

SIENNA feedback to UNICEF Draft Policy Guidance on AI for Children¹

16 October 2020

Coordinated by Trilateral Research based on feedback from:
Nicole Santiago, Rowena Rodrigues, and Anaïs Resseguier (Trilateral Research)
Philip Brey (University of Twente)
Konrad Siemaszko (Helsinki Foundation for Human Rights)

Is the purpose of the guidance clear?

- Not clear at all
- Not clear
- Neutral
- CLEAR**
- Very clear

Is the target audience for the guidance clear?

- Not clear at all
- Not clear
- Neutral
- CLEAR**
- Very clear

Are the format and visual elements of the document easy to understand? If not, please specify what improvements could be made (in the category labeled 'Other').

- Not at all
- No
- Neutral
- YES**
- Yes, very much so
- Other

What are the aspects and/or sections of the guidance that are the most useful to you and your organization and why?

The most useful aspects to the SIENNA project are those concerning data protection, privacy rights, the introduction of a multi-stakeholder approach in government and business, and the *Development Canvas* tool.

One of SIENNA's main concerns is who protects people's data and who can be held accountable in case of infringement of this protection. Section 2.4 provides a useful insight on the problem of children's data protection, and section 3.4 presents a series of tools and principles that should be followed in order to secure children's data privacy.

¹ Available for download at: <https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children>.

Additionally, members of the SIENNA project could make use of these guidelines in order to develop more informed and inclusive ethical analysis. Therefore, the multi-stakeholder approach (presented in section 3.0) represents another relevant aspect for SIENNA. This is in line with SIENNA’s methodology, as it focuses on developing ethical frameworks where the interests of stakeholders play a major role as part of impact assessment.

Closely related, the *Development Canvas* tool is a useful initial framework for an organisation conducting an impact assessment, as it provides valuable introductory guidance with clear and direct indications that can be easily put into practice. We welcome this initiative to present a methodology to transform principles into practical tools.

We also found the summary of the requirements and its statement in child-friendly language and the case studies useful.

Are the terms and definitions in the guidance understandable? If not, what relevant terms could be added or clarified and how? Please limit your response to include no more than 3 key terms in the category labeled 'Other'.

- Not at all
- No
- Neutral
- Yes
- Yes, very much so
- OTHER**

There are two terms we recommend clarifying.

There are two terms we recommend clarifying. First, the requirements should make explicit reference to the specific language concerning rights as outlined in the Convention on the Rights of the Child (CRC), in order to emphasize that there are legal obligations concerning children and AI. Furthermore, this document should contain a list of the rights that States Parties to the CRC are obligated to respect and ensure.

Additionally, the use of the term “digital divide” on paragraph 2.4 is not exhaustive. The guidance focuses on the have/have-not dichotomy (people who have access to AI controlled devices and people who do not), as well as people who have access to them but are unskilled to use them. The latter case should be referred to as “digital inequality”, as divide refers to a dichotomy, while inequality can be characterized as a spectrum. For instance, as already cited, some parents or caretakers may be more or less skilled to provide children with information regarding AI.

Are the use cases presented in chapter 4 relevant for your local context? Are there any other use cases or examples that should be included to further describe the impact of AI systems on children’s rights?

The use cases included are relevant for SIENNA investigation, particularly those that raise privacy protection issues and in the local context of Western Europe where children do interact with AI and have some level of digital literacy. However, there are some critical use cases missing that should be explored, and opportunities to elaborate further on existing ones.

First, there should be more focus on use cases on local contexts in which access to the digital world is more limited, such as in countries in the Global South where access to digital tools is more difficult. Additionally, there should be more discussion on the long-term impacts that AI will have on children's free will, decision-making capacities, intellectual development, socialization, and understanding of what it means to be human.

Some concrete use cases that should be added include:

- The specific use of emotion-detection technology. This should include potential use by parents and guardians and uses where the parent or guardian is unaware that a child is interacting with this type of technology.
- Surveillance technologies in private and public spaces. For example, consideration could be given to the use of AI-powered thermal image processing in school during a pandemic.
- AI-powered assessments tools that influence or dictate children's education curriculum and opportunities, and possible career prospective. The decisions generated from such applications could impact the entire life of a child. One example is AI systems used to determine which children should have access to which schools (e.g. Ofqual's grading algorithm for the A-levels in the UK that was found to advantage students from privileged schools and disadvantage students from unprivileged schools<
<https://www.theguardian.com/education/2020/aug/21/ofqual-exams-algorithm-why-did-it-fail-make-grade-a-levels>).
- Vulnerabilities created when children break or dismantle AI devices, intentionally or unintentionally. For example, a curious child may hack a device and unwittingly disable certain protections or activate services not suitable for children. This is possible not only with AI-power toys, but also with wearable devices, smart appliances, and educational assistants.

