

SNAPSHOT

Generative AI: Risks and Opportunities for Children

Authored by Steven Vosloo, with the support of colleagues from UNICEF Innocenti – Global Office of Research and Foresight

13 July 2023

Current context

Generative AI, now front and centre in our digital experiences, is seeing unprecedented uptake. Until this week, ChatGPT was the [fastest growing](#) digital service of all time (beaten since by the new Meta social app Threads). Some children use it daily, for example when doing their homework or choosing what to wear. Its capabilities and adoption have set off a storm of reactions around the impact and future of AI broadly: from open letters demanding a [pause on AI development](#) to [calls to protect children](#) against its harmful content. In the face of increasing pressure for the urgent regulation of AI, and schools around the world banning chatbots, governments are asking how to navigate this dynamic landscape in policy and practice.

The known issues with AI systems, such as algorithmic bias and a lack of transparency into how they work, are equally applicable to generative AI. It may even amplify some issues – like unpredictable outputs – or introduce new ones. For this snapshot, we don't presume to have the answers. But given the pace of AI development and adoption, there is a pressing need for research, analysis and foresight to begin to understand the impacts of generative AI on children. Here, we explore some of these current opportunities, risks and questions, and propose next steps for UNICEF and others working to empower and protect children.

What is generative AI and how is it being taken up?

[Generative AI](#) is a subset of AI that uses machine learning to discover patterns and structure from training data and then generate new data that have similar characteristics. It has a wide range of uses, such as producing content that mimics human-generated output (including text, images, audio and computer code), completing complex planning tasks, and exploring the development of new medicines.

Data on how children or youth around the world are using generative AI tools is limited, but initial studies indicate it is much more than adults. A small online poll in the US indicated only 30% of parents say they have used ChatGPT compared to [58% of children aged 12–18](#). Many children reported hiding their usage from parents or teachers. In another US survey, young adults (aged 18–29) who had heard of ChatGPT reported using it [more than their older counterparts](#), and they did so for entertainment, to learn something new and for work tasks.

AI is already central to children's digital environment, through recommendation algorithms or automated decision-making systems, for example. Generative AI will no doubt take an even greater hold of their digital experiences, and at a rapid pace. First, it is being embedded as chatbots, digital assistants, personal assistants and search engines into software and apps already used by children. Snapchat has introduced an 'AI friend' chatbot to its 375 million daily active users. Snapchat is very popular with children: [almost 60% of 13–17-year-olds](#) in the US and [almost half of 3–17-year-olds](#) in the UK use it (it is the fourth most popular app in the UK after YouTube, WhatsApp and TikTok). Meta has [announced plans](#) to introduce AI agents into every product used by over 3 billion people every day, including Instagram and WhatsApp.

Second, generative AI systems are being finetuned within specific sectors, potentially enabling more targeted and powerful capabilities. For example, Khan Academy has introduced a [personal tutor in its platform](#) that was honed from a general purpose chatbot to one particularly strong in learning conversations.

What are some key opportunities and risks for children?

Opportunities

With the increasing power and prevalence of AI, it is easy to see the many opportunities that generative AI can potentially offer children and young people directly or indirectly. For example:

- Personalized learning systems that can explain difficult concepts and help children develop skills, tailored to a child's specific needs. Additionally, the systems can adapt according to the learning style and speed to maximize the child's learning experience.
- Tools to support children's play and creativity in new ways, like generating stories, artwork, music or software (with no or low coding skills).
- Enhanced accessibility for children and young people with disabilities who would benefit from new ways to interface and co-create with digital systems.
- In healthcare, early detection of health and developmental issues as children use the systems directly. Indirectly, generative AI systems can [provide insights into medical data to support advances in healthcare](#).

More broadly, the analysis and generative capabilities can potentially be applied in a range of sectors to improve efficiencies and develop innovative solutions that can positively impact children. For example, in the public sector governments could augment citizen engagement channels to allow for additional languages and a mix of text, images or audio exchanges (for people with low levels of literacy).

