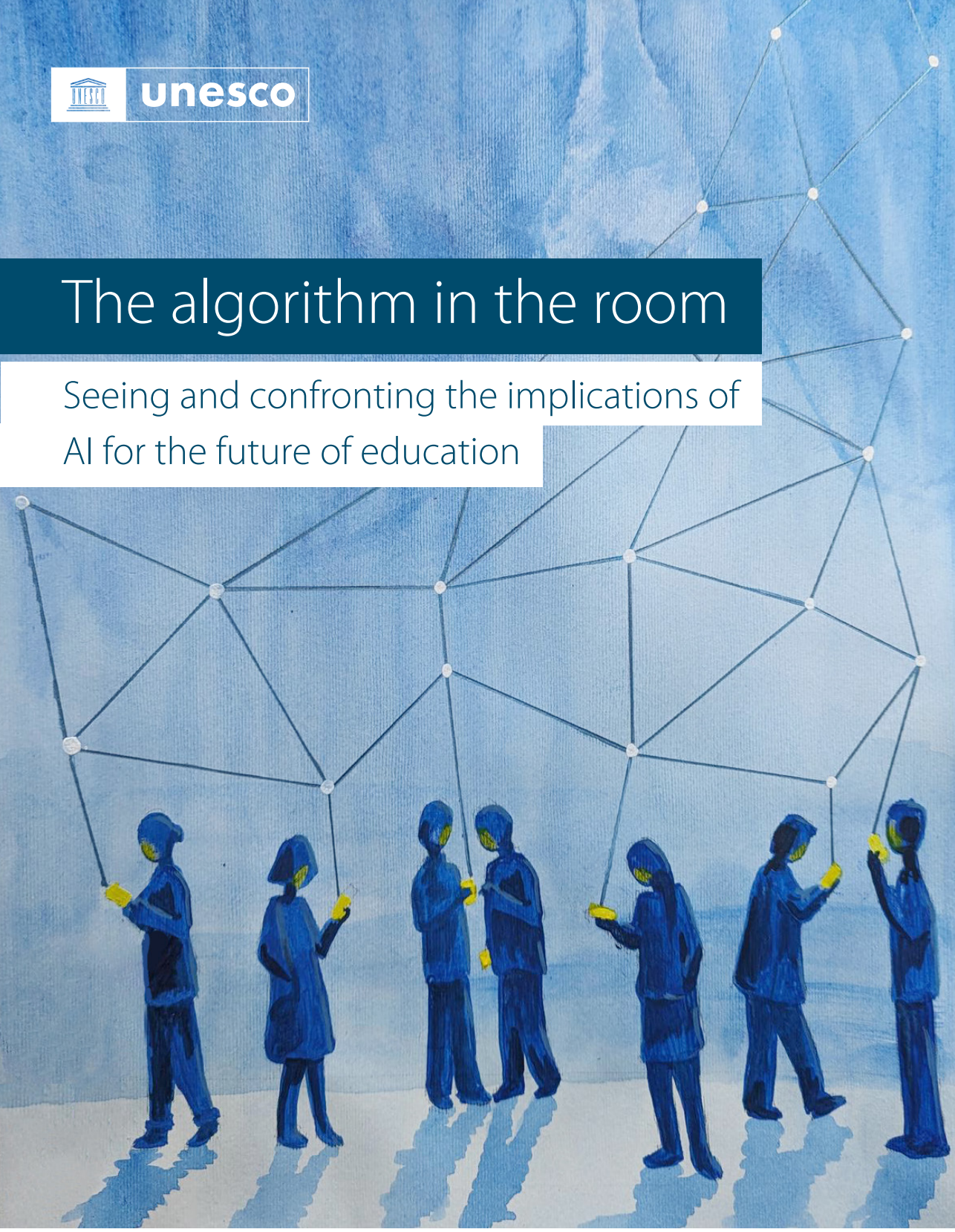
unesco

The algorithm in the room

Seeing and confronting the implications of AI for the future of education



Education 2030

UNESCO – a global leader in education

Education is UNESCO's top priority because it is a basic human right and the foundation for peace and sustainable development. UNESCO is the United Nations' specialized agency for education, providing global and regional leadership to drive progress, strengthening the resilience and capacity of national systems to serve all learners. UNESCO also leads eorts to respond to contemporary global challenges through transformative learning, with special focus on gender equality and Africa across all actions.

Futures of Education

The International Commission on the Futures of Education was established by UNESCO in 2019 to reimagine how knowledge and learning can shape the future of humanity and the planet. The initiative incorporates extensive public and expert engagement and aims to catalyze a global debate on how education needs to be rethought in a world of increasing complexity, uncertainty, and fragility.

unesco

United Nations
Educational, Scientific
and Cultural Organization

Published in 2026 by the United Nations Educational, Scientific and Cultural Organization,
7, place de Fontenoy, 75352 Paris 07 SP, France

© UNESCO 2026

ISBN: 978-92-3-100897-9
DOI: https://doi.org/10.54675/PMRO5023

This publication is available in Open Access under the Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) license (http://creativecommons.org/licenses/by-sa/3.0/igo/). By using the content of this publication, the users accept to be bound by the terms of use of the UNESCO Open Access Repository (https://www.unesco.org/en/open-access/cc-sa).

The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The ideas and opinions expressed in this publication are those of the authors; they are not necessarily those of UNESCO and do not commit the Organization.

Cover credit: Jamillah Knowles & Reset.Tech Australia / https://betterimagesofai.org. Available under CC BY 4.0

Layout: UNESCO/Aida Alhabshi

Printed by UNESCO

Printed in France

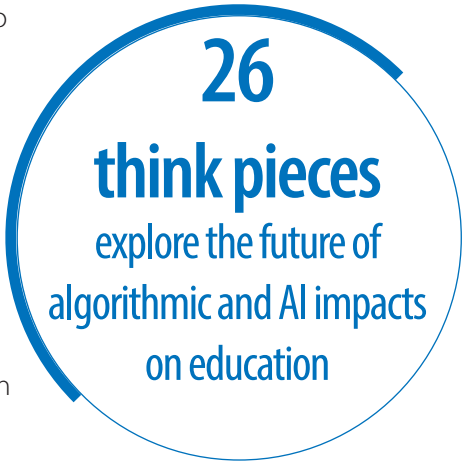
SHORT SUMMARY

Navigating educational priorities in the age of AI

We live in algorithmic times. Algorithmic systems, content and infrastructures that underpin artificial intelligence (AI) now impact education systems across the world.

The title of this volume is a play on the expression 'the elephant in the room', which refers to any massive and far-reaching issue that is largely avoided or unaddressed. AI and other algorithmic technologies are among these issues. They are driving profound changes to education but are often insufficiently acknowledged and discussed.

This collection of 26 think pieces aims to see and confront the 'algorithms in the room' at work in education. The think pieces are organized around six urgent questions for the future:



Who will hold agency, autonomy and oversight? How will young people develop relationally, emotionally and cognitively? What will be meaningful and relevant for pedagogy and assessment? How will we think, question and express ourselves? How will we govern education as a common good that serves public purposes and is accountable to the public? How might AI affect imagination about education's future?

