



The Alan Turing Institute

Case study

Name of initiative

Understanding AI Ethics and Safety for Children

Implementing organization

The Alan Turing Institute

Initiative type

Artificial intelligence (AI) policy

About the initiative

The Alan Turing Institute is expanding its public policy guide *Understanding artificial intelligence ethics and safety*, to give public sector employees a better practical understanding of how to design responsible AI for children. This capacity-building effort aims to formulate ethical considerations to support the development of AI policies that are non-discriminatory and inclusive of and for children.

Alignment with UNICEF Policy Guidance on AI for Children

The initiative has aimed to:

- Empower governments and businesses with knowledge of AI and children's rights
- Ensure inclusion of and for children
- Prepare children for present and future developments in AI

Location

United Kingdom

Launched

Pilot initiated in 2020

Overview



The Turing team found that public sector organizations see existing policies as possible constraints to support and protect children effectively when designing, developing and deploying AI technologies.

In June 2019, the Alan Turing Institute – the United Kingdom’s national institute for data science and artificial intelligence (AI) – published [Understanding artificial intelligence ethics and safety](#), a comprehensive guide conceived to help public sector actors design, assess and implement responsible and accountable AI projects. This guide is now considered a fundamental resource for AI policymaking throughout the UK Government and has been cited frequently in academic publications.

The Alan Turing Institute (hereafter referred to as ‘the Institute’) is updating and expanding the guide, which will now include a series of practice-based workbooks, one of which examines the ethical considerations and implications of using children’s data in the design, development and deployment of AI technologies in the public sector. This workbook will be informed by [UNICEF’s Policy Guidance on AI for Children](#) and will include reference to its practical applications of child-centred requirements. As part of this effort, the Institute consulted with public sector organizations to gauge their views on the challenges of implementing child-centred AI. Interviewees were also questioned on what they think could alleviate future challenges when it comes to providing a robust accountability framework. The Turing team found that public sector organizations see existing policies as possible constraints to support and protect children effectively when designing, developing and deploying AI technologies. The team examined this issue in depth.

Context and project origins

The Institute aims to leverage data science and AI for the public good and divides its mission into three broad goals: advance world-class research and apply it to real-world problems, train the leaders of the future, and lead the public conversation. In the past decade, the world has seen powerful developments and new capacities in AI technology, data analytics and computational power. With this rise comes a host of questions and ethical concerns, such as:

- > In what ways can AI technologies prove harmful or discriminatory?
- > Does the public truly understand the capabilities of AI systems?
- > How can we design these systems ethically?
- > How can they be developed and implemented to serve the public interest?

AI capabilities can be leveraged for transformational opportunities and societal benefits, but they also raise ethical concerns in practice, when not applied responsibly. As the Turing guide [explains](#), a first set of potential harms that could affect vulnerable groups, including children, are AI-driven discrimination and unwanted biases. In the health sector, research has [shown](#) that “datasets composed of electronic health records, genome databases, and biobanks often under sample those who have irregular or limited access to the healthcare system, such as minoritized ethnicities, immigrants, and socioeconomically disadvantaged groups.” In the case of the [2020 Ofqual Exam controversy](#), an algorithmic model for predicting how well students would perform in their secondary school A Level exams tended to reinforce the economic and societal biases built into the schooling system. Another set of potential harms are threats posed by AI systems to young users’ privacy. For instance, AI-enabled Internet of Things (IoT) devices or smart toys might collect data from children in their home environment without their knowledge or consent or that of their parents or caregivers.

