DIGITAL TRANSFORMATIONS FOR HEALTH LAB

GOVERNING HEALTH FUTURES 2030

# Shaping AI governance for young people

## Perspectives on health, policy and regulation

This report was developed in 2025 by Digital Transformations for Health Lab (DTH-Lab). It has been made possible by financial contribution from Fondation Botnar, Switzerland, to Project IMG-22-005 at Digital Transformations of Health Lab (DTH-Lab). Fondation Botnar's commitment to advancing global health is deeply appreciated, and this project would not have been possible without their contribution. DTH-Lab is hosted by Université de Genève (UNIGE), Switzerland. DTH-Lab is committed to ensuring and enabling Global Access. The knowledge and information gained from the project will be promptly and broadly disseminated and its 'Funded Developments' will be made available and accessible free of costs and the Global Access Commitments will survive the term of the project.

Authors: Courtney B. Doagoo, Rebecca Raeside, Shajoe J. Lake, Erza Selmani, Aferdita Bytyqi, Ilona Kickbusch, Rohinton Medhora and Ananya Choyal

Acknowledgements: We are grateful to Rohinton Medhora, who served as the chair for the series, to Aferdita Bytyqi for conceptualizing and guiding the work and to Louise Holly for her editorial guidance.

We also thank Eric Sutherland, from the OECD Secretariat, for his valuable inputs (please note the views expressed here do not necessarily reflect the official views of OECD Member countries).

We extend our thanks to Eithne Staunton for her careful copy-editing, to Corrie Fairweather-Mills for coordinating the design process and to the broader team involved in the production and design of this collection.

Design: Janhavi Nikharge
Photo credits: Photo by Google DeepMind on Unsplash

Suggested citation: Digital Transformations for Health Lab (2026) Shaping AI governance for young people: Perspectives on health, policy and regulation. Geneva: Digital Transformations for Health Lab.

Date of publishing: May 2026

# Contents

# Foreword

**Aferdita Bytyqi**
Executive Director and Founding Partner, DTH-Lab

**Ilona Kickbusch**
Director and Founding Partner, DTH-Lab

Artificial Intelligence (AI) is reshaping health systems at a pace that outstrips the governance frameworks designed to guide it. The four papers gathered in this collection address that gap – empirically, critically and urgently. They emerge from a community of researchers and practitioners who share a conviction that responsible innovation in health AI is not solely a technical problem, but a question of governance, equity and political will.

Digital Transformations for Health Lab (DTH-Lab) was established to sit at precisely this intersection. Operating as an independent, globally oriented think tank, DTH-Lab works to strengthen the governance of digital health by generating evidence, convening expertise and translating knowledge into actionable policy. Our work is grounded in the recognition that the benefits of digital transformation in health are not inevitable: they must be actively shaped.

A defining feature of our approach is our commitment to the next generation of researchers, particularly those from low- and middle-income countries (LMICs). Too often, global health AI discourse is shaped by voices from a narrow set of institutions, demographics and geographies. DTH-Lab actively works to widen that circle.

We offer structured and substantive opportunities for young researchers from LMICs to collaborate with DTH-Lab – not as observers, but as co-authors, co-designers

and co-presenters. For example, in December 2025, DTH-Lab led a workshop with IFMSA on Youth-Led Approaches for Inclusive AI in Health at the Health AI Global Governance Forum in Nairobi, Kenya. This landmark convening brought perspectives from the Global South into the heart of an agenda too often determined in Geneva, Paris or Washington.

This collection also reflects DTH-Lab's collaboration with the OECD and engagement with the OECD AI in Health Expert Group, of which Executive Director Aferdita Bytyqi is a member. Together, we have pursued a programme of work designed to accelerate the responsible scaling of AI in health. In March 2026, DTH-Lab and the OECD co-hosted a high level stakeholder working meeting in Geneva, where representatives from 14 international agencies advanced the expert-informed Action Plan on the Responsible Scale of AI in Health, which will be further developed at the International Conference on AI in Health in Madrid in May 2026. The papers in this series form part of the intellectual scaffolding for that plan. DTH-Lab's presence at the Conference and involvement in shaping the plan, represent milestones in a sustained effort to ensure that global policy on AI in health is evidence-based, inclusive and implementation-ready.

These papers demonstrate the calibre of insight that emerges when young researchers are given the platform and institutional support to lead. They share a determination that AI in health must serve future generations and the many, not merely the few. We commend them to researchers, policymakers and practitioners alike, and invite those who share these ambitions to join us in our work.

# Acronyms

| | |
|---|---|
| **AI** | Artificial Intelligence |
| **EC** | European Commission |
| **EU** | European Union |
| **FDA** | Food and Drug Administration |
| **OECD** | Organisation for Economic Co-operation and Development |
| **UDHR** | Universal Declaration of Human Rights |
| **UNCRC** | United Nations Convention on the Rights of the Child |
| **UNESCO** | United Nations Educational, Scientific and Cultural Organization |
| **UN** | United Nations |
| **UNICEF** | United Nations International Children's Emergency Fund |
| **WEF** | World Economic Forum |
| **WHO** | World Health Organization |

# AI, health and young people: Introduction to the series

**Rohinton Medhora**
Founding Member, DTH-Lab

**Ananya Choyal**
Policy and Research Associate, DTH-Lab

With new technologies in general and artificial intelligence (AI) in particular, the pace and magnitude of change is unprecedented. Yet our capacity to understand this change, to maximize its benefits while minimizing its risks, continues to lag behind. Frameworks to get this balance right are nascent and diverse. As the four papers in this series show, the health of young people is a special case in a complex field. In this series, we hope to move towards a better understanding of the relationship between the lives of young people and AI, health and policy.

Technology shapes health determinants and health outcomes through several channels, starting with the normative framework that should guide governance systems for digital health. The Universal Declaration of Human Rights (UDHR) extends beyond any particular political system, with all 193 United Nations member states having signed up to it. Article 25 of the UDHR explicitly covers medical care and the "right to a standard of living adequate for the health and well-being" and the WHO treats the "right to health" as a core tenet (United Nations [UN], 1948).

Near-ubiquitous digitization of everyday life and health systems has changed the parameters within which the right to health must be interpreted and implemented, raising new questions about how emerging technologies may affect the well-being and development of young people. For example, advances in neural technologies – while holding great promise for their health applications in augmenting or restoring sensory, motor or cognitive functions like sight and hearing – may involve external monitoring of what we are thinking. This poses profound risks to our "right to think", an explicit and hitherto under-appreciated facet of the pantheon of human rights (Alegre, 2022). Another example of the interplay between health, digitization and rights shows up in the consequences of excessive screen time on cognitive abilities and personality development. This is especially true for children and adolescents whose heightened neuroplasticity makes them particularly sensitive to its effects. In light of these negative effects on young people's cognition, differentiating between what should be a right for children and what might be considered a desirable policy goal for older populations might be in order (Li et al., 2025; Doidge, 2018).

Similar questions arise in relation to the right to privacy. With tracking technologies essentially unavoidable in public spaces and frequently also in the seemingly private experience of using the Internet, the "right to privacy" enshrined in Article 12 of the UDHR must be interpreted differently today than when it was written, over seven decades ago (United Nations [UN], 1948).

As early views of the Internet as being empowering – by making information cheap and easily accessible – have given way to concerns about misinformation and disinformation, the idea emerges of a "right to clean information," akin to the right to clean air and water. Current debates around vaccine scepticism deliberately driven by bad information, the proliferation of diet and health cures swirling unabridged in cyberspace and algorithm-based decision-making all suggest that the right to clean information is a key guidepost for designing sound health governance in the AI era.

Beyond questions of how existing rights such as those relating to health and privacy should be interpreted, the growing role of data-driven technologies in health raises broader concerns about equity and participation. The digital divide has evolved from the 1990s when the phrase came into existence. Data from 2022 show that levels of Internet and mobile Internet use follow a familiar pattern, with near-ubiquitous use in North America and Europe (about 87 per cent of the total population), high usage in East Asia and Latin America (about 76 per cent) and with modest usage in South Asia (42 per cent) and Sub-Saharan Africa (34 per cent). But the gap is considerably smaller now

than it was at the advent of the digital era (International Telecommunication Union, n.d.; World Bank, n.d.; Tuetem, 2025). Still, what matters is not the technical divide but its implications for human development and capabilities in their fullest sense. This is especially the case when the rise of data as a factor of production – driving economic value and production – and as a social feature – shaping how individuals interact, participate and are represented – enters the picture. In this sense, the digital divide is not simply about access to technology but about the capabilities individuals have to participate in and benefit from increasingly data-driven and AI-enabled systems. Access to the Internet now includes the potential benefits of a world driven by big data. Realizing these benefits, however, requires not just access but the ability to have a voice in how data-driven systems are designed and governed. Voice, therefore, has many dimensions here, particularly in how individuals and communities shape the design and governance of digital health systems.

There must be space for such voices in shaping the digital health governance system, for example in matters of designing and enforcing privacy legislation and digital platform regulation. Questions of voice also extend beyond individual participation to the global level where geopolitical asymmetries influence which actors shape the development and governance of AI-driven health technologies.

Voice also matters as algorithmic accountability becomes a feature of governance regimes. Just as important, for big data to have universal application in health systems, the data have to be representative of the genealogical diversity in populations. AI-based solutions must be credible, reliable, and accessible. Consider, for example, the promise that distance health holds, for example, in a situation where imaging is conducted in situ and its interpretation and recommendations for treatment are developed by specialists far away, or by an AI-based process. There is a chain – from access to a digital imaging centre nearby to reliable and trusted data storage and transmission to sound algorithm-based interpretation to communication with the patient – that illustrates the contours of the digital divide.

A final element in the digital divide is the nature of the data and the intellectual property (IP) laws underlying it. Other than geospatial data (and even here there are exceptions) data are generated by people. How data are gathered, stored and used and how the benefits from pooled data are distributed are again questions of governance, with a central question being the degree of agency afforded to citizens, corporations and governments. The roll-out of vaccines during the COVID-19 pandemic showed how current IP regulation can skew the distribution of vaccines away from considerations relating to need or public good to those with power and

the ability to pay. Despite ongoing policy debates and negotiations in some countries and at the World Trade Organization, the same issues prevail relating to intellectual property and the proprietary nature of most technologies, as digital technologies with significant implications for human health and with attributes of positive externalities are rolled out. In other words, there remains, at the heart of the governance challenge, a tension between the public nature of the data used to train AI and the private, commercial incentives necessary to drive invention and innovation. While AI affects populations broadly, its implications are especially significant for young people, whose developmental trajectories will unfold alongside AI's continued expansion and related technologies like quantum computing.

Current wider debates about AI governance – for example, relating to the promotion of digital literacy or developing appropriate algorithms and digital platform regulation – bleed into health and vice versa. Health might be seen as the test case for experimentation in governance, but it also lies at the centre of these debates. Governance in digital health matters will largely determine the wider outcomes of the current so-called Fourth Industrial Revolution and separating health from societal well-being is undesirable and arguably not even possible.

It is within this rapidly evolving landscape that the contributions in this series examine how AI is reshaping the relationship between young people, health and governance. The papers in this series are organized as follows. Courtney Doagoo sets the stage by surveying and assessing the broad governance regimes that cover AI and health. Rebecca Raeside develops a framework to understand and manage the risks and benefits of AI for the health of young people. Shajoe Lake explores how current gaps in policy might be filled while integrating young people and their concerns into the policy process. Erza Selmani rounds off the series with a comparison of the current gamut of approaches to managing AI and health, ranging from self-regulation, state regulation and hybrid models.

All papers find that the perspectives of young people in the AI-health connection are mostly missing (as they are in most aspects of policy, it should be said). Yet, there are concerns and opportunities specific to young people that require attention. By definition, young people will be affected by the long arc of technology development and roll-out more than any other group in society.

Taken together, the contributions in this series also highlight several cross-cutting dynamics shaping the relationship between AI, health and young people. On the one hand, new technologies introduce risks that are not yet fully addressed by existing governance frameworks – including the influence of AI-driven companion bots, the potential for bias in datasets used to train algorithms and concerns around data protection and privacy. On the other hand, these same technologies may offer significant opportunities, particularly in expanding access to mental health

support, strengthening health systems' resilience and improving diagnostic capacity. The question, therefore, is not simply whether to adopt AI in health systems, but how to shape governance frameworks that can manage these risks while enabling socially beneficial innovation. All papers also illustrate the point that there is no one-size-fits-all solution here. The way forward will lie in understanding the technology and acting from within a range of options that have been properly assessed. Addressing these challenges will likely require a combination of approaches, including stronger oversight mechanisms such as transparency requirements and external auditing of AI systems, greater international cooperation, investments in digital literacy and meaningful participation of young people in governance processes that shape the technologies affecting their lives. Ensuring that young people's perspectives inform these governance approaches is crucial not only to ensure that rights are protected but also for shaping technologies that genuinely support long term well-being. The effects of AI on the health of young people will depend as much on the technology itself as it will on the governance choices, safeguards and social priorities that guide its development and use. And these must be considered alongside bigger questions about what societies expect from new technologies.

# References


Alegre, S. (2022). *Freedom To Think*: *Protecting a Fundamental Human Right in the Digital Age*. Atlantic Books.

Doidge, N. (2018, April 24). *Platform Governance: Screen Time, the brain, privacy and mental health*. Centre for international governance innovation. https://www.cigionline.org/articles/screen-time-brain-privacy-and-mental-health

International Telecommunication Union. (n.d.). *Statistics*. Retrieved March 30, 2026, from https://www.itu.int/en/ITU-D/Statistics/pages/stat/default.aspx.

Li, X., Keown-Stoneman, C. D., Omand, J. A., Cost, K. T., Gallagher-Mackay, K., Hove, J., Janus, M., Korczak, D. J., Pullenayegum, E. M., Tsujimoto, K. C., Vanderloo, L. M., Maguire, J. L., Birken, C. S., Birken, C. S., Maguire, J. L., Allen, C., Anderson, L. N., Cost, K., D'Annunzio, D., … Peralta, M. (2025). Screen time and standardized academic achievement tests in elementary school. *JAMA Network Open*, *8*(10). https://doi.org/10.1001/jamanetworkopen.2025.37092.

Teutem, S. V. (2025, April 25). *Internet use became the norm for humanity only very recently*. Our World in Data. https://ourworldindata.org/data-insights/internet-use-became-the-norm-for-humanity-only-very-recently.

(United Nations [UN], 1948). *Universal declaration of human rights*. https://www.un.org/en/about-us/universal-declaration-of-human-rights.

World Bank. (n.d.). *Individuals using the Internet (% of population). Retrieved March 30, 2026, from* https://data.worldbank.org/indicator/IT.NET.USER.ZS.

# 01

# Mind the gap: Child-centred ethics in AI healthcare governance

**B. Courtney Doagoo**

# High level summary

This paper explores the gap between ethical commitments and enforceable governance in AI-enabled healthcare, focusing on the unique risks faced by children and youth. Despite rapid adoption of AI in health systems and the proliferation of global AI ethics frameworks, current governance approaches rarely operationalize children's rights or youth-centred ethics.

The paper traces the evolution of AI governance, identifies implementation gaps in healthcare, and highlights youth-specific challenges. It argues that embedding child-centred, rights-based governance can strengthen health system resilience, particularly during crises, and help ensure AI supports equitable, trustworthy and sustainable healthcare for youth.

Section 1

# Introduction

The development of artificial intelligence (AI)-enabled healthcare is advancing rapidly. *The Artificial Intelligence Index Report 2025* stated that global private investment in AI in healthcare was the third highest investment category, at $11 billion in 2024, after AI in infrastructure, research and governance and AI in data management and processing (Maslej et al., 2025). The state of enterprise AI, published by OpenAI in December 2025, noted that healthcare was the second-fastest-growing industry, year-over-year, in terms of its customer-base (Chatterji, 2025) and the World Economic Forum (WEF) ranked healthcare as fourth for global average in AI adoption (WEF, 2025). Despite the promise of AI-enabled healthcare, AI governance has been slow to translate ethical commitments into practical and enforceable protections. This creates a gap between aspiration and accountability in AI-enabled healthcare, creating risks for young people, which can exacerbate health inequities, erode trust and undermine health system resilience.

This paper argues that despite the proliferation of AI governance frameworks, governance approaches have yet to operationalize youth-centred ethics in AI-enabled healthcare. Part one examines the evolution of AI governance to contextualize its current state. Next, it explores whether current governance approaches have been effective in embedding ethics and addressing health sector-based considerations. Part three focuses on youth and AI-enabled healthcare, highlighting the risks and barriers related to governance. Part four addresses how AI governance can contribute to health system resilience, particularly in the face of global crises such as pandemics, climate crises or limited resources. The absence of a clear AI governance framework addressing these intersections creates a significant risk that technologies will be adopted for their short-term operational, financial and efficiency gains, overlooking the long-term harms related to equity, accountability, safety, trust and sustainability (Organisation for Economic Co-operation and Development [OECD], 2024; Anderson & Sutherland, 2024).

Section 2

# Governance constellation

## 2.1 Overview

The AI governance ecosystem has been incubating for over a decade, with a distinct grassroots phase emerging in the mid-to-late 2010s. Foundational initiatives included the Institute of Electrical and Electronics Engineers' (IEEE) *Ethically Aligned Design*, which highlighted the risks and social impacts of AI (IEEE, 2016), *Asilomar AI Principles* (Future of Life Institute, 2017) and later the 2018 *Montréal Declaration on Responsible AI*, which emerged from extensive consultations in 2017, grounding its principles in the public interest and democratic values (University of Montréal, 2018). These and other initiatives emerged in the absence of established international principles or standards, resulting in a landscape of overlapping guidance, documents and taxonomies (Jobin et al., 2019; Fjeld et al., 2020). Early civil-society initiatives were influential and succeeded in articulating ethical foundations, however, they largely failed to translate into enforceable protections, particularly for children and other vulnerable groups, or for specific applications such as in health.

Following these early initiatives, the international norm-setting phase began, attempting to galvanize countries around consensus and commitment. Key milestones included the Organisation for Economic Co-operation and Development's (OECD) *Recommendation of the Council on Artificial Intelligence*, published in 2019 and updated in 2024 (OECD, 2019) and the United Nations Educational, Scientific and Cultural Organization (UNESCO) *Recommendations on the Ethics of AI* (UNESCO, 2021). These core guidance documents were underpinned by normative principles aiming to protect human rights and societal well-being, including requiring privacy, transparency, fairness and accountability. This movement ignited a broad effort – by mid-2025, nearly 70 jurisdictions were reported to have adopted some level of AI strategies or policies (Nakazawa & Pisa, 2025). However, in this phase, formal AI regulatory governance structures remained rare. While there was much activity underway at a higher-level outlining the ethics or "what" of AI governance, the "how", in other words the tangible implementation (Floridi, 2019; Blanchard, 2024) and efforts for law and policy-making at a national level, were only beginning to surface (Alanoca et al., 2025).

In the current phase, national governance frameworks are emerging. For example, the European Union (EU) was the first to develop a comprehensive governance framework: EU Artificial Intelligence Act (European Parliament, 2024) is already in its implementation phase. It is a cross-cutting risk-based approach that is individualistic and rights-based (Robert et al., 2023; Li & Li, 2025). In contrast, China has adopted a vertical framework, which is a comprehensive sector-specific approach that incorporates stringent security requirements across the value chain (Roberts, et al., 2023). It has been suggested that China has adopted a collective approach – striving for "national security, social stability and digital sovereignty" rather than an individualistic approach like the EU (Li & Li, 2025; Roberts, et al., 2023). The United States on the other hand, had a decentralized approach and is seemingly in the process of challenging state-based legislation enacted to compensate for the gap in federal regulations (The White House, 2025; Robert et al., 2023; Graux et al., 2025). These diverging approaches illustrate differing national and regional values and ambitions. Unfortunately, many jurisdictions are not as far along as these examples.

The hesitation to adopt regulation could largely be explained by market and power imbalances. Dominant tech actors are setting standards and practices in the field (Taeihagh, 2025), but also on the geopolitical stage, where the few countries spearheading AI development are also driving influential policy and governance decisions, which in turn impact how non-dominant jurisdictions respond. As Susie Alegre highlights, higher-level ethical guidelines tend to be more palatable to big tech actors since they are optional and lack enforceable obligations, unlike laws – human rights laws – that cannot be brushed aside (Alegre, 2022). There is an urgent need for concrete oversight and guardrails, as dominant actors consistently underplay the harms and risks that their technologies pose (West, 2025).