Education in the age of AI is far from settled. It is vital for all actors in education to acknowledge and speak about the 'algorithms in the room'. This will help them to determine what comes next and respond proactively to build just, equitable, and meaningful futures of education.

unesco

"Since wars begin in the minds of men and women, it is in the minds of men and women that the defences of peace must be constructed"

The algorithm in the room

Seeing and confronting the implications of AI for the future of education

Acknowledgments

This publication stemmed from an open invitation to students, educators, policy-makers, researchers, education leaders and administrators to contribute short think pieces to an ongoing conversation about the implications of AI in education. The thought-provoking online series has coalesced here in a collective intelligence project to help advance what education stakeholders are seeing and understanding about the implications of AI for the future of education in diverse contexts around the world.

Under the guidance of Sobhi Tawil, Director of the Futures of Learning and Innovation Division, the drafting and editing of this volume was led by Elena Toukan, Research Coordinator, in close collaboration with Mark West, Education Research and Innovation Lead, and Shafika Isaacs, Chief of Section for Technologies and AI in Education. Aida Alhabshi, Programme Assistant in communication provided layout and design. Anastasia Christodoulelis, Research Assistant, contributed to the initial review process and assisted with the drafting and copyediting of the volume. We also thank the following colleagues for their review and suggestions: Keith Holmes, Elizaveta Kondrashova and Paula Razquin from the Division for the Future of Learning and Innovation, as well as Valtencir Mendes, UNESCO Office in Santago, and Muriel Poisson, UNESCO International Institute for Educational Planning.

We express our immense gratitude to the following experts for their valuable comments: Bryan Maddox (Digital Education Futures Initiative [DEFI], University of Cambridge), Kevin Martin (DEFI, University of Cambridge), Arathi Sriprakash (University of Oxford) and Ben Williamson (University of Edinburgh).

We thank all authors who contributed their ideas and insights in these swiftly moving conversations from such a wide range of vantage points.

Table of contents

Short summary 1

Acknowledgments 5

Table of contents 6

Introduction 8

Algorithms everywhere 9
Educational purposes and the alignment problem 10
Navigating this collection 11

Part I : Who will hold agency, autonomy and oversight? 13

Are human beings necessary to the life cycle of knowledge?
David Ross 15

Beyond the loop: Reclaiming pedagogy in an AI age
Daniela Hau 21

AI operators or creators? Two visions of agency and learning
Kenneth Y T Lim, Ahmed Hazyl Hilmy and Bryan Zi Wei Kuok 26

Deepfakes and the crisis of knowing
Nadia Naffi 31

Part II: How will young people develop relationally, emotionally and cognitively? 38

Ghost in the chatbot: The perils of parasocial attachment
Clayton Ramsey 40

Outpaced: When the world changes faster than children grow
Andrej Novik 45

'I think we should tell your teacher we used AI'
Alexandra Samur 50

Trauma, education and AI in conflict contexts: Care and control
Mariela Sanchez Salas 54

Part III: What will be meaningful and relevant for pedagogy and assessment? 59

Bespoke or prescribed? The myth of personalized learning
Bianca Farthing 61

What's worth measuring? The future of assessment in the AI age
Hrishikesh Desai 64

When my students told me I 'sound like AI'
Claudia Muellauer 69

Between wood and algorithms: Crafting AI literacy in TVET
Marianne Haarsbjerg 73

Reimagining entrepreneurship education in the AI age
Efe Imiren and Katerina Nicolopoulou 78

AI and accessibility: Abdicating engagement?
Alice Bennett 83

Part IV: How will we think, question and express ourselves? 88

The cultural costs of AI in Africa's education systems
James Maisiri and Solomon Musonza 90

Simulating the soul: AI and the legacy of Kazakh literature
Zhaiyk Sultan 95

AI and the great linguistic flattening
Aristotelis Ioannis Paschalidis 98

The disappearance of the unclear question
Victoria Livingstone and Jeppe Klitgaard Stricker 103

Part V: How will we govern education as a common good? 107

Open education principles: Resisting the metrics of AI black boxes
Tel Amiel, Glenda Cox and Colin de la Higuera 109

Baumol's cost disease: Long-term economic implications where machines cannot replace humans
Guido Appoldt and Zhen Gong 116

Who educates the LLMs? The uncertain role of the nation-state
Camille Fabo 121

Environmental contradictions of large language models in higher education
James O'Sullivan and Mark Kelly 125

Could zero-rating advance meaningful connectivity in education?
Kevin Martin 131

Part VI: How might AI affect imagination about the future? 139

AI labs and the great academic migration: A speculative scenario
James Ransom 141

AI and education for collective intelligence: A futures perspective
Imogen Casebourne and Rupert Wegerif 145

After AI
Serge Pouts-Lajus 151

Conclusion 155

Three horizons to empower education in algorithmic times 155
Horizon 1: See the algorithm 157
Horizon 2: Steer the algorithm 157
Horizon 3: Rewrite the algorithm 158

Bibliography 160

Abbreviations and acronyms 162

Introduction

There is a common refrain: 'AI is reshaping education'. Since the release of mainstream generative AI applications in late 2022, this phrase has been repeated with such regularity that it has become axiomatic. The phrase is often joined by statements like 'The question is not whether AI will transform education but whether this transformation will [pick your choice] be equitable/require as many human teachers/be built around schools/improve learning outcomes/and so forth'. These affirmations paralyse the debate from the outset by announcing their central premise off limits.

This volume confronts this paralysis, asserting that whether and how AI transforms education is indeed worthy of debate. Should children who are at the earliest stages of literacy and reading be having full conversations with machines that pretend to be sentient and caring authority figures? Should students learning to structure their own thinking be using AI tools that will, with virtually no student effort, do their writing for them? Should educational content and assessment be generated by AI models, including those that come with prominent warnings that they often 'make mistakes'? These questions and more are 'on limits', as there is nothing unreasonable about asking whether AI, and algorithms more generally, should or will transform education.

But beyond simply questioning AI's shaping power, the volume also recognizes the irreplaceable agency of those within education: students, teachers, parents, school leaders and more. As AI has begun to crop up in classrooms, schools and systems across the globe, many educational actors are actively responding to its intervention, exploring the expectations and norms concerning the uses, and misuses, of the technology. And while the futures of AI and the futures of education are becoming deeply intertwined, it is still far from clear if or how their aims and methods can be truly aligned. Thus, the pieces featured in this volume do not necessarily assume that AI-enabled education is what inevitably lies ahead, nor do they suggest that education systems resisting AI adoption are inherently 'lacking' or destined to be 'left behind'.

To encourage a wide range of responses to AI's implications, UNESCO invited authors from diverse regions, countries and positions to consider what the presence of algorithmic systems could mean for education now and for educational futures. Authored in 2025 to early 2026, just a few years after the rapid and largely global uptake of commercial large language models (LLMs), the pieces featured in this volume are not just speculative: they often observe and describe unfolding experiences in education, spanning from on-the-ground observations and reflections to hypothetical visions of future scenarios.

The pieces, therefore, help illustrate the necessity and urgency of debate about the future. Entire cohorts of young people are passing through schools as test subjects in grand AI experiments (Burns et al., 2026). The precise results of these experiences remain emergent at best and are far beyond the purview of this publication. What this volume is concerned with is how to better see, acknowledge and confront the substantial implications of AI for education. The imaginaries, narratives and scenarios detailed in the individual contributions

explain decisions and pathways that can be taken now to shape the future, whether through regulation and governance, pedagogical design, algorithmic transparency, safeguarding measures, teacher education or leadership development. The authors invite readers to question dominant narratives and imagine alternative trajectories and futures for education in an AI era.