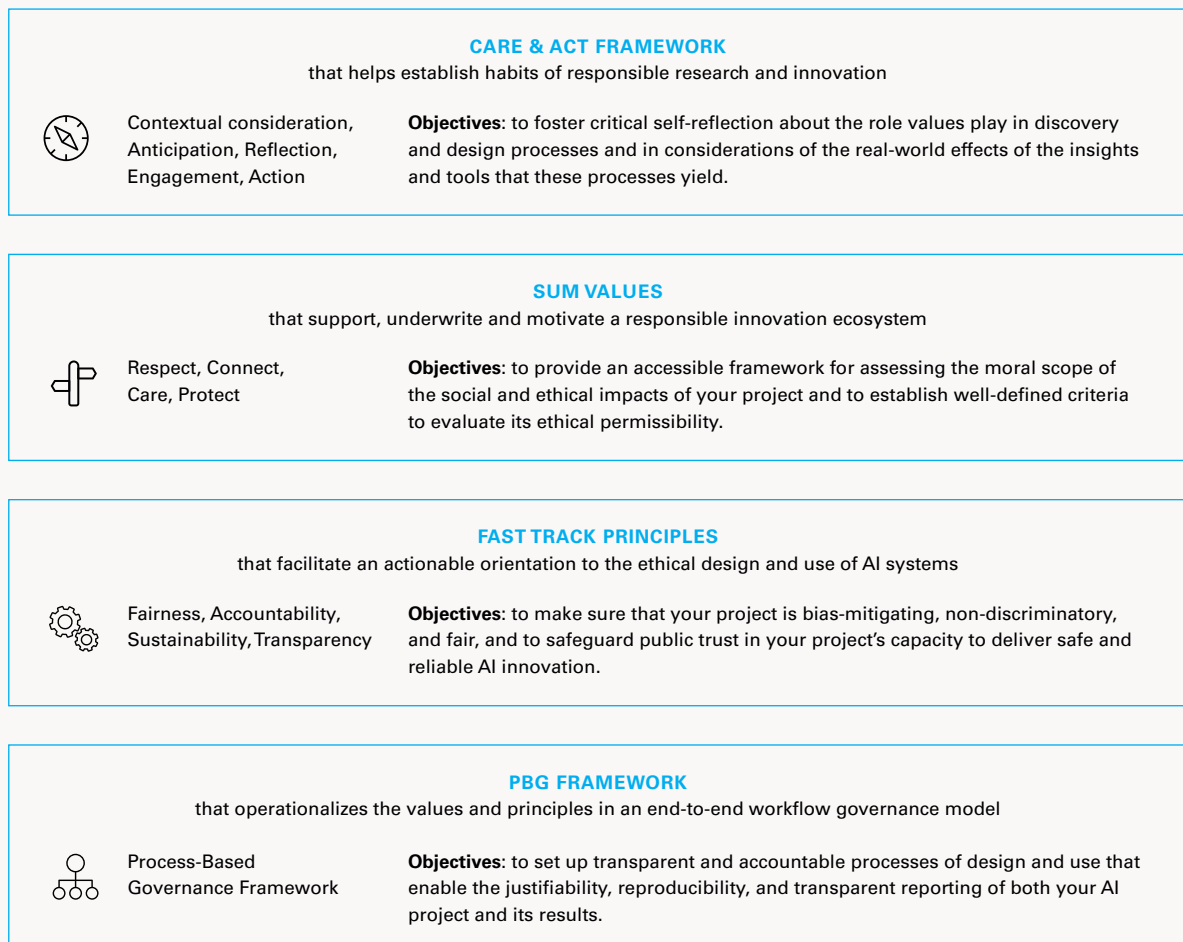
In a context where AI ethics matters more and more, the Institute identified significant gaps in how child-centred requirements were considered and applied to the responsible design and implementation of AI systems. In updating its guide, the Institute will also hold activity-based workshops and dialogue with young users and their guardians. This is to tap into the lived experiences of children to ensure the guidance is contextually sensitive and responsive.

About the AI policy

Understanding AI ethics and safety provides developers and public officials with methodological approaches to anticipate and identify potential AI harms and presents concrete, actionable measures to prevent or defend against these harms. The Institute's guide offers four essential building blocks that form an ethical platform for responsible AI project delivery.