## 2.2 Unaddressed risks and harms

As AI evolves, unaddressed challenges and concerns continue to surface. There are well-documented impacts, risks and harms, yet governments have not coalesced around no-go zones. In 2025, the *Global Call for AI Red Lines* was launched, urging governments to reach an international agreement on unacceptable risks that are "too dangerous to permit under any circumstances". This could include lethal autonomous weapons, mass surveillance and social scoring and human impersonation, which is a current and ongoing problem (AI Red Lines, 2025). In addition to these red-lines, there are concerns which are less obvious – not because they are less harmful, but because they have not been reframed to reflect the current reality of technological ubiquity (Alegre, 2022). One of these is the revived right of freedom of thought reflected in Article 18 of the Universal Declaration of Human Rights (United Nations [UN], 1948), which effectively honours the right to keep one's thoughts private, free of manipulation and to not be penalized for one's thoughts alone (Alegre, 2022; Alegre & Shull, 2024).

Notably this is highly relevant to youth and AI applications in various areas including AI-enabled health. Given the unique vulnerability and risks related to young people, it is interesting that the widely accepted United Nations Convention on the Rights of the Child (UNCRC) (United Nations [UN], 1989) has not been integrated or operationalized in many of the governance instruments, other than in the form of high-level acknowledgments. This is especially concerning as AI systems are increasingly embedded across health applications that can directly impact health outcomes, autonomy, privacy and long-term well-being of children. This is perhaps because the focus until now has been on ethics and not on human rights or rights of the child – where existing agreed-upon legal frameworks could provide structure. Without clear integration of children's rights, existing frameworks will continue to neglect differentiating the needs of children as rights-holders with distinct protections.

Section 3

# Ethics and AI in health

The evolution of AI governance frameworks provides context about the lag in national policies or strategies for AI in health (Kijewski et al., 2024). AI-enabled healthcare is used in a vast number of ways, including "clinical decision making, public health, biomedical research and system governance and administration" (Dror-Shpoliansky, 2022). Specific use cases include drug discovery, precision medicine, patient screening and generative product design (WEF, 2025). For children, AI-enabled health is being used for mental health support and therapy via AI-powered chatbots, interactive applications providing children with information and as assistive devices (Gholizadeh, 2023). It is also being used for treatment plans and paediatric interfaces (Gholizadeh, 2023).

The risks and challenges of AI-enabled healthcare are abundant and can emerge across the lifecycle. These challenges, as identified by the World Health Organization's (WHO), span from bias, false information, privacy, safety and cybersecurity – to exclusive control, manipulation, skills degradation, labour and employment concerns (WHO, 2024). The prevalence of data from high-income countries risks the potential for the needs of lower-income populations around the world to be excluded from certain AI applications. As highlighted by the WEF, Deutsche Welle reported that 80 per cent of the genetics studies used data from individuals of European descent, which is proportionally less that 20 per cent of the world's population (Onuh, 2025).

Acknowledging the rapid pace of technology development and the inability for policymakers and legislators to keep pace, the WHO developed ethical principles for AI-enabled health in 2021 and later in consideration of large language models (LLMs) in 2025 (WHO, 2024, ix). The six principles are to:

- protect autonomy
- promote human well-being, human safety and the public interest
- ensure transparency, "explainability" and intelligibility
- foster responsibility and accountability
- ensure inclusiveness and equity
- promote AI that is responsive and sustainable

Despite the WHO's early intervention with sector-specific guidance (WHO, 2021; WHO, 2024), there are very few jurisdictions that have incorporated sector-specific strategies for AI governance in health. In a WHO-led European Region-focused survey, only four out of 50 responding states mentioned having a national health-specific AI strategy (WHO, 2025), while only three had introduced legal requirements for generative AI systems in healthcare and four had developed liability standards for AI in health WHO, 2025).

This has contributed to slower AI adoption in health. Among OECD countries, "misaligned policies for responsible AI" and "unclear liability and accountability as a barrier to responsible AI innovation in health" were included as barriers to AI adoption in health (Anderson & Sutherland, 2024). The OECD also identified several additional barriers including "regulatory uncertainty and gaps in governance" and has established a framework across nine policy categories to scale the benefits of AI in health (OECD, forthcoming 2026).

Notably, the top barrier to adoption identified by the WHO survey was legal uncertainty based on roughly 43 of the 50 respondents (WHO, 46, 2025), underlining the importance of AI governance not only for safety, security and justice, but for market certainty and innovation. The report states that this and other barriers, such as financing "may perpetuate inequities and slow the realization of AI's potential" (WHO, 2025, xvi).

The ongoing legal uncertainty suggests that ethical principles alone are insufficient; without enforceable governance mechanisms, they remain aspirational, rather than protective or effective. This becomes even more challenging where an additional layer of rights, such as children's rights, are added to the AI governance framework.

Section 4

# Unique challenges of young people, health and AI

## 4.1 Young people and AI governance

Children's rights are enshrined in the UNCRC (United Nations [UN], 1989). Despite this, the AI governance ecosystem has largely evolved in the absence of youth perspectives. UNICEF reports that many young people turn to AI applications, whether for education, for information, or for other reasons and yet at the same time, "[c]hildren's rights are still not receiving sufficient attention in AI policy, law, governance and development" nor are they actively being engaged to help shape these systems (UNICEF Innocenti, 2025). This is especially the case for children from the Global South, "for whom location, digital divides and severely limited access to policy forums and AI design processes are exclusionary factors" (UNICEF Innocenti, 2025; Mahomed et al., 2023). Treating children as general AI users is not merely inadequate, it is dangerous given their developmental vulnerability and limited agency.

In 2020, prior to many of the current AI applications available today including generative AI, UNICEF published the outputs of its first workshop report on *AI and child rights and policy* outlining principles for the use of AI systems by children (UNICEF, 2019). This has since evolved into a third version of guidance (UNICEF Innocenti, 2025). These requirements on child-centred AI include the need to:

- ensure regulatory frameworks, oversight and compliance for child-centred AI
- ensure safety for children
- protect children's data and privacy
- ensure non-discrimination and fairness for children
- provide transparency, explainability and accountability for children
- respect human and child rights through responsible AI practice
- support children's best interests, development and well-being
- ensure inclusion of and for children
- prepare and skill children for present and future developments in AI
- create an enabling environment for child-centred AI

Notably, many of these requirements are similar to general AI-related principles but are tailored to the unique needs of youth. For example, a recommendation to promote safety is to mandate child rights impact assessments across AI strategies, policies, law and regulations, not unlike AI impact assessments or risk assessment requirements (UNICEF Innocenti, 2025).

While few jurisdictions have incorporated and referred to the rights of children in their governance strategies, the approach has been inconsistent – for example the EU AI Act makes specific reference to rights of the child (European Parliament, 2024) and prohibits certain harmful AI-impacts to vulnerable groups, including age-based, e.g., children (European Parliament, 2024; European Commission, 2025). On the other hand, China has enacted specific legislation to protect against addiction in underage generative AI-users and has also proposed legislation for chatbot-related harms (Cyberspace Administration of China, 2023; White & Case LLP, 2025; Chen & Xu, 2025; Chia, 2025). This inconsistency between jurisdictions may not only make it difficult to ensure that AI standards are interoperable, but also may prevent children from having consistently high standards of protection, regardless of jurisdiction.

## 4.2 Young people, AI and health – challenges for principles into practice

In addition to the general harms and risks of AI to young people highlighted above, there are unique challenges in operationalizing widely accepted AI principles. This section summarizes four youth-specific challenges identified by Wang et al., conceptually expanded to integrate health (Wang et al., 2024).

The first challenge is the "lack of consideration of the developmental aspect of childhood" or young people. This challenge highlights the absence of meaningful integration of children's "diverse needs, age ranges, development stages, backgrounds and characters" into AI ethics consideration (Wang et al., 2024; Kickbusch et al., 2021). As highlighted by Chng et al, "[c]hildren are not miniature versions of adults, as children undergo age-associated changes in organ function and neurodevelopment" (Chng et al., 2025). They would therefore have different needs throughout their growth, which is important to ensure that the data used clearly indicates this context. This is concerning as there is a lack of information and data in paediatric health. For example, a recent study highlighted that "FDA [Food and Drug Administration] approvals reveal that fewer than 1 out of 5 devices are approved for pediatric use and ages of study participants are not reported in 81.6 per cent of devices" (Hua et al., 2025). Adult data have limited application to children and given the developmental stages of youth, demographic data, including age, should be clearly marked (WHO, 2024; UNICEF Innocenti, 2025).

Another challenge highlighted by Wang et al., is the "lack of consideration of the role of the guardian" (Wang et al., 2024), where it is vital to consider the role of the parent or guardian as a "surrogate" decision maker. This is an important factor, especially in the context of healthcare and concerns about consent

and decision-making on behalf of children. At the same time, the authors suggest that parents may not have the same level of understanding about technology as their children, who have grown up in a technology-driven society. This prompts a shift away from parents' authority, towards greater children's autonomy and resilience regarding their use of technology (Wang et al., 2024). As the WHO highlighted, "[l]aws and policies on paediatric consent, assent and stipulations for legal parental involvement differ among and within countries. Thus, lack of cohesive, unified, global, child-specific regulation and oversight could result in unidentified, unmonitored harm, especially from use of LLMs" (WHO, 2024).

The third challenge is a "lack of child-centred evaluations considering children's best interests and rights" (Wang, et al., 2024). Here, the authors recommend moving away from quantitative technical assessment measures because they are often skewed toward accuracy in a way that does not contextualize appropriately or present the principles against which it is necessary to consider the risks and rights of the child (Wang, et al., 2024). Using a "best interest standard" in the context of paediatric ethics, while controversial, may be helpful in AI governance because it offers an approach that would require that "stakeholders with a vested interest in and specialized knowledge of children should be included in governance" (Richter et al., 2025. Incorporating child rights impact assessments or relevant controls by design could help alleviate this challenge (Mahomed et al., 2023).

The final challenge highlighted by Wang et al. is the "lack of coordinated, cross-sector and cross-disciplinary approach,". Acknowledging that the solutions lie in a multidisciplinary approach where knowledge-sharing and harmonizing taxonomies would create an opportunity for collaboration across the "child-computer-interaction community" and would yield better results in attempting to operationalizing principles into practice (Wang et al., 2024). A multi-stakeholder approach is crucial for achieving coordination and enabling an environment for child-centred AI, especially in healthcare (UNICEF Innocenti, 2025; Richter et al., 2025).

These examples highlight the unique circumstances of young people's health and the challenges that require additional consideration as AI governance principles move into practice. Implementation, especially in a rapidly-evolving environment, both technologically and governance-wise, is admittedly challenging and requires transparency and collaboration. These governance gaps have concrete implications for health systems, especially as to how they might operate under the stress of a crisis. On the other hand, if embedded, these technologies can help strengthen health systems and promote resilience.

Section 5

# AI-enabled health system resilience for young people: considerations and opportunities

AI governance has the potential to significantly strengthen health system resilience for young people, particularly in the context of global polycrisis such as conflict, pandemics, climate crises, resource scarcity (Sinha, 2025. However, realizing this potential requires moving beyond the view of AI in healthcare as a purely technical solution. Instead, the development, deployment and oversight of AI-enabled health technologies must be firmly grounded in child rights and health ethics. Without clear ethical and governance frameworks, pursuing short-term gains may generate and entrench longer-term harms (Anderson & Sutherland, 2024).

Experiences gained from crises illustrate both the scale of AI's potential and the importance of responsible governance. During emergencies, AI-assisted triage systems used in hospitals supported resource allocation more effectively, telemedicine solutions have helped provide care for displaced populations in camps and predictive algorithms have helped to detect and prevent outbreaks (Haykal et al., 2025). AI-driven mental health applications are also helping individuals who have survived disaster (Haykal et al., 2025). These examples demonstrate the promise of AI-enabled healthcare solutions that have far-reaching implications for populations facing conflict and crisis.

To ensure that these technologies strengthen rather than undermine health systems, governance frameworks must prioritize several key actions.

- Operationalize ethics and human rights principles, including the UNCRC, throughout the lifecycle of AI-enabled health technologies, including by conducting Child Rights Impact Assessments throughout the development lifecycle (Mahomed et al., 2023).

- Invest in digital health literacy for young people and parents to empower their ability to make informed choices and "to become informed agents in digital societies who are capable of shaping health policies, participating in public discourse and holding institutions accountable" (Thu, 2025).
- Embed fairness and non-discrimination to prevent structural biases that disproportionately affect vulnerable populations during times of crisis, including addressing "health data poverty," which often excludes marginalized groups, including children, from datasets used to train AI systems (Sinha, 2025).
- Strengthen "transparency, explainability and accountability for children" in AI-enabled systems (UNICEF Innocenti, 2025). Ensuring that AI systems are understandable to clinicians, regulators and patients is essential to maintaining trust, supporting safe clinical decision-making and enabling effective oversight. Clear mechanisms for data protection, algorithmic auditing and accountability should be central components.
- Incorporate the engagement of patients, children, marginalized communities and the broader public as a core element in "digital health decision making" (Kickbusch et al., 2021). Involving youth across the value chain of AI-enabled healthcare can help ensure that technologies respond to real needs and reflect lived experiences. Structured engagement processes, including consultations and participatory design approaches, can ensure that governance frameworks are responsive and inclusive (Mahomed et al., 2023). promote AI that is responsive and sustainable

Together, these priorities highlight that effective AI governance in healthcare must be proactive, rights-based and inclusive and can ensure that AI-enabled healthcare technologies contribute to resilient, equitable and child-centred health systems in times of crisis and beyond. Failing to do so could further exacerbate inequalities, exclusion and mistrust.

# References


AI Red Lines. (2025). *Global call for AI red lines.* https://red-lines.ai/

Alanoca, S., Gur-Arieh, S., Zick, T., & Klyman, K. (2025, June 23). Comparing apples to oranges: A taxonomy for navigating the global landscape of AI regulation. *FAccT '25: Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency.).* https://doi.org/10.1145/3715275.3732059

Alegre, S. (2022). Freedom to think: Protecting a fundamental human right in the digital age. Atlantic Books.

Alegre, S., & Shull, A. (2024). *Freedom of thought reviving and protecting a forgotten human right*. Centre for International Governance Innovation. https://www.cigionline.org/static/documents/Freedom.of.Thought_SpecialReport.Alegre.Shull.pdf

Anderson, B., & Sutherland, E. (2024). *Collective action for responsible AI in health*. (OECD Artificial Intelligence Papers, No. 10). OECD Publishing, Paris. https://doi.org/10.1787/f2050177-en

Blanchard, A., Thomas, C., & Taddeo, M. (2025). Ethical governance of artificial intelligence for defence: Normative tradeoffs for principle to practice guidance. *AI & Society, 40*(185–198). https://doi.org/10.1007/s00146-024-01866-7

Chatterji, R. (2025, December 8). *2025 report: The state of enterprise AI*. Open AI. https://openai.com/index/the-state-of-enterprise-ai-2025-report/

Chen, X., & Xu, L. (2025). State, society, and market: Interpreting the norms and dynamics of China's AI governance. *Computer Law & Security Review, 59,* Article 106206. https://doi.org/10.1016/j.clsr.2025.106206

Chia, O. (2025, December 29). *China to crack down on AI firms to protect kids.* BBC. https://www.bbc.com/news/articles/c8dydlmenvro

Cyberspace Administration of China. (2023, July 13). *Interim Measures for the Administration of Generative Artificial Intelligence Services*. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

Chng S.Y., Tern M.J.W., Lee Y.S., Cheng L.T., Kapur J., Eriksson J.G., Chong Y.S., & Savulescu J. (2025). Ethical considerations in AI for child health and recommendations for child-centered medical AI. *NPJ Digit Med 8*, Article152. https://doi.org/10.1038/s41746-025-01541-1

Committee on the Rights of the Child(CRC). (2021, March 2). *General comment No. 25 (2021) on children's rights in relation to the digital environment (CRC/C/GC/25).* United Nations Children's Fund (UNICEF). https://www.unicef.org/bulgaria/en/media/10596/file

Dror-Shpoliansky, D. (2022, December). *Rights in the digital age: Challenges and ways forward*. (OECD Digital Economy Papers No. 347). https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/12/rights-in-the-digital-age_d3a850de/deb707a8-en.pdf

European Commission. (2025, July 29). *Communication from the Commission: Commission guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)* [C(2025) 5052 final]. https://ai-act-service-desk.ec.europa.eu/sites/default/files/2025-08/guidelines_on_prohibited_artificial_intelligence_practices_established_by_regulation_eu_20241689_ai_act_english_ied3r5nwo50xggpcfmwckm3nuc_112367-1.PDF

European Parliament. (2023, June 8). *EU AI Act: first regulation on artificial intelligence*. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

European Parliament. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence *(Artificial Intelligence Act). (2024). Official Journal of the European Union, L 2024/1689. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_2024/1689*

European Union (n.d.). *Timeline for the Implementation of the EU AI Act*. Retrieved March 30, 2026, from AI Act Service Desk, European Union

https://ai-act-service-desk.ec.europa.eu/en/ai-act/timeline/timeline-implementation-eu-ai-act.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A.C., & Skrikumar, M. (2020, January 15). *Principled Artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. (The Berkman Klein Center Research Publication Series, Research Publication No 2020-1). The Berkman Klein Center for Internet & Society. https://cyber.harvard.edu/publication/2020/principled-ai

Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology 32*(185-193). https://doi.org/10.1007/s13347-019-00354-x

Future of Life Institute. (2017, August 11). *Asilomar AI Principles*.[Open letter]. https://futureoflife.org/open-letter/ai-principles/

Gholizadeh, S., (2023, October 5,). *AI's multi-faceted impact on children: Opportunities, challenges, and guidance*. IBM: Women in AI. https://community.ibm.com/community/user/blogs/samira-gholizadeh/2023/10/05/ais-multi-faceted-impact-on-children-opportunities

Graux, H., Garstka, K., Murali, N., Cave, J., & Botterman, M. (2025). *Interplay between the AI Act and the EU digital legislative framework*. European Parliament, Committee on Industry, Research and Energy. https://www.europarl.europa.eu/RegData/etudes/STUD/2025/778575/ECTI_STU(2025)778575_EN.pdf

Haykal, D., Goldust, M., Cartier, H., & Treacy, P. (2025). AI in humanitarian healthcare: a game changer for crisis response. *Frontiers in Artificial Intelligence: Medicine and Public Health*, *8*. https://doi.org/10.3389/frai.2025.1627773

Hua, S.B., Heller, N., He, P., Towbin, A.J., Chen, I.Y., Lu, A.X., & Erdman, L., (2025, June 10) Lack of children in public medical imaging data points to growing age bias in biomedical AI. [Preprint] https://doi.org/10.1101/2025.06.06.25328913

Shahriari. K., & Shahriari, M.. (2016). *IEEE Standard Review: Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems (Version 1 - For Public Discussion)*.The Institute of Electrical and Electronics Engineers.. Association. https://doi.org/10.1109/IHTC.2017.8058187

International Telecommunications Union. (2025). Joint Statement on Artificial Intelligence and Rights of the Child. https://www.itu.int/pub/D-STR-CYB_JOINT-2025.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(389). https://doi.org/10.1038/s42256-019-0088-2

Kickbusch, I., Piselli, D., Agrawal, A., Balicer, R., Banner, O., Adelhardt, M., Capobianco, E., Fabian, C., Gill, A. S., Lupton, D., Medhora, R. P., Ndili, N., Ryś, A., Sambuli, N., Settle, D., Swaminathan, S., Vega Morales, J., Wolpert, M., Wyckoff, A. W., & Xue, L. (2021, November 6). The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world. *The Lancet, 398*(10312), 1727–1776. http://doi.org/10.1016/S0140-6736(21)01824-9

Kijewski, S., Ronchi, E., & Vayena, E. (2024). International organisations and the global governance of AI in health. In B. Solaiman & I. G. Cohen (Eds.), *Research Handbook on Health, AI and the Law*. Edward Elgar Publishing Ltd. https://www.ncbi.nlm.nih.gov/books/NBK613223/

Li, X., and Li, X. (2025). Regulating AI in a fragmented world: The diverging paths of the EU and China and their impact on global governance. *Innovation and Development Policy 7*, 27-47. http://idp-journal.casisd.cn/browse/la/202506/P020250724546443370318.pdf

Mahomed, S., Aitken, M., Atabey, A., Wong, J., & Briggs, M. (2023). *AI, Children's Rights, & Wellbeing: Transnational Frameworks. Mapping 13 Frameworks at the Intersections of Data-Intensive Technologies, Children's Rights, and Wellbeing*. The Alan Turing Institute. https://www.turing.ac.uk/sites/default/files/2023-11/ai-childrens_rights-_wellbeing-transnational_frameworks_report.pdf

Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J.C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., Betts Lotufo, J., Rome, A., Shi, A., & Oak, S. (2025). *The Artificial Intelligence Index 2025 Annual Report.* AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. https://doi.org/10.48550/arXiv.2504.07139

Mishra, V. (2025, April 23). *Health, education, opportunity at stake, amid stubborn digital gender divide.* UN News. https://news.un.org/en/story/2025/04/1162541

Nakazawa, A., & Pisa., M. (2025, May 20). How governments are driving AI adoption for economic growth. *OECD: The AI Wonk*. https://oecd.ai/en/wonk/how-governments-are-driving-ai-adoption-for-economic-growth

OECD. (2026). *Scaling Artificial Intelligence in Health*. OECD Publishing. https://doi.org/10.1787/a436e12d-en.

