Challenges and opportunities in translating ethical AI principles into practice for children

Ge Wang[1], Jun Zhao[1], Max Van Kleek[1], & Nigel Shadbolt[1]
[1]Department of Computer Science, University of Oxford

May 13, 2024

Abstract

AI systems are becoming increasingly pervasive within children's devices, apps, and services. The concern for a world where AI systems are deployed unchecked has raised burning questions about the impact, governance and accountability of these technologies. While recent effort on AI ethics has converged into growing consensus on a set of high-level ethical AI principles, the engagement with children's issues is still limited, and even less is known about how to effectively apply them in practice for children. This perspective first maps the current global landscape of existing ethics guidelines for AI and analyses their correlation with children. We then critically assess the strategies and recommendations proposed by current AI ethics initiatives, identifying the critical challenges in translating such ethical AI principles into practice for children. Finally, we tentatively map out several suggestions regarding embedding ethics into the development and governance of AI for children.

1 Main

Artificial Intelligence (AI) systems are fundamentally changing the world and affecting present and future generations of children. Children are already interacting with AI technologies in many different ways: embedded in the connected toys, smart home IoT technologies, apps, and services they interact with on a daily basis [1, 2]. Such AI systems provide children with many benefits, such as enjoyment and convenience from connected devices [3], personalised education and learning from intelligent tutoring systems [4], or online content monitoring and filtering by algorithms that proactively identify potentially harmful content or contexts [5]. Going forward, AI systems will, in all likelihoods, become altogether even more pervasive in children's lives simply due to their unprecedented capability in creating compelling, adaptive, and personal user experiences. Yet, despite its enormous potential, AI presents challenges for children, including biases affecting vulnerable sub-groups [6], unforeseen negative consequences [7], and looming privacy risks from extensive data collection practices [8]. Over recent years, significant efforts have been made to regulate ethical AI [9]. While

there is growing consensus about what the principles require, in general, engagement on children's issues is still largely lacking and limited. Though the AI principles remain valid in cases involving children, the unique characteristics and rights of children necessitate a more nuanced approach. Meanwhile, while various ethical principles have been proposed to safeguard the rights of children, their effective implementations and practical applications remain relatively unexplored. It is thus crucial to address the challenges that emerge when translating these principles into real-world scenarios for children. This ensures that the benefits of AI are maximized while minimizing its potential harms.

In this perspective, we undertook an analysis of existing AI ethics guidelines, considering the unique characteristics, rights, and needs of children in the digital world. From this examination, we developed a synthesised set of terminologies, drawing from both the ethical principles within digital environment and those extending beyond. Recognizing the challenges inherent in tailoring these principles to suit children's distinct needs, we delineated a roadmap for prospective inquiries aiming at establishing child-centred AI.

2 A review of ethical AI principles and how they relate to children

To establish the scope of this paper, we start by asking: How are 'children' characterized in current ethical AI principles, and how are their rights framed? According to the United Nations Convention on the Rights of the Child (UNCRC), a 'child' is defined as anyone under the age of 18 [7]. While this is a broad definition, it is directly adopted by numerous major ethical AI guidelines, often treating 'children' as a singular category without further exploration.

Meanwhile, the UN Committee on the Rights of the Child's endorsement of General Comment 25 in February 2021 provides a milestone guidance on children's rights within the digital domain, outlining four key principles essential for upholding children's rights in a digital context: 1. **non-discrimination**: Ensure all children have equal access to meaningful digital experiences; 2. **best interests of the child**: Prioritize children's welfare in all decisions and actions affecting them; 3. **right to life and development**: Protect children from digital threats like violent content, harassment, and exploitation; emphasize technology's influence during early childhood and adolescence and educate caregivers on safe usage; 4. **respect for the child's views**: Promote children's expression on digital platforms, integrate their input into policies, and ensure service providers respect their privacy and thought freedom. Our reviews of major ethical AI frameworks identified key themes and principles for upholding these children's rights from the General Comment 25. Moreover, we examine the varied terminologies employed within these principles across different disciplines and domains.