FIGURE 1: ETHICAL PLATFORM FOR THE RESPONSIBLE DELIVERY OF AN AI PROJECT

FOUR TIERS OF A RESPONSIBLE PROJECT DELIVERY ECOSYSTEM



Source: https://www.turing.ac.uk/sites/default/files/2019-08/understanding_artificial_intelligence_ethics_and_safety.pdf#page=7



The target audience for the AI guide is anyone who is involved in the design, production and deployment of public sector AI. This, in practice, often refers to two groups: those who create and procure AI products and those who deploy them, including engineers and public sector stakeholders. For public sector stakeholders, such as local and national officials, the Institute's guide helps them understand and anticipate how certain communities, including vulnerable groups and particularly children, may be impacted by the deployment of AI policies. Currently, *Understanding AI ethics and safety* is used across the UK Government, including by the Ministry of Justice, the Home Office, the National Health Service, and many other government organizations. To date, the guide has been read over 10,000 times, downloaded over 4,000 times, and cited in academic literature 115 times.

While the guide has been published and is available online, its updated and expanded content will be released as a series of workbooks during 2021 and 2022. The Institute (together with the UK's Office for AI) plans to help public sector employees use these workbooks to help them gain a practical understanding of AI ethics and safety through activities that build awareness, expertise and capacity. This policy translation and capacity-building effort will focus on many different components of AI design and will specifically include exercises relating to the topic of AI and children.

Application of child-centred requirements

The Institute focused specifically on three child-centred requirements that align with the [UNICEF Policy Guidance on AI for Children](#), namely: to empower governments and businesses with knowledge of AI and children's rights, to ensure the inclusion of and for children, and to prepare children for the present and future developments of AI.

Empowering governments and businesses with knowledge of AI and children's rights

In this case, the Institute sought to empower government and public sector organizations with knowledge of AI and children's rights. The team had conversations with public sector organizations about the existing engagement with children on topics such as AI. The Institute spoke with experts in the field about how they have consulted children on similar topics in the past, what content they used, and how they accessed these communities. This helped inform the content development and research project plan. Through preliminary conversations with public sector organizations, a few gaps in the existing landscape were also identified. The first was that many public sector organizations want to meaningfully engage children, but they do not know the best way to do this; and second, public sector organizations believe that the current state of data literacy presents a large barrier for engagement with children and their families on topics such as the design, development and deployment of AI technologies that use their data.

Through preliminary conversations with public sector organizations, a few gaps in the existing landscape were also identified.

Ensuring the inclusion of and for children

The Institute carried out 14 hour-long semi-structured interviews with public sector organizations that provide services for children and families. The public sector organizations are based across the UK and include health, education, social services and policy bodies.



These semi-structured interviews covered an array of topics, including:

1. the organization's current use of AI technologies that employ children's data (if at all),
2. what they believe to be possible impacts of AI technologies used in the public sector on children and their families,
3. current policies that are used within their organizations to support the design, development, and deployment of AI technologies, and
4. specific questions about existing normative frameworks such as the UNICEF policy guidance and the General Data Protection Regulation (GDPR).

Some of these discussions are still ongoing and critical insights from these interviews are detailed in the following sections of this case study. The Institute received ethics approval for this second phase of research in late September 2021.

Preparing children for the present and future developments of AI

The team is working on making content accessible to children and adolescents for the second phase of their research. This will include the creation of child-friendly infographics and animations to help explain AI systems and ethical challenges of AI to children clearly and easily. The team will also consult children about public sector uses of AI by exploring children's interests, concerns, and current understandings of AI, and how children prefer to learn about applications of AI in the public sector. The Turing team continues to develop its methodology and content through conversations with Scotland's Data Lab and Children's Parliament. Both organizations have significant experience engaging with children across Scotland, and the Institute plans to draw on this expertise.

Results and findings

Most of the interviews and consultations highlighted challenges that the public sector faces when it comes to translating and implementing existing AI guidance with a child-centric perspective. They also raised further questions about how to effectively develop AI policies that can protect and empower children. There were several general categories of concern that the interviewees brought up, including:

- > Public sector organizations' perceptions of low data literacy among the public;
- > Policymakers' lack of understanding about the implementation of GDPR principles;
- > Policymakers' questions about how to make the UNICEF policy guidance actionable;
- > Tensions related to policymakers and the public's growing concerns over the potential negative impacts of AI technologies.

