Concept Note: The AI Revolution and Its Implications for Human Rights (2024)

Łukasz Szoszkiewicz

As we stand on the brink of a technological revolution driven by advancements in Artificial Intelligence (AI), the implications for human rights, particularly for children and future generations, have never been more significant. This concept note aims to explore the intersection of AI, neurorights, and child rights, highlighting the need for a proactive, strategic approach to ensure that technological progress facilitates the protection of human rights worldwide.

Where are we now?

Artificial Intelligence (AI) is machine-based system (usually software) designed to simulate human intelligence in generating outputs such as content, predictions, recommendations, or making decisions influencing the digital or physical environment. AI systems are now capable of performing tasks that, until recently, were thought to be exclusively within the human domain. **Large Language Models**, such as ChatGPT or Google Gemini, exemplify capabilities in content generation (text, pictures, video) which has vast implications, from education[1] and job market to misinformation and child exploitation, enabling personalized interactions at scale. As they provide interactive experience that simulates human-like conversations, they bring systematic challenges to mental health, particularly for children (see recent lawsuit against Meta/Instagram). **Face recognition technology**, another significant AI application, has been widely adopted by law enforcement agencies or personal device access. One of the most controversial companies offering such technology, Clearview AI, has "reportedly collected 20 billion photographs — the equivalent of nearly three per human on the planet", raising critical privacy and ethical questions. AI algorithms are increasingly used in the **administration of justice**, from predicting policing to assisting in sentencing decisions and deportation procedures, highlighting the potential for algorithmic bias and discrimination.

Legislative initiatives around AI aim to strike the balance between harnessing its potential benefits and mitigating its risks. Globally, the OECD Principles for AI (2019) stand as a universal standard for guiding policymakers. Legislation is pending in the European Union and Canada. The Council of Europe will soon finalize the **Convention on AI, Human Rights, Democracy, and the Rule of Law**, however there are serious concerns over its standard of protection due to systematic weakening of its provisions. Although AI is increasingly present in children's lives, their voices are rarely integrated into AI governance (for promising practices see UNESCO). The United States has set forth a Blueprint for an AI Bill of Rights. Japan, Singapore, and South Korea took a non-regulatory approach, worrying that legislation can hamper countries' AI potential. This is also why India faces AI regulation dilemma. The legislative landscape is as varied as the technology itself, reflecting the unique concerns of each jurisdiction.

One of the most far-reaching implications for human rights emerges on the **intersection of AI and neuroscience**. At the heart of this junction lies the development and application of neurotechnologies, which embody devices and procedures used to access, monitor, investigate, assess, manipulate, and/or emulate the structure and function of the neural systems of a human (OECD 2019). Recent years have brought significant developments in this area. Brain-computer interfaces (BCIs) aim to convert electrical signals generated in the brain into computer or mechanical actions, such as moving cursors on the screen or robot arms. One of the most controversial applications was reported in China, where schools introduced monitoring of children's brain activity. In a study published in 2023, scientists have made a leap forward in decoding human language. Using fMRI brain scan and deep neural networks (AI), they **reconstructed continuous language from brain activity scans**, successfully capturing the meaning of imagined speech (Tang et al. 2023). Since speech is central to many cognitive abilities, the study heralds the prospect of decoding thoughts and creates a challenge to privacy.

[1] Some authors suggest that the generative AI has the potential to address the so-called Two Sigma Problem, which reveals that one-on-one tutored students outperform 98% of those in traditional classrooms.

These applications provoke numerous questions on the adequacy of the existing legal framework to address new challenges. An increasing number of **neuroscientists and human rights experts call for the recognition of neurorights**. The postulated catalogs of new rights vary, but several are most commonly defined: the right to mental privacy (protection from decoding of brain activity) and the right to cognitive liberty (Yuste et al., 2017, Farahany 2023).

The debate on neurorights has already been elevated to the international fora. In October 2022, The UN Human Rights Council adopted a resolution on neurotechnology and human rights (A/HRC/51/L.3). The International Bioethics Committee of UNESCO in the report published in late 2021 to "propose the adaptation of existing human rights instruments and the proclamation of new human rights" in the area of neuroscience. The need to recognize new rights and/or draft a new instrument is supported by developments in other international organizations and national jurisdictions (Council of Europe, Inter-American Commission of Human Rights, OECD).

On the national level, the **most far-reaching legislative response emerged in Chile**. In 2021, the Chilean parliament unanimously approved a **constitutional amendment** that explicitly protects mental integrity, brain data and information derived therefrom. Chile's Supreme Court recently dealt with the world's first neurorights case, setting a global precedent. Legislative initiatives are ongoing in **Brazil, Mexico Uruguay, and the United States** (Colorado adopted the relevant law in 2024).

How to think about (and plan for) the future?