Risks

At the same time, there are clear risks that the technology could be used by bad actors, or inadvertently cause harm or society-wide disruptions at the cost of children's well-being and future prospects.

Persuasive disinformation and harmful and illegal content at scale and lower cost. Generative AI can instantly create text-based disinformation that is indistinguishable from, and [more persuasive](#) in swaying people's opinion than, human-generated content. In a test by the [Center for Countering Digital Hate](#), Google's Bard generated misinformation without disclaimers when prompted on 78 out of 100 false and potentially harmful narratives, including climate, vaccines, LGBTQ+ hate and sexism.

AI-generated images are now also indistinguishable from – and, in some cases, deemed more trustworthy – than real faces (see figure 1).

Looking ahead, synthetic content (e.g., misleading deepfake images, audio, video and text) can potentially be personalized to individual users and, as a result, be harder to combat by automated and human moderators. Future generative AI chatbots could be programmed to impersonate humans and adapt in real-time conversations to attempt to persuade real people about issues or urge them to act. A particularly malicious case might be one-on-one chats, where unwitting individuals would be targeted and the chatbot would attend to the person’s concerns or counterarguments directly, thus likely increasing the odds of persuasion.

Figure 1: Mix of real and generated faces

AI-synthesized faces are indistinguishable from real faces. All faces in columns 1 and 3 (from left) are real, those in columns 2 and 4 are synthetic.

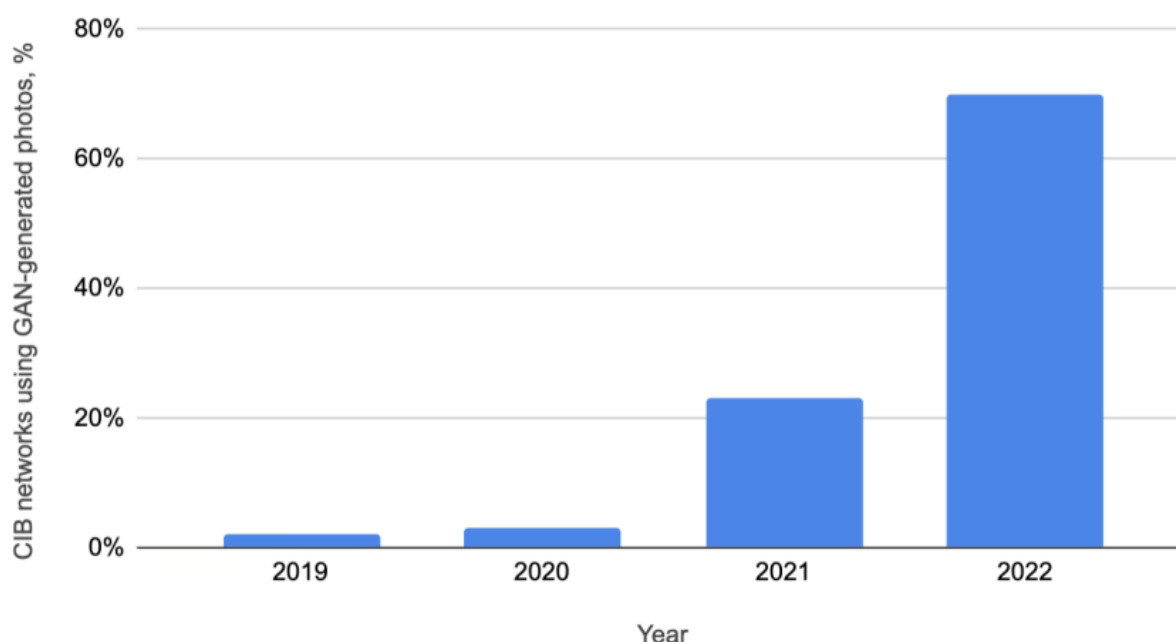


Source: [Sophie J. Nightingale and Hany Farid, 2021](#)

When used by propogandists, generative AI models will likely transform online influence operations, posing a threat to democratic processes. By mimicking human-generated text, generative AI can instantly flood regulatory and legislative systems with automatically composed messages – effectively lobbying for certain interests – and drown out any human input. Meta reported an exponential growth since 2019 in the use of fake profile photos (i.e., generated by GANs (generative adversarial networks)) (see figure 2).

