UNICEF (2017) notes that for businesses, "children can be important targets as sources of data because they influence their friends' and families' consumer decisions". Further, children may be consumers themselves and/or potential customers in the future. The long-term impacts of the development of the child as a consumer have yet to be observed.

Such targeted advertising can be used to manipulate behaviour even outside of explicitly commercial settings, particularly as advertisers "become better at blurring the boundaries between commercial and noncommercial content" (Gasser et al., 2010). Evidence shows that profiling and marketing to children is lucrative for advertisers, but "inherently manipulative of an audience who [may] not have the full mental capacity to understand or resist the techniques used to sell to them" (Global Action Plan, 2020) – thus putting a child's agency at risk. Byrne et al. (2021) suggest that "legal frameworks generally overlook the risks for children of group data profiling", noting that profiling activities typically place children in groups, which generates data that can reveal protected attributes of children, thus risking violating their right to privacy.

6.2.10 The creation of filter bubbles

When PA is used to profile children, algorithms that use their past online behaviour to recommend content can limit the information they are exposed to, and thus affect their right "to reliable information from a variety of sources" as noted in Article 17 of the UNCRC. This phenomenon is known as keeping people in "filter bubbles" (Pariser, 2011) – that is, where information based on past behaviour is used to show users more of the same information, thus isolating them from diverse viewpoints. This is particularly relevant given that online profiling and targeting can give the illusion of agency to children because the creation of filter bubbles is invisible, and children and their carers may not realize they are only being shown limited information. This risks children developing "new expectations of... informational environments" (LaFrance, 2017) in ways that could affect their independence, autonomy, and creativity and inadvertently prevent them from engaging with new or alternate opinions, landscapes, or interests.

6.2.11 Directing attention and resources away from deeper structural problems

As with almost any technology that seeks to address complex social or political problems, PA systems run the risk of directing attention away from deeper structural problems, towards attempts to 'solve' more surface-level issues, described by Morozov (2013) as "solutionism" – an approach that can shift responsibility to individuals rather than engaging in more structural reforms. For Leslie et al. (2020), ML and PA systems are liable to "unethically perpetuate the dynamics of domination and inequity that underlie those very same problems". This risk is particularly specific to the way in which a PA approach is designed and rolled out, however.

Many of the child-specific issues that PA seeks to address are rooted in significant structural problems. This raises serious concerns about what kinds of problems should be prioritized with any kind of technical approach, including ML and PA systems. It also highlights the issue of the need for explicit reflection at the outset that PA deployments could, in the long run, lead to the continued existence of structural systems that have negative impacts on children.

6.3 Model-related risks

6.3.1 Entrenching bias and discrimination

Although bias is often framed solely in relation to training data (Hao, 2019), there are many other stages at which bias can creep in. As Glaberson (2019) outlines, there are multiple "decision points" within the design and implementation of a machine learning model itself at which human prejudices and bias can affect accuracy and efficacy. Indeed, Satell and Sutton (2019) suggest that it is "not realistic" to believe that bias can be eliminated entirely from ML systems given the diverse ways in which bias can enter a system, even with mitigation approaches and strategies.

As predictive analytics inherently "makes accurate out-of-sample predictions by replicating the social and cultural patterns of the past" (Leslie et al., 2020), past value systems, assumptions, and social dynamics embedded within historical data are, by necessity, replicated in forward-facing predictions. However, value systems and culture evolve over time, affecting how children develop, how they act, and how they live (Chan, 2004); but these changes are not reflected in a PA model that has been trained on data from the past. Subsequently, the values embedded in historical data are reflected in the outputs of PA, which may then become part of future datasets used for PA, thus becoming further entrenched. Leslie et al. (2020) note that this can "lead to an amplification of discriminatory configurations".

These issues have been reported in the application of PA models in educational settings. As Ekowo and Palmer (2016) note, PA that relies too heavily on demographic data and ignores population disparities and historical disadvantages will likely entrench disparities in college achievement among certain groups, favouring middle class white students over students from lower income families, especially children of colour. Early-alert systems can, for instance, disproportionately flag low-income students of colour for poor achievement, and thus "steer them away from more challenging and/or economically lucrative majors" by sending the message that "they do not have what it takes" (Ibid.). This phenomenon is described by Vannier Ducasse (2020) as the spread of risk models "set to mistakenly equate socio-economic disadvantage with risks, thus threatening to automatize the discrimination and alienation of the poorest sections of the population".

There is also a level of arbitrariness in how ML models 'learn' from training data. Black and Frederikson (2021) demonstrated this with what they named "leave-one-out unfairness", whereby "unfairness is determined by whether a model's prediction for an individual will change due to the inclusion or removal of a single other person in the model's training data". Although this analysis technique is still emerging and has not yet been tested on PA models used in the humanitarian or development sector, it presents the possibility that the inclusion or removal of a single person can have an outsized impact on generated outcomes. An illustrative example showed that the exclusion of a single person affected 2 per cent of predictions from a model trained to predict creditworthiness, trained on a subset of US census data.

At the time of writing in 2021, the issue of how to fairly address outliers in a way that does not exacerbate discrimination or bias is currently under debate within the machine learning and ethics community, with no established best practice as yet (ODD, 2021).

Case study: *Plataforma Tecnológica de Intervención Social (Technological Platform for Social Intervention) in Argentina*

In 2017, the government of the municipality of Salta, Argentina, partnered with Microsoft to develop a model that might help predict the risk of teenage pregnancy among low-income girls and women (Freuler and Iglesias, 2018; Microsoft, 2018).

The project built on data from household surveys conducted in low income neighbourhoods in previous years by Salta's Ministry of Early Childhood (Freuler and Iglesias, 2018). The surveys collected data on age, ethnicity, nationality, disability status, living conditions, education levels, and more (Sternik, 2018). Of the 296,612 people surveyed, 12,692 were female adolescents between the ages of 10 and 19. Microsoft relied on this particular subset of the data for its prediction algorithms (Freuler and Iglesias, 2018).

The first iterations of the model were developed through an iterative process and made public on GitHub by the Microsoft developers (Ibid.), but subsequent versions were developed privately, behind closed doors. Though, according to the Ministry, the goal behind the model was to "check whether there are members of the family who should receive some type of social assistance and are not receiving it", there are no indications as to what kind of services, if any, these girls and women received.

The project received criticism from a number of angles. Freuler and Iglesias (2018) note that though the inputs (and to a certain extent the outputs) of the model may have been accessible, the algorithm itself was a 'black box' model, making it impossible to explain the relationship between the model and its outputs and limiting the ability of affected populations to learn how the system arrived at its results.

In 2018, the Applied Artificial Intelligence Lab (LIAA in Spanish) published a technical analysis to highlight the shortcomings of the model (LIAA, 2018). The analysis found that the effectiveness of the model was overstated, and the database also oversampled low-income neighborhoods, thus entrenching bias and stigmatization. It concluded that the model was not sufficiently positioned to answer the question at hand, especially since "the conditions which led to a pregnancy in the past will not necessarily lead to pregnancies in the future inasmuch as the other variables affecting the outcome do not remain fixed" (Freuler and Iglesias, 2018).

Civil society organizations also expressed concern, In an open letter, Argentina's Observatory on Violence Against Women pointed out that teenage pregnancy is a highly complex issue caused by deeper issues of structural inequality, such as lack of access to information, poverty, systemic sexism, and high levels of gender-based violence. As such, and echoing the LIAA's findings, the Observatory argued that the model was inadequate for the issue at hand (Ceballos et al., 2018).

Despite the criticism, however, similar models have been proposed in three other provinces in Argentina, in Colombia (Freuler and Iglesias, 2018), as well as in Brazil – a country with one of the largest social welfare databases in the world (Peña and Varon, 2021). These projects point to an increased interest in the region for such initiatives.

6.3.2 Networked effects of PA on children

By design, predictive analytics uses inferences drawn from children who share particular traits, to generate outcomes that may affect other children. But in comparison with adults, children hold relatively little agency to determine their social network and environment – their community, socioeconomic status, and location, for example, are all determined by adults in their lives. As individual-focused PA approaches often include demographic data to train the ML model, this means that a child's social network and demographic are likely to heavily shape recommendations generated. Otherwise put, the behaviour of other children is likely to shape outputs for specific individual children because of these networked effects.

This means that any score or individualized recommendation that comes from an individual-focused PA system will be based not on that specific child's behaviour but rather, on the behaviour of other children with whom they share particular demographic traits (though specifically which traits can be difficult, if not impossible, to know, depending on the level of interpretability of the machine learning model).

A similar effect can be seen through the common practice of using familial or household data to make inferences about children. For instance, in their article on predictive analytics and foster care, Russell and MacGill (2015) use cultural orientation and socioeconomic data to explain geographical differences in foster care rates. Alongside having little agency over their social or physical network, children also have very little input or control over factors like their household income or parent's employment status. This leads to "networked privacy harms", whereby "users are simultaneously held liable for their own behaviour and the actions of those in their networks" (Madden et al., 2017), an approach that has been demonstrated to have particularly negative impacts on poorer people.

As such, PA systems are more likely to produce accurate outcomes for children who fit more neatly into a certain demographic – in other words, who possess more of the traits that define that particular demographic – than those who sit at the margins, or those who might be considered as 'outliers' within a dataset. This may contribute to further marginalizing these children.

This introduces a number of problematic assumptions. Specifically, assuming that a child will behave in a certain way based on the behaviour of others (be that their peers, or those who come from similar family socioeconomic backgrounds) denies a child's freedom of choice and self-determination. It may also contribute towards solidifying societal biases, if certain options are made available or denied to children as a result of a PA system.

Finally, depending on the way in which recommendations from an individual-focused PA system are used, these could potentially limit or affect their long-term life prospects. This approach –whereby factors outside a child's control determine recommendations that affect their future –goes against the UN Convention on the Rights of the Child, which "enshrines the right of children to self-determination".

6.3.3 Inaccurate outputs

The accuracy of currently deployed PA models raises concerns in almost all analyzed areas, particularly when predictive models have been focused on individuals as opposed to populations. The high false negative and false positive rates that stem from poor algorithmic models can result, and have resulted, in unnecessary interventions and harmful decisions in clinical settings (Van Calster et al., 2019), educational programmes (Ekowo and Palmer, 2016), and, most notably, within child protective services (Clayton and Sanders, 2020). Furthermore, the predictive performance of PA models changes significantly over time and across populations, illuminating the importance of context-specific and time-sensitive models, requiring continuous iterations and updates to the underlying algorithms (Parikh et al., 2019).

Studies in the UK showed that predictive analytics in use in children's social care failed to identify four out of every five children at risk; and where the models flagged a child as being at risk, they were incorrect six out of ten times (Clayton and Sanders, 2020). Similarly, a large global collaboration between 160 teams assessing the use of predictive analytics on life trajectories, which used 15 years of high-quality data, also suggested that there are "practical limits to the predictability of life outcomes in some settings", which is particularly illustrative given that the quality and depth of their data is considerably higher than one would expect to find in the humanitarian or development sector (Salganik et al., 2020).

New Zealand worked, from 2013 to 2015, on feasibility studies and a review of the ethical implications of deploying predictive analytics models in child protection risk assessments, but the programme was put on hold by the Ministry of Social Development in 2015 amid fears of unethical experimentation with children (Jones, 2015).

This limited evidence base assessing the accuracy and effectiveness of implemented PA models points to a need for caution, and claims in this area should therefore not, at this point in time, be taken at face value. Indeed, based on the limited evidence available, there is the considerable risk of investment in PA for use on children being a waste of investment and resources which, in the current resource-constrained environment, could be particularly damaging. While this could be mitigated in the mid- or long-term through increased staff investment to ensure better understanding of the limitations of PA, or through the incorporation of PA as triage to support better prioritization of resources (*see Recommendations*), this risk remains particularly valid as long as low capacity levels among involved users of PA are widespread.