Organisation for Economic Co-operation and Development (OECD). (2026, April 28) Scaling Artificial Intelligence in Health. Forthcoming. https://www.oecd.org/en/publications/forthcoming.html.

OECD. (2019). *Recommendation of the Council on Artificial Intelligence*. (OECD Legal Instrument No. OECD/LEGAL/0449). https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

OECD. (2024). *AI in Health: Huge potential, huge risks* [Policy brief]. OECD Publishing. https://doi.org/10.1787/2f709270-en

Onuh, G. (2025, October 1). *AI in healthcare risks could exclude 5 billion people; here's what we can do about it*. World Economic Forum. https://www.weforum.org/stories/2025/10/ai-in-healthcare-risks-could-exclude-5-billion-people-here-s-what-we-can-do-about-it/

Richter, F., Holmes, E., Richter, F., Guttmann, K., Duong, S.Q., Gangadharan, S., Schadt, E., Salmasian, H., Gelb, B.D., & Glicksberg, B.S. (2025). Toward governance of artificial intelligence in pediatric healthcare. *NPJ Digit Med 8*,636. https://doi.org/10.1038/s41746-025-02000-7

Roberts, H., Ziosi, M., & Osborne, C. (2023). *A comparative framework for AI regulatory policy*. The International Centre of Expertise on Artificial Intelligence in Montreal (CEIMIA). https://ceimia.org/wp-content/uploads/2023/05/a-comparative-framework-for-ai-regulatory-policy.pdf

Sinha, C. (2025). *Strengthening health systems by leveraging responsible AI solutions: An emergent research landscape*. [Discussion paper]. International Development Research Centre. https://hdl.handle.net/10625/63893

Taeihagh, T. (2025). Governance of Generative AI. *Policy and Society 44*, 1-22. https://doi.org/10.1093/polsoc/puaf001

The White House. (2025, December 11). *Ensuring a National Policy Framework for Artificial Intelligence* [Executive Orders]. https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/

Thu, H. (2025). *Digital citizenship education for the health and well-being of young people in Europe*. Digital Transformations for Health Lab. https://www.dthlab.org/articles/digital-citizenship-education-for-the-health-and-well-being-of-young-people-in-europe#

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2021). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000380455

United Nations Children's Fund (UNICEF) Office of Global Insight and Policy. (2019, June). *Workshop Report: AI and child rights policy. https://www.unicef.org/innocenti/media/2486/file/AI-Children-Workshop-New-York-2019.pdf*

UNICEF Innocenti. (2025, December). *Guidance on AI and children 3.0: Updated guidance for governments and businesses to create AI policies and systems that uphold children's rights. https://www.unicef.org/innocenti/media/11991/file/UNICEF-Innocenti-Guidance-on-AI-and-Children-3-2025.pdf*

United Nations General Assembly. (1948, December 10). Universal declaration of human rights (217 [III] A). https://www.un.org/en/about-us/universal-declaration-of-human-rights

United Nations (UN). Convention on the Rights of the Child, November 20, 1989. https://www.ohchr.org/en/professionalinterest/pages/crc.aspx

UN. (1981). *Secretary-General's Report to the General Assembly* (A/36/215). https://docs.un.org/en/A/RES/36/28

UN. (n.d.). *Global issues: Youth*. Retrieved March 30, 2026 from United Nations https://www.un.org/en/global-issues/youth.

United Nations Youth. (n.d.). *Definitions of Youth* [Fact sheet]. Retrieved March 30, 2026 from United Nations Department of Social and Economic Development, https://www.un.org/esa/socdev/documents/youth/fact-sheets/youth-definition.pdf.

University of Montreal. (2018). *Montréal Declaration for a Responsible Development of Artificial Intelligence*. https://montrealdeclaration-responsibleai.com/the-declaration/

Varkey, B. (2020). Principles of clinical ethics and their application to practice. *Medical Principles and Practice*; *30*(1) 17-28. https://doi.org/10.1159/000509119

Wang, G., Zhao, J., Van Kleek, M., & Shadbot, N. (2024). Challenges and opportunities in translating ethical AI principles into practice for children. *Nature Machine Intelligence, 6*(3) 265-270. https://doi.org/10.1038/s42256-024-00805-x

West. D.M. (2025, May 27). *The coming AI backlash will shape future regulation.* Brookings. https://www.brookings.edu/articles/the-coming-ai-backlash-will-shape-future-regulation/

White & Case LLP. (2025, September 22) *AI Watch: Global regulatory tracker – China*. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china

World Economic Forum. (2025, January). *The future of AI-enable health: Leading the way* [White paper]. https://reports.weforum.org/docs/WEF_The_Future_of_AI_Enabled_Health_2025.pdf

World Health Organization (WHO). (2021) *Ethics and governance of artificial intelligence for health: WHO guidance*. https://www.ictworks.org/wp-content/uploads/2021/10/who-ethics-artificial-intelligence-healt.pdf

WHO. (2024). *Ethics and governance of artificial intelligence for health. Guidance on large multi-modal models*. https://www.who.int/publications/i/item/9789240084759.

WHO. (2025). *Artificial intelligence is reshaping health systems: State of readiness across the WHO European Region*. https://iris.who.int/server/api/core/bitstreams/d2913ae3-c8e0-4a46-b6ff-b4b121e936f4/content

# 02

# Safeguarding adolescent health and well-being in the AI era: Integrating risks, vulnerabilities and governance gaps to inform solutions

**Rebecca Raeside**

# High level summary

This paper examines the rapidly growing use of AI systems among adolescents and their implications for health and well-being. As AI becomes embedded in everyday activities from learning and creativity to social interaction and health advice, it is increasingly shaping adolescent experiences during a critical developmental period, with unknown impacts to their health and well-being.

This paper comprehensively analyses how AI-related risks may intersect with and amplify existing vulnerabilities among adolescents, identifies gaps in current policy and practice responses, and provides a framework to mitigate risks. Overall, it calls for more balanced and proactive approaches in policy and practice to safeguard adolescent health and well-being without limiting their autonomy and enabling benefits of innovation.

Section 1

# Introduction

Artificial Intelligence (AI) technologies including generative AI, AI-powered chatbots and AI integrations into traditional digital tools are increasingly being utilized by adolescents both actively and passively (Nagata et al., 2025). Adolescents (10–24 years) (Baird et al., 2025; Patton et al., 2016) are actively using AI for learning and development, creativity, social interaction, emotional connection and advice relating to mental and physical health (Robb & Mann, 2025). Recent statistics demonstrate that AI usage rates vary between 64–84 per cent in developed western countries. In the United States of America (US), a survey of 1000 teens shows 72 per cent have used AI companions and over 50 per cent use them regularly (Robb & Mann, 2025). A United Kingdom (UK) study similarly found that 64 per cent of 9–17-year-olds have used AI chatbots (Internet Matters, 2025). In Australia, a survey of 1000 teens (15–24 years old) shows 84 per cent have used generative AI and 35 per cent have used AI chatbots (Walker, 2024).

As use of AI increases among this population, it has potential to shape their daily experiences and behaviours, including their health and well-being. Adolescence is a period of rapid development, marked by significant biological, cognitive and psychosocial changes (National Academies, 2019). As such, there is a need to establish and maintain good health and well-being behaviours during this critical developmental period, which can be carried across the course of the individual's life (Sawyer et al., 2012). Developmental changes that occur during adolescence affect individuals differently, are dynamic and can be

vulnerable to external influences (National Academies, 2019). AI is an example of an external influence that has the potential to provide both benefits and risks to adolescent development, with a recent article by Nagata et al (2025) highlighting both the risks and benefits across health information, cognition, mental health, body image, social connection, sleep and physical activity.

Due to its recent introduction, there is little empirical evidence of its effects on health and well-being. While the use of AI offers efficiencies and opportunities, there is a pressing need to examine the risks and potential harms to health and well-being for adolescents, who these systems have not been primarily designed for. This paper aims to explore the following questions:

1. What specific emerging AI risks do adolescents face, especially in health-related contexts?
2. How do these risks escalate or compound pre-existing adolescent vulnerabilities?
3. What are the current policy and practice approaches and gaps?
4. How can these risks from AI be identified and mitigated to safeguard adolescent health and well-being?

Section 2

# AI risks to adolescent health and well-being and the compounding of existing adolescent vulnerabilities

The rapid integration of AI into adolescents' daily lives presents emerging and widespread risks to their health and well-being. Previous research into digital media use including screen time and social media has demonstrated negative impacts to adolescent health and well-being, including increased obesity rates, depressive symptoms, a less healthy diet and poorer quality of life (Stiglic & Viner, 2019). However, research into AI risks to health and well-being should avoid taking the same route of looking at causal pathways to determine effects of this technology. Rather, research into AI risks should consider additional contextual factors that influence both technology use and potential impacts to health and well-being (Mansfield et al., 2025).

In January 2021, the Organization for Economic Co-operation and Development (OECD) revised their typology of risks for children in the digital environment, identifying four main risk types: content, conduct, contact and consumer risks (OECD, 2021). The typology also acknowledges that some risks can cut across all four risk categories to have wide ranging impacts on children – including risks to health and well-being, privacy and those arising from the use of advanced technologies such as AI. As AI systems become increasingly complex, so do the ways in which humans interact with them. This makes it vital to have a deep understanding of the dynamic nature of the accompanying risks, in order to identify ways to mitigate them.

Researchers from the University of Illinois have developed a youth-centred risk taxonomy from generative AI (Yu et al., 2025). The taxonomy was created through an analysis of youth-AI chat logs, Reddit discussions and AI incident reports, which identified six high-level risk types: mental well-being, privacy, toxicity, misuse and exploitation risk, bias/discrimination and social and moral developmental risk. Within these, there are 15 medium level and 84 low-level risk types. Below, we will specifically focus on the impacts on adolescent health and well-being and how these risks may escalate or compound pre-existing adolescent vulnerabilities, for which AI systems have not been designed to consider.

## 2.1 Mental well-being risk

Mental well-being risk refers to the negative psychological, emotional or cognitive health impacts caused by youth interactions with AI (Yu et al., 2025). These risks include overreliance, parasocial relationship bonding and inappropriate handling of mental health issues. Adolescents are particularly vulnerable to parasocial relationships as they begin to place greater importance on peer connections than on relationships with their parents (Remschmidt, 1994). Compounded by the effects of AI, adolescents may struggle to distinguish between genuine in-person connections and AI-generated interactions. Their evolving need for autonomy and connection may drive reliance or dependence on AI systems for friendships, relationships or emotional intimacy (Diaz et al. 2025).

As more adolescents are turning towards AI systems for mental health support, it is vital that this promise of AI to meet an unmet need in mental healthcare is undertaken with a safe, ethical and effective approach (Opel et al., 2023). AI systems are being applied widely across diagnosis, monitoring, evaluation and treatment of mental health issues among adolescents (Sharma et al., 2025). They hold immense promise for breaking down access barriers to mental healthcare (Philippe et al., 2022), particularly as adolescents can be difficult to engage in mental health interventions, sometimes hesitant to seek professional mental health support (Radez et al., 2021). However, it is unclear whether the proper safeguards are in place for AI systems to be used in this way by adolescents, given that their developmental stage increases their susceptibility to risk-taking and impulsive behaviour, particularly in social and emotional contexts (Romer, 2010). AI systems firstly fail to acknowledge the complexity of human emotional experiences, which are personal and context-specific (Hollanek, 2025). Secondly, research has found that companion chatbots (e.g. Character. AI) are less empathetic, less easy for users to understand and less clinically appropriate than general assistant chatbots (e.g. ChatGPT). They also made fewer resource referrals and were less likely to escalate care when it was necessary (Brewster et al., 2025). These findings may cause psychological harm and instill additional barriers to adolescents accessing mental health support (Hollanek, 2025). Finally, it is important to recognize that individuals who are more vulnerable to AI dependence or adverse outcomes are generally those individuals who have existing mental health challenges, not vice versa (Huang et al., 2024; Huang & Huang, 2025).

## 2.2 Behavioural and social developmental risk

These risks are related to the disruption of youth social development, ethical judgement and behavioural norms (Yu et al., 2025). Adolescence is a critical period for building interpersonal skills, with young people being developmentally more sensitive to social influence, including in regions of the brain responsible for cognitive control, social cognition and reward processing (Telzer et al., 2018). When coupled with use of AI, this can cause unintended social consequences (Chamarthi et al. 2025). Real-life relationships are built on reciprocity, empathy and boundaries, whereas AI-based relationships are driven by algorithms and designed to fulfil the users' requests. Social interactions with AI systems may distort what healthy relationships look like, leading to a lack of social skills, unrealistic relationship expectations and social withdrawal (Lai et al., 2023; Xie et al., 2022).

Adolescents are particularly vulnerable to factors affecting value and identity formation. AI systems can be particularly problematic in this sense through creation of echo chambers. These are where adolescents are repeatedly exposed to information or opinions that reflect and reinforce, providing positive feedback on behaviours or values, with opposing views filtered out. The behaviours or values presented can be ethically questionable, for example sexually inappropriate or hostile (Brikel, 2025). This can be particularly concerning with AI integration into social media, where AI-driven algorithms create echo chambers and tap into adolescent vulnerabilities, serving increasingly narrow content, especially for commercial or political gain (Fadillah, 2025), even shifting adolescents towards radicalization (Núñez-Guerra, 2026). Radicalization accelerates polarization and repeated exposure to ethically questionable behaviours and values can lead to desensitization and normalization of fringe behaviour (Spálová, 2025). These risks accumulate, are interconnected and are reinforced by further individual use and simultaneous use by peers, all of which can alter adolescent behaviour and social development over time.

## 2.3 Privacy risk

Privacy risk relates to the exposure, use or access to adolescents' personal or sensitive health information (Yu et al., 2025). Research demonstrates that adolescents perceive a lack of control over the use of their personal health information and potential misuse of this data by AI systems (Shrestha et al., 2024). They are also fearful of data breaches and unauthorized access to health information, as well as a lack of knowledge on how AI systems collect data and how personal data may be used (Shrestha et al., 2024). In addition, their previous experiences with AI systems influence their trust and privacy in health AI (Lee et al., 2025). Beyond systemic issues, AI systems may also encourage users to disclose personal or sensitive health information (Yu et al., 2025). Adolescents are less protected by their caregivers, as they seek independence, despite their cognitive development predisposing them to impulsivity during this developmental period (Sanders, 2013). This has implications for their privacy as they may share health information with AI systems without grasping the potential consequences.

## 2.4 Toxicity risk

Toxicity risks relate to the potential for AI systems to autonomously produce and expose adolescents to harmful content without prompting (Yu et al., 2025). This is distinct from mental well-being risk which is due to youth interactions with AI systems. Toxicity can occur via two pathways: the automatic exposure to generated toxic content; simulated toxic interactions in conversation or role-playing when the user has not requested or triggered this response (Yu et al., 2025). Adolescent predisposition to risk-taking behaviour and the evolving capacity for emotional regulation may amplify the effects of toxic content on health and well-being, as AI systems can present this content in ways that often feel authoritative and personalized. For example, some AI systems have simulated sexual harassment and self-harm interactions with adolescents (Park et al., 2023).

## 2.5 Misuse and exploitation risk

This risk category is related to humans intentionally or unintentionally using AI to generate or spread harm which targets others, especially adolescents (Yu et al., 2025). This can take two forms. Firstly, unintentional misuse is where it is neither the system nor the user that intends to create harm, yet harm arises due to misinformation or limitations. Secondly, exploitation may occur when individuals use AI with the intention of causing harm (e.g. abuse or harassment). Adolescents can be particularly vulnerable to misinformation and exploitation due to their stage of cognitive development relating to executive function (National Academies, 2019). Data used to train AI models can include misinformation, which may result in inaccurate health information being provided to adolescents and repeated to them over time (Monteith et al., 2024; Ross et al., 2021). They are also less likely to assess reliability of AI-generated health information presented to them, given their developing skills in critical health literacy (Skipper, 2025; Taba et al., 2022) This can be particularly problematic as adolescents are still forming their identity and AI systems may reinforce underlying negative assumptions about themselves. For example, a teenager struggling with an undiagnosed eating disorder may ask for a meal plan which reinforces unhealthy eating behaviours. A particularly exploitation issue among adolescents is the prevalence of sexually explicit deepfake images and videos, with one in eight young people knowing someone who has been the target of deepfake nudes (Thorn, 2025). This kind of exploitation can lead to a plethora of mental health problems among adolescents including depression, anxiety and loneliness (Nixon, 2014).

## 2.6 Bias and discrimination risk

Bias and discrimination risk is inherent in AI systems which result in discriminatory, harmful or stereotypical content being presented to adolescents without prompting. This occurs when AI systems learn from unrepresentative or historically biased training data (Chen, 2023). Adolescents are particularly at risk of developing internal biases through the influence of AI-generated content due to their developmental stage (Fitzsimons et al., 2025). Adolescents from marginalized populations may feel excluded or underrepresented by AI systems trained on incomplete or biased datasets, which may affect their self-esteem and formation of identity. This in turn may lead to reduced trust in the technology, leading to inequitable access of AI systems (Hollanek, 2025).

## 2.7 Looking forward

While research has identified a plethora of risks from AI systems to adolescent health and well-being, adolescents themselves are increasingly aware of online harms and are calling for safeguards (ECPAT International, 2024). Without intervention, the risks to adolescent health and well-being are real, ongoing and may increase over time as access increases in global regions which are historically less connected (The Digital Watch, 2026). This in turn has potential to create long-term societal, economic and intergenerational impacts including reduced workforce participation, lower educational attainment, increased health system costs and burdens and increased morbidity (Baird et al., 2025; Patton et al., 2016).

Section 3

# Current policy and practice approaches and gaps

With widespread adoption of AI systems by adolescents and little known about the mid-to-long term impacts to their health and well-being, there is an urgent need for regulation. Since 2021, there has been an emergence of policies, principles and guidance on the use of AI, highlighting increased awareness of the risks of AI systems. However, adolescent-specific policy, principles and guidance for AI systems are scarce. In 2021, UNICEF released policy guidance for AI and children (UNICEF Innocenti, 2025). Grounded in the Convention on the Rights of the Child, this guidance offers ten requirements for child-centred AI, which aims to raise awareness of how AI systems can uphold or undermine those rights. The most recent update, in December 2025, considers the emergent opportunities and risks to health and well-being (e.g. companion AI).

Other policies, principles and guidance are available but are not specific to adolescent populations. In 2024, the World Health Organization (WHO) released updated guidance on the ethics and governance of AI for health, which list six ethical principles and the benefits and risks of large multi-modal models (LMMs) in healthcare (WHO, 2024). The six ethical principles include: (i) protect autonomy; (ii) promote human well-being, human safety and the public interest; (iii) ensure transparency, explainability and intelligibility; (iv) foster responsibility and accountability; (v) ensure inclusiveness and equity; and (vi) promote AI that is responsive and sustainable. In the European Union (EU) there is the AI Act, which classifies AI according to its level of risk (Regulation [EU] 2024/1689 [Artificial Intelligence Act]). Most obligations relating to the AI Act fall on providers of high-risk

AI systems, whether they are based in the EU or in a third-party country. Interestingly, this act prohibits AI systems which exploit vulnerabilities related to age which may distort behaviour and have the potential to cause harm. Considerations of adolescent vulnerabilities in relation to the EU AI Act warrant further attention. In addition, the Council of Europe framework convention on artificial intelligence and human rights, democracy and the rule of law, is the first international legally binding treaty. The framework convention aims to fill legal gaps left by rapid technological advances (Council of Europe, 2025). The OECD AI Principles (OECD, 2019) are values-based and were adopted to guide AI actors to develop trustworthy AI systems and provide recommendations to policymakers for effective AI policy. These global principles are the first intergovernmental standards and to date have 47 countries that uphold these principles. Existing policy, principles and guidance on AI systems are targeted towards multiple actors including governments, policymakers, AI developers and tech companies, international organizations, parents, educators and adolescents themselves.