The deleterious effect of an internet awash with misleading content is an erosion in trust as the authenticity of all content is questioned, including information from UNICEF. With their cognitive capacities still in development, children are particularly vulnerable to the risks of mis/disinformation and a more uncertain, corroded, and harmful disinformation ecosystem is of great concern.

Figure 2: More than two-thirds of the influence operations removed by Meta last year featured accounts that likely had AI-generated profile pictures



Source: [Meta, 2023](#)

Regarding illegal content, generative AI is being used to create photo-realistic child sexual abuse material (CSAM). The AI image generation models, which are open-source and can thus be operated with no protective guardrails, are trained on existing CSAM and photos of children from public social media accounts. While the volume of such content is still relatively small, given how quickly generative AI tools are developing, researchers predict the number of cases will only grow. This would flood law enforcement with new ‘fake’ CSAM, which would not only detract from their ability to handle CSAM depicting real children, but also complicates victim-identification and rescue operations.

In addition, a recent [FBI alert](#) noted an uptick in reporting of sextortion, including of minors, using AI-generated images. Bad actors use existing photos (e.g., from social media accounts) to generate explicit, sexually themed photos (i.e., “deepfakes”) with which to harass and blackmail victims. Reports of other types of scams are also emerging: synthetic voices models are being used to convictims by impersonating real people, such as relatives, requesting money.

Given the human-like tone of chatbots, where the line between animate and inanimate blurs, what are the impacts on children’s development – and privacy – when they interact with these systems? Research indicates they may influence children’s perceptions and attributions of intelligence, their cognitive development, and social behaviour – especially during different developmental stages.

When generative AI hallucinate (confidently make-up false information), what impact does this have on children’s understanding and education, especially if they become increasingly reliant on chat-enabled tools? Generative AI has produced dangerous content: Amazon Alexa advised a child to stick a coin in an electrical socket. Snapchat’s AI friend gave inappropriate advice to reporters posing as children. Snap, the creator of Shapchat, has since put in tools that attempt to detect “non-conforming” language, including references to violence, sexually explicit terms, illicit drug use, child sexual abuse, bullying and hate speech. However, it remains an issue that many generative AI systems are already live and accessible to children, while continuing to produce misleading and harmful content or interactions. More broadly, as children interact with generative AI systems and share their personal data in conversation and interactions, what are the implications for children’s privacy?

Modification of children’s behaviour and worldviews – intentionally or not. AI systems already power much of the digital experience, largely in service of business or government interests. Microtargeting used to influence user behaviour can limit and/or heavily influence a child’s worldview, online experience and level of knowledge. As UNICEF has noted, “children are highly susceptible to these techniques which, if used for harmful goals, are unethical and undermine children’s freedom of expression,” freedom of thought and right to privacy. Whereas current algorithms often promote sensationalist content to maximize attention, generative AI chatbots posing as real people could first gain children’s trust and, over time, influence them in more subtle ways for commercial or political gains. The online battleground could thus “shift from attention to intimacy,” according to historian and philosopher Yuval Noah Harari. He is concerned that bad actors could use AI bots to manipulate real users in hidden and persuasive ways.

Another professor, Seth Lazar, warns that we can expect the same challenges with search engines and social media platforms shaping people’s worldview, social interactions and experiences, only now pitched as an “agent that really *knows* you, and can engage with you in natural language.” Even when there isn’t bad intent, research on the use of generative AI systems to write essays shows that the biases in the models nudge users towards or against certain views in the writing process. The result is that students might be unconsciously pushed to take on certain views when using AI systems to help them learn.