6.3.4 Inaccuracies arising from PA trained on adult data used for children

Given the physical and neuropsychological differences between adults and children, a key measure to increase the accuracy of predictions made through PA is to ensure that the PA model is trained exclusively on children's data, separate and distinct from the broader population. However, this is often not the case due to a lack of disaggregated data, or a lack of high-quality data available about children.

For example, datasets on the health effects of air pollution commonly exist for adults, although the WHO (2018) has noted that children's health and survival are uniquely impacted by air pollution. Therefore, health interventions and policy recommendations drawing from ML models trained on these datasets will be erroneous for children if they do not take into consideration the physiological differences between children and adults (World Health Organization, 2007 and 2018). In the case of image classification in medical contexts, ML models are often trained with images of adults, which can lead to diagnoses and treatments that may not be appropriate for children (Medical Research Council, 2004). This can lead to what Selbst et al. (2019) name the "portability trap" – where "failure to understand how repurposing algorithmic solutions designed for one social context may be misleading, inaccurate, or otherwise do harm when applied to a different context". The portability trap could then lead to increased false negatives or false positives, and subsequent misallocation of resources, unnecessary work, missed interventions and opportunities, or inappropriate diagnoses or treatments.

6.3.5 Long term environmental impacts of energy-intensive machine learning

As outlined in Crawford (2021), machine learning and artificial intelligence have a significant environmental cost due to the level of computational power required, as well as the costs of building the requisite hardware and infrastructure. Strubell et al. (2019) estimated that training a single state-of-the-art deep neural network model produces carbon emissions almost equal to the lifetime emissions of five cars.

While neural networks and deep learning require far more computational power than most predictive analytics systems, this trend towards using increasing amounts of power and energy is worth considering as a risk, particularly noting that Amodei and Hernandez (2018) state that "the computer used in various large AI training models has been doubling every 3.4 months since 2012". Within these contexts, the increasing power and energy resources are increasing the ecological footprint of AI and need to be explicitly considered within the broader context of cost/benefit assessments, including recognizing that data quantities, the location of analytics, and the hardware adopted can all impact on energy use.

6.3.6 Accountability, transparency, and explainability

The introduction of PA models that have low levels of interpretability into decision-making processes that affect children's lives decreases the possibility for transparency, an important prerequisite for accountability. As put by Sloan and Warner (2018), "[a] decision procedure is transparent if the associated risks and benefits are readily ascertainable ... It is clear that, in general, predictive systems do not currently meet this requirement".

Without transparency in how a PA model works, children and their caregivers do not have the opportunity to understand why a decision is being made. Explainability of PA and its outputs is enshrined within data protection legislation – for example, the GDPR includes the "right to explanation" for algorithmically-driven tools, meaning that the implementer of algorithmic systems must be able to explain how particular outcomes that affect an individual were reached (Kaminski, 2019).

The introduction of PA into decision-making processes formerly made exclusively by humans complicates lines of accountability; if humans are responsible for inaccurate decisions, the lines of recourse are clear about who should be held responsible. However, when humans are 'in the loop' – that is, included as a key part of the machine learning system at play – research shows that accountability tends to land on the human regardless of whether or not they are at fault. Elish (2016) identifies the concept of a "moral crumple zone" to describe "how responsibility for an action may be misattributed to a human actor who had limited control over the behaviour of an automated or autonomous system". This points to a key need for clear lines of accountability and proactive planning for what happens if and when a machine learning system fails.

These considerations are particularly important when it comes to the use of PA for children, as children themselves may not have the capacity to directly demand accountability, nor be able to identify when 'incorrect' decisions have been made that affect their lives.

Case study: ***A-level grade allocation by algorithm in the United Kingdom.***

In the summer of 2020, following the cancellation of exams due to the Covid-19 pandemic, A-level grades in the UK were assigned using a machine learning algorithm developed by the Office of Qualifications and Examinations Regulation (Ofqual). It was subsequently reported that "poorer students were among those more likely to have received lower grades based on algorithmically assigned results" (Duncan et al., 2020), and "students from disadvantaged backgrounds were more likely to be downgraded" in comparison to the grades they were assigned by their teachers (Ibid.).

On the basis of their grades, students were either offered or denied places at universities, which were conditional upon obtaining certain grades. As a result of widespread critiques, Ofqual then withdrew the algorithmically-generated grades and students were left to receive teacher's predicted grades after all – but universities were unable to honour all conditional offers, leaving some young people without university places, despite having received the required predicted grades (Harkness, 2021).

Han (2020) critiqued the above implementation, noting that when decisions that fundamentally shape children's opportunities are left to automated systems "it's vital to clearly explain the results to allow students to contest or fix incorrect decisions". Moreover, determining a child's future based on historical grading patterns as opposed to their individual performance fails to recognize the "premise that education can open doors to opportunity, based on equally recognizing every child's merit and potential" (Ibid.).

In this sense, this particular application of PA went against children's right to self-determination, and also left those who were dissatisfied with their PA-generated results with little to no recourse as there was no redress mechanism set up in advance, and it was unclear who was ultimately responsible for ensuring equitable outcomes from the predictive model used.

6.4 Key points: Child-specific risks

- Risks from PA deployments arise from three angles: risks that arise from implementation, data-related risks, and risks that arise from the PA model design.
- Risks that arise from implementation include directing attention away from structural problems; affecting the behaviour of those impacted directly or indirectly by the implementation; undermining practitioner agency; wasting resources; risks to privacy; limiting children's voices; and risks to children's right to access information.
- Risks that arise from the use of data in PA include not respecting children's evolving capacity to grant consent for use of their data; low data quality and bias within datasets; lack of disaggregated data about children; and security risks.
- Risks that arise from the design of a PA model include entrenching bias and discrimination; networked effects on children; inaccurate outputs; and accountability and explainability risks.

7. RECOMMENDATIONS

Based on evidence gathered through this research, and drawing particularly upon the relevant frameworks, the following recommendations are designed to support the responsible implementation of predictive analytics for children.

Of note, many aspects of predictive analytics are, in some form or another, already in use in more passive ways throughout society – for example, our collective reliance upon PA-driven meteorological data and forecasts, or epidemiological modelling of COVID-19 trends. The recommendations here pertain primarily to explicit and deliberate integration or implementation of PA and are outlined for practical use in a checklist included in the Annex.

7.1 Assessing potential use cases

7.1.1 Consider whether the project should take place or not, and be open to the possibility of not moving forward with it

For the following recommendation to be meaningful, stakeholders must remain open to the possibility that "the best solution to a problem may not involve technology" (Selbst et al., 2019) – noting that some potential harms of PA are so severe that mitigation strategies will not be sufficient. Additionally, drawing on the taxonomy developed in Leslie et al. (2020), it is critical to question the legitimacy of the project from an external point of view to allow for space to examine the justification of the project head-on.

Key considerations to determine whether to proceed with the PA project include:

- **Establishing the effect of intended outcomes on lives of children:** Do the intended outcomes benefit the lives of children?
- **Understanding the project's role within social and political structures**: Is this use of PA allowing "avoidance rather than tackling of the structural reasons giving rise to the demand for this service?" (Leslie et al., 2020)
- **Understanding the appropriate legislative environment:** Is there a legal basis for using the data required for this project, and for using predictive analytics in the way it has been articulated here?
- **Understanding what 'fair' looks like:** Have impacted communities been involved in determining the most appropriate 'fairness benchmark', and can this benchmark be met in terms of fairness of the PA model?

As established in the GDPR, children should not be subject to automated decision-making processes where these have a "legal, or similarly significant effect" upon them (ICO, 2021).

7.1.2 Innovation, never experimentation, for children

The history of human experimentation – that is, the use of emerging technologies on vulnerable populations – has been well documented (Jacobsen, 2010). Imperative in any discussion of PA on children is this: no experimentation should take place on children or young people. In this context, the Willowbrook State School Experiments – based on the infamous and much-criticized hepatitis vaccine experimentation on children with intellectual disabilities – set the precedent for many of the current global ethics frameworks (Fansiwala, 2016). This case made clear that harming vulnerable children is never appropriate and established that "there is no right to risk an injury to one person for the benefit of others" (Beecher 1966, cited in Robinson and Unruh, 2008).

At the same time, innovation as an action in and of itself is valued within many sectors as a demonstration of forward thinking and a commitment towards progress – for example, within the humanitarian sector (Betts and Bloom, 2014). However, critical reflection is required as to the logics and potential impacts on children.

7.1.3 Conduct child risk and child impact assessments of PA

Within the literature, a common tool for assessing and understanding the potential risks of algorithmic decision-making tools, or as here, predictive analytics, are algorithmic impact or risk assessments (see Reisman et al., 2018; Government of Canada, n.d.). Each has its own specificities and may offer useful assessment frameworks for the implementation of PA on children. However, key to this analysis of PA on children is recognizing that the effectiveness of PA models is typically measured in terms of the level of accuracy of the generated predictions and that this is insufficient to assess risk comprehensively.

This approach to assessing effectiveness does not consider harms that lie outside of the technical system at play and are not measured by the technical system. Metcalf et al. (2021) point out that using algorithmic impact assessments introduces the risk of focusing on "organizationally understandable metrics that are nonetheless inappropriately distant from the harms experienced by people, and which fall short of building the relationships required for effective accountability".

As discussed within Section 6 on risks, this could mean that unintended impacts on existing behaviour; opportunities; development and physical health; and consequences, such as loss of trust in institutions and service providers, remain unseen if algorithmic impact assessments are relied upon as the primary way of assessing effectiveness of predictive analytics. Measuring impacts of the implementation of PA – particularly unintended ones – is particularly difficult, and changes may only be observed significantly after implementation. To ensure that risk or impact assessments are conducted with children's needs in mind, child rights specialists should be directly involved in unpacking and exploring potential risks.

Under existing data protection legislation such as the GDPR, a Data Protection Impact Assessment must be conducted if children's personal data will be used to offer an online service to a child. This process could be emulated to provide an appropriate framework for PA targeting children, going beyond data protection to understand broader impacts on children.

7.1.4 Avoid implementations that are safety-critical

As identified in Section 4, individual-focused PA with direct impacts on children has been used to determine the provision of potentially life-altering services to children – for example, within child welfare and child protection, juvenile justice, or education. In some of these cases, "system errors, unreliable performance and lurking biases may have life and death consequences" (Leslie et al., 2020). While risk mitigations (as outlined in the Annex and in recommendation 7.1.5) can go a long way towards alleviating such risks, they cannot ensure that these risks are totally and certainly avoided.

Conducting a risk assessment as a first stage in understanding whether a potential implementation of PA falls in the category of high-impact and safety-critical, and then being open to not moving forward with the implementation of PA based on the outcome of the risk assessment, is essential. In short, avoiding implementations of PA in situations where the risk assessment finds there is even a low likelihood of a negatively life-changing impact to a child is the clearest way of adhering to best-practice principles, thus remaining child-centric, purpose-driven, and professionally accountable (RD4C, 2019).

7.1.5 Only use individual-focused PA with appropriate risk mitigation measures

Given the higher potential and greater risks arising from individual-focused PA, as documented in Section 6, PA should only be implemented as a way of generating outputs about specific individuals with careful mitigation measures in place. These are outlined in further detail in the Checklist Annex, but in summary include:

- Conducting relevant impact assessments, such as a Human Rights Impact Assessment, Privacy Impact Assessment, and potentially (depending on the computational power required of the PA model) an Environmental Impact Assessment.
- Involvement of impacted communities in developing fairness benchmarks, in design, development, monitoring and evaluation of the model throughout its lifecycle, plus capacity development to do so (see: Citizens' Biometric Council as an example of meaningful community capacity strengthening and feedback), and openness to withdrawing implementation based on community feedback.
- Regular and meaningful evaluation with feedback cycles and iterations of the PA implementation within the organization, including assessing not only intended outputs and outcomes, but also attitudes of staff, unintended outcomes, and potential impacts on culture.
- Ensuring transparency and strong data rights as set out in contracts and agreements with relevant parties; for example, ensuring that partners do not use or share data for other purposes, nor share models derived from personal data.
- Clear and accessible redress and feedback mechanisms to ensure that complaints or problems are heard and can be reflected in iterations and changes to the implementation in a timely way.
- Implementation of individual-focused PA only when it significantly adds value as a complement to other robust sources of evidence and knowledge, and never in isolation.

