INTRODUCTION: FORGING AN EQUITABLE AI FUTURE IN AND THROUGH EDUCATION

Moira V. Faul
Executive Director, NORRAG, Switzerland

"Using technology is as essentially human as making ethical decisions; let's lead with ethics."

AI is many things: from spellchecks to text- and image-generation apps to the public and private infrastructure on which these services rely. While AI has been around for decades, the public release of ChatGPT at the end of November 2022 shifted public and policy conversations in important ways. AI is here to stay, and education stakeholders are duty-bound to examine both the opportunities and the pitfalls that AI entails.

Inequities in AI reflect real-world (human-created, non-AI, non-technological) inequalities. AI holds the potential to address certain inequalities. Nevertheless, AI can reinforce and amplify existing inequalities while creating new harms if we do not scrutinise questions of equity in the design and deployment of AI systems now.

Addressing the ethical issues posed by current developments in AI does not mean abandoning it. At NORRAG, we use AI to generate webinar subtitles in multiple languages to increase accessibility. I use AI voice recognition to avoid damaging my wrists with excessive typing. Few of us switch off the spellcheck function when we write.

AI in society is not a binary on/off

Rather, the authors in this collection foreground the ethical challenges that arise with regards to AI use in education whether as a private, public or common good, and invite you to put human and planetary flourishing at the heart of AI decision making, development and deployment.

Different purposes for AI do not need to be incommensurate—AI design, use and monetisation could be oriented to enable individual efficiency and also social effectiveness; to generate reasonable business profits and also human flourishing alongside environmental and labour protections (Radu, 2024; Whittaker, 2021). Nevertheless, as AI development and deployment are currently configured, access to and freedom from exploitation in AI are unevenly distributed in ways that systematically exclude the most vulnerable.

The contributors to this Policy Insights collection provide key takeaways for education stakeholders and decision makers on some of the main challenges concerning inequalities, putting the value of humans and our planet at the centre of our use and governance of AI, and its underpinning value creation models.

AI "evolution": Driven by humans, and human-generated developments in computing power and data availability

AI needs HI (human intelligence). The work of AI researchers and developers can be categorised into distinct generations, with notable shifts in approaches and methodologies in each.

The term "artificial intelligence" (AI) was first coined in 1955 at Dartmouth College, USA, when a group of pioneering academics gathered to explore the idea of creating machines that could mimic human intelligence (McCarthy et al., 1955). Their conjecture was "that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." (p.1). In the first generation (1950s–1980s), human experts encoded their knowledge into a set of explicit rules that govern the behaviour of an AI system. These systems excelled in rule-following and symbolic reasoning tasks but struggled with handling uncertain, complex and ambiguous situations. Progress was slow due to the

complexity of human cognition that these researchers were attempting to describe and simulate.

In contrast to explicit programming, the second generation of AI researchers (1980s-early 2000s) shifted towards developing so-called machine learning algorithms and large datasets. These algorithms use statistical techniques to generalise patterns from massive datasets of examples, and generate predictions of what an appropriate output might be in novel situations. Researchers overcame the limitations on computational power and the size of available datasets that constrained the scalability and effectiveness of these approaches to produce the third and current generation (starting in the mid-2000s). Researchers working within the broader machine learning framework have developed "deep learning". This approach leverages many layers of neural networks inspired by the structure and function of the human brain without attempting to simulate human intelligence as in the 1950s.

In 2022, Microsoft's OpenAI launched ChatGPT (a user-friendly chatbot) and DALL-E (a text-to-image model), and Alphabet's Google launched Bard and then Gemini. Other non-profit models exist: in contrast to these commercial tools, HuggingFace (launched in 2016) is an Open Source and collaborative community for co-creating AI tools.

Narratives of AI "evolution", "learning" and "decision making" tend to hide the fact that humans developed the large datasets and powerful computing resources needed for generative AI, along with the required advancements in neural network architectures and training algorithms. Furthermore, both sides of future-focused narratives (doomers vs. boomers) assume that we need to focus our energy on protecting humans from future harm that may arise. Focusing on the future, however, ignores the actual inequities now that pose as much of a threat if we fail to address them.