Disruption to the future of work upending education provision today. AI combined with robotics first impacted blue collar work in factories through automation of predictable and repetitive tasks. In recent years, the scope of labour market exposure to AI systems has widened to some professional spheres. Now it is predicted that the daily work of teachers, telemarketers and professionals in legal services and securities, commodities and investments – amongst others – will see high exposure to generative AI tools. More research and foresight are needed to anticipate the impacts of generative AI in the workplace. Its potential disruption has great relevance for children’s future work life, as well as how and what education is provided to them today.

Aside from the future, work practices in today's AI supply chain are questionable. The quality of AI systems is dependent on accurately labeled and annotated training data, which is prepared by humans. Gig workers, often based in developing countries, assess video clips for sexual content, rate chatbot responses, identify objects within images, and upload selfies of various facial expressions, amongst other tasks. These jobs are low paying, stressful, unpredictable, unrecognized and shrouded in secrecy to protect Big Tech companies that run AI systems, bringing into question whether they constitute ethical and decent work.

Aggravation of existing inequalities around the digital divide, with some children more at risk from the harmful effects of generative AI and unable to help shape its development and access its benefits. We know that since emerging technologies are not evenly distributed within and between countries, they have the potential to aggravate existing inequalities. Generative AI is no exception. Equitable, inclusive and responsible generative AI needs to cater to the different contexts and developmental stages of all children, especially those from marginalized communities, and be available to every child when beneficial. If children's data is used to shape generative AI systems, the collection and processing of their data must be done responsibly, with clearly defined purposes, and safely, regardless of where the data come from. This will not happen through market forces alone, but only by regulation and actively addressing power imbalances and digital exclusion. UN Secretary General António Guterres believes that guardrails, grounded in human rights, transparency, and accountability, are needed to ensure that AI development "benefits all."

While this list of opportunities and risks is far from exhaustive, it illustrates that the impacts of generative AI will be wide-ranging and demand a pro-active response from those who regulate and develop AI – and related data – systems so that these empower and protect children. It also highlights the critical need for a strong advocacy and education investment to ensure that communities, children, parents, teachers, policymakers and regulators are able to engage in meaningful conversation with tech companies towards child-centred AI.

Children will experience an increasingly high level of exposure to AI systems over the course of their lives, with impacts in childhood having long-term effects. Addressing their needs is critical. Looking further ahead, the way AI is shaped today will have significant bearings on future generations. As noted by Mr Guterres, "present generations have a responsibility to 'halt and prevent developments that could threaten the survival of future generations ... [including] new technologies'". Concerns around the future of AI-enabled warfare are particularly crucial to address.

How should UNICEF approach generative AI?

As AI academic and critic Gary Marcus said, with generative AI, "We have a perfect storm of corporate irresponsibility, widespread adoption, lack of regulation and a huge number of unknowns." UNICEF has a key role to play in meeting this moment to maximize the opportunities and minimize the risks for children.

As a starting point, we can draw on existing resources. In 2021, UNICEF Innocenti – Global Office of Research and Foresight developed a UN-first [policy guidance on AI](#) to promote children’s rights in government and private sector AI policies and practices. The guidance, which is drawn from the Convention on the Rights of the Child and was developed in consultation with experts and [adolescents](#) from around the world, has nine requirements for child-centered AI that still hold true. For example, AI policies and systems should *Prioritize fairness and non-discrimination for children* by minimizing bias. This needs to be applied to new AI image generators that are [reflecting the sexist biases of their training data](#). The systems perpetuate harmful gender stereotypes by, for example, showing mostly men when prompted to represent an engineer, a scientist and an IT expert. Or that AI policies and systems need to *Provide transparency, explainability and accountability for children*. Disclaimers on generative AI systems, warning that they may produce inaccurate, misleading, offensive or biased information (because their product developers cannot control or explain how the systems work), do not adequately meet that requirement.