7.1.6 Identify and clearly articulate why the PA model is needed, and how the potential benefits relate to improving children's lives

Articulating the objectives of a potential PA implementation, as well as how the intended or potential benefits relate to improving children's lives will ensure that any proposed implementation is purpose-driven and centres on children's needs (RD4C, 2019). This is necessary for PA implementations with both direct and indirect impacts on children.

7.1.7 Establish transparency and accountability standards for private sector agreements

Given that many implementations of PA take place in partnership with the private sector, codifying standards of transparency and reporting in advance (Leslie et al., 2020) is an important step in ensuring that data sharing practices can be communicated transparently to children and caregivers.

7.2 Data-related recommendations

7.2.1 Acknowledge tensions between minimizing data to protect children's rights, and designing predictive analytics systems that reflect children's contexts and particularities

As discussed earlier, to improve the accuracy of predictive analytics models, the quality of the data the model is trained on must be improved. Also, children's data is different to adult's data. Specifically, this points to a recommendation of training PA on child-specific data, to ensure the outputs of any model are appropriate for children. However, creating, storing, and using data on children poses serious potential risks, as discussed in Section 6.

Even when data on children is aggregated and non-personal in the form of group data, this data, "if exploited, can reveal characteristics, attributes and locations of children", thus posing immense challenges and amplified risks for children (Children's Data Manifesto, 2021).

Assessing this balance requires careful consideration of context, and balancing of the following issues:

- **Is sufficiently high quality and suitable children's data available to make the use of PA worthwhile?** In order for PA models to produce useful outcomes, accurate data relevant for the context at hand is "essential if data will be used to inform decision-making affecting children" (RD4C, 2019). Additionally, it is not appropriate to use population-level data for decision-making at an individual level (Leslie et al., 2020). Reviewing existing data regularly within a PA project helps to avoid missed uses of data (RD4C, 2019).
- **If sufficient data does not yet exist, are there infrastructures in place to responsibly collect this data from children?** As mentioned in Section 5.6, for ML models to produce accurate outcomes, large amounts of high quality data are required, meaning that collecting data that could be used to train a PA model likely requires long-term investment in data collection infrastructure rather than simply a one-off data collection exercise.
- **Does the very existence of this data put children at risk in the long or shorter term?** Noting that no system is 100 per cent secure against data breaches, leaks, or hacks, consider what would happen if this data were inadvertently accessed by malicious actors.
- **Is it possible to gain meaningful consent from individual children whose data will be processed or used in any way as part of this project?** This is particularly important in cases that require the processing or use of data directly gathered from specific individual children.

7.2.2 Data minimization of children's data

To avoid the risk of longer-term harms following a child, delete children's data once the intended purpose has been met. This approach can mitigate the risk, as outlined in RD4C, that "unforeseen data linkages and re-uses can emerge over time". In order to avoid potential risks that could arise from data breaches, practicing data minimization – that is, only collecting the data necessary for a particular purpose, and storing it for as long as necessary for that purpose, but no longer – is a common responsible data best practice. Both anonymized and non-anonymized data, when disaggregated, can re-identify and in turn, potentially harm children at risk (UNICEF and GovLab, 2019). "The collection and retention of data should be relevant, limited and adequate to what is necessary for achieving intended purposes (RD4C, 2019)".

Additionally, children's data should be shared with the minimum number of actors possible, and only used for specified purposes, which should be clearly outlined in any data sharing agreement at the beginning of the project.

7.2.3 Where data from individual children is being used, adopt consent policies that recognize children's development

As Berman et al. outline (2016), the competencies, analytical capacities, and agency of children change as they develop and recognizing this means that consent processes for younger children should be different to those designed for older adolescents. However, it is worth noting that obtaining meaningful consent from children – particularly at the scale that would likely be required if individual children's data is being used as part of a big dataset to train a PA model – is at times an impossible undertaking (RD4C, 2019). In those cases, where obtaining meaningful consent is practically impossible, it is worth seriously considering the feasibility of the project as a whole and the potential use of alternative data sources.

Under the GDPR, clear privacy notices should be written for children in a way that enables them to understand what happens to their personal data, and what rights they have (ICO, 2021). Further, it notes that consent is required from children above 16 years of age, and from their guardians for children below this age. Respecting an individual's "right to erasure", as enshrined in the GDPR, becomes particularly important in light of respecting children's evolving capacities, noting that children may have consented to particular uses of their data at a certain point but then want to remove that consent at a later date.

New forms of data ownership may offer potential future mitigation strategies with regard to the challenges raised by the use of personal data; notably, data stewardship via data trusts. The term 'data stewardship' itself is currently used in different ways (Gorr and Zawacki, 2017). The RD4C Principles, for example, call on organizations to establish 'data stewards' – "individuals or groups whose duties cut across departments and functions, and whose broad remit is to oversee responsibility and accountability in the way data is handled" – as key to operationalizing responsible data for children.

However, given that at the time of writing, data trusts are "possible models that need testing and piloting" (Ada Lovelace Institute, 2021) more work must be done on implementing data trusts before they can be recommended explicitly for children here. Other approaches, such as data stewardship conducted by a board or group of people representing the interests of children (for example, as part of the role of a data controller, or external to this role), could offer potential routes for exploration but again, there is not enough evidence for how implementation would happen for this to be an explicit recommendation at this time.

7.2.4 Ensure compliance with data protection regulations, but do not rely solely upon legislative regimes to ensure children's rights are protected

Currently, the GDPR offers some guidance on how to protect children's data (Information Commissioner's Office, n.d.), as does the Children's Online Privacy Protection Rule (e-CFR, accessed 2021), but both offer relatively little in terms of how predictive analytics may affect their rights.

Noting that data protection legislation has yet to keep up with emerging uses of data, those wishing to implement predictive analytics in a way that prioritizes and protects children's rights should not rely upon legislation (or lack thereof) to mandate how data could or should be used. In addition to a lack of data protection legislation, there is "virtually no specific acknowledgement in national AI policies of how AI affects children" (Byrne et al., 2021) – meaning that the processing or use of children's data specifically within predictive analytics models is also not adequately legislated. USAID (2018) notes that "the weakness of absence of personal data protection laws is a widespread problem that can create opportunities for malicious actors to surveille and manipulate with impunity", thus emphasizing the need for parallel strong governance structures to ensure the rights of children are protected in any implementation of PA.

7.2.5 Avoid implementations that require the use of datasets with historically embedded inequities

Depending on the sector and use case at hand, potential training data for PA may include data which reflects socio-historic patterns of over- or under-representativeness (Leslie et al., 2020). While there are mitigation approaches that can be taken, in cases where "historically embedded inequities" exist in datasets used to train PA models, Leslie et al. (2020) suggest that "where extant formations of societal discrimination are replicated in the selected sample, there is no clear-cut technical solution available to create optimally representative datasets". As such, in cases where datasets that reflect social inequities are the only option for training PA models, avoiding such use is the best way of ensuring that historically embedded inequities do not negatively impact the lives of children.

7.3 Design and implementation of predictive analytics models

7.3.1 Involve staff from diverse and multidisciplinary backgrounds from the very first stage, ensuring children's rights expertise and technical expertise

In order for PA models to centre children's rights, staff with expertise in advocating on behalf of children should be involved as key parts of any team considering using PA. As Abebe et al. (2020) note , "a holistic analysis of a sociotechnical system must draw from a wide range of disciplines in order to comprehensively identify the issues at stake". In order for risks to be proactively flagged and mitigated, meaningful collaboration between multiple parties needs to take place from the earliest possible point.

Similarly, any team considering using PA should involve technical experts with deep knowledge of the model being used, its weaknesses and strengths, and understanding of the data that has been used to train it. Collaboration between staff who hold such different areas of expertise will be essential to ensuring the PA model is designed and implemented in a way that respects both children's rights and also, technical context.

7.3.2 Consider options for reducing the environmental impact of any PA implementation

Policy recommendations outlined by Dobbe and Whittaker (2019) for mitigating the climate impact of artificial intelligence and machine learning suggest assessing the energy costs and increasing transparency on energy sources can support more climate-friendly implementations of AI. Researchers Schwartz et al. (2019) suggest evaluating ML models with energy efficiency in mind, alongside accuracy and related methods, as well as reporting financial and energy costs of "developing, training, and running models" to provide baselines for moving towards more energy-efficient methods. Lacoste et al. (2019) have produced an online tool to compute the carbon emissions of maintaining and running a machine learning model, which may provide a useful starting point towards assessing energy efficiency of any proposed PA implementation. Environmental impacts of any PA project could be considered in parallel with the privacy and ethical impacts.

7.3.3 Provide training and skill development for staff involved in any PA implementation

As a major identified risk is the misinterpretation and misuse of PA outputs, providing capacity development for any staff interacting with the PA system may be a first and important step to in mitigating this. Specifically, capacity development should highlight the limits of any PA outputs and provide information on how the outputs can be used to strengthen (not replace) any existing human-led decision-making processes.

This could also be framed within the broader picture of supporting staff to critically understand and engage with tech-driven solutions more generally, ensuring that they feel able to take advantage of the affordances of any new tech-driven project, while properly assessing the outputs. This could be a valuable investment in future-proofing any child rights organization, given the opportunities offered by technology to support and strengthen child rights.

7.3.4 Provide clear guidance to staff on how to assess, interpret, or use PA outputs as they affect children

Following from the above, and as outlined by Leslie et al. (2020), child social work practitioners may "focus on information readily available to them rather than looking at all available information and give more weight to memorable, vivid, emotion-arousing information" – meaning that the process by which the outputs of PA are presented and integrated within existing information systems may significantly change the way in which the outputs are interpreted. Additionally, staff whose primary area of expertise is in children's rights may not have a good enough understanding of the limitations of PA models, leading to inappropriate interpretations of PA outputs. Mitigating this requires ongoing training and skill development (*see Recommendation 7.3.3*) but could also in part be addressed with clear and specific guidance tailored for PA systems to flag, for example, areas where the model may be more prone to error, or ways in which outputs could be used to strengthen human decision-making.

7.3.5 Involve children and their communities in the design and validation of PA models wherever possible

To ensure that PA models are best meeting the needs and lived experiences of children, children should be involved in the design of models that affect them – for example, through participatory, collaborative workshops that bring their experience in. While the actual technical design of a model is unlikely to be contributed to by children directly, children, their communities, guardians and/or advocates could, for example, contribute to a deeper understanding of how to collect meaningful data to meet the objectives of a PA project, via participatory workshops. Such activities could also help establish what the 'reasonable expectations' of children and their representatives are with regard to their data and ensure that any model meets these expectations (Byrne et al., 2021).

7.3.6 Use models that are interpretable and explainable to children and/or their caregivers or representatives

Drawing upon Leslie et al.'s (2020) definition of interpretability as being intelligible to human reasoning, and explainability as conveyability of the logic behind its results, PA models used must be interpretable by children, with the use meeting a child's reasonable expectations (Byrne et al., 2021), and explainable to either older or more mature children, or their caregivers or guardians. Specifically, if the PA output "influences or automates a decision that impacts an individual, the model should be more interpretable" and not a 'black box' system (USAID, 2021). Depending on the exact type of model used, this may mean avoiding unsupervised learning, as these algorithms are among the least interpretable.

7.3.7 Ensure children's voices are valued alongside PA outputs by triangulating data

The outputs of predictive analytics – particularly in relation to individual-based PA – should be just one source of data triangulated with children's voices and lived experiences in order to make decisions that affect children either directly or indirectly. PA should be seen not as a solution, but as a tool (Centre for Humanitarian Data, 2019). As PA humanitarian worker Interviewee 4 suggested, an appropriate metaphor for the responsible integration of PA would be to consider PA outputs in the same way that "we think about weather forecasts. It is one source of information. You can also look outside to see the weather and then make your decision about getting your umbrella out."