Fairness, equality, inclusion and access. This theme resonates deeply with the principle of children's right to non-discrimination. By setting a standard of

non-discriminatory harm, it becomes crucial for AI system designers to prevent unfair and unequal consequences across different communities [10]. However, there's a noticeable gap in research focusing on *child-related fairness*. Frameworks, particularly from entities like UNICEF, stand out for their emphasis on advocating for marginalised children and the importance of diversifying datasets to minimise biases [2]. Several resources also champion the cause of universal digital access for every child, ensuring no discrimination based on gender, disability, or ethnicity [7]. The push for diversity in AI system design and the heightened call for active child participation in AI policy-making and design processes is evident [2, 11]. The conceptualisation of fairness within AI is often viewed through different lenses based on professional backgrounds. For educators, fairness is about tailoring learning experiences to meet individual needs and ensuring equitable access to quality education for all children, irrespective of their backgrounds [12]. Child psychologists focus on age-appropriate AI interactions and equal treatment for children with developmental challenges [13]. On the other hand, AI engineers often aim to address fairness technically, developing unbiased algorithms and striving for uniform service quality [14]. Yet, the practicality and common use of such methods in the industry remain debatable. Experts in AI ethics [15, 16] have noted that this technical approach to fairness isn't always feasible or consistent across the industry. Moreover, a purely technical perspective on fairness often overlooks the complex, varied realities of how individuals experience fairness in real-world scenarios.

Transparency and accountability. Transparency and accountability are often brought up together in the literature, and is closely related to supporting children's best interest in the design of systems. Accountability requires an ability to identify a chain of responsibility for system (mis)-behaviours, and justify how the designers and developers of AI systems should be held accountable [10]. Only very few AI frameworks mentioned the importance of extra attention paid to systems that could be accessed by children, in particular, the importance of impact assessments [17]. The AI frameworks proposed by UNICEF and the UN were among the very few ones that urged for constant review, update and refinement to integrate children's rights [2]. References to transparency comprise efforts to improve the understandability of information given [8], enable caretakers of children to understand impact on children [18], as well as making information accessible [8, 13]. It is essential to recognise that the interpretations of transparency and accountability can vary and emphasise different aspects across various sectors. In education, transparency clarifies AI-driven learning, personalization, and data protection [19]. For children's online safety, it involves clear data handling and algorithmic safety communication [8]. In healthcare, it addresses AI's role in diagnostics and handling children's health data [20].

Privacy, manipulation and exploitation. Privacy is a recurring theme in current literature, typically discussed in relation to data protection [8], data security [21], and data trust [22]. For child-specific data privacy, numerous frameworks highlighted the importance of regulating data practices for children, including setting higher default privacy settings on child-accessible systems, retaining minimal personal data, and avoiding sharing such data if detrimental

effects are foreseeable [8, 17, 10]. Privacy is closely linked to preventing data exploitation and manipulation, particularly concerning AI's data-driven personalized targeting methods and their potential harms in various contexts [17]. Current principles emphasize a heightened scrutiny of AI systems, particularly with respect to their impact on children's behavior or emotions, offering concrete examples of behavioral manipulation for practitioners to avoid, such as using personal data to incentivise engagement, nudging children to continue playing by implying potential loss, or exploiting children's vulnerabilities through the profiling of their personal data [7, 8, 23]. Beyond the commercial aspect of children's data, the professional use of their data in areas like paediatric bioethics brings to light the intricate balance between the privacy rights of professionals, parents, caregivers, and the children themselves. A significant challenge arises when whether a child possesses the capacity provide informed consent, either due to their age or cognitive ability. The enduring ethical obligation remains: protecting a child's confidentiality, respecting parental access, and fostering a constructive parent-child relationship.