To support implementation of the guidance, we also produced a [guide for parents; guide for teens and operationalization toolkit](#), and developed [case studies](#) of organizations applying the guidance. Creating additional case studies on how generative AI is being developed and regulated responsibly today could help to provide insights for future developments.

But generative AI tools also bring novel challenges and unknowns, prompting the need to interpret how to apply existing policies in novel contexts, and to develop new guidance and regulations where there may be gaps. To help policymakers and companies make sense of current and future developments and be guided on how to respond appropriately for children we recommend – to UNICEF and others working to protect children and future generations – the following approaches:

- **Help fill the knowledge gaps through critical reflection and research.** Before proposing more solutions, consider: Are we asking the right questions? Do we adequately understand the issues? Are we inviting and listening to different perspectives? UNICEF Innocenti will explore children’s experiences of generative AI through its forthcoming Disrupting Harm surveys, but this is only a starting point.

We should conduct or commission dedicated research to gauge to what extent children are using generative AI, their experiences and how it is impacting their development. Within education, for example, we could research how educators are adapting to the ‘democratization’ of generative AI, how they are incorporating it into their pedagogy (if at all), how they are adjusting to a reality where such powerful tools are in many of their students’ pockets, and what guidance they may need.

- **Engage in foresight – including with children – for better anticipatory governance.** Employing foresight methods to map possible futures of AI, including scenario development and horizon scanning, will inform recommendations for actions today that help us be better prepared for future impacts. As part of looking ahead, we need to meaningfully engage children and youth on the AI-enabled digital environment they want to help policymakers and AI developers think beyond the current digital landscape and shape more child-centred AI futures.
- **Keep pushing for child-centred AI in policy and practice.** This involves engaging the tech sector and calling for greater transparency and responsibility, as well as advocating for children in the AI regulatory and policy process. For example, we have [previously submitted inputs](#) to the

draft EU AI Act to highlight children's rights, opportunities and risks. Leveraging generative AI responsibly within UNICEF as we develop solutions for children and the organization, [supporting startups that do the same](#), applying [data governance](#) and [responsible data use principles](#), and developing a strategy to AI-enable the whole organization will show leadership within the UN and help us learn about what works and what doesn't.

Additional information

- [ChatGPT is going to change education, not destroy it](#) (MIT Technology Review, 2023)
- [How worrying is generative AI?](#) (Babbage podcast) (The Economist, 2023)
- [How will ChatGPT change the way we think and work?](#) (Stanford, 2023)
- [Policy guidance on AI for children](#) (UNICEF, 2021)

Acknowledgements

This paper was written by Steven Vosloo, with review support from Daniel Kardefelt-Winther and Karen Carter, and editorial support from Céline Little and Amanda Marlin. Layout design by Kathleen Edison.

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

About us

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

UNICEF Innocenti – Global Office of Research and Foresight tackles the questions of greatest importance for children, both current and emerging. It drives change through research and foresight on a wide range of child rights issues, sparking global discourse and actively engaging young people in its work.

UNICEF Innocenti equips thought leaders and decision-makers with the evidence they need to build a better, safer world for children. The office undertakes research on unresolved and emerging issues, using primary and secondary data that represents the voices of children and families themselves. It uses foresight to set the agenda for children, including horizon scanning, trends analysis and scenario development. The office produces a diverse and dynamic library of high-level reports, analyses and policy papers, and provides a platform for debate and advocacy on a wide range of child rights issues.

UNICEF Innocenti provides, for every child, answers to their most pressing concerns.

Published by

UNICEF Innocenti – Global office of Research and Foresight

Via degli Alfani 58
50121, Florence, Italy

Tel: (+39) 055 20 330

Email: innocenti@unicef.org

Social media: @UNICEFInnocenti on Facebook, Instagram, LinkedIn,
Twitter and YouTube

Suggested citation

Vosloo, S., *Generative AI: Risks and Opportunities for Children*, UNICEF
Innocenti – Global Office of Research and Foresight, Florence, July 2023

© United Nations Children’s Fund (UNICEF), July 2023

for every child, answers