Inequities in AI

Who currently has access to AI? The question is broad and encompasses: access to AI technology; access to the possibility of gaining benefit from current AI technologies or developing more in the future; access to researching or critiquing the technology; and access to decisions on AI development and deployment, including over the governance, financing and the allocation of benefits. The 'who' here includes people who are marginalised within countries and companies, as well as the majority of countries and companies that are marginalised from participating in and decision making about our AI present and futures.

Who is currently represented in AI? AI training sets encode the values of privileged members of WEIRD (Western/White, educated/English-speaking, Industrialised, Rich, Democratic) corporations who design these systems and profit from them (Birhane et al., 2022; Dixon-Román et al., 2020; Henrich et al., 2010; Raji et al., 2020). This selective inclusion reinforces and promulgates dominant epistemologies, further marginalising other ways of knowing and doing (Mahelona et al., 2023;).

Who and what are exploited by current AI? Students currently cannot give consent to their data being used for profit by platforms that are mandated by their institutions (Boly Barry 2022; Mejias & Couldry, 2024; Williamson, 2019). AI use by all of us and AI companies' data centres and processors divert enormous amounts of electricity and water for cooling away from humans and places that need it (Birch, 2022; Luccioni et al., 2023). Exploitative working conditions abound for data workers in low- and high-income countries, without whom AI tools would not be marketable to schools (BBC, 2021; Luccioni, 2023). Copyright challenges arise where open access or pirated articles and books are fodder for LLMs but are not cited.

What research is currently conducted? AI corporations' enthusiastic—but incomplete—reporting of their work is often uncritically repeated by news sources (Bender & Hanna, 2023). Internal research into the impacts and ethics of companies using and developing algorithms and training sets are suppressed, and ethics teams are disbanded (Financial Times, 2023; MIT Technology Review, 2020). The value creation models of AI—and other technology—companies are opaque, particularly regarding the monetisation of users' attention and data (Faul, 2023; Montag et al., 2019; O'Reilly et al., 2023; Pidoux & Dehaye, 2022). Without rigorous, independent research and transparency, developers cannot be held accountable for the experiments they undertake, for the effects of their products and for taking corrective action if necessary.

AI futures for human and planetary flourishing

Much of our current narrative personifies AI, imbuing it with human characteristics while presenting humans in more mechanistic terms. Nevertheless, generative AI generates (Tucker, 2022); it does not think, predict, create, decide, hallucinate, understand or make meaning (Bender & Koller, 2020). A secondary effect of personifying AI is to diminish the possibility of humans to act—the agency that we will need to use if we are to seize this key moment to address AI's digital inequities. Developing and deploying AI requires infrastructure and software that is developed and provided by humans; humans make decisions about what kind of AI we will develop.

According to many technologists, we have entered the phase in the Gartner (2023) hype cycle that represents a coming decade of AI experimentation and deployment. These experiments are human experimentation; therefore they require the same guardrails as any other human experiments (Wired, 2021). AI that improves individual efficiency and reasonable profit-making can function within guardrails for societal effectiveness and human and planetary flourishing. Humans can decide to take action to achieve that.

How can we put equity at the core of AI development? This collection brings together 29 authors from 5 continents who provide key takeaways for decision makers, educators and students seeking to support more equitable and ethical design and deployment of AI in education across the full ecology of ethical concerns (Figure 1).

How do we get the next ten years right?

In November 2021, 193 states adopted UNESCO's *Recommendation on the Ethics of AI* (UNESCO, 2023d), and in February 2024, eight global technology companies[1] committed to implementing this global standard in developing and deploying AI technology. UNESCO's (2023c) *Guidance for generative AI in education and research* applies these insights to education, and the Global Monitoring Report (UNESCO, 2023a) demands "tech on our terms". Learning the lesson from unregulated social media in the 2010s, the European Parliament demanded an end to the "addictive design of online services" in June 2023 and 41 of the 50 United States of America are suing Meta (the parent company of Instgram and Facebook) for exploiting young people's vulnerabilities. More over, the European Union adopted the AI Act on 2 February 2024, which imposes bans on or limits the use of high-risk technologies and requires the stress-testing and transparency of legal AI technologies. Thus, in addition to exhortations for individuals to "click wisely," humans can change how technology industries develop and deploy AI more wisely.