Safety and safeguarding. The term 'harm' is central to safety and safeguarding in AI usage, with the aim to prevent harm to users and to shield users from harmful effects [24]. Broad references to safety include general pleas for safety and security, or an expectation that AI should avoid causing predictable or unintentional harm. This concept is closely related to children's right to life and development, necessitating more nuanced considerations by considering the biological and psychological distinctions between children and adults, as well as recognizing that children may interact with digital services and apps in unforeseenable ways. Meanwhile, interpretations of safety can also vary across different domains. In the educational sector, this relates to educational disparities caused by AI tools, misinformation affecting learning outcomes, or the emotional distress caused by biased AI assessment. Within healthcare, safety can encompass erroneous AI-generated diagnoses, flawed treatment recommendations, or the misuse of personal health data that result in privacy breaches [20]. When considering media interaction, safety entails protecting children from encountering inappropriate content, cyber bullying, online grooming, excessive screen time, and exploitative information [8, 25, 26]. This underscores the necessity for AI systems that can effectively shield children from harmful and unreliable content, while simultaneously upholding their right to freedom of expression.

Age-appropriateness and voice of own. This theme is most closely aligned with children's rights to be heard. The term "developmental stage" is frequently brought up in related references [7, 11, 17], encouraging stakeholders to respect the evolving capacities of children as an enabling principle that addresses the process of their gradual acquisition of competencies, understanding and agency [7]. Statements have also been made around how special attention should be paid to the effects of technology in children's earliest years of life, and to support relationships with parents and caregivers, which is crucial for shaping children's cognitive, emotional and social development [7, 19]. This sentiment often extends to fostering children's voices, appropriate to their age and developmental stages. AI design codes for children have underlined the significance of

designing not just for children, but with them [27]. Practical implementations of these developmental considerations could include age-adapted transparency [28] and control mechanisms for children at different developmental stages [2]. In addition, adopting methodologies that actively involve children and encourage their contributions could also cater to their developmental needs [27].

3 The challenges of translating ethical AI principles into practice for *children*

Translating principles into practice is a well-known challenge, yet the distinct context of doing so for children introduces specific complexities. Children are inherently different from adults in their moral significance. Irrespective of their intellectual capabilities, children do not possess the same level of autonomy and responsibilities as adults. As such, AI principles for children should not be treated merely as an extension of general user guidelines or a subcategory of guidelines for socially vulnerable groups. Given children's distinct attributes and circumstances, we underscore four major challenges in adapting ethical AI principles for their benefits:

Lack of consideration on developmental aspect of childhood. The key challenge in translating AI principles to practice, as highlighted by numerous studies, is the absence of consistent professional codes and norms for AI applications, due to the vast amount of different forms of applications and technologies aiming at different goals with different stakeholders [29, 30, 31]. Incorporating children into the AI ethics conversation introduces a new layer of complexity due to their diverse needs, age ranges, development stages, backgrounds, and characters. Their unique physical and psychological traits necessitate special care in the deployment of AI systems that shape their information, services, and opportunities.

As previously indicated, current ethical AI frameworks either overlook the distinct role of children or categorize them as a homogeneous group. The integration of children into these principles seems more like a superficial gesture of compliance, failing to truly address their distinctive needs and viewpoints. In fact, in many instances, the term 'children' could be swapped out for 'socially vulnerable groups' without significant change to the content or context. To contemplate the ethical nuances associated with the involvement of children and adolescents in research, we must delve deeper into the wide-ranging notion of 'childhood'. A critical feature distinguishing childhood from general users is its developmental progression – transitioning from the utter dependence of infancy to the comparative self-reliance of youth. Implementing straightforward age categories is insufficient due to the vast differences in children's intellectual abilities, pace of growth, maturity, and experiences. Therefore, to ensure ethical considerations match the unique requirements and circumstances of each child, it's crucial to adopt a more nuanced approach beyond mere age-specific classifications.