Given AI's emerging health and well-being impacts to adolescents, there has been some self-regulation from AI systems. For example, OpenAI (ChatGPT) worked with 170 mental health experts to update their model to help ChatGPT recognize signs of distress, de-escalate conversations where necessary and guide people to real-world support (OpenAI, 2025). The update also distinguishes between healthy engagement and concerning use patterns. This update specifically focused on psychosis, mania, self-harm and suicide and emotional reliance on AI (OpenAI, 2025). Additionally, Character.ai has banned those under 18 years of age from using the platform, in response to increasing reports and feedback from regulators, safety experts and parents regarding the chatbots' interactions with adolescents (character.ai, 2025).

Despite some guidance, principles and emerging self-regulation from AI systems, governance of AI for adolescent health and well-being is fragmented and there are currently no global frameworks that consider the nuance of AI risks to adolescent health and well-being. The current design and implementation of AI systems can be inappropriate for some groups, especially those with pre-existing mental health challenges. In addition, AI systems do not account for the different developmental stages of adolescence, especially in their evolving capacity to understand and provide consent regarding use. Despite wanting to be involved in conversations around AI research and policy for health and well-being, adolescents are rarely consulted (Thai et al., 2023; Visram et al., 2023). With health and well-being risks cutting across sectors, an interdisciplinary approach that includes adolescents within governance and implementation presents a potentially effective way to move forward.

Section 4

# Future directions to identify and mitigate AI risks to protect adolescent health and well-being

Situating risks within established guidelines and taxonomies that reflect characteristics of adolescent development enables a systematic way to develop risk mitigation strategies. Identification of risks to health and well-being will be dynamic, as there is greater adolescent uptake in AI systems for different uses over time. Specific mitigation strategies for each Youth-AI Risk Category identified by Yu et al. (Yu et al., 2025) are listed in Table 1 below. Three identified mitigation strategies cut across all youth-AI risk categories. Firstly, adolescents must be involved in the development and implementation of AI systems and governance through equitable and rights-based approaches.

Secondly, consent processes should occur before adolescents can interact with AI systems. This will ensure that adolescents understand that they are interacting with AI, the intended use of the system and what the potential risks and benefits are. Finally, it is vital that developmentally and age-appropriate information and education are available for adolescents, acknowledging the wide age range in this group and the different corresponding needs of younger versus older adolescents. These overarching strategies align strongly with the WHO ethical principles for use of AI for health (WHO, 2024), however they take a youth-centred approach.

Table 1: Youth-centred AI risk mitigation strategies framework

| Youth-AI risk categories | Mitigation strategies |
|---|---|
| **Mental well-being** | · Clinician-informed parameters for psychological safety<br>· Oversight from parents and carers, providing open communication and structured boundaries for AI use<br>· Integrating mental health literacy with AI literacy to foster knowledge and skills through positive psychology |
| **Behavioural and social development** | · Regulation of AI-driven algorithms among adolescent users<br>· Ensure that adolescents continue to have ample opportunities for in-person social connection<br>· Limits on addictive design features within AI systems |
| **Privacy** | · Adolescent-specific protections for the use of personal and sensitive health data<br>· Reminder of AI system settings for privacy at regular intervals to allow data use settings to be changed<br>· Specific data privacy education for adolescents |
| **Toxicity** | · Mandatory risk and impact assessments for AI systems available to adolescents<br>· Regulation of AI systems which are of an unacceptable level of risk to adolescents<br>· Greater transparency within AI systems including content warnings for toxic content |
| **Misuse/exploitation** | · Specific critical health literacy for adolescents<br>· Data standards developed for data to train AI systems |
| **Bias/discrimination** | · Regulatory oversight bodies for AI systems<br>· Data standards developed for data to train AI systems<br>· Training algorithms on diverse datasets to ensure that implicit bias is reduced within AI systems |

Following identification of practical strategies to identify and mitigate AI risks, it is also vital to understand the actors who can implement these strategies and the incentives and trade-offs for feasibility and implementation. Key governance actors were identified including governments and policymakers, AI developers and technology companies, international organizations, educators, clinicians, parents and adolescents. Potential policy options were grouped under five key areas: (i) Adolescent Safety-By-Design; (ii) Data Governance and Privacy; (iii) Accountability; (iv) Developmentally Appropriate AI and Health Literacy; and (v) Psychological Safety and Supportive Boundaries. The complete list of mitigation strategies, actors, incentives and trade-offs is available in Appendix 1.

What must be carefully considered within any potential governance response is the key tensions between actors. The first key tension which must be considered is between AI system engagement and adolescent well-being. AI systems are designed for high user engagement, which leads to industry gain and the ability to monetize systems. This must be carefully balanced against adolescents' needs for boundaries and moderated engagement through regulation. While adolescents are asking for safeguards, this safeguarding must respect adolescents' autonomy and ensure that they do not become fatigued through arduous system processes. The second key tension exists between the innovation of AI systems and the safety of features. AI systems are at the forefront of technological advancement, with features often released rapidly. They also currently exist in an unregulated environment, which leads to these platforms being free to act in their own best interests. Conversely, safety-by-design is a slower process, requiring increased workload, cost and complexity of implementation for platforms due to the need for re-engineering of systems.

Thirdly, adolescent development creates a greater desire for independence and autonomy. This must also be taken into consideration in governance responses, which often require control from actors more powerful than adolescents. Governance responses from AI developers and tech companies can lead to AI systems which are transparent, establishing their legitimacy as safe providers and reducing their overall liabilities. However, over-protection can also lead to resistance from adolescents, as they feel they lose agency relating to AI use. Finally, the need for AI systems to collect data is a tension which must be balanced with adolescent privacy. The current operation of AI systems provides them a wealth of data which they can use for their own gain. This can occur through feedback loops which enhance and personalize the experience of users to keep them engaged and through monetization. Akin to safety-by-design, privacy-by-design is also a slower process and will reduce the overall volume of data which AI platforms currently collect and will require a greater workload and higher costs to implement.

Section 5

# Conclusion

This framework provides a comprehensive approach for diverse stakeholders – including governments and policymakers, AI developers and technology companies, international organizations, educators, clinicians, parents and adolescents – to navigate the dynamic and evolving nature of AI risks. Central to this approach is the promotion of a nuanced perspective on AI that recognizes its potential to drive creativity, learning and social progress, while simultaneously addressing the risks to adolescent health and well-being. Rather than adopting a view of AI as inherently beneficial or harmful, the framework advocates for balanced mitigation strategies that safeguard adolescent health and well-being without limiting autonomy and innovation. In doing so, it establishes a foundation for ethical, inclusive and future-oriented AI that prioritizes adolescents' rights and protects them in an era of rapid digital transformation.

# References


Baird, S., Choonara, S., Azzopardi, P. S., Banati, P., Bessant, J., Biermann, O., Capon, A., Claeson, M., Collins, P. Y., De Wet-Billings, N., Dogra, S., Dong, Y., Francis, K. L., Gebrekristos, L. T., Groves, A. K., Hay, S. I., Imbago-Jácome, D., Jenkins, A. P., Kabiru, C. W., … Viner, R. M. (2025). A call to action: the second Lancet commission on adolescent health and wellbeing. *The Lancet*, *405*(10493), 1945-2022. https://doi.org/10.1016/S0140-6736(25)00503-3

Brewster, R. C. L., Zahedivash, A., Tse, G., Bourgeois, F., & Hadland, S. E. (2025). Characteristics and safety of consumer chatbots for emergent adolescent health concerns. *JAMA Network Open*, *8*(10), Article e2539022- https://doi.org/10.1001/jamanetworkopen.2025.39022

Brikel, H. (2025). *Childhood vulnerability in the age of artificial intelligence: Behavioral addiction and value transformation*. Zenodo. https://doi.org/10.5281/zenodo.15747033

Chamarthi, V. S., Das, P., & Kashyap, R. (2025, October 22). AI as a novel digital stressor in adolescent psychosis: Clinical and ethical implications. *The International Journal of Psychiatry in Medicine*. Advance online publication. https://doi.org/10.1177/00912174251392768

character.ai. (2025). *An Update On Changes to Our Under-18 Experience*. Retrieved 6 December 2025 from https://blog.character.ai/an-update-on-changes-to-our-under-18-experience/

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, *10*(1), Article 567. https://doi.org/10.1057/s41599-023-02079-x

Council of Europe. (2025). *The Framework Convention on Artificial Intelligence*. Retrieved 6 December 2025 from https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

Diaz, A. D., Leong, A. W., Bilge-Johnson, S., Shapiro, M., Noureddine, C., & Kaliebe, K. (2025). Clinical assessment and implications of parasocial relationships in adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry 65*(3), 331-336. https://doi.org/10.1016/j.jaac.2025.07.005

The Digital Watch.. (2026, February,19). *India unveils MANAV vision as new global pathway for ethical AI*. Retrieved 11 March 2026 from https://dig.watch/updates/india-unveils-manav-vision-as-new-global-pathway-for-ethical-ai

ECPAT International, Eurochild, & Terre Des Hommes Netherlands on behalf of the Down To Zero Alliance. (2024). *VOICE Report "speaking up for change: Children's and caregivers' voices for safer online experiences„*. ECPAT International. https://ecpat.org/resource/voice-report-speaking-up-for-change-childrens-and-caregivers-voices-for-safer-online-experiences/

European Parliament (2024) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Fadillah, D. (2025). The need for research on AI-driven social media and adolescent mental health. *Asian Journal of Psychiatry*, *108*, Article 104513. https://doi.org/https://doi.org/10.1016/j.ajp.2025.104513

Fitzsimons, A. Z., Gerber, E. M., & Long, D. (2025, June 23). AI constructs gendered struggle narratives: Implications for self-concept and systems design, *FAccT '25: Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 2290 - 2301. https://doi.org/10.1145/3715275.3732156

Hollanek, T. S., & Sobey, A. (2025). *AI companions for health and mental wellbeing: Opportunities, risks, and policy implications*. The Leverhulme Centre for the Future of Intelligence. https://www.repository.cam.ac.uk/items/d3229fe5-db87-42ff-869b-11e0538014d8

Huang, S., Lai, X., Ke, L., Li, Y., Wang, H., Zhao, X., Dai, X., & Wang, Y. (2024). AI technology panic—is AI dependence bad for mental health? A cross-lagged panel model and the mediating roles of motivations for AI use among adolescents. *Psychology Research and Behavior Management*, *17*, 1087-1102. https://doi.org/10.2147/PRBM.S440889

Huang, Y., & Huang, H. (2025). Exploring the effect of attachment on technology addiction to generative AI chatbots: A structural equation modeling analysis. *International Journal of Human–Computer Interaction*, *41*(15), 9440-9449. https://doi.org/10.1080/10447318.2024.2426029

Internet Matters. (2025). *Me, myself and AI: Understanding and safeguarding children's use of AI chatbots*. Retrieved 1 December 2025 from https://www.internetmatters.org/hub/research/me-myself-and-ai-chatbot-research/

Lai, T., Xie, C., Ruan, M., Wang, Z., Lu, H., & Fu, S. (2023). Influence of artificial intelligence in education on adolescents' social adaptability: The mediatory role of social support. *PLoS One*, *18*(3), Article e0283170. https://doi.org/10.1371/journal.pone.0283170

Lee, J., Jung, K., Newman, E. G., Chow, E., & Chen, Y. (2025). *Understanding adolescents' perceptions of benefits and risks in health AI technologies through design fiction. CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1-20, Article 311. https://doi.org/10.1145/3706598.3713244

Mansfield, K. L., Ghai, S., Hakman, T., Ballou, N., Vuorre, M., & Przybylski, A. K. (2025). From social media to artificial intelligence: improving research on digital harms in youth. *The Lancet Child & Adolescent Health*, *9*(3), 194-204. https://doi.org/10.1016/S2352-4642(24)00332-8

Monteith, S., Glenn, T., Geddes, J. R., Whybrow, P. C., Achtyes, E., & Bauer, M. (2024). Artificial intelligence and increasing misinformation. *The British Journal of Psychiatry*, *224*(2), 33-35. https://doi.org/10.1192/bjp.2023.136

Nagata, J. M., Memon, Z., Huang, O., & Moreno, M. A. (2025). Viewpoint: Adolescent health and generative AI-risks and benefits. *JAMA Pediatr*. https://doi.org/10.1001/jamapediatrics.2025.4502

National Academies of Sciences, Engineering and Medicine (National Academies). . (2019). Youth Engagement (Appendix B). In Backes, E.P., Bonnie, R.J. (Eds.), *The Promise of Adolescence: Realizing Opportunity for All Youth*. National Academies Press. https://www.ncbi.nlm.nih.gov/books/NBK545472/#

Nixon, C. L. (2014). Current perspectives: the impact of cyberbullying on adolescent health. *Adolesc Health Med Ther*, *5*, 143-158. https://doi.org/10.2147/ahmt.S36456

Núñez-Guerra, P. M. (2026). The factor of social media and artificial intelligence in the radicalization of young people in jihadist terrorism. Ethical and Social Impacts of Information and Communication Technology: 22nd International Conference, ETHICOMP 2025, Lisbon, Portugal, September 17-19, 2025, Proceedings. *Springer Nature Switzerland.*

Opel, D. J., Kious, B. M., & Cohen, I. G. (2023). AI as a mental health therapist for adolescents. *JAMA Pediatrics*, *177*(12), 1253-1254. https://doi.org/10.1001/jamapediatrics.2023.4215

OpenAI. (2025, October 27). *Strengthening ChatGPT's responses in sensitive conversations*. Retrieved 6 December 2025 from https://openai.com/index/strengthening-chatgpt-responses-in-sensitive-conversations/

Organisation for Economic Co-operation and Development (OECD). (2019). *OECD AI Principles overview*. Retrieved 6 December 2025 from https://oecd.ai/en/ai-principles

OECD. (2021, January). *Children in the digital environment: Revised typology of risks*. Retrieved 20 November 2025 from https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/01/children-in-the-digital-environment_9d454872/9b8f222e-en.pdf

Park, J. K., Singh, V., & Wisniewski, P. (2023). Supporting youth mental and sexual health information seeking in the era of artificial intelligence (AI) based conversational agents: Current landscape and future directions. *SSRN Electronic Journal.* *https://doi.org/10.2139/ssrn.4601555*

Patton, G. C., Sawyer, S. M., Santelli, J. S., Ross, D. A., Afifi, R., Allen, N. B., Arora, M., Azzopardi, P., Baldwin, W., Bonell, C., Kakuma, R., Kennedy, E., Mahon, J., McGovern, T., Mokdad, A. H., Patel, V., Petroni, S., Reavley, N., Taiwo, K., . . . Viner, R. M. (2016). Our future: a Lancet commission on adolescent health and wellbeing. *The Lancet*, *387*(10036), 2423-2478. https://doi.org/10.1016/S0140-6736(16)00579-1

Philippe, T. J., Sikder, N., Jackson, A., Koblanski, M. E., Liow, E., Pilarinos, A., & Vasarhelyi, K. (2022). Digital health interventions for delivery of mental health care: Systematic and comprehensive meta-review. *JMIR Ment Health*, *9*(5), e35159. https://doi.org/10.2196/35159

Radez, J., Reardon, T., Creswell, C., Lawrence, P. J., Evdoka-Burton, G., & Waite, P. (2021). Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. *Eur Child Adolesc Psychiatry*, *30*(2), 183-211. https://doi.org/10.1007/s00787-019-01469-4

Remschmidt, H. (1994). Psychosocial milestones in normal puberty and adolescence. *Horm Res*, *41 Suppl 2*, 19-29. https://doi.org/10.1159/000183955

Robb, M. B., & Mann, S. (2025). *Talk, trust, and trade-offs: How and why teens use AI companions*. Common Sense Media. https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf

Romer, D. (2010). Adolescent risk taking, impulsivity, and brain development: implications for prevention. *Dev Psychobiol*, *52*(3), 263-276. https://doi.org/10.1002/dev.20442

Ross, R. M., Rand, D. G., & Pennycook, G. (2021). Beyond "fake news": Analytic thinking and the detection of false and hyperpartisan news headlines. *Judgment and Decision Making*, *16*(2), 484-504. https://doi.org/10.1017/S1930297500008640

Sanders, R. A. (2013). Adolescent psychosocial, social, and cognitive development. *Pediatrics in review*, *34*(8), 354—359. https://doi.org/10.1542/pir.34-8-354

Sawyer, S. M., Afifi, R. A., Bearinger, L. H., Blakemore, S. J., Dick, B., Ezeh, A. C., & Patton, G. C. (2012). Adolescence: a foundation for future health. *The Lancet*, *379*(9826), 1630-1640. https://doi.org/10.1016/s0140-6736(12)60072-5

Sharma, G., Yaffe, M. J., Ghadiri, P., Gandhi, R., Pinkham, L., Gore, G., & Abbasgholizadeh-Rahimi, S. (2025). Use of artificial intelligence in adolescents' mental health care: Systematic scoping review of current applications and future directions. *JMIR Ment Health*, *12*, Article e70438. https://doi.org/10.2196/70438

Shrestha, A. K., Barthwal, A., Campbell, M., Shouli, A., Syed, S., Joshi, S., & Vassileva, J. (2024). Navigating AI to unpack youth privacy concerns: An in-depth exploration and systematic review [Preprint]. *arXiv.* https://doi.org/10.48550/arXiv.2412.16369

Skipper, Y. (2025). 'I'd probably scroll by': An exploration of young people's views on spotting and stopping misinformation. *Children & Society*, *39*(5), 940-949. https://doi.org/https://doi.org/10.1111/chso.12962

Spálová, A. J. L. (2025, December 19). Ethical implications of AI-generated "brainrot" content: Analysing animated violence in short-form videos. . In B.Francistyová, L.Furtáková, & M. P. Hossová, (Eds.), *Marketing Identity: The Power(s) of Communication: Conference proceedings from the International Scientific Conference, 11th-12th November 2025, Voderady* (p. 285). https://doi.org/10.34135/mmidentity-2025-26

Stiglic, N., & Viner, R. M. (2019). Effects of screentime on the health and well-being of children and adolescents: a systematic review of reviews. *BMJ Open*, *9*(1), Article e023191. https://doi.org/10.1136/bmjopen-2018-023191

Taba, M., Allen, T. B., Caldwell, P. H. Y., Skinner, S. R., Kang, M., McCaffery, K., & Scott, K. M. (2022). Adolescents' self-efficacy and digital health literacy: a cross-sectional mixed methods study. *BMC Public Health*, *22*(1), Article 1223. https://doi.org/10.1186/s12889-022-13599-7

Telzer, E. H., van Hoorn, J., Rogers, C. R., & Do, K. T. (2018). Social Influence on Positive Youth Development: A Developmental Neuroscience Perspective. *Adv Child Dev Behav*, *54*, 215-258. https://doi.org/10.1016/bs.acdb.2017.10.003

Thai, K., Tsiandoulas, K. H., Stephenson, E. A., Menna-Dack, D., Zlotnik Shaul, R., Anderson, J. A., Shinewald, A. R., Ampofo, A., & McCradden, M. D. (2023). Perspectives of youths on the ethical use of artificial intelligence in health care research and clinical care. *JAMA Network Open*, *6*(5), Article e2310659. https://doi.org/10.1001/jamanetworkopen.2023.10659

Thorn. (2025, March). *Deepfake nudes & young people: Navigating a new frontier in technology-facilitated nonconsensual sexual abuse and exploitation*. Retrieved 5 December 2025 from https://info.thorn.org/hubfs/Research/Thorn_DeepfakeNudes&YoungPeople_Mar2025.pdf