7.3.8 Establish redress mechanisms that are accessible and understood by children and/or their parents, caregivers, or guardians

Children's interests must be represented within administrative processes (Byrne et al. 2021), which includes ensuring that any redress mechanism – where people who face harms as a result of a PA system can seek resolution and have their grievances addressed – can be used by children themselves, or their parents, caregivers, or guardians. This will likely require participatory design with children or their guardians/advocates involved, as well as directed communication and outreach to ensure children and their caregivers/advocates are aware of the redress mechanism and feel comfortable making use of it.

7.3.9 Consider the potential risks of PA to the agency and personhood of children

As articulated in Section 6, given that PA outcomes sit at the intersection of issues of data ownership; data sharing; identity; autonomy; self-determination; voice; privacy; and consent, an intersectional approach that considers all potential risks and benefits in relation to children's agency and personhood may offer a more helpful framing than considering each of those issues separately. Further, in order to adequately balance these rights, careful consideration of all potential and actual rights that may be impacted is necessary, acknowledging that rights are non-hierarchical, interdependent, and indivisible (UN General Assembly, 1948).

7.3.10 Keep humans involved and empowered in any PA system

In addition to keeping humans involved in a system – known as keeping a human in the loop – the people involved must "feel empowered to override results if they suspect an error" (USAID, 2021). Encouraging the people involved to "keep a healthy skepticism of model results" can also help in ensuring vigilance and avoiding unintended negative consequences (USAID, 2021). Feedback loops between human expertise and machine prediction must be "closely examined and refined" (Elish and Watkins, 2020) to avoid over-reliance, deskilling, or misinterpretation of results.

7.3.11 Choose a PA model that fits the objectives, available data, and intended context

In order to choose the type of PA model best suited to the project at hand, the availability; accuracy; data gaps; intended objective; and domain of application must all be considered in tandem to inform the choice of PA model used (Leslie et al., 2020). Each type of ML model brings with it benefits and risks, and human assumptions, biases, or errors can be baked in at different stages (see Leslie et al. [2020] for a detailed discussion on how these errors can appear in different ML models).

Assessing the effectiveness of a PA model with children's needs in mind may require a multidisciplinary team with child rights specialists involved throughout. If such expertise is not available in-house, engaging in a peer-review framework process – such as the approach devised and run by the Centre for Humanitarian Data (2021) – could be a valuable way of engaging diverse expertise, assuming that specialists could be brought in who hold expertise in children's rights. In this process, models would undergo technical and ethical reviews, alongside reviews of the implementation plan. These types of reviews are performed by volunteer specialists who are part of the Centre's roster (Centre for Humanitarian Data, n.d.) and could either be utilised or the process replicated.

7.3.12 Incorporate community opinions and consistently iterate based on ongoing feedback

The opinions and perceptions of potentially impacted communities must be incorporated from the earliest possible stage – for example, in assessing the fairness of potential models, and determining what criteria will be optimized in order to reach the intended outcome. Given the ongoing impact of PA, community feedback must be sought on an ongoing basis, both on the intended impacts and outputs, but also on broader perceptions, engagement with the institutions, and potential unintended socio-cultural impacts. In order for this feedback to be meaningfully used, feedback mechanisms must be built that allow for it to shape iterations of the model in an ongoing way.

7.3.13 Continue to build a knowledge base on the potential impacts – negative and positive – of PA for children

As this paper has identified, emerging uses of PA for children are either in early stages or thus far implemented as pilot initiatives. In order for lessons learned from these initiatives to contribute to strengthening a broader body of work and best practices on the use of PA for children, these must be documented, with evaluations conducted at regular intervals. This is necessary to ensure that future implementations of PA are evidence-based and rooted in the most up-to-date knowledge and research. Publishing these research evaluations will contribute to sector-wide shared learning on the topic.

8. FUTURE DIRECTIONS

Given the severe risks involved in using individual-based predictive analytics on children, as outlined here, the area of population-based predictive analytics seems to hold the most potential for implementation in a way that has benefits to children's lives. Areas of particular potential and high impact for the implementation of population-based PA include the field of public health – for example, addressing high impact initiatives particularly impacted by the Covid-19 pandemic, such as national immunization programmes for children – and climate, using population-based predictive analytics to anticipate climate shocks and emergencies.

Generally speaking, as predictive analytics is an area of growing interest, but noting that the potential risks to children are numerous, **pilot projects and notably individually-focused PA could and should be initiated 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. The exploration of what this looks like in practice offers an area of particular potential for future work on the issue.

There is currently a lack of evidence on how predictive analytics for children affects efficiency and resource allocation **therefore, gathering such evidence in the future will be essential to ensuring organizations can make informed decisions about whether or not predictive analytics would strengthen their work with children**. Another area of potential future work arises from the need to consider the impact of predictive analytics beyond accuracy or efficacy of outputs and instead, account for their impact as broader sociotechnical systems with potential unintended impacts within the broader social and political system, as well as the environment in which they are implemented.

Long-term investment towards staff training and capacity development is also essential if the implementation of PA is to be done in a way that properly reflects the strengths and limitations of the technology. This will require development of proper training processes and materials aimed at building their technical intuition, as well as adequate investment and resource allocation of staff time. In the short term, beginning to develop the technical intuition of staff could be an important step towards future-proofing institutions for the incorporation of emerging technologies.

On the level of PA models, more work could be done in terms of identifying or creating algorithms and predictive models and accompanying policies and systems that meet the recommendations identified here. None of the existing implementations of PA identified through this paper have established redress mechanisms accessible to children, nor were any identified that involved children in the development or design of the PA initiative itself. This introduces the opportunity of **child-centred innovation on predictive analytics** – using participatory and human-centred design methods to ensure that children's needs and, wherever possible and appropriate, voices contribute to the earliest stages of a PA project, and developing engagement methods that allow for children to understand how their lives may be impacted by PA.

In order for predictive analytics to be implemented in ways that benefit children's lives, **frameworks to guide this implementation in ethical and responsible ways** must be developed and operationalized on a policy and programmatic level. This requires **the allocation of appropriate organizational resources to ensure policies can be properly and appropriately implemented and embedded in operational systems and practices.**

The field of predictive analytics is under increased scrutiny, and new ethical issues and approaches are rapidly being uncovered. This is inclusive of both potential harms and unintended impacts, as well as mitigation strategies and new ways of leveraging predictive analytics for positive societal impact. As more robust tools for privacy, fairness and equity are developed, these may offer opportunities to address some of the risks outlined here – though it is unlikely, in the near future, that these risks can be entirely mitigated. To ensure that any potential applications of PA are cognisant of these emerging critiques and approaches, **a regular landscape scan (at least every two to three years) is necessary**, as best practices and mitigation strategies continue to be developed.

9. CONCLUSION

While predictive analytics brings many benefits, as highlighted in this paper, there is frequently limited evidence of its effectiveness in real world implementation – particularly in cases with high power asymmetry and already vulnerable communities. This raises concerns around deployment of this technology, particularly as it relates to children's rights.

These risks are likely, in many instances, to be significantly higher in the case of individual-focused predictive analytics used to generate outputs that may affect specific children's lives. Hence, the strong recommendation that extensive mitigation measures are put into place if and when individual outcome-based PA is considered or implemented. These risks are heightened by a general lack of understanding on the limitations of PA, and point also to a significant need for staff training and ongoing capacity development on the limitations and challenges of PA.

Above all, it is imperative that organizations and agencies considering PA that may impact children carefully consider and balance the potential benefits of PA with the severity of potential harms and risks. This approach, as well as the broader environment in which the sector finds itself, has guided the focus of this report, reflecting a position that exploration and the uptake of PA must be done thoughtfully and cautiously to ensure children's rights are respected and protected from design through to implementation and beyond.

10. REFERENCES

5Rights Foundation, 'But How Do They Know It Is a Child? Age assurance in the digital world', 5Rights Foundation, London, March 2021a, 5rightsfoundation.com/uploads/But_How_Do_They_Know_It_is_a_Child.pdf, accessed July 2021.

5Rights Foundation, 'Pathways: How digital design puts Cchildren at risk', 5Rights Foundation, London, July 2021b, www.5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf, accessed July 2021.

Abebe, R., et al., 'Roles for Computing in Social Change', paper presented at Conference on Fairness, Accountability and Transparency, Barcelona, Spain, July 2020, www.arxiv.org/pdf/1912.04883.pdf, accessed April 2021.

Ada Lovelace Institute, 'Exploring Legal Mechanisms for Data Stewardship', London, March 2021, www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/, accessed July 2021.

African Declaration on Internet Rights and Freedoms Coalition, 'Privacy and Personal Data Protection in Africa: A rights-based survey of legislation in eight countries', May 2021, www.apc.org/sites/default/files/PrivacyDataProtectionAfrica_CountryReports.pdf, accessed May 2021.

African Union, 'African Union Convention on Cyber Security and Personal Data Protection', 2014, au.int/sites/default/files/treaties/29560-treaty-0048__african_union_convention_on_cyber_security_and_personal_data_protection_e.pdf&sa=D&source=editors&ust=1622479035290000&usg=AOvVaw2xxgh8K-BA7KGBagaz_ueQx, accessed May 2021.

Alston, P., 'Report of the Special Rapporteur on Extreme Poverty and Human Rights', U.N. Human Rights Council, New York, 18 October 2019, digitallibrary.un.org/record/3883131, accessed April 2021.

Amodei, D., and D. Hernandez, 'AI and Compute', OpenAI blog post, 16 May 2018, www.openai.com/blog/ai-and-compute/, accessed July 2021.

Angwin, J., et al., 'Machine Bias', ProPublica, May 2016, www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, accessed April 2021.

Barocas, S., and A. D. Selbst, 'Big Data's Disparate Impact', *104 California Review 671*, posted August 2014, reviewed September 2016, www.papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899, accessed April 2021.

Benjamin, R., 'Race after Technology: Abolitionist tools for the new Jim Code', Cambridge: Polity Press, 2019.

Berman, G., et al., 'What We Know about Ethical Research Involving Children in Humanitarian Settings: An overview of principles, the literature and case studies', Innocenti Working Papers, UNICEF Office of Research, Florence, 2016, www.unicef-irc.org/publications/849-what-we-know-about-ethical-research-involving-children-in-humanitarian-settings-an.html, accessed April 2021.

Berman, G., and K. Albright, 'Children and the Data Cycle: Rights and ethics in a Big Data world', Innocenti Working Papers, UNICEF Office of Research, Florence, June 2017, www.unicef-irc.org/publications/907-children-and-the-data-cyclerights-and-ethics-in-a-big-data-world.html, accessed April 2021.

Betts, A., and L. Bloom, 'Humanitarian Innovation: The state of the art', Occasional Policy Paper, No. 009, OCHA Policy and Studies Series, United Nations Office for the Coordination of Humanitarian Affairs, New York, November 2014, www.unocha.org/sites/dms/Documents/OP9_Understanding%20Innovation_web.pdf.

Black, E., and M. Fredrikson, 'Leave-One-Out Unfairness', paper presented at ACM Virtual Conference on Fairness, Accountability, and Transparency, Canada, 3-10 March 2021, dl.acm.org/doi/pdf/10.1145/3442188.3445894, accessed April 2021.

Brigden, A., et al., 'How Are Behavioural Interventions Delivered to Children (5-11 Years Old): A systematic mapping review', *BMJ Paediatrics Open*, vol. 3, no. 1:e000543, 2019, doi.org/10.1136/bmjpo-2019-000543, accessed July 2021.

Brody, R. G.,'The New Victims of Identity Theft: Children', *Journal of Forensic & Investigative Accounting,* vol. 3, no. 2, 2011, www.web.nacva.com/JFIA/Issues/JFIA-2011-2_5.pdf, accessed April 2021.

Brownlee, J., '14 Different Types of Learning in Machine Learning', Machine Learning Mastery, 11 November 2019, machinelearningmastery.com/types-of-learning-in-machine-learning/, accessed April 2021.