The current deficiency in diverse considerations further reflects the lack of concrete conceptualization around what "children's best interests" truly signify. Often referenced in literature, the term is generally employed to denote a child's overall well-being, development, and protection in a somewhat vague manner. The traditional interpretation of this term may not entirely capture or may need modification when applied to children in diverse circumstances. Essentially, this suggests that our usual understanding of certain terms may be inadequate or necessitate refinement, particularly when considering the specific contexts and characteristics of children.

Lack of consideration on the role of guardians in childhood. One aspect that distinctly separates 'children' from other general users is the presence of parents or guardians. They hold significant legal and ethical roles in making decisions for their children. It's crucial to acknowledge the meaningful moral distinction between 'competent children' and 'adults'. Despite their intellectual abilities, children do not possess the same rights or responsibilities as adults. Ethically and legally, parents bear this responsibility. Consequently, it's essential to examine the roles of parents in this context, considering their unique interests, which may differ from those of other 'adults'. Yet, existing references on ethical AI principles reveal a significant gap in addressing surrogate or substitute decision-making for children, such as those decisions made by parents. The fact that many ethical AI frameworks scarcely address the roles of children makes it unsurprising that the interactions between children and their guardians are infrequently discussed. Such lack of oversight in examining parental decision-making may inadvertently increase risk exposure for children.

On the other hand, the few frameworks that do address parents and families, such as those by UNICEF, often adhere to a traditional assumption and portray their roles as possessing superior expertise and skills to navigate their children through the digital landscape and facilitate their learning. The topic of effectively bolstering children's personal resilience development, however, is rarely addressed. While parents undoubtedly play a pivotal role in protecting their children in the digital world, children, born into the digital era, interact with the AI environment with an ease that is almost instinctive. In contrast, their parents may lack the same depth of understanding in this domain. [32]. This potential expertise shift underscores the need to transition from a parent/teacher-led to a child-centered approach [33]. This involves moving from *instruction* to *support* in children's experiences, fostering self-determination, values, and self-identity. This aligns with the child-computer interaction community's trend [34] to promote children's autonomy and resilience in AI interactions, including critical comprehension of digital environments and informed choices. Guardians (parents, teachers, or carers) and children should collaborate, yet current ethical AI principles lack guidance on enhancing joint consent and decision-making processes.

Lack of child-centred evaluations considering children's best interests and rights. One key challenge of translating ethical AI principles into practice for children is the difficulty in translating the high-level principles to quantifiable outcomes or technical standards. A recent survey on 188 AI systems developed

for children showed that almost all of them relied solely on technical evaluations to measure their performance, such as through the accuracy, precision and recall of the results generated by the systems [31]. In fact, quantitative measurements have long been considered the gold standard for performance assessment in the fields of AI, algorithms, and related disciplines. This approach has provided a seemingly objective and standardized method for evaluating the effectiveness of various systems and models. However, as these fields continue to advance, it has become increasingly clear that relying solely on quantitative metrics can present challenges in meeting certain empirical requirements. For instance, for the principle of safety and safeguarding, it is not surprising to see that within the current AI community, such a principle/requirement on safeguarding would be just directly translated into identifying online inappropriate content, evaluated by the accuracy of its classification. Another example is the principle of sustainability and age-appropriateness – how the developmental needs and long-term well-being of children could be even evaluated by technical measures remains questionable.

This is not to say that technical evaluations are not important – they are, but translating principles into practice requires more than that. The current trend of relying solely on quantitative measurements may steer the AI community away from prioritizing human-centred factors when designing for children. As a result, critical needs, such as understanding how users desire to be treated by the developed systems, may be overlooked. This mandates a more balanced approach that prioritizes both empirical variables and quantitative measurements; and more fundamentally, it necessitates a paradigm shift towards cultivating a more human-centred approach within the AI community.