Algorithms everywhere

Today, the predictive powers of the algorithms that steer AI have vast implications for nearly all facets of society. The explosion of algorithmically driven applications and technologies in recent years sees automation applied more expansively to private and public domains. Individuals listen to music and watch videos on platforms that queue what's next based on what they have selected in the past. Travelers walk through airport scanners that use algorithms to identify expected patterns and flag anomalies as potential risks. And increasingly, algorithms mine citizens' personal data, enabling new form of surveillance (Fontes et al., 2022). As more personal activity moves to digital environments, its data points are being run through algorithmic processes to make determinations automatically and, often, with little human oversight. These determinations increasingly carry major consequences and can affect the trajectory of people's lives (O'Neil, 2016). In the field of education, algorithms in use today decide whether an applicant is a good fit for a job or a study programme, whether a student is likely to graduate, or whether an entire class or school is likely to rank well in a standardized exam (Engler, 2021).

Many experts warn, however, that both the rules of algorithms and the datasets that AI models are constructed upon are often not well understood. A model, simply put, is an abstract representation of a given process, instructing the algorithm. Models rely on assumptions about what data points best approximate human behaviours, and how statistical weights, values and constraints should be calibrated accordingly. It can be immensely difficult to get these assumptions right, however, and both algorithms and the data sets they are trained on, then, are far more socially constructed and subjective than presented (O'Neil, 2016). Factors including the high costs of quality assurance, limited resources, reliance on shortcuts, and limited accountability greatly contributes to the complexity of understanding the underlying biases, gaps and premises of what models are fundamentally trained on (Orr and Crawford, 2024).

Algorithmic decisions are also replete with blind spots that reflect the priorities and judments of their creators. Researchers have illustrated how algorithms often entrench harmful stereotypes of women and girls, ethnic and cultural minorities, and low-income populations, reinforcing longstanding historical discrimination and injustices in their outputs (Noble, 2020).

When applied at scale to a population in any field, algorithmic logics can be more prone to punish outliers and exceptions. Even if an answer or action is not inherently 'wrong', algorithmic systems commonly flag people and their outputs as deviations from the statistical norm. And because of the opacity of the decision-making of these systems and

the feedback loops that reinforce them, they can be very difficult to challenge or reverse (Orr and Crawford, 2024). In such a context, education is not separate or detached from a world in which the patterns that algorithms reinforce permeate all aspects of society.

Early on, UNESCO recognized many of the ethical risks to societies of using digital algorithms at new scales and for new applications. It established the first global standard-setting instrument on the ethics of AI, adopted by 193 countries (UNESCO, 2022). It asserts the centrality of human rights and dignity; it opposes discrimination, exclusion, and the spread of misinformation or violence; it calls actors to assume responsibility for AI's environmental impacts and to foster peace and respect. The framework also condemns AI use for systems of mass surveillance and social scoring (UNESCO, 2021). It recognizes that the risk of disguising human biases with technology must be mitigated and prevented, and that the purpose of contributing to human flourishing must be at its core.

Educational purposes and the alignment problem

One of the most significant challenges facing AI in any sector or domain is the degrees to which algorithmic systems align or diverge from human aims, values and intentions (Yudkowsky, 2016). In AI research, this 'alignment problem' is well-known, but only partially understood. While ensuring that AI systems pursue outcomes that correspond with humans' objectives might seem like a basic requirement, it is not as straightforward as it might seem (Christian, 2020). AI developers must often find workarounds and train algorithms towards the best approximation of what users would expect, for example, by constraining mentions of certain topics, prioritizing courses of logic, or blocking them from pursuing harmful courses of action (ibid). AI researchers are paying increasing attention to the value systems, biases, and preferences that various AI algorithms demonstrate and their ethical implications. Yet, there is less research on the educational values within AI systems and how consistent their algorithms are with pedagogical best practices (Autenrieth, 2026). The most weighty alignment problems are rarely technical. Humans regularly disagree on purposes, objectives, and what is reasonable, acceptable, expected and ethical. As such, humans can find it difficult to give clear and detailed instructions with the specificity that the algorithms powering and directing AI require.

In education, human beings have long debated what and how we should teach and learn and for what purposes. Schools, universities, teachers and education systems must navigate complex tensions simultaneously, without one goal completely subsuming the other. For example, educational systems often aim to foster learners' appreciation for culture and critique, while also building skills that hold value in the labour market. Education systems are also expected to measure academic performance while encouraging the intrinsic and broader benefits of learning throughout life. The sophistication required to navigate seemingly conflicting goals effectively and with nuance is a constant challenge and defies formulaic solutions.

When AI algorithms enter educational contexts, the alignment problem poses risks of losing sight of how the many vital purposes of education can be mutually pursued. Several think

pieces in this collection demonstrate instances where AI algorithms miss what is important in education entirely, for example by subverting the process of arriving at a meaningful research question (Livingstone and Stricker, pp. 103-106) or optimizing efficiency in filling apparent 'gaps' rather than deeply exploring a topic (Farthing, pp. 61-63). Some point out that the goal of learning is itself a deeply challenging objective to effectively measure or assess. Even before AI, education relied on imperfect proxies such as grading assignments, participation and exams. When AI algorithms aim to optimize these proxies, they can move yet another inferential step away from many essential and broader goals such as knowledgeability, curiosity, justice and capability (Desai, pp. 64-68).

To help enable education systems, policymakers, students, educators and leaders to respond proactively in these rapidly evolving debates, UNESCO has released several publications including 'Guidance for generative AI in education and research', (2023), AI competency frameworks for teachers and students (2024a; 2024b), and 'AI and the future of education: disruptions, dilemmas and directions' (2025). This volume builds on these publications by extending their implications into the futures of education.

Navigating this collection

This volume comprises twenty-six short think pieces that are organized around pressing questions with sweeping implications for the futures of education in a world of algorithms and AI. The six questions are:

1. **Who will hold agency, autonomy and oversight?** Decisions and actions are increasingly being described as 'co-created' by humans and AI systems. Does this anticipated repositioning of human agency align with education's human aims?

2. **How will young people develop their relationships, emotions and cognition?** AI chatbots and 'companions' are increasingly standing in for the human peer and mentorship relationships that across history have been essential for healthy social, emotional and intellectual development. What could it mean if such commercial algorithmic products exert a growing influence on childhood and adolescent development in and beyond their education?

3. **What will be meaningful and relevant for pedagogy and assessment?** Algorithmic learning tools are becoming prolific. AI tutors, educational games and testing systems offering promises of personalization, real-time data and fine grained feedback to improve learning and performance. But are the aims encoded into algorithms aligned with what is most beneficial and meaningful for learners?

4. **How will we think, question, and express ourselves?** While education has long grappled with the central yet intangible nature of thought and knowledge, AI tools reduce many of the frictions of generating text and arriving at ideas. Will education continue to nurture students' own distinct thought processes and unique styles of

expression when algorithms can arrive at plausible-sounding outputs and answers instantaneously?

5. **How will we govern education as a common good?** While UNESCO and others have long asserted that education is a public endeavor and a common good, the expansion of AI products has introduced a growing array of private and profit-making interests into educational governance and decision-making. Will AI systems serve as vectors for accelerated privatization, data extraction and the reorganization of education infrastructures around commercial and market-driven logics? Or can they be realigned to support humanistic values and participation in a shared knowledge commons?

6. **How might AI affect imagination about the future?** Technological revolutions across time have made great promises for the future. Proponents predict mass upheavals across society driven by AI, including in education. But this is only one possible scenario, and how education responds to AI, instead of simply being 'shaped by AI' should expand imagination, rather than limit it.

The concluding chapter proposes a response of what has been seen, acknowledged and explained. It recommends action in three temporal horizons: today, the near future and the long-term.