In addition, the following use cases should be better developed:

- The use case on facial recognition is too generic and is not tailored to the child-context. Applications that could be discussed include: the use of AI surveillance systems schools for security purposes; and in law enforcement for identification of child abuse victims.
- The section on AI enabled toys may be enriched with insights regarding children behavior conformity to robot behavior when put in contact and discussion with AI robots. As a study from University of Plymouth has shown, children – as opposed to adults – tend to trust robots without questioning the truth of their statements (*Children conform, adults resist*, reference provided in question 17). This case study could be implemented into the AI enabled toys section or be presented as another section to the use cases.

Our aim is to provide practical guidance that can be used by government agencies, companies and organizations. Do you consider the guidance practical enough to use in these settings? If not, please specify what improvements could be made (in the category labeled 'Other').

- Not at all
- No
- Neutral
- Yes
- Yes, very much so
- OTHER**

The guidance is useful for implementing the principles in practical scenarios. Nonetheless, it is very high-level and sections regarding marginalized communities or minorities are too generic. For

instance, efforts such as “capacity-building projects by governments” (section 3.3) are hardly applicable in communities where the government is corrupted and has a strong and bounding influence over society.

Additionally, the guidance could take a harder stance on red lines in deployment and use of AI in children's settings. Part of this is ensuring we are addressing concerns with the best available solutions, not the easiest or cheapest. For example, even though an AI-powered robot could assist a child with schoolwork, is that the type of educational experience we desire for children? When AI systems impact children, we should be asking ourselves whether it is a necessary application of AI, and, if so, whether the use is appropriate and proportionate.

Are the nine requirements for child-centred AI understandable?

- Not at all
- No
- Neutral
- YES**
- Yes, very much so

Is a requirement missing or does a requirement need to be expanded further? If so, which one and why?

We recommend adding (or incorporating) three requirements:

- Respect human rights. AI must be developed, deployed, and used in a way that does not violate (or enable violation) of children’s human rights. Additionally, AI should be developed in a way that is consistent with ethical principles, particularly of non-maleficence.
- Take into account the results of a cost-benefit analysis. The costs and benefits should include the impacts for children, communities and the environment (including at the stage of the development of the AI system itself).
- Be sustainable. Sustainability includes, but should not be limited to, environmental sustainability. Considering the severe threats due to global warming and the environmental impact of AI, it is essential to call for sustainable AI as children will increasingly have to face these challenges. Engaging with this is also essential considering the children and youth’s concern for environmental challenges.

We also recommend making the following changes:

- Requirement 3 should also reference culture, religion and nationality. This is very important, for example as migrant children are particularly vulnerable in automated decision making. Additionally, the fact that bias is referred to as something that needs to be “eliminated” is problematic. As it has been shown by Veronica Barassi in her work “The Human Error in AI and Children’s Rights”, algorithmic bias is something that cannot be eliminated, and the conception of “unbiased data” is dangerous. Therefore, when talking about non-discrimination it is important to remember that no AI algorithm will ever be fully non-discriminatory.
- Requirement 5 should lay more stress on conformity assessment and certification. For children’s products and services, this is a highly useful and time-tested approach.

Are concrete recommendations missing from any of the nine requirements? Do any recommendations need further elaboration? If so, please propose additional recommendations or edits and explain why they are needed.

As a general note, there should be more elaboration on complaint and redress mechanisms, beyond merely monitoring of impacts of AI systems. There could also be more calls for studies on the concrete impacts of AI on children in the short, medium, and long term. We must recognize that we don't yet know all the impacts of AI on children, and we should therefore display more caution in the development and deployment of AI.