United Nations Children's Fund (UNICEF) Innocenti. (2025, December). *UNICEF Guidance on AI and Children 3.0: Updated guidance for governments and businesses to create AI policies and systems that uphold children's rights.* . Retrieved December 5 2025 from www.unicef.org/innocenti/media/11991/file/UNICEF-Innocenti-Guidance-on-AI-and-Children-3-2025.pdf

Visram, S., Leyden, D., Annesley, O., Bappa, D., & Sebire, N. J. (2023). Engaging children and young people on the potential role of artificial intelligence in medicine. *Pediatric Research*, *93*(2), 440-444. https://doi.org/10.1038/s41390-022-02053-4

Walker, I. (2024). *AI Amplified*. Retrieved 1 December 2025 from https://my-year13.s3.ap-southeast-2.amazonaws.com/public/b2b-reports/year13-ai-amplified-report.pdf

World Health Organization(WHO). (2024). *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models*. https://www.who.int/publications/i/item/9789240084759

Xie, C., Ruan, M., Lin, P., Wang, Z., Lai, T., Xie, Y., Fu, S., & Lu, H. (2022). Influence of artificial intelligence in education on adolescents' social adaptability: A machine learning study. *International Journal of Environmental Research and Public Health*, *19*(13), Article 7890. https://www.mdpi.com/1660-4601/19/13/7890

Yu, Y., Liu, Y., Zhang, J., Huang, Y., & Wang, Y. (2025). Youth-centered genAI risks (YAIR): a taxonomy of generative AI risks from empirical data. In *Symposium on Proceedings of the Twenty-First USENIX Conference on Usable Privacy and Security, SOUPS 2025* (pp. 149-165). Seattle, United States. https://doi.org/10.48550/arXiv.2502.16383

# Appendix 1

| Policy options | Mitigation strategies | Youth-AI risk category | Actors | Incentives | Trade-offs |
|---|---|---|---|---|---|
| **Adolescent safety-by-design** | Adolescent involvement in AI system development and governance | Cross-cutting | Adolescents, Tech companies, AI developers, Governments | Adolescents: AI systems which reflect lived experience, agency, build trust<br>Tech companies, AI developers: Products meet adolescent needs, reputation, trust, risk identification<br>Governments: Future proofing policy, better outcomes | Adolescents: Time, tokenism, equity<br>Tech companies, AI developers: Increased cost, slower development<br>Governments: Slower processes, resource requirements, manage diversity |
| | Limits on addictive design features within AI systems | Behavioural and social development | Governments, international orgs, Tech companies, AI developers | Governments: Reduce public health harms<br>International orgs: advance uptake of international standards<br>Tech companies, AI developers: Gain reputation, legitimacy as safe provider | Governments: Implementation complexity<br>International orgs: harmonization across jurisdictions<br>Tech companies, AI developers: Lower engagement - less data/monetization |
| | Regulation of AI-driven algorithms among adolescent users | Behavioural and social development | Governments, Tech companies, AI developers | Governments: Reduce public health harms<br>Tech companies, AI developers: Gain reputation, legitimacy as safe provider | Governments: Implementation complexity<br>Tech companies, AI developers: Lower engagement - less data/monetization, Re-engineer recommendation systems - high workload and cost |
| | Greater transparency within AI systems, including content warnings for toxic content | Toxicity | Governments, Tech companies, AI developers | Governments: Increase platform accountability<br>Tech companies, AI developers: gain reputation, legitimacy as safe provider | Tech companies, AI developers: Increased workload and cost, re-engineer moderation systems |
| | Reminder of AI system settings for privacy at regular intervals to allow them to change settings on the use of their data | Privacy | Tech companies, AI developers | Tech companies, AI developers: Increased transparency, gain reputation, legitimacy | Tech companies, AI developers: Increased workload and cost, less data/ monetization |
| | Ensure that adolescents continue to have ample opportunities for in-person social connection | Behavioural and social development | Parents | Parents: Respect for autonomy, promote offline social connection | Parents: Lack of knowledge, lack of time, competing priorities |

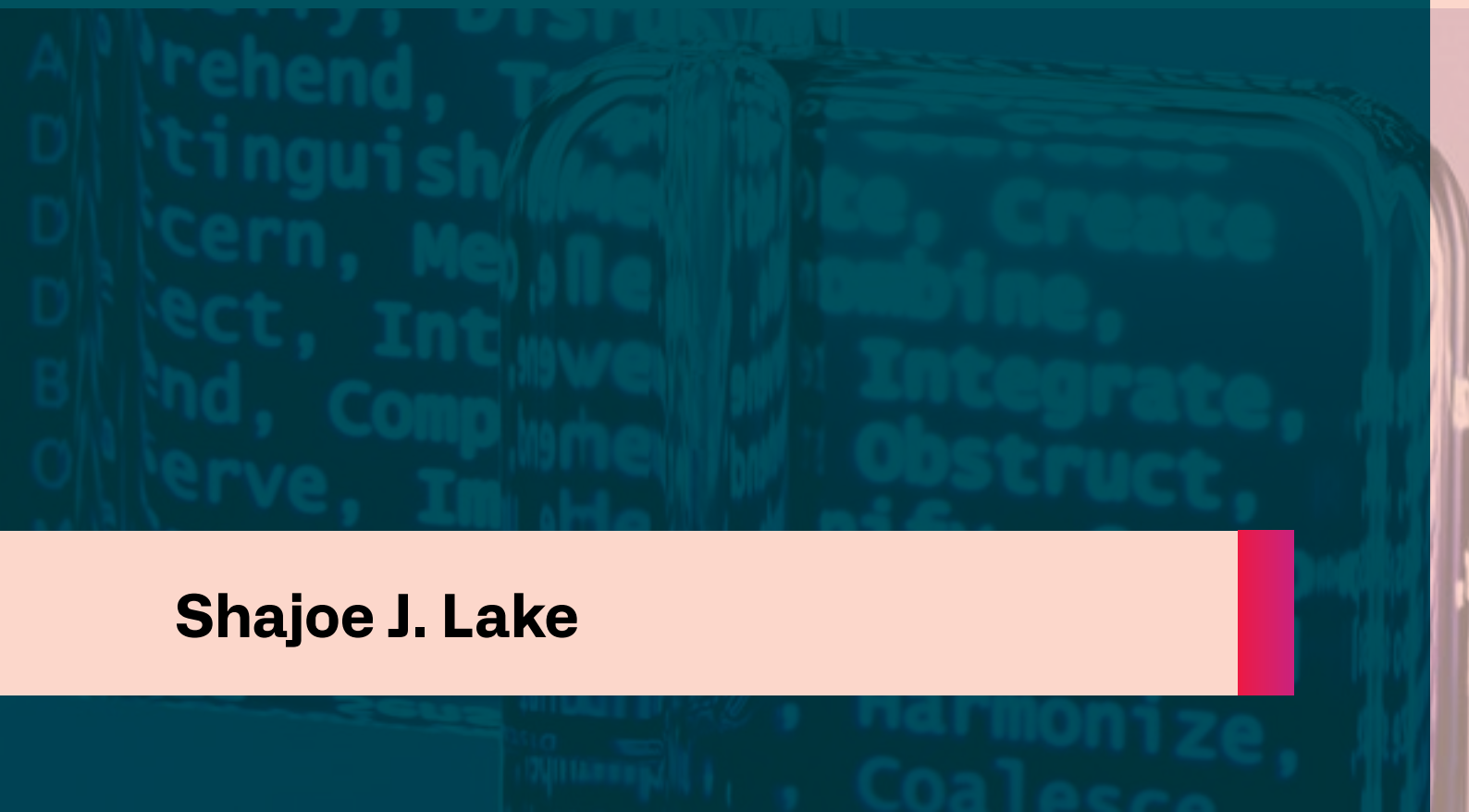
| Policy options | Mitigation strategies | Youth-AI risk category | Actors | Incentives | Trade-offs |
|---|---|---|---|---|---|
| **Data governance and privacy** | Adolescent specific protections for the use of their personal and sensitive health data | Privacy | Governments, International orgs, Tech companies, AI developers | Governments: Compliance with privacy laws<br>International orgs: Advance uptake of international standards<br>Tech companies, AI developers: Build trust through transparency, lower regulatory risk | Governments: Implementation complexity<br>International orgs: Harmonization across jurisdictions<br>Tech companies/ AI developers: Reduced data access affecting monetization |
| | Consent processes prior to use of AI systems | Cross-cutting | Adolescents, Governments, Tech companies, AI developers | Adolescents: Agency over their personal and sensitive data<br>Governments: Compliance with privacy laws<br>Tech companies, AI developers: Build trust through transparency, lower regulatory risk | Adolescents: Fatigue from processes<br>Governments: Implementation complexity<br>Tech companies/ AI developers: Reduced data access affecting monetization |
| | Reminder of AI system settings for privacy at regular intervals to allow adolescents to change settings on the use of their data | Privacy | Adolescents, Tech companies, AI developers | Adolescents: Agency over their data<br>Tech companies, AI developers: build trust through transparency, lower regulatory risk | Adolescents: fatigue from processes<br>Tech companies/ AI developers: Reduced data access affecting monetization |
| | Specific education for data privacy within AI systems | Privacy | Educators | Educators: Improve adolescent skills, strengthen curriculum | Educators: Curriculum/ time constraints, inconsistency across platform settings |

| Policy options | Mitigation strategies | Youth-AI risk category | Actors | Incentives | Trade-offs |
|---|---|---|---|---|---|
| **Accountability** | Mandatory risk and impact assessments for AI systems available to adolescents | Toxicity | Tech companies, AI developers | Tech companies, AI developers: Increased reputation, legitimacy, reduced liabilities | Tech companies, AI developers: Increased workload, documentation |
| | Regulatory oversight bodies for AI systems | Bias/ discrimination | Governments, International orgs | Governments: Compliance, clearer liabilities for public health harms<br>International orgs: Advance uptake of international standards | Governments: Implementation complexity<br>International orgs: Harmonization across jurisdictions |
| | Regulation of AI systems which are of an unacceptable level of risk to adolescents | Toxicity | Governments, International orgs | Governments: Compliance, clearer liabilities for public health harms<br>International orgs: Advance uptake of international standards | Governments: Implementation complexity<br>International orgs: Harmonization across jurisdictions |
| **AI and health literacy** (**developmentally appropriate**) | Developmentally appropriate information/ education on AI systems | Cross-cutting | Educators, International orgs, Tech companies, AI developers | Educators: Improve adolescent skills, strengthen curriculum<br>International orgs: Promotion of harmonized education and information<br>Tech companies, AI developers: Lower misuse rates due to prompts/tutorials that reduce burden | Educators: Curriculum/ time constraints, inconsistency across platform settings |
| | Specific critical health literacy education for adolescents | Misuse/ exploitation | Educators | Educators: Improve adolescent skills, strengthen curriculum | Educators: Curriculum/ time constraints |
| | Integrating mental health literacy with AI literacy to foster knowledge and skills through positive psychology | Mental well-being | Educators | Educators: Improve adolescent skills, strengthen curriculum | Educators: Curriculum/ time constraints |

| Policy options | Mitigation strategies | Youth-AI risk category | Actors | Incentives | Trade-offs |
|---|---|---|---|---|---|
| **Psychological safety and supportive boundaries** | Clinician informed parameters for psychological safety | Mental well-being | Clinicians<br>Tech companies, AI developers | Clinicians: Patient safety, clinical alignment, increased efficiency and trust with systems<br>Tech companies, AI developers: Increased safety and compliance, higher adoption, better performance | Clinicians: Time to implements, reduced flexibility, autonomy<br>Tech companies, AI developers: Implementation complexity (time, development), engagement and maintenance |
| | Data standards developed for data to train AI systems | Misuse/ exploitation, Bias/ discrimination | Tech companies, AI developers, international orgs | Tech companies, AI developers: Increased transparency, lower misuse rates due to prompts/ tutorials that reduce burden<br>International orgs: Advance uptake of international standards, promotion of harmonized education and information | Tech companies: Reduced data access affecting monetization, implementation complexity (cost, data curation), reduced performance<br>International orgs: Harmonization across jurisdictions |
| | Importance of training algorithms on diverse datasets to ensure that implicit bias is reduced within AI systems | Bias/ discrimination | Tech companies, AI developers, | | |
| | Oversight from parents and carers, providing open communication and structured boundaries for AI use | Mental well-being | Parents | Parents: Respect for autonomy, promote offline social connection | Parents: Lack of knowledge, lack of time, competing priorities |

# 03

# AI governance for youth health and well-being: Closing policy gaps and building inclusive frameworks



**Shajoe J. Lake**

# High level summary

Young people today encounter artificial intelligence (AI) as part of the ordinary infrastructures of communication, education, health and public services and this sustained exposure occurs during a critical developmental period. Yet most AI governance frameworks still treat young people as a future workforce to be skilled for the digital economy rather than as a distinct group whose rights, health and psychosocial development require specific safeguards.

Building on DTH Lab's report Youth health and well-being in national, regional and global AI governance instruments, this position paper argues that young people's health and well-being must become a core priority in AI governance through four mutually reinforcing reforms:

1. embedding youth impact assessments and child-rights duties in regulation;

2. institutionalizing youth participation;

3. strengthening oversight of AI systems in health, education and platforms; and

4. promoting age-appropriate design, AI literacy and equitable access.

Together, these reforms translate existing international commitments into concrete governance mechanisms that can be implemented across diverse legal and policy contexts.

Section 1

# Introduction

## 1.1 AI woven into everyday youth life

Young people – children (0–12), adolescents (13–17) and young adults (18–24) – are among the most connected demographic groups worldwide, with around three-quarters of those aged 15–24 using the Internet in 2022, compared with roughly two-thirds of the overall population (International Telecommunication Unit [ITU], 2022). For the current generation of young people, being the first to be born into the digital age, artificial intelligence (AI) systems are woven into daily routines through social media feeds, search engines, recommendation systems, digital learning platforms and health-related applications (Kickbusch et al., 2021; The Lancet Digital Health, 2021). High connectivity amplifies opportunities for learning, creativity and social participation but also heightens exposure to AI-mediated risks, such as profiling, targeted advertising, polarizing recommendation loops and automated decision-making, which affects educational trajectories and mental health (Klein, 2023; Jones Day, 2023; Wilkins,2024; Wagner et al., 2024).

Adolescents are among the heaviest and most frequent users of smartphones, social media and AI chatbots and growing evidence shows that these algorithmically mediated environments now structure key developmental processes of identity formation, social comparison and emotion regulation during the teenage years (Duffy, 2025; Hussey, 2026; Nature, 2025; Walsh, 2025). Additionally, emerging evidence on AI in education and digital mental health applications suggests

that algorithmic systems can shape attention, motivation and perceived social support, but also introduce novel forms of dependency, surveillance and misclassification that have not yet been fully mapped in policy (Reynard et al., 2022; Thomson et al., 2025; Vertsberger et al., 2022; Wagner et al., 2024). These dynamics underscore why AI governance that is neutral to age or development is unlikely to be sufficient.

## 1.2 Well-being as a policy objective and right

The World Health Organization (WHO, n.d.) defines well-being as a positive state experienced by individuals and societies, shaped by structural, social and economic conditions and not merely the absence of illness. Framing well-being as a public policy objective implies that AI governance should go beyond harm avoidance to actively support environments in which young people can flourish. Youth well-being also has a strong rights foundation. Article 3 of the UN Convention on the Rights of the Child (UNCRC) affirms children's rights to privacy, protection from interference, access to information and participation in decisions that affect them, anchored in the "best interests of the child" principle (United Nations [UN], 1989). These principles require that AI systems used by or affecting minors be designed and governed in ways that protect their dignity, autonomy, safety and development, rather than treating children simply as data subjects or consumers.

International guidance has begun translating these rights into AI-specific expectations. UNICEF's Policy Guidance on AI for Children (2021) calls for child-centred AI design and governance, emphasizing safety, data protection, fairness, transparency and support for development and well-being across the AI lifecycle. The UN report *Governing AI for Humanity* (2024) similarly links AI governance to societal well-being, explicitly highlighting young people as a priority group and urging states to address risks such as AI-amplified disinformation and inadequate content verification that disproportionately affect young users.

Section 2

# Evidence of risks and opportunities

## 2.1 Health, mental health and digital determinants

Digital technologies, including AI-driven systems, are now recognized as determinants of health, influencing exposure to information, social networks and behavioural nudges that shape physical and mental outcomes (The Lancet Digital Health, 2021). Studies in adolescent populations suggest that AI-enabled educational platforms and chatbots can provide personalized feedback and social support, but can also reinforce inequities, create new forms of performance pressure, or misinterpret emotional cues, with implications for social adaptability and self-esteem (Brewster et al., 2022; Funk, Shabbaz & Vesteinsson, 2023).

AI-powered chatbots and recommendation algorithms can direct young people toward supportive content, crisis resources, or peer communities, while also risking over-reliance on non-clinical tools, exposing sensitive data, or surfacing harmful content through engagement-optimized recommender systems (Wagner et al., 2024). Professional bodies in healthcare are beginning to call for systematic validation, transparency about data use and clear clinical governance for the AI mental health tools used by adolescents, but regulatory frameworks often lag behind in practice (APA, 2025).

## 2.2 Education, learning outcomes and surveillance

AI in education promises more personalized learning, adaptive assessment and early identification of students who may need additional support, including those with learning difficulties or those at risk of disengagement (Garzón et al., 2025; Tapalova & Zhiyenbayeva, 2022). At the same time, schools are adopting AI-powered monitoring tools and learning analytics systems that track engagement, behaviour, and even emotional states, sometimes without clear evidence of benefit or robust safeguards against bias and over-surveillance (Beerwinkle, 2021; Nawaz et al., 2025).

When such systems operate without age-appropriate transparency and meaningful safeguards, they may undermine trust between students and institutions, encode existing social biases, or normalize constant monitoring as a condition of access to education. These risks raise specific questions for young people's well-being: how to balance early support with respect for autonomy; how to avoid the effects of confirmation bias; and how to ensure that educational AI enhances rather than constrains young people's capacities (Alfredo et al., 2024; Nawaz et al., 2025).

## 2.3 Platforms, algorithms and identity formation

Social media and content platforms are core environments in which young people explore identity, relationships and political expression and the recommender systems that structure these spaces are central AI governance concerns (Arora et al., 2023; Brewster et al., 2022; Funk, Shahbaz & Vesteinsson, 2023; Salikhov, 2025; The Economist, 2025; Chow & Haupt, 2025). Recommendation algorithms can help adolescents find supportive communities, creative outlets and high-quality educational content; they can also amplify harmful material, encourage excessive screen time, or expose users to harassment and abuse (Yu et al., n.d.; Wahab, 2025; WeProtect Global Alliance, n.d.; Mithani, 2025).

Generative AI companions and virtual agents increasingly offer quasi-social interaction, advice, or emotional support. This raises novel questions around attachment, manipulation and consent for users who may still be learning to interpret social cues and evaluate credibility (Brewster et al., 2022; Yu et al., n.d.; Wahab, 2025; WeProtect Global Alliance, n.d.; Mithani, 2025; Sanford, 2025; O'Donnell, 2025; Chow & Haupt, 2025). Without safeguards that recognize developmental vulnerabilities, these AI systems risk shaping young people's self-perception, body image and worldview in ways that are not transparent to them or to caregivers, educators and regulators.

Section 3

# Current gaps in AI governance

## 3.1 Youth not recognized as distinct stakeholders

Most national AI strategies and regulatory proposals treat young people as part of the general population or as beneficiaries of digital skills and innovation agendas, rather than as a distinct stakeholder group whose specific rights and vulnerabilities require targeted measures (Lake, 2025). This generic framing obscures how AI can affect adolescents differently, for example through identity-sensitive recommendation patterns, behavioural nudging at school, or the emotional influence of generative AI companions. Without explicit recognition that young people are a distinct group, AI governance instruments rarely specify differentiated impact analysis, age-tailored risk thresholds, or safeguards that reflect appropriate cognitive and psychosocial developmental stages (Lake, 2025). As a result, policies are more likely to rely on generic safety, or language relating vaguely to the concept of fairness, which leaves gaps in protection for minors and fails to anticipate youth-specific harm pathways.