Büchi, M., et al., 'Chilling Effects of Profiling Activities: Mapping the issues', *SSRN Electronic Journal,* 7 May 2019, https://ssrn.com/abstract=3379275 or http://dx.doi.org/10.2139/ssrn.3379275 , accessed April 2021.

Buckingham, D., 'Kids for Sale?', pp. 242-257, in *Thatcher's Grandchildren,* part of the Palgrave Studies in the History of Childhood series, 2014, link.springer.com/chapter/10.1057/9781137281555_13, accessed April 2021.

Byrne, J., E. Day, and L. Raftree, 'The Case for Better Governance of Children's Data: A manifesto', UNICEF, New York, May 2021, www.unicef.org/globalinsight/media/1741/file/UNICEF%20Global%20Insight%20Data%20Governance%20Manifesto.pdf, accessed May 2021.

Cadwalladr, C.,, and E. Graham-Harrison, 'Revealed: 50 million facebook profiles harvested for Cambridge Analytica in major data breach', The Guardian, 17 March 2018, www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election, accessed May 2021.

Calvert, C.E., 'Developing a Model and Applications for Probabilities of Student Success: A case study of predictive analytics', *Open Learning: The Journal of Open, Distance and e-Learning*, July 2014, www.tandfonline.com/doi/abs/10.1080/02680513.2014.931805, accessed April 2021.

Carroll, W., 'Predicting Severe Maternal Morbidity and Mortality: An informatics opportunity', *Online Journal of Nursing Informatics,* November 2018, www.researchgate.net/profile/Whende-Carroll/publication/338111621_Predicting_Severe_Maternal_Morbidity_and_Mortality_-_An_Informatics_Opportunity/links/5dff85d792851c836492403e/Predicting-Severe-Maternal-Morbidity-and-Mortality-An-Informatics-Opportunity.pdf, accessed April 2021.

Centers for Disease Control and Prevention, 'Caring for Children in a Disaster: How are children different from adults?' CDC, Atlanta, Ga, 2 August 2019, www.cdc.gov/childrenindisasters/differences.html, accessed May 2021.

Ceballos, M. P., et al., "¿Inteligencia Artificial o Artificios Patriarcales?' Cuarto Poder Salta, April 2018, cuartopodersalta.com.ar/inteligencia-artificial-o-artificios-patriarcales/, accessed May 2021.

Centre for Humanitarian Data, 'Catalogue Of Predictive Models In The Humanitarian Sector', Catalogue of predictive models in the humanitarian sector – The Centre for Humanitarian Data (humdata.org), accessed April 2021.

Centre for Humanitarian Data, 'Predictive analytics in humanitarian response', Workshop report, The Hague, 11-12 April 2019, centre.humdata.org/wp-content/uploads/2019/05/Workshop-Report-Predictive-Analytics-April-2019.pdf, accessed April 2021.

Centre for Humanitarian Data, 'Peer Review Framework for Predictive Analytics in Humanitarian Response', The Hague, May 2021, data.humdata.org/dataset/2048a947-5714-4220-905b-e662cbcd14c8/resource/76e488d9-b69d-41bd-927c-116d633bac7b/download/peer-review-framework-2020.pdf, accessed May 2021.

CGTN Africa, 'UNOCHA Faces Funding Shortfall to Meet Humanitarian Needs in Ethiopia', 4 August 2020, www.africa.cgtn.com/2020/08/04/unocha-faces-funding-shortfall-to-meet-humanitarian-needs-in-ethiopia, accessed April 2021.

Chandir, S., et al., 'Using Predictive Analytics to Identify Children at High Risk of Defaulting from a Routine Immunization Program: Feasibility study', *JMIR Public Health Surveill,* vol. 4, no. 3, September 2018, www.pubmed.ncbi.nlm.nih.gov/30181112/, accessed April 2021.

Chaudron, S., R, Di Gioia and M. Gemo, "Young Children (0-8) and Digital Technology: A qualitative study across Europe", JRC Science For Policy Report, European Commission, Brussels, 2018 www.core.ac.uk/download/pdf/159629895.pdf, accessed April 2021.

Children's Hospital Association, 'Using Predictive Analytics to Assist Mental Health Providers', November 2020, www.childrenshospitals.org/Newsroom/Childrens-Hospitals-Today/Articles/2020/11/Using-Predictive-Analytics-to-Assist-Mental-Health-Providers, accessed April 2021.

Church, C. E., and A. J. Fairchild, 'In Search of a Silver Bullet: Child welfare's embrace of predictive analytics', *Juvenile & Family Court Journal*, vol. 68, no. 7, pp. 67-81, 2017, www.doi.org/10.1111/jfcj.12086, accessed July 2021.

Clayton, V., and M. Sanders, 'Machine Learning in Children's Services', Technical Report, What Works for Children's Social Care, London, September 2020, www.whatworks-csc.org.uk/wp-content/uploads/WWCSC_technical-_report_machine_learning_in_childrens_services_does_it_work_Sep_2020.pdf, accessed April 2021.

Council of Europe, 'Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data: Children's data protection in an education setting', 2020, https://rm.coe.int/educational-settings/16809f3ba3 (coe.int), accessed May 2021.

Crawford, K., '*The Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*', Yale University Press, New Haven, CT, 2021, www.doi:10.2307/j.ctv1ghv45t, accessed July 2021.

Dare, T., and E. Gambrill, 'Ethical Analysis: Predictive risk models at call screening for Allegheny County', Allegheny County Analytics, April 2017, www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Ethical-Analysis-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-2.pdf, accessed April 2021.

Dobbe, R., and M. Whittaker, 'AI and Climate Change: How they're connected, and what we can do about it', AI Now Institute, 17 October 2019, www.medium.com/@AINowInstitute/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it-6aa8d0f5b32c, accessed July 2021.

Doleck, T., et al., 'Predictive Analytics in Education: A comparison of deep learning frameworks', *Education and Information Technologies,* vol. 25, pp. 1951-1963, November 2019, www.link.springer.com/article/10.1007/s10639-019-10068-4, accessed April 2021.

Döring, M., 'Prediction vs Forecasting', Data Science blog, 9 December 2018, www.datascienceblog.net/post/machine-learning/forecasting_vs_prediction/, accessed May 2021.

Drouin, M., 'Predicting Recidivism Among Internet Child Sex Sting Offenders Using Psychological Language Analysis', *Cyberpsychology, Behavior, and Social Networking,* vol. 21, no. 2, February 2018, www.liebertpub.com/doi/abs/10.1089/cyber.2016.0617, accessed April 2021.

Duncan, P., et al., 'Who Won and Who Lost: When A-levels meet the algorithm', The Guardian, 13 August 2020, www.theguardian.com/education/2020/aug/13/who-won-and-who-lost-when-a-levels-meet-the-algorithm, accessed April 2021.

Ekowo, M., and I. Palmer, 'The Promise and Peril of Predictive Analytics in Higher Education', *New America,* October 2016, The Promise and Peril of Predictive Analytics in Higher Education (newamerica.org), accessed May 2021.

Elish, M.C., 'Moral Crumple Zones: Cautionary tales in human-robot interaction', *Engaging Science, Technology, and Society,* vol. 5, March 2019, www.doi.org/10.17351/ests2019.260, accessed May 2021.

Elish, M.C., and E.A. Watkins, 'Repairing Innovation: A study of integrating AI in clinical care', Data and Society Research Institute, New York, September 2020, www.datasociety.net/wp-content/uploads/2020/09/Repairing-Innovation-DataSociety-20200930-1.pdf, accessed May 2021.

Eubanks, V., *Automating Inequality: How high tech tools profile, police and punish the poor,* St. Martins, New York, 2017.

European Commission: Directorate General for Justice and Consumers, *The EU Data Protection Reform and Big Data',* Publications Office of the European Union, Luxembourg, 2016, www.data.europa.eu/doi/10.2838/190200, accessed September 2021.

European Union: European Commission, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance)*, Official Journal of the European Union, Luxembourg, 2016, https://eur-lex.europa.eu/eli/reg/2016/679/oj, accessed September 2021.

Executive Office of the President of the United States, 'Big Data: A report on algorithmic systems, opportunity, and civil rights', The White House, Washington, D.C., 2016, www.obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf, accessed May 2021.

Fansiwala, K., 'The Duality of Medicine: The Willowbrook State School experiments', The Review and Debates at NYU, 20 February 2016, www.thereviewanddebatesatnyu.com/all/2016/2/20/the-duality-of-medicine-the-willowbrook-state-school-experiments, accessed May 2021.

Ford, J. D., et al., 'Opinion: Big Data has big potential for applications to climate change adaptation', *Proceedings of the National Academy of Sciences of the United States of America,* vol. 113, no. 39, pp. 10,729-32,2016, www.pnas.org/content/113/39/10729.short, accessed May 2021.

Fossheim, H., and H. Ingierd, eds., *Internet Research Ethics*, 1st ed., Cappelen Damm Akademisk, Norway, 2015, www.library.oapen.org/viewer/web/viewer.html?file=/bitstream/handle/20.500.12657/31153/637311.pdf?sequence=1&isAllowed=y, access April 2021.

Fumo, D., 'Types of Machine Learning Algorithms You Should Know', Medium, August 2017, www.towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861, accessed May 2021.

Furlong, A., *Handbook of Youth and Young Adulthood: New perspectives and agendas*, 1st ed., Routledge International Handbooks, London and New York, 2009.

Gillingham, P., 'Predictive Risk Modelling to Prevent Child Maltreatment and Other Adverse Outcomes for Service Users: Inside the "Black Box" of machine learning', *British Journal of Social Work*, vol. 46, no. 4, pp, 1,044–58, 2016, doi.org/10.1093/bjsw/bcv031, accessed June 2021.

Glaberson, S. K., 'Coding Over the Cracks: Predictive analytics and child protection', *Fordham Urban Law Journal,* vol. 46, no. 2, 2019, www.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.

Global Action Plan, 'Kids for Sale: Online advertising and the manipulation of children', September 2020, www.globalactionplan.org.uk/files/kids_for_sale_v2.pdf, accessed April 2021.

Goic, M., 'COVID-19: Short-term forecast of ICU beds in times of crisis', *Plos One*, January 2021, www.journals.plos.org/plosone/article?id=10.1371/journal.pone.0245272, accessed April 2021.

Goldhaber-Fiebert, J.D., and L. Prince, 'Impact Evaluation of a Predictive Risk Modeling Tool for Allegheny County's Child Welfare Office', Allegheny County Analytics, March 2019, www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Impact-Evaluation-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-6.pdf, accessed April 2021.

Gorr, K., and K. Zawacki, 'Data Stewardship: What is it and why does it matter?', Mozilla Foundation, 17 February 2017, www.foundation.mozilla.org/en/blog/data-stewardship-what-it-and-why-does-it-matter/, accessed April 2021.

Government of Canada, 'Algorithmic Impact Assessment Tool', Ottawa, Canada, www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html, accessed April 2021.

Green, B., and Y. Chen, 'Algorithmic Risk Assessments Can Alter Human Decision-Making Processes in High-Stakes Government Contexts', Proceedings of the ACM on Human-Computer Interaction 5, Article 418, ArXiv:2012.05370 [Cs], 9 December 2020, www.arxiv.org/abs/2012.05370, accessed May 2021.

Han, H. J., "An Algorithm Shouldn't Decide a Student's Future', Politico, August 2020, www.politico.eu/article/an-algorithm-shouldnt-decide-students-future-coronavirus-international-baccalaureate/, accessed May 2021.

Hao, K., 'This is How AI Bias Really Happens - and Why It's So Hard to Fix', MIT Technology Review, February 2019, www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/, accessed April 2021.

Hao, K., 'A Radical New Technique Lets AI Learn with Practically No Data', MIT Technology Review, October 2020, www.technologyreview.com/2020/10/16/1010566/ai-machine-learning-with-tiny-data/, accessed May 2021.