One of the chief concerns was public sector stakeholders' perceptions that levels of AI and data literacy are low among public audiences. Interviewees mentioned several times that the public perception of AI is not always aligned with the technological data-analytics capabilities of AI. Most mentioned that 'killer-robot' or 'Terminator-style' AI fears were often raised in public discourse. Counteracting these fears would prove to be an obstacle. Improving data literacy would need to go beyond simply reflecting, discussing and deconstructing public narratives in participatory settings. Public sector organizations would have to unpack over-simplified or over-hyped, negative AI narratives and help public audiences understand and develop their own informed perspectives of what AI is, how it works and what it is capable of.

There is a need for further engagement with the public to better understand opinions on these topics.

One of the chief concerns was public sector stakeholders' perceptions that levels of AI and data literacy are low among public audiences.

Further, the challenge of improving AI and data literacy is not simply an issue at the childhood level, but it is pervasive at all ages and across many different backgrounds.

Therefore, developing and implementing AI technologies must focus on building trust with diverse segments of society to help ensure that AI policies can be debated inclusively and openly in the public sphere.

Interviewees also stressed a lack of understanding and clarity about GDPR principles, particularly compliance with principles that aim to minimize the amount of data about children that is collected and stored (e.g. data-minimization and data-storage limitation). GDPR was viewed as an imperfect starting point for child-centred AI design. Many interviewees mentioned that the interpretation of GDPR requirements will not be standardized in all countries, leading to significantly different AI policies.

Interviewees also suggested that current age verification measures are ineffective, meaning that it is difficult to know when a child is accessing their AI systems. They asked: if an AI system is meant for an adult, but is being accessed by a child, should the system be designed to take that into account? Several interviewees mentioned that GDPR guidelines were not clear enough to provide a confident answer.

The UNICEF Policy Guidance on AI for Children, on the other hand, was well received by interviewees, who believed that its child-centred design requirements were a great starting point for AI development. One interviewee stated, “We’re really glad that the UNICEF guidance makes reference to development... [to] the whole lifecycle; at the moment we’re thinking about how children should be engaged or considered at each of those stages in the development lifecycle.” Another interviewee said, “This is an excellent set of principles, because ... children will be subjected to those technologies in a similar way that we all are, but they have less impact on how these technologies are created, and they have maybe less understanding, less agency to ... challenge some of those choices or situations.”

However, interviewees did mention that these requirements, as written, had shortcomings in their implementation and their use, including by vulnerable groups such as children. Several requests were also voiced to make the UNICEF policy guidance more accessible, but also to cover in more depth the following gaps, and concerns:

- > Break down recommendations by sector and include examples and resources for how to include children in participatory research and policy;
- > Make the language of the guidance more accessible to all audience and users, including children;
- > Offer more details about the risks of not abiding by child-centric requirements.



These concerns and requests represent a broad overview of the takeaways from the interviews, but again stress that interpretation of principles, along with perceived low rates of general AI and data literacy, are still significant barriers to their effective deployment.

Participants frequently suggested that it would be helpful to have more practical guidance on how to translate UNICEF child-centred requirements into an actionable set of methodologies and normative mechanisms. Importantly, interviewees also expressed concerns about how to navigate the ethical and policy landscape of child-centred AI, as they must understand and apply a growing set of responsible AI frameworks. Specific questions focused on how to use the UNICEF policy guidance along with other normative tools relevant to responsible use of children's data, such as GDPR and the UK Information Commissioner's Office's Age Appropriate Design Code.¹

Within the UNICEF policy guidance, three child-centred requirements were cited as being particularly vexing for public sector organizations: providing transparency, explainability, and accountability to children; preparing children for present and future developments in AI; and ensuring the inclusion of and for children. Interviewees mentioned that AI developments are often difficult for adults to understand, let alone children, but making AI policies understandable to children is a difficult goal to achieve. Recommendations on how to make this easier included making UNICEF and GDPR guidelines more specific and tailored to different fields. Many of the interviewees recommended providing examples of how to explain the issues clearly – in effect giving designers a set of boilerplate texts they could customize to explain their projects to the public.