In the present, education leaders and other stakeholders need to **see the algorithms** and their implications all around them. In the medium term, they must be given sufficient means and capabilities to **steer the algorithm**, confronting its impacts and disarming possible harms. In the long-term, students, educators and societies must be enabled to **rewrite the algorithm** to ensure alignment with the aims of education.

Education's transformative power is vitally needed to steer many more empowering and human-centred futures for AI. Taken together, the pieces in this volume are a call to action. The collection contains numerous examples of people seeing, confronting, and rewriting the algorithms that enable more empowering present and future visions for education in algorithmic times.

Part I

Who will hold agency, autonomy and oversight?

AI companies have been racing towards their anticipated 'agentic era', a period when computer systems can autonomously perform a wide range of complex tasks, such as making decisions and executing multiple-step processes, even when no human is prompting them to do so (Sukharevsky et al., 2025). Companies tout agentic systems with the promise of exponential efficiency gains. Agency, however, has long been an essential human capacity, and with the emergence of generative AI, there are growing concerns that autonomy is being taken away from people, often without them fully noticing.

In the context of education, the claims of autonomous AI algorithms are perplexing. A central purpose of education is fostering agentic capabilities in *students,* including creative expression, unique problem solving and effective decision-making. Handing over decision-making and action to a machine appears to contradict the importance of requiring students to practice decision-making and action on their own, or in collaboration with other people. And yet, as several authors point out, there is a growing tendency for students, teachers and administrators to willingly delegate creative choices and executive functions to LLMs. While this often involves routine or monotonous tasks, it increasingly extends to cognitively demanding activities.

An example of how these pressures to perform can accelerate the willingness to turn agency over to a machine is explored in the piece **'Are human beings necessary to the life cycle of knowledge?'** by David Ross. Ross's scenarios put readers in the shoes of teachers and students turning to AI as they undergo the repetitive cycles of knowledge production and performance that schools are placing on their time and labour. Moreover, Ross asks us whether sufficient human agency was even involved in these cyclical 'loops' in education – producing assignments, completing assignments, grading assignments, etc. – to begin with. He observes that a Kafkaesque 'human-AI doom loop' can be understood not as an anomaly or worst-case development, but as a natural evolution of the increasing standardization, bureaucratization and high pressures of today's education systems.

Daniela Hau addresses questions of agency from the point of systems design in **'Beyond the loop: Reclaiming pedagogy in an AI age'.** The notion of 'human-in-the-loop' originated in computer science as a means of ensuring human oversight of automated systems, and its use has since permeated other industries employing AI tools, such as engineering and business. Even so, Hau protests its seepage into education, arguing that education should not be considered a 'loop' to begin with – it is, instead, open-ended, relational and responsive to human needs and capacities rather than something guided by pre-determined, closed-system logics.

Two speculative scenarios composed by Kenneth Y T Lim, Ahmed Hazyl Hilmy and Bryan Zi Wei Kuok consider the different trajectories AI design could follow depending on their assumptions of agency, pedagogy and purpose in '**AI operators or creators: Two visions of agency and learning'.** In the first, the year 2035 finds instruction shackled to the monetized tiers of LLM platforms, stratifying achievement according to what one can pay. In the second, a more optimistic alternative illustrates AI tools constructed for open-ended, collective use, with students creating projects that leverage these tools towards community priorities, such as minority language reconstruction and birdsong classification. These two visions demonstrate how: 'when AI systems are granted more autonomy over decisions, human autonomy decreases proportionally, unless the technology is deliberately designed to enhance rather than diminish human agency.' (p. 23).

An unsettling question about agency concerns whether humans will retain cognitive independence and epistemic grounding in an age where AI-generated images and videos increasingly blur the line between truth and falsehood. This line was already becoming distressingly porous before the widespread use of generative AI (Rauch, 2021). In '**Deepfakes and the crisis of knowing',** Nadia Naffi argues that we are no longer facing a crisis of disinformation but a crisis of knowing itself, as deepfakes and synthetic media are eroding the very foundations by which societies construct shared understanding. Naffi argues that education requires new tools to navigate such a 'post-evidentiary' world and ensure that our agency and autonomy remain intact.

At a time when discernment and judgment are central capacities to preserving and strengthening human agency, we do not yet know the long-term opportunities and costs of synthetic knowledge on human learning and the world (Stricker, 2025). We see signs in research that humans and LLMs are increasingly co-constructing epistemic error and this rebuts assumptions that AI will correct and improve human-made mistakes and logical missteps (Oliveira et al., 2025). It is no longer only 'to err is human'. Our tools are erring with us. And this condition makes it doubly important that education assure the human agency to course correct ourselves and our machines when necessary.

Are human beings necessary to the life cycle of knowledge?

David Ross

A teacher warns of a looming AI crisis in education: without clear guidance, learning risks falling into an uncontrollable doom loop.

A few nights ago, I spent 90 minutes online with the staff of an elite public school in Shenzhen, China. I gave an hour-long talk about generative AI and its uses in education then engaged my audience in a 30-minute question and answer period.

Near the end of our chat, one of the teachers made a passing comment that has made me think about a potential crisis in our implementation of AI in schools. I call this foreseeable crisis, with very little hyperbole, the classroom AI doom loop.

The teacher in question strode to the front of the room to grab the microphone and then explained the tremendous pressure he and his colleagues face. Students, parents, administrators, government officials, industry leaders and the public were demanding that they increase students' AI literacy while training learners to use the tools in ethical and effective ways.

This implementation was, somehow, to occur in a learning environment bereft of guidelines, policy, laws, regulations, research, models or proven strategies. He and his teaching colleagues were left on their own to face the most invasive disruption to education since, well, paper.

Students in his high school history class were turning in 25-page reports, perfectly written, clearly reasoned and adorned with pages of citations. Did he really want to tell influential parents that their cherished offspring had cheated by using AI? Even if he did, could he prove it? He had no clue how to assess the work.

The teacher glared at me – the expert – and asked a two-part question: 'Where do you see this going? Best-case scenario and worst-case scenario?'

Worst-case scenario: A screenplay

I, and all the other experts in this field, have answered the best-case scenario question countless times, but I had never really pondered the idea of a worst-case scenario. Three days later, I have one. I'll call it *Johnny and Mr. Bennett Enter the Classroom AI Doom Loop,* with apologies to the team who wrote the introduction for the classic American television programme Twilight Zone ...

You're traveling through another dimension – a dimension not only of sight and sound but of mind, both human and artificial. A journey into a wondrous land whose boundaries are that of imagination. There's a signpost up ahead: You are now entering the Classroom AI Doom Loop…

> I, and all the other experts in this field, have answered the best-case scenario question countless times, but I had never really pondered the idea of a worst-case scenario.

Mr Bennett, an American 10th grade history teacher and the father of a newborn, asks ChatGPT to create five standards-based writing prompts for the topic of a compare-and-contrast essay on the root causes of the French and American Revolutions. ChatGPT pops out five prompts a moment later. Mr Bennett skims the options and quickly chooses prompt No. 2:

'Examine how economic hardship – especially unfair taxation and national debt – fueled the American and French Revolutions. How did the similarities and differences in each country's financial pressures affect the direction of revolt?'

Mr Bennett pastes the prompt into a new assignment in Google Classroom and schedules it for release the following morning. He goes back to caring for his new baby.