While it is impossible to predict the future of science and technology, we can assume that these fields will continue to evolve rapidly. For example, Large Language Models, which were first introduced in 2018 (BERT and GPT-1), are now being used by 180 million people worldwide (e.g., ChatGPT). In the next five years, we can witness similar acceleration in biotech. Assuming current trends continue, AI will be in the future:

- **More powerful** – the number of transistors on a microchip has been increasing exponentially since 1970s, allowing us to build more powerful computers and smartphones. the computational requirements for artificial neural networks or training Large Language Models were once prohibitive, but we're now approaching a moment where cutting-edge models developed by large corporations are merely few weeks ahead of open-source counterparts. This trend is a double-edge sword as it promises wider access to the benefits of AI but impacts climate.
- **More white-collar** – Moravec's Pardox is a phenomenon where AI systems can perform high-level cognitive tasks such as reasoning or playing chess better than humans but struggle with low-level sensorimotor tasks such as perception or locomotion that humans do effortlessly. This is why AI-generated image can win a prestigious international photography competition, but cannot replace car drivers or food deliveries. According to Goldman Sachs, 18% of work globally could be computerized, with the effects felt more intensely in advanced economies than in emerging markets. It has implications for the right to work and the right to social security.
- **More divisive among individuals and groups** – the adoption of an innovation doesn't occur all at once within a society but gradually, with certain individuals inclined to embrace new ideas or behaviors earlier than others. In the AI era, these individuals will be more productive both in educational settings and in their professional lives (see what can be done by AI in 59 seconds)[2]. The need to develop skills for collaborating with AI for all segments of society, presents a challenge for (life-long) education. Moreover, due to the uneven digitalization of entire segments of society, certain groups may not benefit as much as others. Bias in training data leads to decreasing performance of AI systems when dealing with less common scenarios – a phenomenon known as the "long tail problem in AI". Obtaining training data from the long tail of less common scenarios is usually difficult and costly for business actors, which disproportionately affect group underrepresented in digital environment – women, refugees, or indigenous peoples.

[2] In the large-scale study carried out in 2023, the authors discovered that consultants utilizing the standard version of Chat GPT-4 managed to complete an average of 12.2% more tasks, accomplished tasks 25.1% faster, and achieved 40% better quality outcomes compared to their counterparts who did not use the AI assistant.

- **More disproportionate among regions** – Currently, AI is based on the paradigm of accumulation of data and computing power, which favors centralization and provides systemic advantages to the largest corporations located primarily in developed countries. There is a threat of 'digital colonization' and new forms of exploitation (e.g., OpenAI used Kenyan workers to train their AI filter for harmful content). At the moment, most of the world's population falls into the "long tail", which will affect how effectively AI can address their unique needs and challenges. For instance, according to the study from 2022, 86% of genetic studies is based on data collected from individuals of European descent (an increase from 81% in 2014). Similarly, our knowledge on human brain is based mostly on data coming from brain scans of adult students studying at US-based universities Next to the right to health, the unequal pace of digitalization will also impact cultural rights, preservation of indigenous knowledge and languages.
- **More accessible –** the technology diffuses faster than ever and its costs decreases rapidly. It took more than 15 years to bring personal computers to 100 million households, mostly in developed countries. For the Internet, it took 7 years to achieve 100 million users. The same for Facebook took 4,5 years. ChatGPT, the most popular general-purpose AI, made it in only two months. In the context of neurotechnologies, nearly one out of every five Americans using a wearable neurotech device, e.g., earbuds or VR headsets that monitor brain activity (Farahany 2023). One of these devices have been used for strategic litigation purposes in the world's first neurorights case (Chile).

How to act?

To effectively navigate the future shaped by AI and neurotechnologies, we must adopt a forward-thinking and multifaceted approach to action. This necessitates not merely reacting to the challenges of the moment but shaping the trajectory of technology in a way that safeguards and enhances human rights, especially those of children and future generations:

- **Strategically – building resilience and adaptability:** The rapid evolution of AI technologies requires a strategy that is not overly fixated on the immediate challenges but is resilient and adaptable to the unforeseen shifts in the technological landscape. One the one side of spectrum there is a commercial application of quantum computing, one other side – another "AI winter", a decade(s) of technological slowdown. Adaptation also means facing the challenge of AI brain drain. Although far from easy, this dynamics can be utilized as an engine for innovation in some regions. One example is Albania's pioneering use of ChatGPT to accelerate the integration of EU law into the national legal framework – developed in collaboration with the CEO of OpenAI, Mira Murati.
- **Capacity building – experimentation, individualization, looking for weak ties:** To keep pace with technological advancements, there's a critical need for constant experimentation. It is particularly vital in navigating the jagged technological frontier of AI adoption, where AI can easily accomplish some tasks while struggling with others that appear similarly challenging. This disparity underscores the importance of fostering a culture that values individualized approaches and the exploration of weak network ties (this is how innovation diffuse).
- **Education – elevating digital literacy and understanding:** Human rights education should include digital and data literacy to deepen the understanding of how AI technologies work. This includes basic programming skills, not for the purpose of creating programmers out of everyone, but to ensure a broad-based understanding of the technological world, empowering individuals to engage in meaningful discourse about the direction of AI development. In the long term, education can be also perceived as a means to influence the trajectories of future technologies.
- **Inclusive consultation and meaningful child participation:** Drawing inspiration from innovative participatory models like vTaiwan, which crowdsources legislative and policy ideas, we should strive for inclusive consultation processes. It is particularly important for children, as the decision-making processes around AI have been predominantly adult-centric. This approach must be reconsidered to prioritize the participation of children, ensuring that their best interests are central, not only in the present but in shaping a future that they will inherit. The concept of children's best interests should be expanded to include their digital rights and interests, recognizing that excluding children's data from AI training datasets perpetuates an adult-centric bias in AI development.