Lack of a coordinated, cross-sector and cross-disciplinary approach. Meanwhile, a pertinent challenge arises from the absence of supportive resources, theories, and empirical evidence that can facilitate the translation from principle to practices. Ethical AI principles pertaining to children's rights, such as those championed by organisations like UNICEF and UK ICO [2, 8], predominantly focus on domains like education, health, privacy, and online safety. Inadequate discussions have been presented concerning potential risks, disadvantages attributed to fairness and non-discrimination oversights, or the deficiency of robust legal and professional accountability mechanisms.

Conversely, it is interesting to see that while there have been limited shared resources on developing AI for children, experts from other domains (e.g., law, psychology) often have their own established codes of practice and substantial knowledge bases to draw on. It is even more interesting to observe that sometimes experts from different domains work on analogous issues but use completely different vocabularies and methodologies. The previous discussion in section 2 illustrated how diverse terminologies can convey disparate meanings and contexts, depending on the domain of expertise. The crux of the challenge lies in the adaptability of these terms and methodologies across different AI principles. This highlights the need for strengthening collaboration within the child-computer-interaction community, to harmonize multidisciplinary domains and encourage knowledge transfer to avoid duplicate efforts. Ultimately, this

cross-sector and cross-disciplinary cooperation could play pivotal role in safeguarding children's interests, rights, and well-being in AI system development.

4 Future directions on ethical AI for children

While there is a rising sense of urgency to focus on child-centred AI, marked by significant initiatives like 'Exploring Children's Rights and AI' at the Alan Turing Institute, and 'Child Rights by Design' by the Digital Futures Commission; our approach uniquely aims to bridge the gap between theoretical ethical AI principles and their real-world application in contexts centered on children. Within this larger, imperative framework of setting the agenda, we present preliminary recommendations that explicitly link ethical AI considerations to actionable guidelines in the realm of child-centred AI technologies.

Increased stakeholders' involvement. Our analysis indicates that the current principled approach of AI for children fails to take into account critical considerations regarding the best interests of children, and is ambiguous with regard to the actual needs and empirical requirements of both children and their families as they were rarely consulted. On the other hand, research indicates that stakeholders, including children, parents, and other caretakers, strongly desire a say in defining how they should be treated by AI systems [35]. We recommend future designers and developers on AI for children to take a more participatory and more inclusive approach, to include stakeholders including parents, schools and teachers, practitioners, and most significantly, children themselves from diverse backgrounds to better understand what is actually needed by different stakeholders. In recent years, participatory methods involving stakeholders have gained popularity in the field of human-computer interaction (HCI). These approaches include organizing focus groups with parents, collaborating with educators to develop AI tools that align with curriculum needs and enhance learning, and consulting child psychologists to ensure AI technology is age-appropriate [36]. Additionally, there has been significant effort in specialized fields like healthcare to incorporate the perspectives of children and young people regarding the role of AI in medicine [37] and clinical care [38]. Specifically, the Child-Computer Interaction community has increasingly emphasized methodological approaches that involve children directly in the design process [39, 40, 41]. This movement advocates for empowering children by giving them a voice in these processes, thus supporting their autonomy and fostering resilience development. However, such effort and methodologies were typically utilized primarily in the HCI field; as discussed before, the prevailing convention still revolves around pure technical and quantitative evaluation metrics within the AI and algorithm community – which leads us to the second recommendation.

Direct support for industry designers and developers. To ensure the ethical AI principles actually get implemented in practice, we must recognise that it is crucial to *directly* involve designers and developers in the process of developing these ethical AI principles and building associated best practice

guidelines to transform these principles into practice. Recent research has shown that the lack of guidance in navigating the often abstract landscape of AI principles for children, as well as the treacherous landscape of existing tools such as third-party services [42] remain a critical bottleneck for industrial practitioners in the journey of creating ethical technologies - industry support is lacking for developers to interpret the guidelines, and supporting resources (e.g., existing libraries and tools) are scarce in translating such principles. One way to address this open challenge is to create mechanisms to incentivise community building amongst practitioners, designers and developers, facilitating their sharing and building of knowledge and even the development of momentum for fundamental changes. Furthermore, we suggest that ethical AI practitioners and organizations should increase their collaboration with developers and designers in order to take a bottom-up approach and create a shared foundation for industry standards and best practices. By bringing together the knowledge and expertise of these multiple stakeholders, we expect this will not only catalyse practical and workable guidelines and resources but also promote a culture of accountability and continuous improvement.