Several interviewees concluded that the UNICEF policy guidance could theoretically prove effective in reducing the potential negative implications of AI – which they found to dominate public perception of the technology.

Participants frequently suggested that it would be helpful to have more practical guidance on how to translate UNICEF child-centred requirements into an actionable set of methodologies and normative mechanisms.

1. The Children's code (or the [Age appropriate design code](#)) is a framework of 15 standards that online services need to follow to fulfil their obligations under data protection law to protect children's data online.



Next steps

Consultations and interviews identified that enabling and promoting sustained public dialogue about AI was an important next step to overcoming the rising challenges that the public sector faces with implementing child-centric AI. Improving the guidelines and adding examples may help increase explainability to a degree, but the public and policymakers must first gain a working comprehension of AI. The sustained public dialogue must tackle prevailing negative narratives about AI. Beyond this, the public must have the opportunity to learn and better grasp what AI can and cannot do. Essentially, the public must gain a balanced base-level understanding of AI – one that is not driven by sensationalism or narratives that focus purely on the negative aspects.

Another critical next step for the Institute's research is to engage with children and their families directly. Their goal is to use this engagement to elucidate findings from the interviews with public sector professionals. By looking at AI from an end user standpoint, researchers will gain a better understanding of how children and guardians interact with AI technologies, how they comprehend it, and how they value data privacy and wish to be involved in conversations about the public sector's use of AI technologies that use their data. These insights will be incorporated into the Institute's child-centred workbook that will be published alongside the revised guidance.

About this case study

This case study is part of [UNICEF's Artificial Intelligence for Children project](#), which aims to better understand how AI-powered technologies can protect, provide for and empower children. Through a partnership with the Government of Finland, the project developed a draft [Policy Guidance on AI for Children](#), which contains practical recommendations and principles for governments and industry to create child-centred AI policies and systems. In order to inform and improve future versions of the guidance and inspire others, organizations were selected to pilot the guidance and share their diverse experiences to illustrate how AI-based initiatives could be designed to be more child-centred. The approaches taken, lessons learned and insights gathered will be one contribution to the global effort towards AI policies and systems that support children's development. View all of the case studies on our [website](#).

Contact

Morgan Briggs
mbriggs@turing.ac.uk

Acknowledgements

This case study was produced by the Office of Global Insight and Policy (OGIP) under the guidance of Melanie Penagos and Steven Vosloo, and was authored by Eleonore Pauwels. UNICEF is grateful to David Leslie, Morgan Briggs, and Mhairi Aitken for their support in developing this case study. It was copy-edited by Allison McKechnie, art directed by Kathleen Edison, and designed by Grace Leong. Cover: Photography by Åaker / Unsplash; artwork by MacWell.

UNICEF works in the world's toughest places to reach the most disadvantaged children and adolescents – and to protect the rights of every child, everywhere. Across 190 countries and territories, we do whatever it takes to help children survive, thrive and fulfill their potential of early childhood through adolescence. And we never give up.

The Office of Global Insight and Policy serves as UNICEF's internal think-tank, investigating issues with implications for children, equipping the organization to more effectively shape the global discourse, and preparing it for the future by scanning the horizon for frontier issues and ways of working. With dedicated expertise in seven policy areas – digital technology, human capital, governance, the environment, society, markets, and finance – the Global Insight team assists the organization in interpreting, and engaging in, a rapidly changing world. Visit us online to learn more: unicef.org/globalinsight

Office of Global Insight and Policy
United Nations Children's Fund
3 United Nations Plaza, New York, NY, 10017, USA

© United Nations Children's Fund (UNICEF), November 2021

This is a working document. It has been prepared to facilitate the exchange of knowledge and to stimulate discussion. The text has not been edited to official publication standards and UNICEF accepts no responsibility for errors. The statements in this publication are the views of the author(s) and do not necessarily reflect the policies or the views of UNICEF. The designations in this publication do not imply an opinion on the legal status of any country or territory, or of its authorities, or the delimitation of frontiers.



This document is interactive and designed for digital viewing.



Please consider the environment and refrain from printing.