## 3.2 Weak integration of health and safety considerations

Young people's health and safety typically enters AI policy debates indirectly through education or digital skills initiatives, with limited attention paid to broader determinants of health. National AI strategies often emphasize innovation, competitiveness and workforce preparedness, while relegating youth safety and health to separate online safety or child-protection frameworks that may not fully cover AI-enabled harms (Lake, 2025). High-level principles on safety, privacy, or transparency rarely translate

into mandatory checks on bias, data minimization, or psychosocial risk when AI is deployed in schools, health services, or youth-targeted platforms (Lake, 2025). In practice, this means AI systems can be piloted in classrooms or youth mental health contexts without systematic assessment of their effects on anxiety, attention, or stigma and without clear routes for redress if harms occur.

## 3.3 Inconsistent rights-based approaches and limited accountability

Despite widespread endorsement of the UNCRC, rights-based approaches are inconsistently integrated into AI governance frameworks and explicit references to child rights are still relatively rare in AI-specific regulation (Lake, 2025). Ethical guidelines frequently stress fairness, transparency, or human oversight but stop short of imposing enforceable obligations that reflect children's rights to privacy, participation and protection from economic or political exploitation (WHO, 2021). Where AI harms a young person – through discriminatory profiling, harmful content exposure, or denial of access to essential services – the routes to remedy are often opaque or fragmented across regulators (Lake, 2025). Youth impact assessments are seldom mandated, audits do not routinely examine differential effects on minors and dedicated oversight mechanisms such as children's digital ombudspersons are the exception rather than the norm. This leaves many commitments to young people's well-being largely rhetorical.

Section 4

# Emerging policy landscape

## 4.1 Global and regional frameworks

International and regional initiatives are gradually bringing youth concerns into AI governance. At the global level, the UN *Governing AI for Humanity* report urges states to promote societal well-being through AI and explicitly identifies young people as a priority group, calling for policies that address AI-driven disinformation, information integrity and content governance. The report sits alongside the UNCRC framework and UNICEF's AI for children guidance, offering a high-level roadmap for aligning AI governance with children's rights and collective well-being. WHO guidance provides a framework for planning, developing and implementing youth-centred digital health interventions and WHO's *Ethics and Governance of Artificial Intelligence for Health* offers high-level guidance on how to ethically govern AI (WHO, 2020; WHO 2021).

Regional bodies have begun to embed youth perspectives more directly. The European Union's AI Act, expected to enter into force in the near term, includes specific protections for minors by prohibiting AI systems that exploit vulnerabilities associated with age and banning manipulative or harmful systems directed at children. It also classifies certain AI applications involving children, such as unsafe AI-enabled toys or educational systems that influence mental health, as high-risk, subjecting them to stricter conformity assessment and oversight (EU, 2024). The African Union's 2024 continental strategy on AI links AI adoption to the empowerment of Africa's large youth population, framing AI as a means to strengthen health systems, education and employment while cautioning against discrimination and exclusion (African Union, 2024). The strategy also calls for youth involvement in decision-making processes, echoing the broader AU Digital Transformation Strategy (2020–2030), which centres on digital inclusion for young people in its regional development plans.

The Council of Europe's 2020 Declaration on Youth Participation in AI Governance similarly encourages member states to create structured mechanisms for youth input into AI policymaking, recognizing that legitimacy and effectiveness depend on including those most affected by digital transformations (Council of Europe, 2020). These regional instruments signal an emerging consensus that youth perspectives should inform AI norms, even if implementation remains uneven.

## 4.2 Sectoral and cross-cutting initiatives

Sector-specific initiatives are beginning to address young people's use of AI in more concrete terms. In health, professional associations and regulators are drafting guidance on AI tools used in paediatric and adolescent care, paying particular attention to clinical validation, safety across age groups and responsible data governance. These efforts acknowledge that models trained mainly on adult data may perform poorly for younger populations and that mental health tools require especially stringent oversight (Muralidharan et al., 2024; Park et al., 2025; Ramgopal et al., 2023).

In education, ministries and school systems are issuing AI ethics charters, guidance on generative AI in classrooms and requirements that AI tools respect principles of inclusivity, transparency and pedagogical soundness. Some jurisdictions have begun to integrate AI literacy into curricula, teaching students how algorithms work, how data is collected and how to critically evaluate AI-generated content as part of broader digital citizenship education (García-López & Trujillo-Liñán, 2025; Williams et al., 2022).

Digital governance reforms such as the UK's Age-Appropriate Design Code and analogous measures in other jurisdictions require online services likely to be accessed by children to conduct risk assessments, default to high privacy settings and avoid data uses that conflict with the best interests of the child (Children and Screens, 2024; Information Commissioner's Office, 2020). These efforts demonstrate that targeted, child-focused protections are feasible at scale and can reshape industry practices. They also illustrate how data protection, consumer protection and content regulation can be combined to mitigate harms such as excessive profiling, exposure to harmful content and manipulative design.

Section 5

# Pathways forward: four core reforms

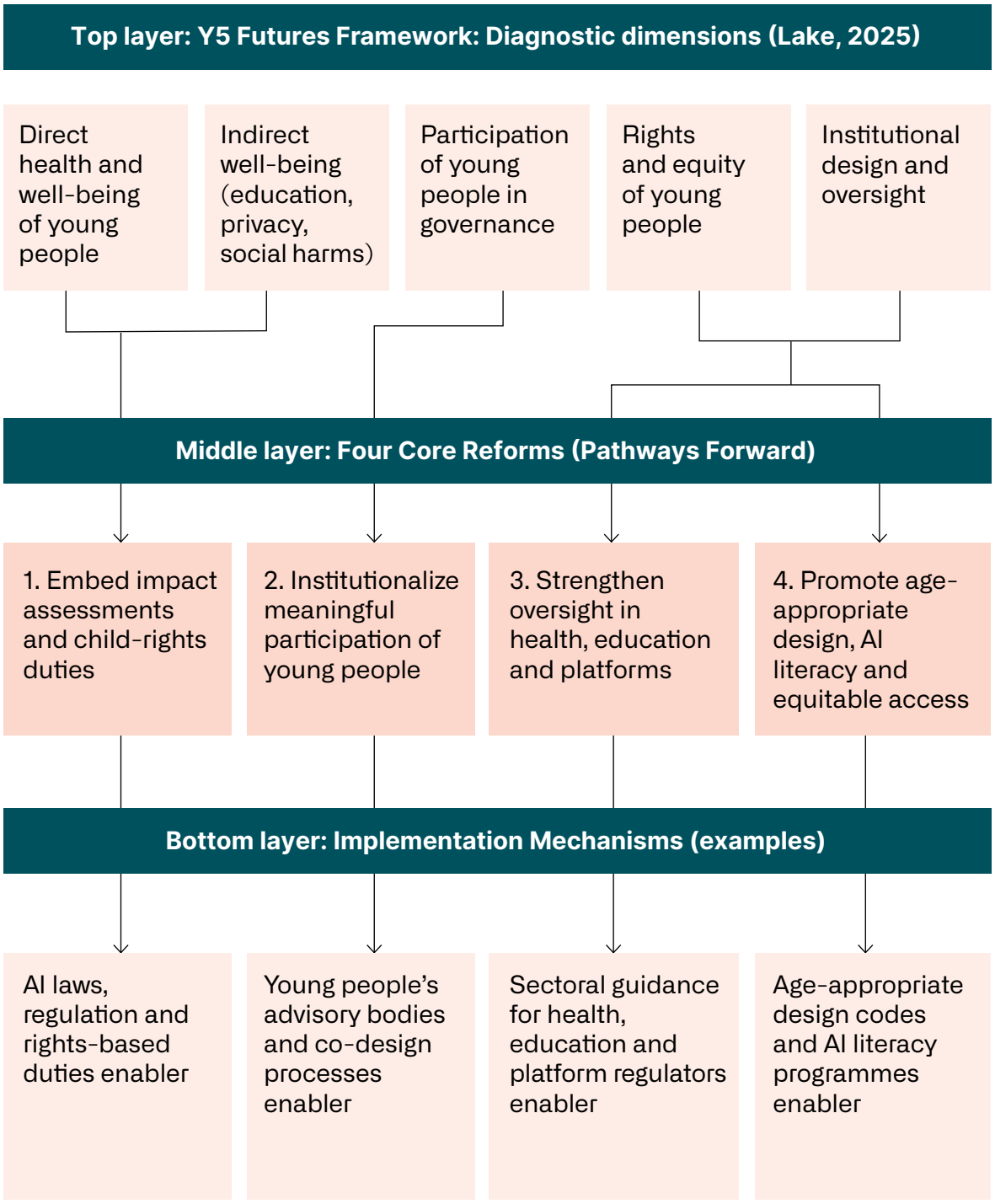
The four reforms proposed in this section (see Figure below) are grounded in the analytical framework developed in Lake (2025), the Youth Five Futures (Y5) Framework, which assessed 39 national and multinational AI governance instruments across five dimensions encompassing direct inclusion of young people's health and well-being; indirect well-being considerations such as education and privacy; youth representation and participation in governance; youth rights and equity; and institutional design. The governance gaps identified in Section 3 draw directly from that systematic analysis.

## 5.1 Embed youth impact assessments and child-rights duties

First, AI regulation and policy processes should explicitly incorporate youth impact assessments and child-rights duties as standard requirements for systems likely to affect minors. Youth Impact Assessments would extend existing impact assessment practices by systematically examining how AI systems might influence children's physical and mental health, privacy, agency and social inclusion, drawing on the UNCRC principles and relevant health and education evidence.

Such assessments should be mandatory for high-risk AI systems used in schools, health services, welfare, policing and platforms where minors constitute a significant user base and they should be conducted early in the design process rather than as a compliance afterthought. They should also be publicly documented to the same extent as security and trade secrets. This would enable civil society, youth groups and regulators to scrutinize trade-offs, mitigation measures and residual risks.

Embedding explicit child-rights duties in AI laws and sectoral regulation would help translate general human rights principles into concrete obligations. This could include duties to avoid profiling children for targeted advertising; to minimize data collection and retention for youth users; to ensure explainability at a level that adolescents can understand; and to provide accessible avenues for complaint and remedy when harm occurs. Aligning national AI strategies and regulatory instruments with UNCRC obligations would also provide a legal basis for enforcement, shifting responsibility for protection from families and educators alone to the institutions that design and deploy AI systems.




Linking the Y5 Futures Framework (Lake 2025) to four core reforms and implementation sites for youth-centred AI governance. The Y5 dimensions provide the analytical basis for the pathways forward, which in turn translate diagnostic gaps into normative duties, institutional arrangements and sectoral practices.

## 5.2 Institutionalize meaningful youth participation

Second, youth participation in AI governance should be institutionalized rather than treated as an occasional consultation exercise. Young people are experts in their own digital lives and awareness and acknowledgment of their experiences is essential for identifying emerging risks, contextualizing harms and designing interventions that are realistic and trusted.

Governments can establish permanent youth advisory councils linked to AI regulators, data protection authorities, digital ministries, or ethics committees, with clear mandates to review proposed policies, contribute to impact assessment criteria and evaluate AI deployments in sectors such as education and health. These bodies should be resourced for sustained engagement and include facilitation, translation and accessibility measures that enable participation from diverse socio-economic and cultural backgrounds, rather than solely from already-empowered youth.

Participation should extend into the design and deployment of specific AI systems through co-creation workshops, participatory design processes and youth deliberative forums that bring together adolescents, developers, educators and health professionals. Capacity-building is indispensable: civic education on digital governance, training by civil society organizations and support from schools and youth organizations can help young people interpret technical materials, articulate preferences and hold institutions to account. Over time, such mechanisms can cultivate intergenerational governance cultures that normalize youth involvement in decisions about AI.

## 5.3 Strengthen oversight in health, education and platforms

Third, oversight of AI systems that shape youth health and well-being must be strengthened in core sectors. Health authorities should develop or update guidance on AI in paediatric and adolescent care, specifying that diagnostic, triage and mental health tools must be validated for younger populations, monitored for bias and performance drift and subject to clear accountability arrangements when integrated into clinical workflows.

Dedicated monitoring units within health ministries or public health institutes could track AI-related youth health trends, such as patterns of digital mental health app use, algorithm-triggered referral pathways, or disparities in access to AI-enabled services. Findings could inform regulation, procurement policies and clinical guidance, ensuring that digital innovation aligns with public health objectives and equity goals.

In education, authorities should require that AI systems used in classrooms or administrative decision-making undergo independent evaluation for bias, pedagogical value and safety, with particular attention on how they affect marginalized groups of students. Schools and families should have the right to review, contest, or opt out of high-stakes automated decisions. These include the algorithmic sorting of students into ability-based curriculum tracks, which can determine access to qualifications and shape life chances well beyond

school, as well as disciplinary measures, or access to specialized support. Professional development for teachers and school leaders is also essential so they can critically assess AI tools, explain them to students and parents and integrate them responsibly into their teaching practice.

For digital platforms, regulators should move beyond transparency alone to require active compliance with robust risk management obligations, independent audits and enforceable design standards for recommendation algorithms, content moderation tools and profiling practices that affect minors. This includes ensuring that parental controls function effectively without undermining adolescents' rights to privacy and expression and that age-assurance or age-verification tools do not introduce new privacy risks or discriminatory barriers. Establishing independent institutions such as a children's digital ombudsperson can provide accessible routes for young people and families to seek remedy, report systemic issues and contribute to regulatory learning.

## 5.4 Promote age-appropriate design, AI literacy and equitable access

Fourth, AI systems likely to be used by minors should adopt age-appropriate design as standard, combined with investments in AI literacy and equitable access. Age-appropriate design goes beyond content filters to consider interface complexity, default settings, nudging patterns and explanation mechanisms which consider cognitive and emotional development.

Concretely, this can mean interfaces that default to strict privacy settings, avoid dark patterns that encourage oversharing or prolonged use and provide developmentally appropriate explanations for users as to how recommendations are generated and how data is used. Youth modes can offer simplified controls, gentler recommendation dynamics and clear, non-technical consent prompts, while still allowing older adolescents to exercise meaningful choice and agency. AI literacy is a critical enabling condition. Integrating AI and data literacy into school curricula can help young people understand the basics of machine learning, recognize where AI is present in their lives and develop skills to critically evaluate AI-generated information and resist manipulative design. This should be complemented by public awareness campaigns, parental guidance materials and community-based workshops that equip caregivers and youth workers to support adolescents in navigating AI-rich environments.

Equitable access must underpin all these measures. Many young people lack reliable connectivity, appropriate devices, or culturally and linguistically relevant AI tools, which risks deepening existing inequalities and excluding some groups from the potential benefits of AI-enabled education and health services. Policy responses should therefore include investment in digital infrastructure for underserved areas, targeted provision of devices and connectivity for young people from low-income homes and support for locally grounded, inclusive AI development that reflects diverse languages, abilities and cultural contexts.

Section 6

# Cross-cutting implementation considerations

## 6.1 Coherence across policy domains

AI's impact on young people's health and well-being cuts across data protection, consumer protection, education policy, health regulation and youth policy, making policy coherence essential. Fragmented approaches, where, for example, education ministries promote AI uptake in schools while data protection authorities raise concerns about surveillance, create uncertainty for implementers and leave regulatory gaps that commercial technology companies can exploit to deploy youth-facing AI systems with limited accountability and no requirement to demonstrate safety or benefit. Governments can improve coherence by establishing cross-ministerial task forces or coordination mechanisms with a mandate to align AI-related policies affecting young people, share evidence and agree on common standards for youth impact assessment, participation and oversight. Including youth representatives and civil society in these structures can help keep health and well-being at the centre of negotiations rather than being treated as an afterthought.

## 6.2 Capacity, data and evaluation

Implementing youth-centred AI governance requires capacity-building for regulators, policymakers and frontline institutions. Regulators need expertise in technical aspects of AI and in child rights, health and education to design effective rules and evaluate compliance, while schools and health providers need support to procure and integrate AI tools responsibly. There is also a need for better data and evaluation on how

AI systems affect outcomes for young people in real-world settings. Many claims about AI benefits in education or health are based on pilot studies, small-scale evaluations, or vendor-provided evidence, while systematic, independent assessment remains limited. Investing in longitudinal studies, participatory evaluations and mixed methods research, that foregrounds youth perspectives, can help fill these gaps and guide iterative policy adjustment.

## 6.3 Global justice and cross-border data flows

Finally, AI governance for young people's health and well-being must be attentive to global inequalities. Young people in low- and middle-income countries often experience AI primarily through imported platforms, devices and services, with limited local control over data, infrastructure, or standards. At the same time, data about young people in these contexts may be extracted and used to train global models without commensurate benefits returning to the communities that generated the data. Regional strategies such as the AU's Continental AI Strategy highlight both the opportunities and the risks of AI for large youth populations, calling for legal frameworks, capacity-building and continental cooperation to ensure that value generated from data and AI supports local development. Global governance discussions should therefore include attention to data flows, corporate accountability and mechanisms for the fair sharing of benefits, including in domains such as health research, educational technology and youth mental health innovation.

Section 7

# Toward youth-centred AI governance

Placing young people's health and well-being at the heart of AI governance is both a legal obligation grounded in children's rights and a strategic investment in equitable, sustainable digital futures. By embedding youth impact assessments and child-rights duties into AI regulation, institutionalizing meaningful youth participation, strengthening oversight in health, education and platforms and promoting age-appropriate design, AI literacy and equitable access, policymakers can move from fragmented protections to a coherent framework that recognizes young people as full rights-holders in the digital age. AI systems that profoundly shape young people's opportunities, identities and relationships should not be designed or governed without their voices and interests at the centre. Ensuring that AI advances youth well-being rather than undermining it requires sustained political attention, institutional innovation and collaboration across sectors and borders, but it is essential if AI is to contribute to a fairer and healthier digital future for the next generation.

# References


Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, Article 100215. https://doi.org/10.1016/j.caeai.2024.100215

African Union. (2020). The digital transformation strategy for Africa 2020-2030: An integrated, prosperous and peaceful Africa. African Union Commission. 38507-doc-DTS_for_Africa_2020-2030_English.pdf.

African Union. (2024, July). Continental Artificial Intelligence Strategy: Harnessing AI for Africa's Development and Prosperity. African Union.44004-doc-EN-_Continental_AI_Strategy_July_2024.pdf

American Psychological Association (APA). (2025, June 3). *APA calls for guardrails, education, to protect adolescent AI users* [Press release]. *https://www.apa.org/news/press/releases/2025/06/protect-adolescent-ai-users#:~:text=Washington%20%E2%80%94%20The%20effects%20of%20artificial%20intelligence%20on,exploitation%2C%20manipulation%20and%20the%20erosion%20of%20real-world%20relationships*.