Establishment of legal and professional accountability mechanisms. Recent regulatory efforts such as the Online Safety Bill [26], the EU Artificial Intelligence Act [43], and the Algorithmic Accountability Act [44] have taken the first steps towards addressing the challenges of regulating AI. While these regulations made a good start in promoting the responsible and accountable use of AI technology, the importance of child-specific legalisation is often underestimated. Although certain legislation such as FTC COPPA and Online Safety Bill mentioned aspects around children, such legislation typically focused more on children's general well-being online and less on AI-specific impact and harms. Meanwhile, existing AI regulatory effort often focuses on specific sectors or applications of AI, and often underscores the importance of collaboration between different stakeholders, including policymakers, industry players, and civil society groups. Finally, we often hear statements suggesting that regulations may never fully catch up with the rapid pace of technological development. This sentiment is sometimes used as an excuse for the absence of up-to-date legislation and professional standards. Future legislative efforts could use the UNCRC's position on children's digital rights as a foundation. These initiatives could then explore how to adapt children's basic human rights within the context of the digital world.

Increased multidisciplinary collaboration around a child-centred approach. One of the challenges in converting abstract principles into a reliable set of guidelines for children is mainly due to the inadequate availability of resources. For instance, while all working towards designing for a better experience for children, researchers from HCI and design domains may typically focus on the interaction between children and AI, along with their user experience and perceptions on a specific topic [45]; whereas researchers from education domain may focus more on children's learning performance and long-term behavioural change [46]; likewise, the work from researchers in policy guidance domain may be more heavily oriented around how AI for children could be as-

sociated with greater societal impact [2]. By fostering collaborations amongst experts from various domains such as HCI, design, algorithms, policy guidance, data protection law, and education, along with various practices from within the AI and related communities, we emphasize the importance of uniting voices that might deploy different terminologies. Ultimately, we assert that there's a compelling necessity to forge a transformative discipline as we've noticed a disconnect in the knowledge and methodologies employed by researchers and practitioners across various disciplines. This necessitates a radical rethinking of our disciplinary approaches, transcending boundaries and integrating the strengths and perspectives of numerous disciplines. In particular, this new interdisciplinary sphere should strongly advocate for a child-centred approach, wherein the needs, experiences, and perspectives of individuals, particularly children, are at the forefront of design and implementation. This integrated focus will enable us to devise future ethical AI systems that are equipped to address and be well-prepared for the unique socio-technical challenges pertinent to creating AI systems for children.

References

[1] Li, J.-P. O. *et al.* Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective. *Progress in retinal and eye research* **82**, 100900 (2021).

[2] Sims, A. *et al.* Unicef public consultation on draft policy guidance on ai for children. `https://www.unicef.org/globalinsight/media/2356/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf`. (UNICEF, 2022).

[3] Strengers, Y., Kennedy, J., Arcari, P., Nicholls, L. & Gregg, M. Protection, productivity and pleasure in the smart home: Emerging expectations and gendered insights from australian early adopters. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. `https://doi.org/10.1145/3290605.3300875`. (ACM, 2019).

[4] Fadhil, A. & Villafiorita, A. An adaptive learning with gamification & conversational uis: The rise of cibopolibot. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, 408–412. `https://doi.org/10.1145/3099023.3099112`. (ACM, 2017).

[5] Kumaresamoorthy, N. & Firdhous, M. An approach of filtering the content of posts in social media. In *2018 3rd International Conference on Information Technology Research*, 1–6. `https://doi.org/10.1109/ICITR.2018.8736152`. (IEEE, 2018).