Johnny sees the assignment in the morning, copies it, pastes it into Claude.ai and asks Claude to write a response in the style of a 10th grade student. Claude generates a passable essay, which Johnny scoops off the screen and pastes into a blank email in an attempt to remove any digital watermarks. He then copies the scrubbed text and pastes it into the AI humanizer Twixify. The Humanizer makes it read even more like a 10th grader by eliminating the mechanical tropes of AI-generated text.

Johnny, who is thinking only about his overloaded daily schedule, doesn't even read the output from the humanizer before downloading it as a Word doc. and sending it via email attachment to Essay Grader AI for automatic grading.

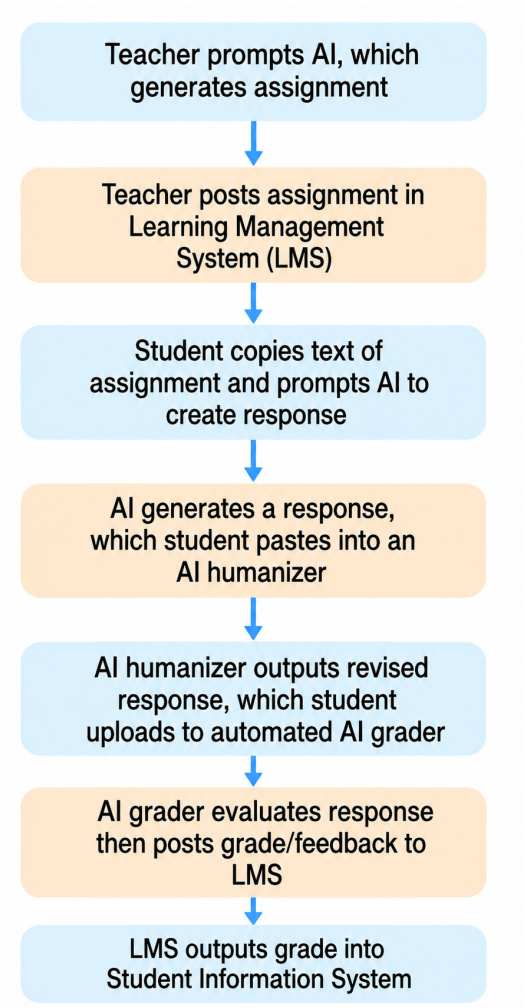
> We exist, of course, but how, in what way? As we believe, as flesh-and-blood human beings, or are we simply parts of an AI's automated workflow?

Essay Grader AI assesses the essay using a system-generated rubric, giving Mr Bennett advice for improving the assignment and Johnny for improving his writing. The AI outputs the score back into the Google Classroom grade book, which ports that grade into the district grade book on PowerSchool.

Mr Bennett, consumed by his fatherly duties, should really look at the feedback from Essay Grader AI, but he doesn't have time. He supplements his salary by coaching the girls' basketball team after school. For understandable reasons, he trusts the AI to relieve him of the onerous task of grading 125 essays every week.

We know that a scenario can be real, but who ever thought that reality could be a scenario? We exist, of course, but how, in what way? As we believe, as flesh-and-blood human beings, or are we simply parts of an AI's automated workflow? Think about it, and then ask yourself, do you live here, in this classroom, in this school, or do you live, instead, in the Classroom AI Doom Loop?

The mechanics of the doom loop



© Adapted from David Ross

Let's follow the workflow:

- Teacher prompts AI, which generates assignment
- Teacher posts assignment in Learning Management System
- Student copies text of assignment and prompts AI to create response
- AI generates a response, which student pastes into an AI Humanizer
- AI Humanizer outputs revised response, which student uploads to automated AI grader
- AI grader evaluates response then posts grade/feedback to LMS
- LMS outputs grade into Student Information System

No humans were harmed in this process because humans were only ancillaries to the process. And this is today's technology. By the beginning of the next school year, agentic AIs such as Manus, Convergence or Responses API will be able to eliminate humans from any involvement in the knowledge transmission cycle. If the last 100 years of technological innovation have taught us anything, it's that if something can be automated, it will be automated.

Is this scenario really that far-fetched? Students and parents are busy and stressed. They hate homework because they have to give up their evenings and weekends to do it or monitor it. Teachers are busy and stressed. They hate homework because they have to give up their evenings and weekends creating it and then grading it. There is no conspiracy here, but humans will all choose to use AI for similar reasons.

Wouldn't it be ironic if the solution to the industrial model of knowledge transmission is in fact automation?

Beyond blaming the tech sector

The tech sector did not create AI so that students could cheat on history homework or teachers could relieve themselves of the drudgery of grading assignments. These are unintended consequences of AI's functionality. However, the training data and processes used to build these tools play a significant role in how a student or teacher uses them. One also can't ignore the reality of the market: Companies must create products that generate revenue as well as text and images. Form and function are close to an ideal match. Generative AI tools are LLMs. Teaching and learning are even larger language models. They fit together naturally.

By the beginning of the next school year, agentic AIs will be able to eliminate humans from any involvement in the knowledge transmission cycle. If the last 100 years of technological innovation have taught us anything, it's that if something can be automated, it will be automated.

In 1951, the philosopher Bertrand Russell wrote an essay entitled '*Are Human Beings Necessary*?' in response to the new field of cybernetics. He pondered the consequences of an automated society long before generative AI as we know it was even imagined. Russell concluded that an automated society might not necessarily require human beings to function, but that probably wasn't a good idea.

We have come to the inflection point where we can automate most elements of knowledge transmission. I'm not sure if that is a good idea. But before we take up residence in the Classroom AI Doom Loop, we should have a serious policy discussion about the purpose of education. If a major function of education can be automated, it's probably not human enough.

I can illustrate this point by advising you to walk through the aisles of your local megamart. You will see families, which almost always include a toddler who is sitting in the shopping cart, eyes glued to a tablet, oblivious to the chaotic bounty of capitalism that surrounds

them. We have trained a generation of children to be entertained and taught by staring at a screen.

The current iteration of AI tutors is really good at teaching content to students, though, as Jeremy Knox (2025) points out, in a manner that currently breaks no new ground pedagogically. If children prefer to learn that way, let them. For thousands of years teaching has consisted of equal parts transmission of knowledge and human development. Let the machines handle the mundane transmission of low-level knowledge per Bloom's Taxonomy (Derek Bok Center for Teaching and Learning, n.d.).

There is not a school on earth that doesn't have a poster or plaque in the office that says something like the following: 'Our mission is to develop lifelong learners who effectively communicate, critically think, collaborate and use their creativity to become successful in college and career while improving their community as active citizens'. These competencies are the components of education that humans excel at modelling, explaining, coaching and teaching.

> But before we take up residence in the Classroom AI Doom Loop, we should have a serious policy discussion about the purpose of education. If a major function of education can be automated, it's probably not human enough.

Let's live up to those promises by letting the humans focus on human development and the array of skills that are described in frameworks as varied as the UNESCO Competency Framework (UNESCO, 2016), 21st Century Skills (Partnership for 21st Century Skills, 2009), Socio-Emotional Learning (CASEL, 2020), Employability Skills (United States of America, OCTAE, 2018), Character Education (Character.org, n.d.), the Whole Child Approach (ASCD, n.d.), and Global Competence (OECD, 2018). They are all based on lived experience, and that's something humans currently hold a monopoly on.

References

Association for Supervision and Curriculum Development (ASCD). n.d. *The whole child approach to education*.

Character.org. n.d. *11 Principles: An Evidence-based Framework*.

Collaborative for Academic, Social, and Emotional Learning (CASEL). 2020. *CASEL'S SEL framework: What Are the Core Competence Areas and Where Are They Promoted?*.