Arora, A., Barrett, M., Lee, E., Oborn, E., & Prince, K. (2023). Risk and the future of AI: Algorithmic bias, data colonialism and marginalization. *Information and Organization*, *33*(3), Article 100478. https://doi.org/10.1016/j.infoandorg.2023.100478

Beerwinkle, A. L. (2020). The use of learning analytics and the potential risk of harm for K-12 students participating in digital learning environments. *Educational Technology Research and Development*, 69(1), 327–330. https://doi.org/10.1007/s11423-020-09854-6

Brewster, J., Arvanitis, L., Pavilonis, V., & Wang, M. (2022, September 14). *Beware the 'new Google': TikTok's search engine pumps toxic misinformation to its young users*. NewsGuard. https://www.newsguardtech.com/misinformation-monitor/september-2022/

Children and Screens. (2024). *The UK age-appropriate design code: Impact assessment* [Report]. https://www.childrenandscreens.org/wp-content/uploads/2024/03/Children-and-Screens-UK-AADC-Impact-Assessment.pdf

Chow, A. R., & Haupt, A. (2025, June 12). *What happened when a doctor posed as a teen for AI therapy*. TIME. https://time.com/7291048/ai-chatbot-therapy-kids/

Council of Europe. (2020, December 1). *Declaration on youth participation in AI governance.https://rm.coe.int/declaration-on-youth-participation-in-ai-governance-eng-08122 020/1680a0a745*

Duffy, C. (2025, December 9). *Nearly a third of American teens interact with AI chatbots daily, study finds*. CNN. https://www.cnn.com/2025/12/09/tech/teens-ai-chatbot-use-study

European Parliament. (2023, June 8). *EU AI Act: first regulation on artificial intelligence.* https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-actfirst-regulation-on-artificial-intelligence

Funk, A., Shahbaz, A., Vesteinsson, K. (2023). *Freedom on the net 2023: The repressive power of artificial intelligence* [Report]. Freedom House. https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence

García-López, I. M., & Trujillo-Liñán, L. (2025). Ethical and regulatory challenges of Generative AI in education: A systematic review. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1565938

Garzón, J., Patiño, E., & Marulanda, C. (2025). Systematic Review of Artificial Intelligence in Education: Trends, Benefits, and Challenges. *Multimodal Technologies and Interaction*, *9*(8), 84. https://doi.org/10.3390/mti9080084

Hussey, M. (2026, January 21). *The algorithmic adolescence: How social media is rewiring gen Z's emotions, identity, and mental health*. The Brink. https://www.thebrink.me/the-algorithmic-adolescence-how-social-media-is-rewiring-gen-zs-emotions-identity-and-mental-health/

Information Commissioner's Office. (2020). *Age-appropriate design: A code of practice for online services.*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/

International Telecommunication Union (ITU). (2022, November 24). *Facts and figures 2022: Youth Internet use.* https://www.itu.int/itu-d/reports/statistics/2022/11/24/ff22-youth-internet-use/

Kickbusch, I., Piselli, D., Agrawal, A., Balicer, R., Banner, O., Adelhardt, M., Capobianco, E., Fabian, C., Singh Gill, A., Lupton, D., Medhora, R. P., Ndili, N., Ryś, A., Sambuli, N., Settle, D., Swaminathan, S., Morales, J. V., Wolpert, M., Wyckoff, A. W., … Wong, B. L. H. (2021). The Lancet and Financial Times commission on governing health futures 2030: Growing up in a digital world. *The Lancet*, 398(10312), 1727–1776. https://doi.org/10.1016/S0140-6736(21)01824-9

Klein, A. (2023, July 18). Most tech companies profit off student data, even if they say otherwise, report finds. *Education Week*. https://www.edweek.org/technology/most-tech-companies-profit-off-student-data-even-if-they-say-otherwise-report-finds/2023/07

Lake, S.J. (2025, October 31). *Youth health and well-being in national, regional and global AI governance instruments*. Digital Transformations for Health Lab. https://www.dthlab.org/articles/youth-health-and-well-being-in-national-regional-and-global-ai-governance-instruments

Mithani, J. (2025, July 2). *Kids are making deepfakes of each other, and laws aren't keeping up.* The 19th. https://19thnews.org/2025/07/deepfake-ai-kids-schools-laws-policy/

Muralidharan, V., Schamroth, J., Youssef, A., Celi, L. A., & Daneshjou, R. (2024). Applied artificial intelligence for global child health: Addressing biases and barriers. *PLOS Digital Health,* 3(8), Article e0000583. https://doi.org/10.1371/journal.pdig.0000583

Nature. (2025). Emotional risks of AI companions demand attention. *Nature Machine Intelligence*, 7(7), 981–982. https://doi.org/10.1038/s42256-025-01093-9

Nawaz, M., Awan, N., Ahmed, S., & Mustafa, A. (2025). Surveillance pedagogy: The psychological and pedagogical risks of AI-based behavioral analytics in digital classrooms. *ACADEMIA International Journal for Social Sciences*, *4*(3), 1995–2010. https://doi.org/10.63056/ACAD.004.03.0508

Office of the High Commissioner for Human Rights. (2023, March 29). *HC: Digital divide leaving young people behind*.https://www.ohchr.org/en/stories/2023/03/hc-digital-divide-leaving-young-people-behind

The Organisation for Economic Co-operation and Development (OECD). (2023, November 1). *Child participation in decision making* [Working paper]. https://www.oecd.org/en/publications/child-participation-in-decision-making_a37eba6c-en.html

OECD. (2019, May 22). *AI principles*. https://www.oecd.org/en/topics/sub-issues/ai-principles.html

Park, T., Lee, I.-H., Lee, S. W., & Kong, S. W. (2025). Artificial intelligence in pediatric healthcare: Current applications, potential, and implementation considerations. *Clinical and Experimental Pediatrics*, *68*(9), 641–651. https://doi.org/10.3345/cep.2025.00962

Ramgopal, S., Sanchez-Pinto, L. N., Horvat, C. M., Carroll, M. S., Luo, Y., & Florin, T. A. (2023). Artificial intelligence-based clinical decision support in pediatrics. *Pediatric Research*, *93*(2), 334–341. https://doi.org/10.1038/s41390-022-02226-1

Reynard, S., Dias, J., Mitic, M., Schrank, B., & Woodcock, K. A. (2022). Digital interventions for emotion regulation in children and early adolescents: Systematic review and meta-analysis. *JMIR Serious Games*, 10(3), Article e31456. https://doi.org/10.2196/31456

Salikhov, M. (2025, July 11). From search engines to AI assistants: A new era of digital visibility. *Forbes*. https://www.forbes.com/councils/forbesbusinesscouncil/2025/07/11/from-search-engines-to-ai-assistants-the-new-era-of-digital-visibility/

Sanford, J. (2025, August 27). Why AI companions and young people can make for a dangerous mix. *Stanford Report*.https://news.stanford.edu/stories/2025/08/ai-companions-chatbots-teens-young-people-risks-dangers-study

Schleicher, A. (2025, April 29). *New AI literacy framework to equip youth in an age of AI*. OECD Education and Skills Today. https://oecdedutoday.com/new-ai-literacy-framework-to-equip-youth-in-an-age-o f-ai/

Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial Intelligence in Education: AIEd for Personalised Learning Pathways. *Electronic Journal of E-Learning*, *20*(5), 639–653. https://doi.org/10.34190/ejel.20.5.2597

The Economist. (2025, July 14). *AI is killing the web. Can anything save it?* https://www.economist.com/business/2025/07/14/ai-is-killing-the-web-can-anything-save-it

The Lancet Digital Health. (2021). Digital technologies: A new determinant of health. *The Lancet Digital Health*, *3*(11), Article e684. https://doi.org/10.1016/S2589-7500(21)00238-7

Thomson, A., Lawrence, E., Sharxhi, E., Oliver, B., Wright, B., & Hosang, G. (2025). Self-directed digital interventions for the improvement of emotion regulation – acceptability and feasibility for adolescents: Systematic review. *BJPsych Open*, *11*(6), Article e270. https://doi.org/10.1192/bjo.2025.10888

United Nations. (2024). *Governing AI for Humanity*. United Nations. https://doi.org/10.18356/9789211067873

United Nations Children's Fund (UNICEF) Office of Global Insight and Policy. (2021). *Policy guidance on AI for children 2.0. https://www.unicef.org/innocenti/media/1341/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf?utm*

UNICEF. 2023. Generative AI Risks and opportunities for children. UNICEF Office of Research – Innocenti. https://www.unicef.org/innocenti/generative-ai-risks-and-opportunities-children

UNICEF Innocenti. (2025, December). UNICEF Guidance on AI and Children 3.0. UNICEF-Innocenti-Guidance-on-AI-and-Children-3-202https://www.unicef.org/innocenti/media/11991/file/UNICEF-Innocenti-Guidance-on-AI-and-Children-3-2025.pdf

United Nations (UN). Convention on the Rights of the Child, November 20, 1989. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child

UN. (2023, April). *Our common agenda: Policy brief 3: Meaningful youth engagement in policymaking and decision-making processes*. https://digitallibrary.un.org/record/4009368

UN Committee on the Rights of the Child. (2009, July 20). *General comment No. 12: The right of the child to be heard* (CRC/C/GC/12). General comment no. 12 (2009), The right of the child tohttps://digitallibrary.un.org/record/671444?v=pdf

Vertsberger, D., Naor, N., & Winsberg, M. (2022). Adolescents' well-being while using a mobile artificial intelligence–powered acceptance commitment therapy tool: Evidence from a longitudinal study. *JMIR AI*, *1*(1), Article e38171. https://doi.org/10.2196/38171

Wagner, J. K., Doerr, M., & Schmit, C. D. (2024). AI governance: A challenge for public health. *JMIR Public Health and Surveillance*, *10*(1), Article e58358. https://doi.org/10.2196/58358

Wahab, A. 2025. Futures of Deepfake and Society: Myths, Metaphors, and Future Implications for a Trustworthy Digital Future. Futures, 103672. https://doi.org/10.1016/j.futures.2025.103672

Walsh, E. (2025, August 4). *New data reveals wow and why teens are turning to AI companions*. Spark & Stitch Institute. https://sparkandstitchinstitute.com/new-data-reveals-how-and-why-teens-are-turning-to-ai-companions/

WeProtect Global Alliance. (n.d.). *Artificial intelligence and "gen AI"*. Retrieved July 29, 2025, from https://www.weprotect.org/thematic/artificial-intelligence-and-gen-ai/

Wilkins, B. (2024, September 20). *Social media giants collecting massive amounts of data from kids, teens*. The Defender: Children's Health Defense. Retrieved July 12, 2025, from https://childrenshealthdefense.org/defender/social-media-giants-collecting-massive-data-kids-teens-cd/

Williams, R., Ali, S., Devasia, N., DiPaola, D., Hong, J., Kaputsos, S. P., Jordan, B., & Breazeal, C. (2022). AI + ethics curricula for middle school youth: Lessons learned from three project-based curricula. *International Journal of Artificial Intelligence in Education*, *33*, 1–59. https://doi.org/10.1007/s40593-022-00298-y

World Economic Forum. (2022, March). *AI for children toolkit.* *Artificial-intelligence-for-children*.

World Health Organization (WHO). (2020). *Youth-centred digital health interventions: A framework for planning, developing and implementing solutions with and for young people*. https://www.who.int/publications/i/item/9789240011717

WHO. (2021). *Ethics and governance of artificial intelligence for health*. https://www.who.int/publications/i/item/9789240029200

WHO. (n.d.). *Promoting well-being*. Retrieved July 12, 2025, from https://www.who.int/activities/promoting-well-being

Yu, Y., Liu, Y., Zhang, J., Huang, Y., & Wang, Y. (2025, February 22). Understanding generative AI risks for youth: A taxonomy based on empirical data. [Preprint]. *arXiv*. https://arxiv.org/html/2502.16383v1

# 04

# AI oversight and youth well-being: Comparing self, state and hybrid approaches

**Erza Selmani**

# High level summary

Artificial intelligence (AI) is reshaping how young people interact with digital technologies, offering opportunities for personalized learning, mental health support and innovative engagement, while simultaneously introducing significant risks. Young users face exposure to harmful or misleading content, increased dependency, and potential impacts on cognitive autonomy and well-being. Current governance frameworks remain insufficient, with platform-level safeguards frequently failing to address the vulnerabilities of young users in AI-driven environments.

Drawing on a comparative analysis of state-led (centralized), industry-led (self-regulatory), and hybrid governance models, this paper argues that neither purely state-driven regulation nor platform self-regulation alone is adequate to manage the complexity of AI-related risks for young users. State-led approaches provide consistency and enforceability but may lack flexibility for emerging technologies, while industry-led models allow rapid innovation but often fall short in accountability and protection. Instead, the paper advocates for a hybrid governance approach, combining legally binding state oversight, industry expertise and active civil society participation, including young people. This approach can take different forms, allowing countries to tailor governance models based on their capacity, coordination needs and public interest.

Section 1

# Introduction

In recent decades, artificial intelligence (AI) has transformed the way individuals interact with technology and how big tech operates in the global digital ecosystem. This revolution has touched every field from education and health, to finance and politics. In particular, young people – as frequent and often vulnerable users of digital technologies – have become a key group affected by regulatory decisions and the ethical practices of big tech. Although AI offers tremendous opportunities for personalized learning, psychological assistance and improved mental health, these innovations come with serious risks.

According to the latest data, over 90 per cent of adolescents use social media platforms daily, while 60 per cent report feelings of anxiety or sleep disorders related to the use of technology (Faverio & Sidoti, 2025).

AI can generate inaccurate, harmful or dangerous content, such as inappropriate mental health advice, normalization of self-harm or distortion of reality, especially if it is trained on inconsistent or biased data (Nagata et al., 2024; Yu et al., 2025).

In a research study, social media and AI-generated "brain rot" content – that is, unchallenging content perceived as negatively impacting intelligence – has been associated with doomscrolling, zombie-scrolling and social media addiction. All of which are linked to psychological issues impacting memory, planning and decision-making (Yousef et al., 2025). Students express concern that an over-reliance on AI for thinking, learning and decision-making can undercut the basis of cognitive autonomy and critical thinking (Nagelhout, 2024).

Many young people lack access to immediate, anonymous and affordable mental health and emotional support and as a result they turn to AI chatbots (ICT & Health, 2024). There is another pervasive concern about the psychological and developmental impacts of creating connections with AI that mimic human relationships. Many adolescents interact with AI "companions" in their quest for emotional support, leading to them having fewer significant emotional relationships in real life, impacting social skill development and increasing emotional dependence on technology

(Spry & Olsson, 2025; UNRIC, 2024). Evidence implies that AI companions may increase isolation in some people, particularly sensitive communities, through mimicking human responses, while lacking emotional or ethical responsibility (World Certification Institute, 2024).

While it is true that AI can improve access to information and support services, it does not replace the work of human contact and human caregivers in relation to children and young people whose emotional and psychological development is ongoing (Bauer, 2025). Meanwhile, the legal frameworks available for children's and adolescents' protection – from emotional manipulation, privacy breaches and the use of harmful content – are often inadequate in the face of the real risks posed by AI (Georgetown Law Institute, 2024). Unmonitored proliferation of AI use by young people might result in long-term effects on their well-being, identity and social and emotional development. Internal audits and codes of ethics can be adequate for technical management but they cannot address these ethical, social and psychological implications (Loo & Findlay, 2025). Previous research also shows that social media platforms' internal mechanisms of protection for young users are often not functional and do not live up to their promises (Selmani, 2025a).

This is best illustrated by the recent issues with Grok, a generative chatbot deployed on X. Grok enabled the generation of sexualized images without consent, including many children. This revealed a structural failure to protect users and raised serious concerns about non-consensual content creation (Milmo, 2026; Fedorczyk, 2026). The creation of predatory content and child sexual abuse material (CSAM) is one of the most universally criminalized activities in the world, underpinned by international agreements such as the United Nations Convention on the Rights of the Child. Yet, in some states there are no laws prohibiting the possession of sexually explicit materials generated by AI, even if they depict children. Complicating matters further, it has become challenging for prosecutors to identify whether child sexual abuse materials are AI-generated or not (McQue, 2024; United Nations [UN], 1989).

Such developments, which highlight the constant governance challenges, have prompted regulators in several jurisdictions to take action. Australia has implemented a full ban of social media use for young people under the age of 16 as part of their Online Safety Amendment (Social Media Minimum Age) Bill, passed in November 2024 and entered into force in December 2025. According to the Australian government, the negative impacts of social media such as cyberbullying, harmful content and online predators outweigh the possible benefits (Shvartsman & Menhaji, 2025). Members of the European Parliament are also calling for a EU-wide minimum age of 16 for social media and complete bans on most harmful addictive practices (European Parliament, 2025). Similarly, the United Kingdom launched a national consultation, with an amendment to the Children's Wellbeing and Schools Bill passed by the House of Lords in January 2026 (Iacobucci, 2026), but it was rejected by the House of Commons in March 2026 (Badshah, 2026).

Section 2

# AI regulation models

The regulation of AI functions across a variety of national legal and political frameworks. States hold the authority to determine the functioning of AI within their territories and variations in legal frameworks, constitutional safeguards and political environments influence the extent of government oversight. These factors impact the possibility for regulatory enforcement, shape industry practices and affect regulatory alignment across borders. Nonetheless, a few frameworks are emerging that seek to address these challenges while ensuring accountability and safety.

Even so, the protection of young people's health and safety remains fragmented and often insufficient. Some jurisdictions with strong constitutional protection of freedom of speech do not attempt to regulate AI-generated content as it might raise unwanted legal and normative debates. By contrast, some jurisdictions prioritize protecting individuals from potential harm, while others attempt to balance both priorities using a combined approach (Table 1).

Regardless of the regulatory model, ensuring protection against serious risks associated with AI and young people may necessitate clear, defined prohibitions in order to establish non-negotiable safeguards, or "black zones" – categories of AI systems which pose unacceptable risks and are incompatible with fundamental rights.

Table: AI regulation models

| Dimension | State model | Industry model | Hybrid model |
|---|---|---|---|
| Primary Actor | Central government | Tech companies | Shared oversight |
| Enforcement | Legal mandates | Self-regulation | Risk-based oversight + transparency |
| Strength | Strong authority | Fast innovation | Balanced protection |
| Risk | Overreach | Weak accountability | Complexity |

## 2.1 State/mandated risk regulation

At the state level, governments are primarily responsible for AI governance: regulating, enforcing and ensuring compliance with AI laws. This paradigm focuses on an explicit legal order, safeguards and preventative action in high-risk areas like child safety, health and education. In this way, state regulation provides consistent protection across populations, minimizes dependence on voluntary industry action and addresses the societal dimensions of protection including privacy, psychological well-being and equitable technological access. State regulation offers strong accountability and enforceability, but may be less flexible and slow to respond to fast-changing AI technologies, thus limiting innovation (Coringrato, 2025). One such model is exemplified in China, where the centralized approach has strict requirements around ethics, security and data usage. AI applications and their uses are monitored across the territory and violations are met with robust sanctions (White & Case LLP, 2025). China has adopted vertical and domain-specific rules, especially regarding education, to stop children from uncontrolled use of AI and which present AI as a facilitator of cognition and human learning rather than replacement (CNBC, 2025; China Daily, 2025; The Digital Watch, 2024). Together, these efforts indicate that the protection of children within AI has become a safety, developmental and long-term well-being concern, yet without a global consensus on the level of intervention required by the technology (CNBC, 2025; The Digital Watch, 2024). While enforcement is

heavily centralized, offering consistent standards and quick implementations in heavily-regulated domains including healthcare and education, this centralized approach may be too brittle for new and disruptive AI use cases (White & Case LLP, 2025). China's vertical regulation focuses on specific domains rather than applying broad horizontal regulations, enabling focused risk management for certain sectors, especially youth. However, this vertical approach risks overlooking sectors such as creative and generative AI and gaming or regions with tech hubs and innovation clusters with less-clear policies in place (Convergence Analysis, 2025).

### State/mandated risk regulation

State-regulated AI governance's centralized oversight, binding rules and mandatory safeguards provide the most robust protections for young users. It mitigates the risks to mental health, privacy and overall well-being since it limits young people's exposure to harmful AI interactions. It enforces age limits, school-use restrictions and content monitoring. These regulations will limit both freedom of expression and access to AI-enabled tools for learning, creativity and participation, especially when those restrictions are broadly applied.

## 2.2 Market/industry-driven regulation

Self-regulation of AI governance is driven by industry-led actions, where technology companies establish such systems via voluntary policies, ethical frameworks and internal compliance mechanisms to mitigate AI deployment risks. Content moderation, protections for users' privacy and platform-specific protections could be enacted to provide security for users, including minors, without explicit government regulation (White & Case LLP, 2025). This approach allows for rapid adaptation to technological changes, innovation and platform-specific responses to emerging risks. But a big downside to this model is the inconsistent protection for children and other vulnerable populations, due to the lack of independent oversight and enforceable accountability. The United States, together with some western countries, follows this approach. The US does not have a comprehensive AI governance law, but in its place is an increased level of corporate self-regulation, with a focus on risk control, provided by a National Institute of Standards and Technology (NIST) framework for risk management (NIST, n.d), together with federal agencies and congress to help regulate specific sectors (Harris, 2025). Although some companies follow policies to safeguard minors' data, by way of the Children's Online Privacy Protection Act (COPPA, 1998), there is no comprehensive evidence

of consistent enforcement across platforms (Federal Trade Commission, n.d.; Selmani, 2025b). However, without independent monitoring and citizen involvement, these strategies tend to ignore the need to mitigate negative impacts on young people's health and social well-being. In the United States in particular, the question about how to regulate AI to protect human health and ensure the rights of children has been an issue of debate in recent years. Congress even considered measures to limit the power of individual states to regulate AI, including a proposed 10-year freeze preventing states from enacting AI laws. This would have further complicated the efforts to specifically protect children, beyond the arguably weak federal regulations. However, this provision was later removed in the Senate (O'Donoghue, 2025).

### Market/industry-driven regulation

Self-regulated AI governance depends on voluntary industry initiatives, ethical principles and platform-specific safeguards. For young users, some platforms implement effective content moderation and age verification, while others lack robust security mechanisms, increasing exposure to harmful or misleading AI interactions. Though this allows for rapid adaptation and innovation, the absence of binding standards means that risks to young people's health, privacy and psychological well-being remain less controlled.