[6] Corbett-Davies, S., Pierson, E., Feller, A., Goel, S. & Huq, A. Algorithmic decision making and the cost of fairness. In *Proceedings of the*

23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 797–806. https://doi.org/10.1145/3097983.3098095. (ACM, 2017).

[7] General comment no. 25 (2021) on children's rights in relation to the digital environment. https://www.ohchr.org/EN/HRBodies/CRC/Pages/GCChildrensRightsRelationDigitalEnvironment.aspx. (UN-CRC, 2020).

[8] Age appropriate design code. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/. (ICO, 2020).

[9] Jobin, A., Ienca, M. & Vayena, E. The global landscape of ai ethics guidelines. *Nature Machine Intelligence* **1**, 389–399 (2019).

[10] Leslie, D. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of ai systems in the public sector. Preprint at https://arxiv.org/pdf/1906.05684.pdf (2021).

[11] Guidance for regulation of artificial intelligence applications. https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf. (The White House, 2020).

[12] Adams, C., Pente, P., Lemermeyer, G. & Rockwell, G. Artificial intelligence ethics guidelines for k-12 education: a review of the global landscape. In *Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14–18, 2021, Proceedings, Part II*, 24–28 (Springer, 2021).

[13] Adams, C., Pente, P., Lemermeyer, G. & Rockwell, G. Ethical principles for artificial intelligence in k-12 education. *Computers and Education: Artificial Intelligence* 100131 (2023).

[14] Artificial intelligence and machine learning: Policy paper. https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/. (Internet Society, 2017).

[15] Greene, D., Hoffmann, A. L. & Stark, L. Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. *Proceedings of the 52nd Hawaii International Conference on System Sciences* https://hdl.handle.net/10125/59651. (2019).

[16] McCradden, M. D., Joshi, S., Mazwi, M. & Anderson, J. A. Ethical limitations of algorithmic fairness solutions in health care machine learning. *The Lancet Digital Health* **2**, e221–e223 (2020).

[17] Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206. (EUR, 2021).

[18] Touretzky, D., Gardner-McCune, C., Martin, F. & Seehorn, D. Envisioning ai for k-12: What should every child know about ai? In *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, 9795–9799. https://doi.org/10.1609/aaai.v33i01.33019795. (ACM, 2019).

[19] Goralski, M. A. & Tan, T. K. Artificial intelligence and sustainable development. *The International Journal of Management Education* **18**, 100330 (2020).

[20] Ethics and governance of artificial intelligence for health: Who guidance. https://www.who.int/publications/i/item/9789240029200. (World Health Organization, 2021).

[21] Roumate, F. *Ethics of Artificial Intelligence, Higher Education, and Scientific Research*, 129–144. https://doi.org/10.1007/978-981-19-8641-3_10. (Springer Singapore, 2023).

[22] Aayog, N. Discussion paper: National strategy for artificial intelligence. *New Delhi: NITI Aayog. Retrieved on January* **1**, 2019–01 (2018).

[23] Gupta, M. & Sharma, A. Fear of missing out: A brief overview of origin, theoretical underpinnings and relationship with mental health. *World Journal of Clinical Cases* **9**, 4881 (2021).

[24] Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar, M. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for ai. *Berkman Klein Center Research Publication* (2020).

[25] Lorenz, B., Kikkas, K. & Laanpere, M. Comparing children's e-safety strategies with guidelines offered by adults. *Electronic Journal of e-Learning* **10**, 326–338 (2012).

[26] Online safety bill. https://commonslibrary.parliament.uk/research-briefings/cbp-9579/. (UK Parliament, 2022).

[27] Child rights by design. https://childrightsbydesign.digitalfuturescommission.org.uk/page/childrens-voices. (5RightsFoundation, 2023).