Harkness, T., 'A Level Results: Why algorithms aren't making the grade', *BBC Science Focus Magazine*, 11 January 2021, www.sciencefocus.com/future-technology/a-level-results-why-algorithms-arent-making-the-grade/, accessed May 2021.

He, L., et al., 'Predictive Analytics Machinery for STEM Student Success Studies', *Applied Artificial Intelligence*, vol. 32, no. 4, pp. 361-87, 2018, doi.org/10.1080/08839514.2018.1483121, accessed June 2021.

Hernandez, K., and T. Roberts, 'Predictive Analytics in Humanitarian Action: A preliminary mapping and analysis', K4D Emerging Issues Report No. 33, Institute of Development Studies, University of Sussex, U.K., 2020, www.ids.ac.uk/publications/predictive-analytics-in-humanitarian-action-a-preliminary-mapping-and-analysis/, accessed April 2021.

Hickey, K., 'Florida Takes Aim at Juvenile Recidivism with Predictive Analytics', GCN, 31 July 2015, gcn.com/articles/2015/07/31/juvenile-predictive-analytics.aspx, accessed May 2021.

Hirabayashi, K., 'The Impact of COVID-19 on Routine Vaccinations', UNICEF, New York, 24 April 2020, www.unicef.org/eap/stories/impact-covid-19-routine-vaccinations, accessed April 2021.

Hoffman, M.K., and N. Ma, 'A Machine Learning Algorithm for Predicting Maternal Readmission for Hypertensive Disorders of Pregnancy', *American Journal of Obstetrics & Gynecology MFM*, vol. 3, issue 1, January 2021, www.sciencedirect.com/science/article/pii/S2589933320302184, accessed April 2021.

Hong, B., 'Applications of Machine Learning Methods to Predict Readmission and Length-of-Stay for Homeless Families: The case of win shelters in New York City', *Journal of Technology in Human Services,* January 2018, www.researchgate.net/publication/322848763_Applications_of_Machine_Learning_Methods_to_Predict_Readmission_and_Length-of-Stay_for_Homeless_Families_The_Case_of_Win_Shelters_in_New_York_City, accessed April 2021.

Horwood, C., and W. Powell, 'Mixed Migration Review 2019', Mixed Migration Centre, 2019, www.drc.ngo/media/hphp35uo/mixed-migration-review-2019.pdf, accessed May 2021.

Information Commissioner's Office, 'Applications: Children and the GDPR', The General Data Protection Regulation, May 2018, www.ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/children-and-the-gdpr-1-0.pdf, accessed May 2021.

Inter-Agency Standing Committee, 'Operational Guidance: Data responsibility in humanitarian action', United Nations, New York, February 2021, www.interagencystandingcommittee.org/system/files/2021-02/IASC%20Operational%20Guidance%20on%20Data%20Responsibility%20in%20Humanitarian%20Action-%20February%202021.pdf, accessed April 2021.

IOM's Global Migration Data Analysis Centre (GMDAC), 'Child and Young Migrants', Migration Data Portal, IOM's GDMAC, Berlin, 6 May 2021, www.migrationdataportal.org/themes/child-and-young-migrants, accessed July 2021.

Jacobsen, K, L., 'Making Design Safe for Citizens: A hidden history of humanitarian experimentation', *Citizenship Studies,* vol. 14, no. 1, 2010, doi.org/10.1080/13621020903466399, accessed May 2021.

Jacobsen, K.L., 'Experimentation in Humanitarian Locations: UNHCR and biometric registration of Afghan refugees', *Security Dialogue,* vol. 46, no. 2, April 2015, doi.org/10.1177/0967010614552545, accessed May 2021.

Jepson, R.G., et al., 'The Effectiveness of Interventions to Change Six Health Behaviours: A review of reviews', *BMC Public Health,* vol. 10. No. 1, p. 538, 2010, www.doi.org/10.1186/1471-2458-10-538, accessed July 2021.

Jones, N., 'Anne Tolley Scraps 'Lab Rat' Study on Children', NZ Herald*,* 30 July 2015, www.nzherald.co.nz/nz/anne-tolley-scraps-lab-rat-study-on-children/C7GIGYW2467HG327FKXFRJDPEM/, accessed April 2021.

Kaminski, M.E., 'The Right to Explanation, Explained', *Berkeley Technology Law Journal,* vol. 34, no.1, May 2019, lawcat.berkeley.edu/record/1128984, accessed April 2021.

Kaplan, J., and S. Morgan, 'Predicting Displacement: Using predictive analytics to build a better future for displaced children', Save the Children International, London, 2018, resourcecentre.savethechildren.net/library/predicting-displacement-using-predictive-analytics-build-better-future-displaced-children, accessed April 2021.

Kelly, J., 'Florida's New Predictive Risk Tool Likely to Drive Down Juvenile Incarceration', The Imprint, 23 June 2015, www.imprintnews.org/featured/floridas-new-predictive-risk-tool-likely-to-drive-down-juvenile-incarceration/10505, accessed May 2021.

Krishnan, A., 'Decolonial Humanitarian Digital Governance', Berkman Klein Center Collection for Internet and Society at Harvard, 15 April 2021, medium.com/berkman-klein-center/decolonial-humanitarian-digital-governance-48b35b05b110, accessed April 2021.

Lacoste, A. et al., 'Quantifying the Carbon Emissions of Machine Learning', Paper submitted to Cornell University, 21 October 2019, ArXiv:1910.09700 [Cs], arxiv.org/abs/1910.09700, accessed July 2021.

LaFrance, A., 'The Algorithm That Makes Preschoolers Obsessed With YouTube', The Atlantic*,* 25 July 2017, www.theatlantic.com/technology/archive/2017/07/what-youtube-reveals-about-the-toddler-mind/534765/, accessed April 2021.

Lanier, P., et al., 'Preventing Infant Maltreatment with Predictive Analytics: Applying ethical principles to evidence-based child welfare policy', *Journal of Family Violence*, vol. 35, no. 1, pp. 1-13, 2020, accessed April 2021.

Leslie, D., et al., 'Ethics Review of Machine Learning in Children's Social Care', What Works for Children's Social Care, The Alan Turing Institute, University of Oxford, January 2020, www.whatworks-csc.org.uk/wp-content/uploads/WWCSC_Ethics_of_Machine_Learning_in_CSC_Jan2020.pdf, accessed May 2021.

LIAA, 'Sobre la Predicción Automática de Embarazos Adolescentes – LIAA', 2018, liaa.dc.uba.ar/es/sobre-la-prediccion-automatica-de-embarazos-adolescentes/.

Livingstone, S., J. Carr, and J. Byrne, 'One in Three: Internet governance and children's rights', Innocenti Discussion Paper (2016-01): 37, UNICEF Office of Research, Florence, 2016, www.unicef-irc.org/publications/pdf/idp_2016_01.pdf, accessed July 2021.

Livingstone, S., M. Stoilova, and R. Nandagiri, 'Children's Data and Privacy Online: Growing up in a digital age', December 2018, www.lse.ac.uk/media-and-communications/assets/documents/research/projects/childrens-privacy-online/Evidence-review-final.pdf, accessed April 2021.

Locey, K.J., et al., 'An Interactive Tool to Forecast US Hospital Needs in the Coronavirus 2019 Pandemic', *JAMIA Open,* vol. 3, no. 4, pp. 506-512, December 2020, academic.oup.com/jamiaopen/article/3/4/506/5907917, accessed April 2021.

Lupton, D., and B. Williamson, 'The Datafied Child: The dataveillance of children and implications for their rights', *New Media and Society,* 23 January 2017, journals.sagepub.com/doi/abs/10.1177/1461444816686328, accessed April 2021.

Madden, M., et al., 'Privacy, Poverty, and Big Data: A matrix of vulnerabilities for poor Americans', *Washington University Law Review,* vol. 5, no. 1, 2017, openscholarship.wustl.edu/law_lawreview/vol95/iss1/6, accessed May 2021.

Magalhães, R., et al., 'Geographical Analysis of the Role of Water Supply and Sanitation in the Risk of Helminth Infections of Children in West Africa', *PNAS,* vol. 108, no. 50, 13 December 2011, www.pnas.org/content/pnas/108/50/20084.full.pdf, accessed April 2021.

Mansour, Z., et al., 'Impact of COVID-19 Pandemic on the Utilization of Routine Immunization Services in Lebanon', Plos One, vol. 16, no. 2, e0246951, 2021, doi.org/10.1371/journal.pone.0246951, accessed July 2021.

Marder, B., et al., 'The Extended Chilling Effect of Facebook: The cold reality of ubiquitous social networking', *Computers in Home Behaviour*, no. 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.

Mascheroni, G., and D. Holloway, *The Internet of Toys: Practices, affordances and the political economy of children's smart play,* Palgrave Macmillan, London, 2019.

Medical Research Council, 'MRC Ethics Guide: Medical research involving children', 27, 2004, https://mrc.ukri.org/documents/pdf/medical-research-involving-children/, accessed May 2021

Meijer, A., and M. Wessels, 'Predictive Policing: Review of benefits and drawbacks'. *International Journal of Public Administration*, vol. 42, no. 12, pp, 1,031-39, 2019, doi.org/10.1080/01900692.2019.1575664, accessed May 2021.

Merriam-Webster, (n.d.,), 'Definition of Algorithm', Merriam-Webster Dictionary, www.merriam-webster.com/dictionary/algorithm, accessed April 2021.

Metcalf, J., and K. Crawford, 'Where are Human Subjects in Big Data Research? The Emerging Ethics Divide', *Big Data & Society,* January - June 2016, journals.sagepub.com/doi/pdf/10.1177/2053951716650211, accessed April 2021.

Metcalf, J., et al., 'Algorithmic Impact Assessments and Accountability: The co-construction of impacts', paper presented ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), virtual conference, Canada, March 3-10, 2021, ssrn.com/abstract=3736261, accessed April 2021.

Michard, F., and J.L. Teboul, 'Predictive Analytics: Beyond the buzz', *Annals of Intensive Care,* vol. 9, no. 1, p. 46, 2019, doi.org/10.1186/s13613-019-0524-9, accessed July 2021.

Microsoft, 'Avanza el Uso de la Inteligencia Artificial en la Argentina con Experiencias en el Sector Público, Privado y ONGs', Microsoft News Center Latinoamérica, April 2018, news.microsoft.com/es-xl/avanza-el-uso-de-la-inteligencia-artificial-en-la-argentina-con-experiencias-en-el-sector-publico-privado-y-ongs/, accessed May 2021.

Morozov, E., *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems That Don't Exist*, Allen Lane, London, 2013.

Morris, E., 'Children: Extremism and online radicalization', *Journal of Children and Media*, vol. 10, no. 4, pp, 508-14, October 2016, doi.org/10.1080/17482798.2016.1234736, accessed May 2021.

Munro, E., 'Improving Practice: Child protection as a systems problem', *Children and Youth Services Review*, special edition, 'Decision Making in Child Welfare', vol. 27, no. 4, pp, 375-91, 2005, doi.org/10.1016/j.childyouth.2004.11.006, accessed June 2021.

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.

Nasser, A., 'Yemen's Donors Need to Tackle Aid Shortfalls', Human Rights Watch, 25 February 2021, www.hrw.org/news/2021/02/25/yemens-donors-need-tackle-aid-shortfalls, accessed April 2021.

Newport, E.L., 'Children and Adults as Language Learners: Rules, variation, and maturational change', *Topics in Cognitive Science*, vol. 12, no. 1, pp. 153-69, 2020, onlinelibrary.wiley.com/doi/abs/10.1111/tops.12416, accessed May 2021.

Noble, S.U., *Algorithms of Oppression: How search engines reinforce racism*, New York University Press, New York, 2018, ebookcentral.proquest.com/lib/londonschoolecons/detail.action?docID=4834260, accessed May 2021.

Nwafor, G., 'Africa to Harmonise Laws for Data Protection, Digital Economy', The Guardian Nigeria News, 9 October 2020, guardian.ng/appointments/africa-to-harmonise-laws-for-data-protection-digital-economy, accessed April 2021.