Derek Bok Center for Teaching and Learning. n.d. *Taxonomies of learning*.

Knox, J. 2025. *Impact, inequality, and imagination: envisioning a more ambitious education in the era of AI*. Paris, UNESCO. https://doi.org/10.54675/SAKE2705

OECD. 2018. *Preparing our youth for an inclusive and sustainable world: the OECD PISA global competence framework*. Paris, OECD Publishing.

Partnership for 21st Century Skills. 2009. *P21 Framework Definitions*. Washington, DC, Partnership for 21st Century Skills.

Russell, B. 1951. Are Human Beings Necessary? A. G. Bone (ed.), *The Collected Papers of Bertrand Russell,* Vol. 26. London, Routledge.

UNESCO. 2016. *UNESCO Competency Framework*. Paris, UNESCO.

United States of America, Office of Career, Technical, and Adult Education (OCTAE). 2018. *Employability skills framework*. Washington, DC, Department of Education.

About the author

David Ross is a writer and AI consultant and is the former CEO of the Partnership for 21st Century Learning.

Beyond the loop: Reclaiming pedagogy in an AI age

Daniela Hau

Rethinking 'human-in-the-loop' design logic reveals the need to reassert pedagogy as the foundation of education, not a fallback to automation.

The loop we didn't design: Rethinking human-in-the-loop in education

The term 'human-in-the-loop' (HITL) has become a familiar feature of conversations about AI in education today. Originally from technical fields such as machine learning and robotics, the term refers to systems that rely on human expertise to guide, correct or oversee automated processes. In this framework, humans provide judgement where algorithms encounter ambiguity, such as labelling data, refining model predictions or making context-sensitive decisions (Natarajan et al., 2025). HITL strategies are employed in fields such as medical diagnostics, autonomous driving and adaptive robotics, where complete automation may pose risks or raise ethical concerns. The idea is that by incorporating human oversight into the system, we can achieve an optimal balance between machine efficiency and human discernment (Wu, 2022; Benedikt et al., 2020).

HITL has evolved into a blanket reassurance: as long as a human is 'in the loop', the system is assumed to be safe, ethical and trustworthy. However, when it comes to education, this metaphor obscures more than it reveals. It implies that AI is the primary actor – the centre of gravity – while humans act as assistants, reviewing feedback, nudging outcomes or approving decisions made by the system.

Even worse, it establishes a subtle hierarchy of agency. The machine leads and the human follows. Pedagogical judgement is treated as a failsafe rather than a foundation.

When automation thinks pedagogy: Why 'human-in-the-loop' is misleading

HITL systems were never designed for educational purposes. They emerged in the fields of machine learning and robotics, and even there, the loop is often unidirectional. Machines do not understand humans; they simply adapt based on input. There is no shared context, no reciprocity (Holzinger, 2016; Shi et al., 2023).

In education, this metaphor is misleading and masks a shift in power. Matthew Rascoff's speech to the UN (2025), he warns that many AI systems promote 'intellectual isolationism' by focusing on one-to-one learning. AI tutors, AI writing assistants, AI feedback systems promise to personalize learning, speed up grading and reduce teacher workload. But most are built around a closed loop. In such models, human teachers are not designers of learning or facilitators but

Education is not an engineering pipeline. This isn't a workflow to be optimized, a dashboard to be monitored or a feedback loop designed by machines with humans ready to step in if necessary.

are end-users of data dashboards. Students become data producers whose behaviour drives optimization. Pedagogy is in danger of becoming an afterthought.

Let us be clear: education is not an engineering pipeline. This isn't a workflow to be optimized, a dashboard to be monitored or a feedback loop designed by machines with humans ready to step in if necessary. Education is a relational, meaning-making act between people anchored in language, context, curiosity and care. And AI, no matter how sophisticated, enters this space not as a neutral calculator but as a co-agent that reframes what it means to learn, teach and know.

So, rather than redesigning the loop, let's step outside of it altogether. Education doesn't need a better feedback mechanism; it requires language that reflects its complexity, questions that challenge its assumptions and priorities that serve learners – not systems. We must change our mindset from systems that humans supervise to pedagogies that humans define. AI should only enter the picture after we've asked: When, where and how – if at all – can AI meaningfully contribute to human learning? The issue is not whether or not to include AI in the loop, but rather to reclaim pedagogy as the primary frame.

Pedagogy-first AI: What it looks like in practice

Moving beyond the loop does not leave us with a void; it leaves us with better questions. How can we design learning environments in which AI plays a supporting role rather than taking the lead? What does it mean to build classrooms, tools and practices where educational values, rather than technological capabilities, set the terms of engagement?

Learning is messy. It's relational, contextual and human. When we reduce the role of the teacher – or the student – to that of an add-on within an AI-powered workflow, we oversimplify the complexity of education. The goal is to make learning more meaningful, not more efficient. This requires placing pedagogy, not automation, at the centre of design.

This isn't just semantics. It's a shift in perspective. Pedagogy-first design asks not how to insert AI, but rather whether it belongs at all. And if it does, how can it be used to expand – rather than constrain – learning, agency and relational depth? In this model, AI is not the driving force. It's the co-agent. It is a contingent actor invited into the learning process – occasionally helpful, always secondary and never neutral.

We must change our mindset from systems that humans supervise to pedagogies that humans define.

The following illustrations of what pedagogy-first AI can look like in practice do not offer best practices or technological blueprints. Instead, they demonstrate a mindset in which teachers and learners remain in charge, not just in the loop.

- **Collaborative writing, not automated essays:** Students are not using AI to write essays, but to refine them. After writing a first draft by themselves, they use AI to identify synonyms or smoother transitions. Working in pairs, they evaluate the AI's suggestions and consider what could strengthen their ideas. The teacher supports this process by encouraging metacognitive thinking rather than simply correcting grammar. Here, AI amplifies student thinking rather than replacing it.
- **Language access without displacement:** In multilingual classrooms, AI translation tools are used to help students with different language backgrounds understand each other. While the translations are not perfect, they are good enough to encourage participation and reduce isolation. Teachers remain central to the process, adjusting pacing, providing context and co-creating glossaries with learners. In this setting, AI extends access without replacing the relational work of teaching.
- **AI as a mirror, not a master:** An AI system analyses classroom interaction to identify patterns. For example, it can determine who is being challenged and how students' ideas are being handled. Teachers review this information not to meet benchmarks, but to gain a better understanding of their own teaching. It's a mirror, not a metric. Here, AI is a lens for self-inquiry, not a system for performance monitoring.

Beyond the metaphor: Reclaiming the language of learning

Words do more than just describe systems; they also shape them. Part of the problem is that we keep using language that doesn't belong to education: 'Artificial intelligence', 'Optimization', 'Human-in-the-loop'. These are terms from engineering, from automation, from systems thinking. They carry assumptions about how knowledge works, who controls it and what constitutes success. If left unexamined, they pull pedagogy towards logics that education never asked for.

Take 'artificial intelligence': The term suggests cognition, autonomy and even creativity – qualities traditionally associated with human minds. However, most AI systems are pattern recognizers, not thinkers. In education, the term is even more misleading. Intelligence has never been neutral; it is cultural, emotional, social and contextual. Treating it as a mechanical function not only flattens its meaning but also narrows its possibilities.

HITL only compounds this confusion. It anthropomorphizes machines while reducing humans to safety features. In education, this framing is particularly dangerous. It brings educators and learners in late, limits their agency and assumes that AI's role is set in stone, rather than being open to public and pedagogical debate.