[28] Markopoulos, P., Read, J. C. & Giannakos, M. Design of digital technologies for children. *Handbook of human factors and ergonomics* 1287–1304 (2021).

[29] Mittelstadt, B. Principles alone cannot guarantee ethical ai. *Nature machine intelligence* **1**, 501–507 (2019).

[30] Prunkl, C. E. *et al.* Institutionalizing ethics in ai through broader impact requirements. *Nature Machine Intelligence* **3**, 104–110 (2021).

[31] Wang, G., Zhao, J., Van Kleek, M. & Shadbolt, N. Informing age-appropriate ai: Examining principles and practices of ai for children. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3502057. (ACM, 2022).

[32] Fletcher, A. C. & Blair, B. L. Maternal authority regarding early adolescents' social technology use. *Journal of Family Issues* **35**, 54–74 (2014).

[33] Langford, R. Critiquing child-centred pedagogy to bring children and early childhood educators into the centre of a democratic pedagogy. *Contemporary Issues in Early Childhood* **11**, 113–127 (2010).

[34] Wang, G., Zhao, J., Van Kleek, M. & Shadbolt, N. 12 ways to empower: Designing for children's digital autonomy. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3544548.3580935. (ACM, 2023).

[35] Wang, G., Zhao, J., Van Kleek, M. & Shadbolt, N. 'treat me as your friend, not a number in your database': Co-designing with children to cope with datafication online. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3544548.3580933. (ACM, 2023).

[36] Steen, M. Co-design as a process of joint inquiry and imagination. *Design issues* **29**, 16–28 (2013).

[37] Visram, S., Leyden, D., Annesley, O., Bappa, D. & Sebire, N. J. Engaging children and young people on the potential role of artificial intelligence in medicine. *Pediatric Research* **93**, 440–444 (2023).

[38] Thai, K. *et al.* Perspectives of youths on the ethical use of artificial intelligence in health care research and clinical care. *JAMA Network Open* **6**, e2310659–e2310659 (2023).

[39] Druga, S., Yip, J., Preston, M. & Dillon, D. The 4as: Ask, adapt, author, analyze-ai literacy framework for families. *Algorithmic Rights and Protections for Children* (2021).

[40] Lee, K. J. *et al.* The show must go on: A conceptual model of conducting synchronous participatory design with children online. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, 1–16. https://dl.acm.org/doi/10.1145/3411764.3445715. (ACM, 2021).

[41] Woodward, J. *et al.* Using co-design to examine how children conceptualize intelligent interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. https://dl.acm.org/doi/10.1145/3173574.3174149. (ACM, 2018).

[42] Ekambaranathan, A., Zhao, J. & Van Kleek, M. "money makes the world go around": Identifying barriers to better privacy in children's apps from developers' perspectives. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. https://dl.acm.org/doi/fullHtml/10.1145/3411764.3445599. (ACM, 2021).

[43] Kop, M. Eu artificial intelligence act: The european approach to ai, transatlantic antitrust and ipr developments. *Transatlantic Antitrust and IPR Developments* (2021).

[44] MacCarthy, M. An examination of the algorithmic accountability act of 2019 http://dx.doi.org/10.2139/ssrn.3615731. (2019).

[45] Zimmerman, J. & Forlizzi, J. Research through design in hci. *Ways of Knowing in HCI* 167–189. (Springer, 2014).

[46] Sandoval, W. Conjecture mapping: An approach to systematic educational design research. *Journal of the learning sciences* **23**, 18–36 (2014).

Acknowledgments

This work was supported by the Ethical Web and Data Architectures project (grant number EWADA).

Competing Interests

The authors declare no competing interests.

Correspondence

Correspondence and requests for materials should be addressed to Ge Wang (email: ge.wang@cs.ox.ac.uk).

Contributions

G.W. conceptualized and wrote the manuscript. J.Z., M.V.K., and N.S. provided critical feedback. All authors read and approved the final manuscript.