ODD 2021, '6th Outlier Detection and Description Workshop', 15 August 2021, www.oddworkshop.github.io/, accessed 6 September 2021.

OHCHR, Convention on the Rights of the Child, United Nations, New York, 20 November 1989, www.ohchr.org/en/professionalinterest/pages/crc.aspx, accessed May 2021.

OHCHR, General Comment on Children's Rights in Relation to the Digital Environment, No. 25, United Nations Committee on the Rights of the Child, 86th session, New York, www.ohchr.org/EN/HRBodies/CRC/Pages/GCChildrensRightsRelationDigitalEnvironment.aspx, accessed May 2021.

Ong, A.G., 'WFP Launches HungerMap Live', World Food Programme, Rome, Italy, 20 January 2020, www.wfp.org/stories/wfp-launches-hungermap-live, accessed May 2021.

Ongsulee, P., et al., 'Big Data, Predictive Analytics and Machine Learning', paper presented at 16th International Conference on ICT and Knowledge Engineering, 21-23 November 2018, Bangkok, Thailand, doi.org/10.1109/ICTKE.2018.8612393, accessed June 2021.

Onuoha, M., and D. Nucera, *A People's Guide to AI*, August 2018, alliedmedia.org/wp-content/uploads/2020/09/peoples-guide-ai.pdf, accessed April 2021.

Ortiz Freuler, J., and C. Iglesias, 'Algorithms and Artificial Intelligence in Latin America: A study of implementation by governments in Argentina and Uruguay', World Wide Web Foundation, Geneva, September 2018, webfoundation.org/docs/2018/09/WF_AI-in-LA_Report_Screen_AW.pdf, accessed July 2021.

Parikh, R.B., Z. Obermeyer and A.S Navathe, 'Regulation of Predictive Analytics in Medicine', *Science,* vol. 363, no. 6429, February 2019, doi.org/10.1126/science.aaw0029.

Pariser, E., *The Filter Bubble: What the internet is hiding from you*, Penguin, New York, May 2011.

Peña, P., and J. Varon, 'Teenager Pregnancy Addressed Through Data Colonialism in a System Patriarchal by Design', Not My A.I., 3 May 2021, notmy.ai/2021/05/03/case-study-plataforma-tecnologica-de-intervencion-social-argentina-and-brazil/.

Perry, W. L., et al., *Predictive Policing: The role of crime forecasting in law enforcement operations*, RAND Corporation, Santa Monica, CA., 2013, www.rand.org/pubs/research_reports/RR233.html, also available in print form, accessed May 2021.

Predict Align Prevent, 'The PAP Program', www.predict-align-prevent.org/implementation, accessed July 2021.

Privacy International, '2020 Is a Crucial Year to Fight for Data Protection in Africa', 3 March 2020, Privacy International, privacyinternational.org/long-read/3390/2020-crucial-year-fight-data-protection-africa.

Qazi, S., M. Usman, and A. Mahmood, 'A Data-Driven Framework for Introducing Predictive Analytics into Expanded Program on Immunization in Pakistan', *Wiener Klinische Wochenschrift*, vol. 133, no. 13-14, pp. 695-702, September 2020, doi:10.1007/s00508-020-01737-3, accessed May 2021.

Raj, A.,et al., 'Application of Machine Learning to Understand Child Marriage in India', *SSM – Population Health*, vol. 12, December 2020, www.sciencedirect.com/science/article/pii/S2352827320303244, accessed April 2021.

Raymond, N. A., *Beyond 'Do No Harm' and Individual Consent: Reckoning with the Emerging ethical challenges of civil society's use of data*, Part of the Philosophical Studies Series, Springer, New York, December 2016, link.springer.com/chapter/10.1007/978-3-319-46608-8_4, accessed April 2021.

Redden, J., L. Dencik, and H. Warne, 'Datafied Child Welfare Services: Unpacking politics, economics and power', *Policy Studies*, vol. 41, no. 5, pp. 507-26, September 2020, doi.org/10.1080/01442872.2020.1724928, accessed May 2021.

Reischmann, J., 'Lifelong and Lifewide Learning: a perspective', in *Lifelong Education and Lifelong Learning in Thailand*, edited by Suwithida Charungkaitt, Bangkok, pp. 286-309, core.ac.uk/download/pdf/268763228.pdf, accessed August 2021.

Reisman, D., et al., 'Algorithmic Impact Assessments: A practical framework for public agency accountability', AI Now, April 2018, www.ainowinstitute.org/aiareport2018.pdf, accessed April 2021.

Reliefweb, 'Global Humanitarian Overview 2020 Monthly Funding Update', 31 December 2020, OCHA Services, reliefweb.int/report/world/global-humanitarian-overview-2020-monthly-funding-update-31-december-2020, accessed April 2021.

Robinson, W.M., and B.T. Unruh, 'The Hepatitis Experiments at the Willowbrook State School', in *The Oxford Textbook of Clinical Research Ethics*, Oxford University Press, New York, 2008, pp. 80-85, repository.library.georgetown.edu/handle/10822/960687, accessed May 2021.

de Ruigh, E.L., et al., 'Predicting Quality of Life during and Post Detention in Incarcerated Juveniles', *Quality of Life Research*, vol. 28, no. 7, pp. 1813-23, 2019, doi.org/10.1007/s11136-019-02160-6, accessed April 2021.

Russell, J., and S. Macgill, 'Demographics, Policy, and Foster Care Rates: A predictive analytics approach', *Children and Youth Services Review,* vol. 58, November 2015, doi.org/10.1016/j.childyouth.2015.09.009, accessed May 2021.

Salganik, M.J., et al., 'Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration', *Proc Natl Acad Sci USA*, vol. 117, no. 15, pp. 8398-8403, April 2020, pubmed.ncbi.nlm.nih.gov/32229555/, accessed April 2021.

Sanchez, L., 'A Practitioner Survey Exploring the Value of Forensic Tools, AI, Filtering, and Safer Presentation for Investigating Child Sexual Abuse Material (CSAM)', *Digital Investigation* vol. 29, pp. S124–142, July 2019, doi.org/10.1016/j.diin.2019.04.005, accessed July 2021.

Satell, G, and J. Sutton, 'We Need AI That Is Explainable, Auditable, and Transparent', Harvard Business Review, 28 October 2019, hbr.org/2019/10/we-need-ai-that-is-explainable-auditable-and-transparent, accessed May 2021.

Save the Children International, 'Save the Children's Migration and Displacement Initiative (MDI)', Save the Children, London, 2021, resourcecentre.savethechildren.net/node/19123/pdf/mdi_editable_flyer_v330_read-only.pdf, accessed May 2021.

Seidel, E., K., Salah, 'Using Predictive Analytics to Target and Improve First Year Student Attrition', *Australian Journal of Education*, vol. 61, no. 2, 2017, journals.sagepub.com/doi/abs/10.1177/0004944117712310?journalCode=aeda, accessed April 2021.

Selbst, A. D., et al., 'Fairness and Abstraction in Sociotechnical Systems', *Association for Computing Machinery*, presented at FAT conference, Atlanta, Ga., 29-31 January 2019, dl.acm.org/doi/pdf/10.1145/3287560.3287598, accessed April 2021.

Sentas, V., and C. Pandolfini, 'Policing Young People in NSW: A study of the suspect targeting management pPlan', Report of the Youth Justice Coalition, Marrickville, New South Wales, October 2017, apo.org.au/node/116176, accessed May 2021.

Schwartz, R., et al., 'Green AI', report submitted to Cornell University, Ithaca, NY, 22 July 2019, arxiv.org/abs/1907.10597, accessed July 2021.

Shea, K. M., 'Children are Not Little Adults', WHO Training Package for the Health Sector: Children's Health and the Environment, World Health Organization, Geneva, July 2008, www.who.int/ceh/capacity/Children_are_not_little_adults.pdf, accessed February 2021.

Siegel, E., *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die,* John Wiley & Sons, Hoboken, N.J, 2016.

Singh, H., et al., 'iNICU – Integrated Neonatal Care Unit: Capturing neonatal journey in an intelligent data way', *Journal of Medical Systems*, vol. 41, no. 132, 26 July 2017, doi.org/10.1007/s10916-017-0774-8.

Spaeder, M.C., et al., 'Predictive Analytics in the Pediatric Intensive Care Unit for Early Identification of Sepsis: Capturing the context of age', *Pediatric Research*, vol. 86, no. 5, pp. 655-661, July 2019, doi.org/10.1038/s41390-019-0518-1, accessed May 2021.

Sphere Association, *The Sphere Handbook: Humanitarian Charter and Minimum Standards in Humanitarian Response,* Geneva, 2018 www.spherestandards.org/handbook, accessed April 2021.

Sternik, I., 'La Inteligencia Que No Piensa, PAGINA12, 21 April 2018, www.pagina12.com.ar/109080-la-inteligencia-que-no-piensa, accessed May 2021.

Strubell, E., A. Ganesh, and A. McCallum, 'Energy and Policy Considerations for Deep Learning in NLP', paper submitted to Cornell University, Ithaca, NY, 5 June 2019, arxiv.org/abs/1906.02243, accessed July 2021.

Talabis, M., et al., *Information Security Analytics: Finding security insights, patterns, and anomalies in Big Data*, 1st edition, Syngress, Waltham, M.A., 2014,

Taxman, F. S., *Handbook on Risk and Need Assessment: Theory and practice*, Taylor & Francis, Baingdon, U.K., 2016, accessed May 2021.

The Engine Room, 'Responsible Data in Development Toolkit', in *The Handbook of the Modern Development Specialist*, Responsible Data, 2016, the-engine-room.github.io/responsible-data-handbook/chapters/chapter-00-introduction.html#why-responsible-data, accessed May 2021.

Ting, M.H., et al., 'Predicting Recidivism among Youth Offenders: Augmenting professional judgement with machine learning algorithms'. *Journal of Social Work*, vol. 18, no. 6, pp. 631-649, 2018, doi.org/10.1177/1468017317743137, accessed May 2021.

Tufekci, Z., 'You Can't Be Discreet Online', *The New York Times*, Late Edition-Final edition, sec. Editorial Desk; SECTA, 22 April 2019, global.factiva.com/redir/default.aspx?P=sa&an=NYTF000020190427ef4m0000a&cat=a&ep=ASE, accessed May 2021.

UNDESA, 'Definition of Youth', United Nations Dept. of Economic and Social Affairs, New York, un.org/esa/socdev/documents/youth/fact-sheets/youth-definition.pdf, accessed May 2021.

UN General Assembly, Convention on the Rights of the Child, 20 November 1989, www.refworld.org/docid/3ae6b38f0.html, accessed February 2021.

United Nations General Assembly, Universal Declaration of Human Rights, New York, 10 December 1948, www.un.org/en/about-us/universal-declaration-of-human-rights, accessed July 2021.

United Nations CERF, Anticipatory Action Update, CERF Advisory Group Meeting, 1 June 2020 https://cerf.un.org/sites/default/files/resources/Anticipatory%20Action%20Update.pdf, accessed April 2021.

UNESCO, 'Policy Brief: Learning analytics', United Nations Educational, Scientific and Cultural Organization, Geneva, November 2012, iite.unesco.org/pics/publications/en/files/3214711.pdf, accessed April 2021.

UNHCR, 'Global Trends: Forced Displacement in 2019', New York, 18 June 2021, unhcr.org/uk/statistics/unhcrstats/5ee200e37/unhcr-global-trends-2019.html, accessed April 2021.

UNHCR Innovation Service, 'Jetson Technical Specifications', New York, jetson.unhcr.org/tech.html, accessed April 2021.

UNHCR Innovation Service, 'Is it possible to predict forced displacement?', New York, 13 May 2019, medium.com/unhcr-innovation-service/is-it-possible-to-predict-forced-displacement-58960afe0ba1, accessed April 2021.

UNICEF, 'The State of the World's Children 2017: Children in a digital world', New York, December 2017, www.unicef.org/media/48581/file/SOWC_2017_ENG.pdf, accessed April 2021.