After all, metaphors aren't just rhetorical; they're architectural. They determine what is built, who it is built for and how it will be used. Yet in pedagogy, the human isn't a fail-safe. The human is the starting point.

Metaphors aren't just rhetorical; they're architectural. They determine what is built, who it is built for and how it will be used.

If we want AI to serve education and not the other way around, we need better metaphors. At the very least, we need to return to language that reflects what education actually is: a collective act of meaning-making. It is a space for ambiguity, dialogue and care. Not a loop to be closed, but a relationship to be cultivated.

This shift requires more than different language. It requires a different design approach. This means considering the social, shared nature of learning rather than focusing on the individual learner. It means creating systems that encourage collaboration rather than personalization. It means giving teachers real agency, not as overseers of algorithms, but as designers of pedagogical ecologies. It means being transparent not just about how AI works, but also about how it shapes relationships, authority and responsibility. Above all, it means resisting the urge to optimize every aspect of learning. Because not everything that matters in education can be measured – or even fully understood.

Education is not a loop to be monitored. It's a world to be co-constructed.

When we treat uncertainty as part of the learning process rather than as failure, we open the door to forms of knowledge that AI cannot model, such as intuition, empathy, improvisation and trust. These are not bugs in the system. They are at the heart of it.

A final thought: Not in the loop, but in the lead

AI in education is advancing rapidly – faster than most education systems are prepared for. We are at a moment where pedagogical principles could easily be overwritten by technological momentum.

If we uncritically adopt the language of systems, loops and optimization, we risk designing tools that undermine the very conditions that enable learning. However, if we prioritize pedagogy, context, care and curiosity, we can ensure that AI plays a meaningful, limited and constructive role.

This is not about rejecting technology. It's about reclaiming education as a process that is too complex to be simplified, too relational to be standardized and too important to be left to metaphors that do not belong.

Education is not a loop to be monitored. It's a world to be co-constructed.

References

Benedikt, L., Joshi, C., Nolan, L. A., HenstraHill, R., Shaw, L. and Hook, S. 2020. Human-in-the-Loop AI in government: A case study. *IUI' 20: Proceedings of the 25th International Conference on Intelligent User Interfaces.* New York, Association for Computing Machinery, pp. 488–97. https://doi.org/10.1145/3377325.3377489

Holzinger, A. 2016. Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Informatics,* Vol. 3, pp. 119–31.

Natarajan, S., Mathur, S., Sidheekh, S., Stammer, W. and Kersting, K. 2025. *Human-in-the-loop or ai-in-the-loop? Automate or collaborate?* arXiv preprint, arXiv:2412.14232.

Shi, F., Zhou, F., Liu, H., Chen, L. and Ning, H. 2023. Survey and tutorial on hybrid human-artificial intelligence. *Tsinghua Science and Technology,* Vol. 28, No. 3, pp. 486–99. https://doi.org/10.26599/tst.2022.9010022

United Nations. 2025. *International Day of Education 2025 - "Artificial Intelligence and Education: Challenges and Opportunities"* [video]. 0:34:47-0:49:01.

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T. and He, L. 2022. A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems,* Vol. 135, pp. 364–81. https://doi.org/10.1016/j.future.2022.05.014

About the author

Daniela Hau is the head of the Innovation Department at SCRIPT (Service de Coordination de la Recherche et de l'Innovation Pédagogiques et Technologiques). She specialises in AI, data literacy, media literacy and digital game-based learning. She holds two master's degrees – one in Economics and the other in Game and Media Pedagogy – and has 16 years' experience as a secondary school teacher. She is currently pursuing a Ph.D. in Psychology on the topic of AI in education.

AI operators or creators? Two visions of agency and learning

Kenneth Y T Lim, Ahmed Hazyl Hilmy and Bryan Zi Wei Kuok

When seeking to understand how the notion of agency might evolve as learners interact with AI, scenarios play a useful part.

Vision 1: The AI operators

Let us imagine a conversation between two high school students in the city-state of Singapore in the year 2035:

'Whoa, is that GPTutor Pro?' Chen Wei asks, pointing at the PRO icon at the corner of his friend's tablet screen. 'How did you convince your parents to pay the $200 a month subscription?'

'Oh, they would never!' Adam exclaims. 'My school got some kind of sponsorship deal with Edu AI. Someone from the company is coming down to make a speech next week.'

The two 16-year-olds sit across each other in a food court, tablet screens propped on their table next to half-finished bubble teas.

'Ugh, I'm stuck with GPTutor Basic', Chen Wei sighed, tapping his own tablet. 'It crashes whenever I ask it anything complicated, and the token limit is so low. I need to wait… eight more hours before I can finish my geography assignment.'

He brings up a dense table of weather data. 'How am I supposed to analyse this myself? The metrics dashboard tells me my "manual analysis efficiency" is below-standard.'

'That's rough,' Adam replies. 'If it makes you feel better, even GPTutor Pro has usage limits too. That's why some of my classmates got their parents to upgrade to the Max subscription. It costs $2000 a month but they say it's totally worth it, because they can basically use it 24/7, no limits on image, video or sound generation either!'

'My teacher asked us to continue analysing the themes in some short story *manually,* she said all we needed was pen, paper and our brainpower. Half the class just took a nap.'

'That's insane!' yells Chen Wei. 'They'll have such an advantage on the upcoming "AI literacy" exams. Ugh, what are we graded on again?'

'Let me check…' Adam taps his screen. 'Prompt optimization and response evaluation.'

'Haha, they expect us to evaluate GPTutor's responses? Why? It's basically smarter than us,' replies Chen Wei.

'Well, there was that network outage last month,' Adam says. 'I was in class when it happened. My teacher asked us to continue

analysing the themes in some short story *manually*, she said all we needed was pen, paper and our brainpower. Half the class just took a nap.'

'Same here,' laughs Chen Wei. 'Mr Lazar just said we needed to "think for ourselves", and went on and on about how he didn't have AI tutors when he was in school. At first, it was funny, but I got jealous when he told us about his learning metrics.'

'Oh, what did they use back then?' asks Adam.

'Nothing!' says Chen Wei, taking a sip of his bubble tea. 'No metrics at all! No penalties for "spending excessive time solving single problems" or anything.'

'Well, that was a different world,' Adam frowns. 'A slower world, yeah? It's different now. We don't need to understand things in detail; AI can do that better than us. We just need to learn how to work with AI effectively. That's what my parents always tell me.'

'Yeah, I guess that makes sense.' Chen Wei looks down at his tablet and sees the notification:

Reminder: *Your GPTutor Basic access will refresh in 7 hours 50 minutes. Consider upgrading to Pro ($200/month) for an extra 4 hours of access time per day.*

Possible future of AI in education

The above scenario illustrates a possible trajectory for the future of AI in education by extrapolating current trends. Much mainstream discourse around AI presents a narrow, technocentric approach that primarily benefits commercial interests, including calls for a version of AI literacy in which the focus is on mastering prompt engineering to optimize outputs from AI systems. This is frequently combined with fearmongering about how workers will lose their jobs if they do not train in such AI skills, effectively pushing them to become more efficient AI consumers. By treating tech as an inevitability and buying into narratives presented by tech companies (in service of their market expansion goals), such approaches risk turning students into passive consumers of black-box technologies, and ignore deeper questions about when AI should be used at all, how it impacts cognitive processes and who benefits from its deployment.

The concept of the 'autonomy budget', coined by Balzan et al. (2025), offers a useful framework.