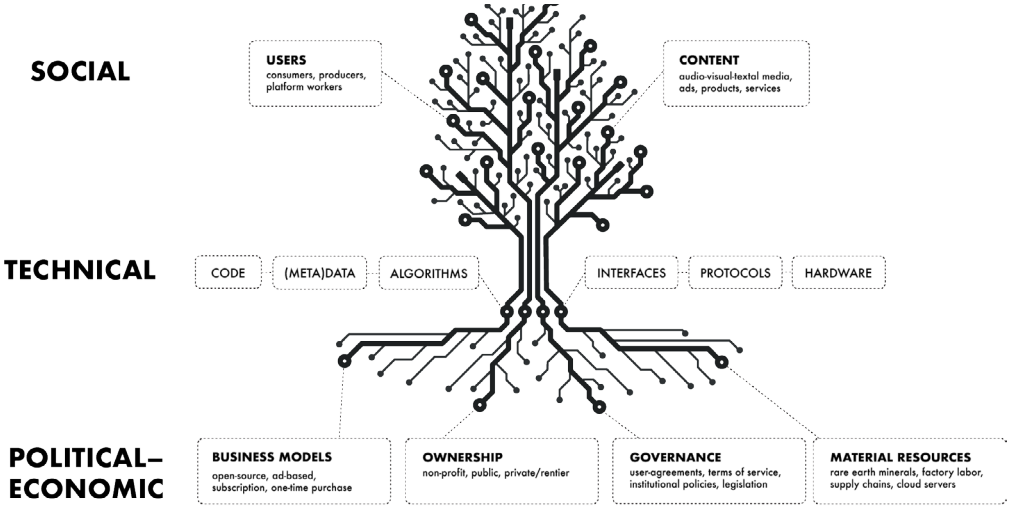
These recent examples show that older human technologies (such as state regulation, corporate governance, collective action and legal challenges) can be used to govern this technocosm ethically (Runciman, 2023) and in the common interest to overcome the allocational and representational harms that are built into current AI development (Bonini & Treré, 2023). AI governance must include decisions and decision makers that safeguard human, social and environmental ecosystems, and ensure that human and planetary wellbeing guide the development and deployment of the algorithms, training sets and energy-hungry processors on which AI depends.

> *"Imagine and craft the worlds you cannot live without, just as you dismantle the ones you cannot live within"*
> Ruha Benjamin

In preparing this introduction, I asked ChatGPT (3.5) the question "Does AI improve equality?" Part of the answer generated was: "Policymakers, technologists, and society as a whole play crucial roles in shaping the impact of AI on equality." That is the challenge the contributors take up in this collection. It is also the challenge they pass onto you: to take action in your spheres of influence early enough to make a difference.

Note: Parts of this introduction were first published in the Geneva Graduate Institute's Globe magazine (Faul, 2023). During the preparation of this work, the author used ChatGPT-3.5 to generate an answer to a specific question, which is reported at the end of the introduction. After using this free tool, the author reviewed and edited the content and takes full responsibility for the content of the publication.

Figure 1
Visualisation of a platform ecology



Source: Nichols and Garcia (2022).

Footnote

1. GSMA, INNIT, Lenovo Group, LG AI Research, Mastercard, Microsoft, Salesforce and Telefonica.

Table 1

Governing AI for humanity: Interim report of the UN AI Advisory Body, convened by Secretary-General António Guterres

GUIDING PRINCIPLES	
Inclusivity	all citizens, including those in the Global South, should be able to access and meaningfully use AI tools.
Public interest	governance should go beyond the do no harm principle and define a broader accountability framework for companies that build, deploy and control AI, as well as downstream users.
Centrality of data governance	AI governance cannot be divorced from the governance of data and the promotion of data commons.
Universal, networked and multistakeholder	AI governance should prioritize universal buy-in by countries and stakeholders. It should leverage existing institutions through a networked approach.
International Law	AI governance needs to be anchored in the UN Charter, International Human Rights Law, and the Sustainable Development Goals.

Source: UN Secretary-General AI Advisory Body (2023)