UNICEF, 'UNICEF Programme Guidance for the Second Decade: Programming with and for adolescents', New York, October 2018, reliefweb.int/sites/reliefweb.int/files/resources/UNICEF_Programme_Guidance_for_the_Second_Decade_0.pdf, accessed April 2021.

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

UNICEF Style Book, New York, September 2018, accessed April 2021.

UNICEF, 'Policy Guidance on AI for Children: Draft 1.0', New York, September 2020, United Nations Global Pulse, Big Data for Development: A Primer, 2013, unicef.org/globalinsight/media/1171/file/UNICEF-Global-Insight-policy-guidance-AI-children-draft-1.0-2020.pdf, accessed May 2021.

UNICEF, 'Executive Summary: Memorandum on Artificial Intelligence and child rights', UNICEF Innovation and Human Rights Center, UC Berkeley, CA, 2019, www.unicef.org/innovation/reports/memoAIchildrights, accessed April 2021.

United Nations, Secretary-General's Report to the General Assembly (UN SG), 1981, A/36/215.

USAID, 'Reflecting the Past, Shaping the Future: Making AI work for international development', USAID, Washington, D.C., September 2018, usaid.gov/digital-development/machine-learning/AI-ML-in-development, accessed April 2021.

USAID, 'Managing Machine Learning Projects in International Development: A practical guide', USAID, Washington, D.C., January 2021, www.usaid.gov/sites/default/files/documents/Vital_Wave_USAID-AIML-FieldGuide_FINAL_VERSION_1.pdf, accessed April 2021.

Uskov, V.L., et al., 'Machine Learning-Based Predictive Analytics of Student Academic Performance in STEM Education', paper presented at the 2019 IEEE Global Engineering Education Conference (EDUCON), 8-11 April, 2019, Dubai, U.A.E, pp. 1370–76, doi.org/10.1109/EDUCON.2019.8725237, accessed May 2021.

Vaithianathan, R., et al., 'Allegheny Family Screening Tool: Methodology, version 2', Allegheny County Analytics, April 2019, www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Methodology-V2-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-7.pdf, accessed April 2021.

Van Calster, B., et al., 'Calibration: The Achilles heel of predictive analytics', *BMC Medicine,* vol. 17. no. 230, 2019, doi.org/10.1186/s12916-019-1466-7, accessed May 2021.

Velluet, Q., 'Can Africa Salvage Its Digital Sovereignty?' The Africa Report, April 2021, www.theafricareport.com/80606/can-africa-salvage-its-digital-sovereignty/, accessed May 2021.

Viola de Azevedo Cunha, M. 'Child Privacy in the Age of Web 2.0 and 3.0: Challenges and opportunities for policy', Innocenti Discussion Paper 2017-03, UNICEF IRC, Florence, December 2017, www.unicef-irc.org/publications/pdf/Child_privacy_challenges_opportunities.pdf.

Von Wachter, T., et al., 'Predicting and Preventing Homelessness in Los Angeles', California Policy Lab, September 2019, www.capolicylab.org/wp-content/uploads/2019/12/Predicting_and_Preventing_Homelessness_in_Los_Angeles.pdf.

Walker, J., et al., 'Youth Inclusion and Support Panels: Preventing crime and anti-social behaviour – Executive Summary', Department for Children, Schools and Families', University of Newcatle upon Tyne, U.K., 2007, assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/222368/DCSF-RW018_20executive_20summary.pdf, accessed July 2021.

Walker, S.C., et al., 'A Real-Time Risk-Prediction Model for Pediatric Venous Thromboembolic Events', *Pediatrics*, vol. 147, no. 6, June 2021, doi:10.1542/peds.2020-042325.

Wang, T., 'You Are Not Your Data But Your Data is Still You', Deep Dives, 7 August 2020, deepdives.in/you-are-not-your-data-but-your-data-is-still-you-b41d2478ece2, accessed April 2021.

Weber, J., M. Umpierre, and S. Bilchik, 'Transforming Juvenile Justice Systems to Improve Public Safety and Youth Outcomes', Georgetown University Center for Juvenile Justice Reform, Washington, D.C., May 2018, csgjusticecenter.org/wp-content/uploads/2020/02/Transforming-Juvenile-Justice-Systems.pdf, accessed May 2021.

WHO, 'Recognizing Adolescents', World Health Organization, Geneva, 2014, Adolescence: a period needing special attention - recognizing-adolescence (who.int), accessed May 2021.

WHO, 'WHO and UNICEF Warn of a Decline in Vaccinations During Covid-19', World Health Organization, Geneva, July 2020, www.who.int/news/item/15-07-2020-who-and-unicef-warn-of-a-decline-in-vaccinations-during-covid-19, accessed April 2021.

WHO, 'Children Are Not Little Adults: Children's health and the environment: Training Package for the health sector', World Health Organization, Geneva, 2007, apps.who.int/iris/bitstream/handle/10665/331237/WHO-CED-PHE-EPE-19.12.07-eng.pdf?ua=1, accessed June 2021.

WHO, 'Air Pollution and Child Health: Prescribing clean air', World Health Organization, Geneva, 2018, www.who.int/ceh/publications/air-pollution-child-health/en/, accessed May 2021.

World Bank, 'Low and Middle Income: Country data", The World Bank, Washington, D.C., June 2021, data.worldbank.org/country/XO, accessed May 2021.

World Wide Web Foundation, 'Artificial Intelligence: The road ahead in low and middle-income countries', Geneva, 2017, webfoundation.org/docs/2017/07/AI_Report_WF.pdf, accessed May 2021.

Young, A., 'Responsible Group Data for Children', Good Governance of Children's Data project, no. 4, UNICEF Office of Global Insight and Policy, New York, October 2020, www.unicef.org/globalinsight/reports/responsible-group-data-children, accessed May 2021.

Završnik, A., 'Algorithmic Justice: Algorithms and big data in criminal justice settings', *European Journal of Criminology*, 18 September 2019, doi.org/10.1177/1477370819876762, accessed May 2021.

Zialcita, P., 'Facebook Pays $643,000 Fine For Role In Cambridge Analytica Scandal', *NPR,* 30 October 2019, www.npr.org/2019/10/30/774749376/facebook-pays-643-000-fine-for-role-in-cambridge-analytica-scandal, accessed May 2021.

Zerilli, J., et al., 'Algorithmic Decision-Making and the Control Problem', *Minds and Machines,* vol. 29, no. 4, December 2019, doi.org/10.1007/s11023-019-09513-7, accessed April 2021.

Zewain, A., 'Machine Learning and Differential Privacy: Overview', 2021.AI, Copenhagen, April 2019, 2021.ai/machine-learning-differential-privacy-overview/, accessed April 2021.

Zuboff, S., 'Big Other: Surveillance capitalism and the prospects of an information civilization', *Journal of Information Technology*, vol. 30, no. 1, pp. 75-89, 2015, doi.org/10.1057/jit.2015.5, accessed May 2021.

ANNEX 1: INTERVIEW QUESTIONS

1. Knowledge and work of interviewee with predictive analytics;
2. What are the benefits of using predictive analytics in policy/services for children?
3. What are some of the risks? How are they usually mitigated?
4. Are communities that will be affected by the use of the technology brought to the table in any way when projects are developed? Who is responsible if users are harmed by this product? Are there/what are the reporting process and process for recourse?
5. A lot of initiatives of predictive analytics are done jointly, in partnership with contractors/private sector. What is your experience with that and how would you evaluate the impact of partnerships?
6. Is ethics considered when rolling out this type of project? How so?
7. Can you talk about the future applications you see emerging (especially for policy/use directed at children)? What specific considerations should be considered in deployment for children?

ANNEX 2: CHECKLIST FOR PREDICTIVE ANALYTICS (PA) FOR CHILDREN

Goal: Understand and assess the ethical, social, privacy-related, and cultural implications of the proposed use case

	Yes	No/Insufficient	Comments
Have you reviewed the role of the proposed implementation to ensure that PA is used as a source of information, to be triangulated with other robust sources of evidence and knowledge (including children's voices), and not in isolation?			
Have you clearly and explicitly articulated the objectives of the project in terms of how it potentially benefits and improves children's lives?			
Have you conducted an ethical review of the proposed use case?			
Have you considered the environmental impact of PA, and is the most energy efficient approach being used?			

Goal: Mitigate risks that may arise from the use of data within the proposed PA implementation

	Yes	No/Insufficient	Comments
Have you ensured that centralization of the data does not put children at risk, both in the long and short term?			
Have you put in place clear data protection (security and confidentiality) protocols?			
Are you practicing data minimization, and sharing data with the minimum number of actors possible?			
Have you ensured that the datasets proposed for use do not contain data that will reinforce and exacerbate historical inequities?			
Have you established a clear understanding of the parameters, limitations, bias, and robustness of the datasets used?			

Goal: Mitigate risks that may arise from the design and implementation of predictive analytics models

	Yes	No/ Insufficient	Comments
Have you clearly articulated the approach adopted or objective function, and a reasonable and justifiable explanation of the purpose?			
Is the methodology and approach used to design the model, and used within the model itself to generate outputs, explainable and transparent?			
Did the PA model selected fit the objectives, availability and accuracy of data, and intended context?			
Have you determined the criteria for computational fairness of the model, in consultation with the impacted community?			

	Yes	No/ Insufficient	Comments
Have you assessed the fairness of the model, based on these criteria?			
Have measures been implemented, where appropriate, to reduce the impact on the environment?			
Have you ensured that private sector partnerships include standards of transparency, and of reporting and sharing, with clear data clauses to ensure partners do not use or share data for purposes not contained in the agreement?			
Have you planned ongoing monitoring and feedback points in order to monitor bias throughout the project?			

Goal: Put measures in place to ensure that staff and impacted communities can participate as meaningful stakeholders whose experience is valued

	Yes	No/Insufficient	Comments
Have you established capacity development to support communities to meaningfully understand how the PA implementation is being conducted?			
Have you provided staff with clear guidance on how to assess, interpret, and use outputs as well as how to reflect on the limitations of the model, mitigate over-dependence, and understand the need to triangulate outputs with other data sources?			
Have you incorporated direct stakeholder engagement in the design and development of the model?			
Is there regular and ongoing monitoring and evaluation of the impacts and accuracy of the model, and is there space for reflecting on these findings through regular updates and adjustments to the model?			

	Yes	No/Insufficient	Comments
Are there regular evaluations of the culture, implementation, and attitudes within implementing institutions in early to mid- stages of implementation?			
Have you established redress mechanisms that are accessible to children and/or caregivers and that are directly linked to monitoring and evaluation processes?			
Is there space to halt implementation at any point if the implementation is perceived as unacceptable by impacted communities?			

Population-based models

	Yes	No/Insufficient	Comments
Are there meaningful mechanisms in place for feedback and course correction?			
Have you determined that you know the context well and your team or implementing partner is regionally-based and familiar with the context?			
Is there disaggregated data available that is adequately large and representative of the population under consideration?			
Does this disaggregated data respect the privacy of the population it is representing?			
Is the data you plan to use an up-to-date reflection of the phenomenon and populations you are trying to model?			

Individual-based models

	Yes	No/Insufficient	Comments
If children's personal data will be used, are you going to conduct a Data Protection Impact assessment?			
Have you reviewed the relevant legislative environment to ensure the use case is appropriate and legal?			
Is there sufficiently high quality and suitable children's data available, or are there infrastructures in place to collect this data responsibly and safely from children?			
Have you gained meaningful consent from individual children whose data is processed or used and have you made sure that your consent policies recognize children's development (and right to erasure) and are tailored to the relevant age group?			
If consent is not feasible – have clear privacy notices been provided at relevant sites?			

for every child, answers

UNICEF Office of Research – Innocenti
Via degli Alfani, 58 – 50121
Florence, Italy
Tel: (+39) 055 20 330
Fax: (+39) 055 2033 220
florence@unicef.org
www.unicef-irc.org
@UNICEFInnocenti
facebook.com/UnicefInnocenti

unicef
Office of Research-Innocenti