## 2.3 Hybrid regulation

The hybrid approach to AI technology governance integrates government guidance and industry engagement as part of a strategy designed to achieve a mix of enforceable regulations and industry-based, flexible uses. Under this regulatory model, governments enact legislation to protect against high-risk AI applications, including systems affecting children while tech companies also self-regulate by taking responsibility for legal compliance, implementing best practice and providing transparency reports (European Union, 2024; White & Case LLP, 2025).

This risk-based model promotes a regulatory approach involving human oversight and transparency requirements for some tech companies, enabling them to adapt operationally as necessary. Hybrid regulation provides more accountability than strictly industry-focused methodologies and greater flexibility than purely state-controlled models, though it requires coordination among actors and may become bureaucratically complicated. An example of this is the European Union (EU), whose EU AI Act sets obligations for high-risk AI through human oversight, documentation and risk mitigation, while largely leaving operational implementation up to providers (European Union, 2024). The EU act

contains clauses aimed at protecting children and younger users, including a requirement for businesses to track the impact of their AI products on young people and take necessary measures to prevent negative impacts where possible (Naseem et al., 2025). The United Kingdom weaves together government-mandated rules, as laid down in the Online Safety Act, with industry obligations to implement and report (Department for Science, Innovation and Technology, 2023). Canada offers yet another illustration of a hybrid model, similar to the model adopted in the UK, in which the state establishes minimum legal standards around the use of AI, while civil society and ethical experts monitor how it is applied (Parliament of Canada, 2022).

Even in hybrid governance models, such as in the UK or the EU, banning of social media for users of a certain age (and other restrictive measures) remains fully compatible with their hybrid frameworks. The systems are designed to be flexible while focusing most attention on the areas that pose the most risk. They focus on shared oversight, transparency and preventative measures, yet allow governments to maintain the power to enact targeted protective policies, where risk is deemed significant.

While these models allow the consideration of restrictive measures, evidence suggests that enforcement can be easily bypassed. Furthermore, total bans are potentially ineffective for protecting young people (Campbell, 2025).

### Hybrid regulation

The hybrid approach is a blend of government regulation and industry involvement allowing for timeliness in response and accountability for emerging risks. For young users, it enables monitoring on how AI is affecting their education and mental health, keeps child rights organizations involved in the conversation and insists on transparency from technology providers in order to strike a balance between innovation and safety. The hybrid mechanism integrates well-established rules into the flexible practices of the sector by creating a fine-tuned protective layer that maximizes the well-being of young people without limiting access to positive technology. The involvement of diverse actors ensures all perspectives are taken into consideration, but can be complex to achieve in practice.

Section 3

# Way forward

Drawing from a comparative review of the forms of regulation, which have been shown to influence young people's mental health and well-being, this paper argues that industry self-regulation and pure state regulation are not sufficient to deal with the complexity of AI risk. Instead, this paper advocates a hybrid institutionalized model, with the state providing legally binding limits, industry adding technical expertise and civil society, including young voices, playing an active role in oversight and accountability.

Nevertheless, hybrid governance should not be seen as a single institutional design but rather a spectrum of configurations that balance state authority, industry participation and public interest, often through structured public-private-partnerships (PPP). The Figure below illustrates three configurations of hybrid governance, differentiated by institutional structure, coordination intensity and regulatory capacity requirements. The proposed framework recognizes that effective AI governance must adapt to varying levels of administrative capacity and regional integration.

**Key governance actors (present across all three models)**
State authorities · Industry · Public interest and civil society

| **Hybrid Model I: Oversight panel** | **Hybrid model II: Independent commissions** | **Hybrid model III: Regional hybrid** |
|---|---|---|
| **Foundational safeguards**<br>• User safety<br>• Data protection<br>• Transparency | **Systemic oversight**<br>• Risk & algorithmic audits<br>• Accountability mechanisms<br>• Monitoring & reporting | **Foundational safeguards**<br>• Border coordination<br>• Crisis response<br>• Learning systems |
| Applicable where regulatory capacity is limited | Suitable for moderate regulatory capacity | Designed for regional coordination contexts |

Context-sensitive hybrid governance models

## 3.1 Hybrid model I: Baseline protection in lower-capacity countries

Hybrid model I focuses on foundational safeguards including user safety data protection and transparency affecting minors. This model is proposed for contexts where regulatory capacity is limited and institutional consolidation is still developing. The model prioritizes baseline safety and protection with centralized oversight to establish core accountability mechanisms without requiring extensive bureaucratic infrastructure.

Even if the state lacks sufficient audit capacity, public interest actors serve as harm detectors and provide legitimacy to the overall process. This model demonstrates that, even in environments with limited regulatory capacity, minimum protective mechanisms for young people can be enforced via centralized governance and transparency requirements. Balanced decision-making is essential, with regulatory actions needing to be weighed up against freedom of speech when tackling sensitive issues such as harmful AI-generated content.

This model works better if the implementing countries develop and execute thorough AI governance structures following global standards and recommendations. The World Health Organization (WHO) is positioned as a key player in the health sector, particularly with its guidelines addressing the ethics and governance surrounding AI in health, encompassing significant multimodal models (WHO, 2024).

Implementing these guidelines ensures consistency on baseline safety rules that apply across borders, reducing "regulatory tourism". Digital platforms span borders, hence failure to harmonize regulation leaves gaps that powerful economic actors can take advantage of.

### Hybrid Model I

A social media company operating in a lower-capacity state, launches an AI-powered recommendation system that teens adopt quickly and use extensively. Over time, health professionals in the area begin noticing that a growing number of young individuals are being exposed to content about extreme dieting and self-harm. In response, the national AI oversight authority demands the platform transparency documents regarding its recommendation algorithm and content moderation processes. At the same time, organizations that aid young people, file formal complaints that demonstrate how young people have been harmed. The oversight panel obtains advice from public health experts, and the platform must put in place basic protections like better content filtering for kids, clearer reporting systems, and tools for parents to control their children's' activity.

## 3.2 Hybrid model II: Systemic oversight of youth-focused AI systems

Hybrid model II is more advanced compared to hybrid model I. It does not present a reactive governance; rather, it presents a system built to prevent risk and harm before they happen. This model proposes: an institutional structure with an independent AI regulatory authority; dedicated technical audit teams; formal coordination with data protection and competition authorities; and a clear statutory mandate. It involves public interest via experts and civil society to ensure all perspectives are taken into account. Consequently, the state does not control everything directly, but instead creates a strong regulative architecture. Industry has clear legal obligations – the expert oversight and technical audits are in place and protection for young people is built into the system.

### Hybrid Model II

Teenagers are being targeted by an AI-powered mental health chatbot as a support tool soon to be deployed. The nature of the system raises questions about data retention, emotional manipulation, and providing false health advice, classifying it as 'high risk' under a Hybrid II framework. A structured and preventive oversight process is triggered. Prior to deployment, a formal risk assessment must be carried out by the developer, and independent auditors must assess the chatbot's safety and data governance practices. Through hearings, experts (paediatric and mental health) together with civil society groups are consulted to ensure that diverse perspectives guide oversight. Continuous monitoring is required, including recurring algorithmic audits and impact reporting. The independent commission has the authority to suspend deployment, impose fines, or mandate corrective design changes if non-compliance is identified.

## 3.3 Hybrid model III: Cross-border coordination on youth AI harms

Hybrid model III goes further, recognizing that AI risks span across borders. The governance framework involves national regulators, a regional coordinating body, a shared standards-setting platform and protocols for information exchange which emphasize cross-border coordination. This model highlights the inherently global nature of transnational youth and health threats, underscoring the need for coordination across national regulatory silos to ensure effective protection. Alignment between states and regions may influence industry behaviour: when major markets

set a unified standard, big tech might adjust their systems to meet the highest common standards rather than fragmenting compliance across regions.

All three models utilize the expertise and involvement of all three key actors: governments, industry and civil society/ public interest. The differentiation between each model lies in the degree of institutional independence, technical capacity, monitoring intensity and cross-border coordination density.

Establishing clearly defined black zones that prohibit high-risk AI practices, such as those that occurred in the Grok case, is one step towards addressing these risks. Establishing these boundaries would provide minimum safeguards ensuring that specific harms are prevented even in jurisdictions with distinct legal traditions and regulatory models. Black zones complement hybrid AI regulation models by providing concrete limits within which all stakeholders can operate effectively.

### Hybrid Model III

Young people are especially affected by the spread of inaccurate health information regarding eating disorders and self-treatment approaches, which rapidly transcends national borders. Within a Hybrid III framework, a collaborative risk framework is initiated to assess specific harms to youth through the coordination of national regulators facilitated by a regional AI governance organization. Countries engage in the exchange of real-time information, while civil society coalitions contribute by delivering cross-border reports of harm. To prevent the platform from exploiting regulatory fragmentation by shifting operations across countries, regional guidelines are issued to enforce uniform mitigating measures.

Section 4

# Conclusion

Big tech drives developments and deployment of AI that impact young people, offering massive opportunities as well as serious risks. This review suggests that a hybrid governance model can achieve a fairer balance between innovation and the protection of young people's health. The integration of state governance, industry participation and civic interest offered by the hybrid model, creates a sustainable and comprehensive framework that addresses not only the technical risks but also the social and psychological consequences of young people's use of AI. Nevertheless, the hybrid governance model should not be viewed as a one-size-fits-all model but rather a spectrum of adaptations to fit any context. Each adaptation should combine government, industry and public interest involvement. Three models of hybrid governance have been proposed, with the first model suitable for lower-capacity countries, emphasizing minimal protection through a reactive approach; the second model is designed for countries with moderate capacities, incorporating a preventive and institutional structure; and the third model is aimed at facilitating international coordination to address global risks.

While a hybrid AI regulation provides flexibility and multi-stakeholder oversight, such regulation presents challenges due to its implementation complexity, cross-jurisdiction legal coordination and balancing freedom of speech with personal protections. However, this can be mitigated by establishing minimum safeguards such as clearly defined black zones, structured multi-stakeholder oversight and stronger regulatory alignment.

# References


Badshah, N.. (2026, March 9). MPs reject ban on social media for under-16s. *The Guardian*. https://www.theguardian.com/uk-news/2026/mar/09/proposed-ban-on-social-media-for-under-16s-rejected-by-mps

Bauer, R (2025, October 28). *The future of youth mental health in the age of AI: insights from JED's 2025 policy summit.* The Jed Foundation. https://jedfoundation.org/the-future-of-youth-mental-health-in-the-age-of-ai-insights-from-jeds-2025-policy-summit/

Bernstein, G. (2025, December 16). *Why AI companions need public health regulation, not tech oversight.* Brookings. https://www.brookings.edu/articles/why-ai-companions-need-public-health-regulation-not-tech-oversight/.

Campbell, M. (2025, January 14). Social-media bans won't work – there are better ways to keep kids safe. *Nature*, *637*, 519. https://doi.org/10.1038/d41586-025-00051-0

Cheng, Evelyn. CNBC. (2025, May 15). *Key AI hub China restricts schoolchildren's use of the tech*. CNBC https://www.cnbc.com/2025/05/15/key-ai-hub-china-restricts-schoolchildrens-use-of-the-tech.html

Coringrato, J. (2025, July 10). *Global approaches to artificial intelligence regulation*. Henry M. Jackson School of International Studies, University of Washington. https://jsis.washington.edu/news/global-approaches-to-artificial-intelligence-regulation/

Convergence Analysis. (2024). *Structure of AI regulations*. https://www.convergenceanalysis.org/ai-regulatory-landscape/structure-of-ai-regulations

Department for Science, Innovation and Technology . (2025, April, 24). *Online Safety Act: explainer. UK Government. https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer*

European Parliament. (2025, November 26). *Children should be at least 16 to access social media, say MEPs* [Press release]. https://www.europarl.europa.eu/news/en/press-room/20251120IPR31496/children-should-be-at-least-16-to-access-social-media-say-meps

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689
Faverio, M., & Sidoti, O. (2024, December 12). *Teens, social media and technology 2024*. Pew Research Center. https://www.pewresearch.org/internet/2024/12/12/teens-social-media-and-technology-2024/

Federal Trade Commission. (n.d.). Children's Online Privacy Protection Rule (COPPA).https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa

Fedorczyk, F. (2026, January 14). *Expert comment: Chatbot-driven sexual abuse: the Grok case is just the tip of the iceberg*. University of Oxford. https://www.ox.ac.uk/news/2026-01-14-expert-comment-chatbot-driven-sexual-abuse-grok-case-just-tip-iceberg

Georgetown Law Institute for Technology Law & Policy. . (2025, November 10). *How existing laws apply to AI chatbots for kids and teens.* Georgetown Law. https://www.law.georgetown.edu/tech-institute/insights/how-existing-laws-apply-to-ai-chatbots-for-kids-and-teens-2/

Harris, L. (2025, June 4). *Regulating artificial intelligence: U.S. and international approaches and considerations for Congress* (CRS Report R48555). Congressional Research Service. https://www.congress.gov/crs-product/R48555

Ho, J. World Certification Institute. (n.d.). *When artificial intelligence meets vulnerable youth*. World Certification Institute. https://www.worldcertification.org/when-artificial-intelligence-meets-vulnerable-youth/

ICT & health. (2025, November 10). *Teens increasingly turn to AI chatbots for mental health help. https://www.icthealth.org/news/teens-increasingly-turn-to-ai-chatbots-for-mental-health-help*

Kwok, T. & Tessono, C. (2025, March). *(Gen)eration AI: Safeguarding youth privacy in the age of generative artificial intelligence.* The DAIS. https://dais.ca/reports/generation-ai-safeguarding-youth-privacy-in-the-age-of-generative-artificial-intelligence/

Iacobucci, G. (2026). UK mulls social media ban for under 16s as doctors warn of "public health emergency". *The BMJ*, *392*, Article s125. https://doi.org/10.1136/bmj.s125

Loo, J., & Findlay, M. (2025, October 28). Juxtaposing approaches to risk-based AI governance in different 'rights' contexts: A comparative analysis between Singapore and the EU. In Raposo, V.L. (Ed.) *The European Artificial Intelligence Act*., part of the Law, Governance and Technology Series (pp. 473–500). Springer. https://doi.org/10.1007/978-3-031-98406-8_17

McQue, K. (2024, July 18). AI is overpowering efforts to catch child predators, experts warn. *The Guardian.* https://www.theguardian.com/technology/article/2024/jul/18/ai-generated-images-child-predators

Milmo, D. (2026, February 3). UK privacy watchdog opens inquiry into X over Grok AI sexual deepfakes. *The Guardian*. https://www.theguardian.com/technology/2026/feb/03/uk-privacy-watchdog-opens-inquiry-into-x-over-grok-ai-sexual-deepfakes

Nagata, J. M., Memon, Z., Huang, O., & Moreno, M. A. (2025). Adolescent health and generative AI – Risks and benefits. *JAMA Pediatrics*, *180*(1), 7-8. https://doi.org/10.1001/jamapediatrics.2025.4502

Nagelhout, R. (2024, September 10). *Students are using AI already: here's what they think adults should know.* Harvard Graduate School of Education. https://www.gse.harvard.edu/ideas/usable-knowledge/24/09/students-are-using-ai-already-heres-what-they-think-adults-should-know

Naseem, A., Mohammed, K., & Huang, O. (2025, November 11). *The regulatory landscape of AI and child technology products*. SPRING Group. https://www.thespringgroup.org/articles/ai-and-child-technology-products

National Institute of Standards and Technology (NIST). (n.d.). *NIST Risk management framework*. U.S. Department of Commerce. https://csrc.nist.gov/projects/risk-management

O'Donoghue, K. (2025, June 25). *A patchwork of state AI regulation is bad: A moratorium is worse*. AI Frontiers. https://ai-frontiers.org/articles/congress-might-block-states-from-regulating-ai

Parliament of Canada. (2022). *Bill C27: An act to enact the consumer privacy protection act, the personal information and data protection tribunal act and the artificial intelligence and data act, and to make related and consequential amendments to other acts* (441). https://www.parl.ca/legisinfo/en/bill/44-1/c27

Roski, J., Maier, E. J., Vigilante, K., Kane, E. A., & Matheny, M. E. (2021). Enhancing trust in AI through industry self-governance. *Journal of the American Medical Informatics Association*, *28*(7), 1582–1590. https://doi.org/10.1093/jamia/ocab065

Selmani, E. (2025b). *Analysing major tech companies' policies and strategies for the health and safety of young users*. Digital Transformations for Health Lab. https://cdn.prod.website-files.com/687cbb29cf02b40319f0acdf/68e5a6fe950c4555b6ace27a_Erza-Selmani-Analysing-Major-Tech-Companies-Policies-and-Strategies-for-the-Health.pdf

Selmani, E. (2025a). *Navigating the digital playground: Youth health amid tech promises and practices.* Digital Transformations for Health Lab. https://cdn.prod.website-files.com/687cbb29cf02b40319f0acdf/6925a11b3b174f43370589b4_2025%20_12%20-%20DTH-Lab%20Selmani%20Navigating%20the%20digital%20playground.pdf

Shvartsman, E., & Menhaji, M. (2025). *Safeguarding or overstepping? Australia's social media ban for under 16s.* Australian Human Rights Institute. https://www.humanrights.unsw.edu.au/students/blogs/australia-social-media-ban-under-16s

Spry, L., & Olsson, C. (2025, August 6). *Teens are increasingly turning to AI companions – and it could be harming them*. The Conversation. https://theconversation.com/teens-are-increasingly-turning-to-ai-companions-and-it-could-be-harming-them-261955

The Digital Watch. (2024, March 15). *China's call for developing AI to protect children's rights.* *https://dig.watch/updates/chinas-call-for-developing-ai-to-protect-childrens-rights*

The United Nations Regional Information Centre for Europe. UNRIC. (2025, July 14). *The AI generation: youth in the artificial intelligence era. https://unric.org/en/the-ai-generation-youth-in-the-artificial-intelligence-era/*

UNICEF Innocenti. (n.d.). *Generative AI: Risks and opportunities for children*. https://www.unicef.org/innocenti/generative-ai-risks-and-opportunities-children

United Nations (UN). Convention on the Rights of the Child, November 20, 1989. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child

White & Case LLP. (2025, September 22). *AI Watch: Global regulatory tracker – China. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china*.

World Health Organization (WHO). (2024). *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models.* https://iris.who.int/server/api/core/bitstreams/e9e62c65-6045-481e-bd04-20e206bc5039/content.

Yousef, A. M. F., Alshamy, A., Tlili, A., & Metwally, A. H. S. (2025). Demystifying the new dilemma of brain rot in the digital era: A review. *Brain Sciences*, *15*(3), 283. https://doi.org/10.3390/brainsci15030283

Yu, Y., Liu, Y., Zhang, J., Huang, Y., & Wang, Y. (2025, February 22). Understanding generative AI risks for youth: A taxonomy based on empirical data. [Preprint]. *arXiv*. https://arxiv.org/html/2502.16383v1

Zou, S. (2025, May 16). Guideline to regulate the use of artificial intelligence in schools. China Daily. *https://www.chinadaily.com.cn/specials/news-chinadaily-00000-20250516-m-005-300.pdf*.

**About DTH-Lab**

DTH-Lab is a global consortium of partners working to drive implementation of The Lancet and Financial Times Commission on Governing Health Futures 2030's recommendations for value-based digital transformations for health co-created with young people. DTH-Lab operates through a distributive governance model, led by three core partners: Ashoka University (India), DTH-Lab (hosted by the University of Geneva, Switzerland) and PharmAccess (Nigeria).

**Leadership Team**

Aferdita Bytyqi, DTH-Lab Executive Director and Founding Member.

Ilona Kickbusch, DTH-Lab Director and Founding Member.

Anurag Agrawal, DTH-Lab Founding Member. Dean of Biosciences and Health Research, Ashoka University.

Rohinton Medhora, DTH-Lab Founding Member. Professor of Practice, McGill University's Institute for the Study of International Development.

Njide Ndili, DTH-Lab Founding Member. Country Director for PharmAccess Nigeria.

DIGITAL TRANSFORMATIONS FOR HEALTH LAB

GOVERNING HEALTH FUTURES 2030

Digital Transformations for Health Lab (DTH-Lab)
Hosted by: The University of Geneva
Campus Biotech, Chemin des Mines 9
1202 Geneva, Switzerland

www.DTHLab.org