In the recommendations under Requirement 3 (Prioritize fairness and non-discrimination for children), more attention should be given to the risk of the growing digital divide and growing inequalities between children on the one hand who will have access to AI and be AI literate, and those on the other hand who will not have access and who will be impacted by AI without being able to shape AI and request redress when necessary. The risk here is that we will make privileged children even more privileged thanks to AI and disadvantaged children even more disadvantaged because of AI (we already know that AI harmful impacts disproportionately impacts the most vulnerable part of the population). Additionally, in the context of this requirement, it would be important to consider Veronica Barassi's solutions to child profiling, as discussed above. Some of these include labelling all AI profiling algorithms as high-risk, deleting all children's data collected through adult profiles, and making citizens aware of AI profiling when these mechanisms are in action.

In the recommendations under Requirement 4 (Protect children's data and privacy), we recommend elaborating further, especially in the context of profiling and digital advertising, as these issues are only briefly mentioned. Although they are discussed more broadly in UNICEF's discussion paper "Children and Digital Marketing", the importance of this topic calls for moving some of the recommendations from the discussion paper to the main body of the Guidance. With the advent of the 'Internet of Toys', in which children could be monitored, profiled and influenced by targeted advertising via 'smart toys', effective protection of children's rights – of privacy, but also freedom of thought and opinion – requires far more decisive recommendations than the ones that are in the current version of the Guidance. This could include a general prohibition of behavioural marketing targeting children and profiling children for marketing purposes. The EU Article 29 Data Protection Working Party already has made similar recommendations, specifying that "data controllers should not process children's data for behavioural advertising purposes, neither directly nor indirectly, since this will be outside of the scope of the child's understanding and therefore exceed the boundaries of lawful processing" (The WP29 Opinion 02/2013 on apps on smart devices, p. 26) and that "Because children represent a more vulnerable group of society, organisations should, in general, refrain from profiling them for marketing purposes" (The WP29 Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, p. 29).

In the recommendations under Requirement 8 (Prepare children for present and future developments in AI), it is crucial to ensure that digital and AI literacy includes also educating about privacy and data protection in the digital environment and teaching critical consumption of new media (e.g., aware of its commercial, surveilling nature).

Are there resources, materials or evidence that could be used to further support the guidance? If yes, please specify: (1) the material type (i.e. report, toolkit, guidance, initiative, etc.), (2) the name of the resource, (3) URL, and (4) what specific section of the guidance it relates to.

- SIENNA Project report, State-of-the-art Review: AI and robotics, https://www.sienna-project.eu/digitalAssets/787/c_787382-l_1-k_sienna-d4.1-state-of-the-art-review--final-v.04-.pdf. This relates to the definitions of AI in section 1 (What do we mean by AI?).
- SIENNA Project report, Ethical Analysis of AI and Robotics Technologies, https://www.sienna-project.eu/digitalAssets/801/c_801912-l_1-k_d4.4_ethical-analysis--ai-and-r--with-acknowledgements.pdf. This relates in particular to sections 2.2 (How children are impacted by AI systems) and 2.4 (Key risks and concerns).
- SIENNA Project report, Recommendations for the enhancement of the existing legal frameworks for genomics, human enhancement, and AI and Robotics, https://www.sienna-project.eu/digitalAssets/894/c_894270-l_1-k_sienna_d5.6_recommendations-for-the-enhancement-of-the-existing-legal-frameworks-for-genomics--human-enhancement--and-ai-and-robotics_www.pdf. This relates to section 2.1 (What are children’s rights) and section 3 (Requirements for child-centred AI).
- Academic paper, “From the 'Digital Divide' to 'Digital Inequality’”, https://culturalpolicy.princeton.edu/sites/culturalpolicy/files/wp15_dimaggio_hargittai.pdf. This paper relates to paragraph 2.4, where the concept of “digital divide” is introduced.
- Academic paper, “The Human Error in AI and Children’s Rights”, http://childdatacitizen.com/cdc/wp-content/uploads/2020/06/The-Human-Error-in-AI-and-Children-Rights_Prof.-Barassi_Response-to-AI-White-Paper-.pdf. This paper relates to requirement number 4, where it reads “Seek to eliminate any prejudicial bias against children, or against certain groups of children, that leads to discrimination and exclusion “.
- Academic paper, “Children conform, adults resist: A robot group induced peer pressure on normative social conformity”, <https://robotics.sciencemag.org/content/3/21/eaat7111/tab-pdf>. This paper relates to the use cases section in chapter 4.
- EU Advisory Body Opinion, Article 29 Working Party (WP29) Opinion 02/2013 on apps on smart devices, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp202_en.pdf. This relates to Requirement 4 (Protect children’s data and privacy).
- EU Advisory Body Opinion, Article 29 Working Party (WP29) Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=61205. This relates to Requirement 4 (Protect children’s data and privacy).

Are the ‘Tools to Operationalize the Policy Guidance’ (roadmap for policymakers and a development canvas for AI software teams) practical for you and your organization? Please describe.

The *Roadmap* for policymakers applies to SIENNA’s ethical, legal and social analysis, as it provides a clear and straightforward guidance to implement in ethical assessment and development of ethical frameworks. We particularly appreciated the recommendation to “Incentivize scientific research on the impact of AI on children”, as it is important that governments fund research on understanding impacts as much as they fund research for AI.

The *Development Canvas*, even if conceived for AI software teams, also informs policymakers about what kind of questions designers ask themselves in the design process. This can therefore be useful to understand and regulate AI software teams' approach.

To improve these tools, particularly the *Development Canvas*, a non-exhaustive list of possible negative impacts should be included. Additionally, the role of parents and guardians should be taken into account and explicitly referenced.

Is anything else missing from the policy guidance that UNICEF should consider for inclusion in the next version?

UNICEF's mandate is "to reach the most disadvantaged children and adolescents – and to protect the rights of every child, everywhere". We are concerned that this document does not do enough to "reach the most disadvantaged children and adolescents" in the world. Indeed, considering how much technical resource and competence AI requires, it is a technology that is primarily in the hands of the most privileged part of the society and whose negative impacts are mainly born by the vulnerable population, including especially, women, minorities and the poor. As such, it is a technology that is already, and risks to further, exacerbate already existing inequalities, within national territories and internationally by benefitting further children from wealthy background and/or countries, and negatively impacting children from deprived backgrounds and/or countries. Considering UNICEF's role to engage with "the most disadvantaged children and adolescents", it is essential that it presses more strongly on the issue of the increased inequalities between children. If one is to read this document with the concern of rising inequalities between children nationally and internationally, this document seems overly optimistic in terms of the benefits AI can bring to children, and fails to account for the severe risks. We recognise that the issue of rising inequalities is mentioned p. 20 as "Exacerbation of the digital divide" and requirement No. 3. However, it should be given more attention.

Additionally, we recommend the following modifications:

- The document should encourage more prudence on the part of governments and businesses regarding the development and use of AI for children. There still is a lot of uncertainty with regards to the impact of AI on children and this uncertainty should be recognized and taken into account when designing policies.
- In order to lift some of the uncertainty regarding impact of AI on children, the document should press the need to conduct impact assessment of the short, medium, long term impact of AI on children
- The Guidance states, "children's lives and well-being are also indirectly impacted by automated decision-making systems that determine issues as varied as welfare subsidies, quality of health care and education access, and their families' housing applications." This should read: there is DIRECT impact on children's lives.
- Risks to autonomy should be better addressed and made more explicit, not just in privacy context. As part of this, whether children should be able to say 'No' or have a choice in the use of AI should be addressed.
- The Guidance should take a harder line on problematic technologies. It should recognise that AI-based harms in children may extend to even suicide and death – and this should be taken into account where sensitive scenarios are mentioned.
- The Guidance states "These techniques, largely driven by business interests, can limit a child's worldview, online experience and level of knowledge, and as such, the child's right to

freedom of expression and opinion”. But these technologies can also be used to ‘control’ and ‘construct’ a child’s worldview. This is very dangerous, especially in authoritarian states; even children in democracies face this threat.

- The consultations should have also reflected children aged 8 and over who are increasingly users (even if this is a grey zone) - this is particularly relevant for “Participation = { include all children }.
- The vulnerability of children of colour should be better recognized, as they are not used in testing, but systems are deployed on them.
- A more in-depth analysis of human error in AI (as already mentioned in question 15) should be included in algorithmic bias assessment.
- Lastly, at some point the document should explain how the guidance fits into the broader context of children’s rights and digital technologies, in particular the draft General Comment under consultation with the CRC on children’s rights in relation to the digital age. For example, will this policy guidance be updated to reflect specific and relevant recommendations there?