Recent research findings on the impact of AI technologies on learning provide insights into the risks of uncritical AI adoption. Lee et al. (2025) found that generative AI use significantly reduces users' perceived cognitive effort in tasks requiring critical thinking, which is key to meaningful learning. Zhai, Wibowo and Li's (2024) systematic review identified several studies which highlight the risk of overreliance on AI systems, and how this can undermine students' critical thinking and problem-solving skills and reduce their media literacy skills. Mei et al.'s (2025) study on generative AI use in creative writing found that AI assistance can lead to a homogenization of ideas, potentially limiting diversity of views in educational settings.

The concept of the 'autonomy budget', coined by Balzan et al. (2025), offers a useful framework. Drawing on self-determination theory, they posit that within any human–AI system, there is a finite 'budget' of decision-making power. When AI systems are granted more autonomy over decisions, human autonomy decreases proportionally, unless the technology is deliberately designed to enhance rather than diminish human agency. This suggests that the way forward is to carefully consider how educational technologies should be designed to augment rather than supplant human agency.

When AI systems are granted more autonomy over decisions, human autonomy decreases proportionally, unless the technology is deliberately designed to enhance rather than diminish human agency.

This approach is supported by UNESCO's 2024 AI Competency Frameworks for teachers (UNESCO, 2024a) and students (UNESCO, 2024b). Rather than training efficient AI consumers, these frameworks emphasize the development of critical understanding and creative agency. The student framework, in particular, outlines twelve competencies across four dimensions (human-centred mindset, ethics of AI, AI techniques and applications, and AI system design) with three progression levels (Understand, Apply, Create) which offer a pathway to a more genuine AI literacy rather than mere AI dependency.

Building on UNESCO's frameworks and a more equitable distribution of the 'autonomy budget', we can imagine a different future, in which students are creators rather than consumers of technology.

Vision 2: The AI creators

'How's your community AI project coming along?' Adam asked, taking a sip of his bubble tea.

'Here, let me show you,' Chen Wei grins, swiping his tablet to show lines of code. 'My team managed to customize an open-source model to translate old Hokkien-language radio broadcasts. We're applying for a grant from the National Archives to create a data repository which can preserve this part of our language heritage.'

'Cool!' Adam exclaims. 'Most commercial models ignore this kind of "economically insignificant" language, yeah? I'm impressed you and your team managed this!'

'Yeah, we couldn't have done it without the help of family members who still speak the language. Their input was essential,' Chen Wei replies. 'How about your big urban ecology project?'

'So, we've set up these acoustic sensors in green spaces across the city, right?' Adam brings up a map of Singapore on his tablet, with points scattered across. 'We're using a community-trained AI model which can identify local bird calls to keep track of which birds our sensors detect in the different spaces. It's part of a citizen science initiative to monitor local biodiversity, see how it changes over time due to climate change and urban development.'

'Wow, that's really cool! Could I set up sensors in my school garden? I hear bird calls there every morning, I just have no idea what kind of bird,' Chen Wei laughs.

'Sure, the source code and sensor schematics are available here, I'll send you the link,' Adam replies.

'Nice, I'll share this with my school's AI Tinkering Club,' says Chen Wei. 'Oh, check this out. Last week, our teacher showed us how to create our own custom language models. We modified the dataset to introduce biases and see how it affected the outputs. Guess what we did with this one!'

'Oh, it really likes ancient memes, huh?' Adam smiles. 'Send me the dataset sometime. I'll browse it when I'm bored. Oh, I just remembered, there's a community Hackathon this weekend! Have you signed up?'

'The one about coming up with solutions to help the elderly?' says Chen Wei. 'Not yet, can I join your team?'

'Yeah, of course!' smiles Adam.

'Nice. I love working on projects like this. It's so cool to use tech to make something which can help the community,' says Chen Wei.

'Couldn't agree more,' says Adam.

It is crucial to move beyond the current chatbot-centric model towards approaches which support metacognition, facilitate deeper engagement and foster critical thinking while respecting ethical principles.

Conclusion

Where the first scenario created passive operators dependent on commercial AI systems, this alternative vision nurtures active creators who harness technology for community benefit.

To realize a vision like this for AI in education, it is crucial to move beyond the current chatbot-centric model towards approaches which support metacognition, facilitate deeper engagement and foster critical thinking while respecting ethical principles (including environmental sustainability).

By distributing the 'autonomy budget' in ways that empower rather than diminish the cognitive capacities of human learners, we can create learning environments in which technology serves rather than shapes educational goals. This vision of AI in education, in which students are creators rather than consumers, is intended to be a hopeful alternative to current trajectories, an ideal to strive for as we navigate the integration of AI technologies into education systems across the globe.

References

Balzan, F., Angeli, L., Meulenbroeks, R. and Russo, F. 2025. Tools or crutches? Budgeting human and machine autonomy when introducing GenAI in education. *Artificial Intelligence in Education,* Vol. 2, No. 1, pp. 18–32.

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R. and Wilson, N. 2025. The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *CHI '25: CHI Conference on Human Factors in Computing Systems***.** Yokohama, Special Interest Group on Computer-Human Interaction, pp. 1–22.

Mei, P., Brewis, D. N., Nwaiwu, F., Sumanathilaka, D., Alva-Manchego, F. and Demaree-Cotton, J. 2025. If ChatGPT can do it, where is my creativity? Generative AI boosts performance but diminishes experience in creative writing. *Computers in Human Behavior: Artifical Humans, Vol. 4.*

Siu, A. and Fok, R. 2025. *Augmenting expert cognition in the age of generative AI: Insights from document-centric knowledge work*. arXiv preprint, arXiv:2503.24334. https://doi.org/10.48550/arXiv.2503.24334

UNESCO. 2024a. *AI competency framework for teachers*. Paris, UNESCO. https://doi.org/10.54675/JKJB9835

––. 2024b. *AI competency framework for students*. Paris, UNESCO. https://doi.org/10.54675/JKJB9835

Zhai, C., Wibowo, S. and Li, L.D. 2024. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments***,** Vol. 11, No. 28. https://doi.org/10.1186/s40561-024-00316-7

About the authors

Ahmed Hazyl Hilmy works with a research team at Singapore's National Institute of Education that explores how technologies such as generative AI, machine learning and open-source sensors can empower students to be self-directed learners, enabling them to investigate their own well-being and the world around them.

Bryan Zi Wei Kuok is an Association of Southeast Asian Nations (ASEAN) Undergraduate Merit Scholar and an incoming freshman at the National University of Singapore. He is also a student developer exploring the intersection of spatial computing and AI for pedagogical applications.

Kenneth Y T Lim is a Senior Research Scientist at the National Institute of Education, Singapore. He leads a small team operating at the intersection of neuroergonomics, the Science of Learning, Data Science, and AI.

Deepfakes and the crisis of knowing

Nadia Naffi

> As deepfakes blur reality, education must go beyond detection, teaching students to navigate truth, knowledge and AI-mediated uncertainty.

As synthetic media and AI-generated and disseminated disinformation proliferate, educational institutions rush to develop technical detection tools and media literacy programmes. Simply put, synthetic media is any content, such as audio, images or video, created by AI. This includes 'deepfakes', which are digital forgeries so realistic they can convincingly mimic a person's voice or likeness (Canadian Digital Regulators Forum, 2025).

Deepfakes differ fundamentally from traditional disinformation – they are convincing, scalable and increasingly accessible. Suspicions of AI generation alone sow doubt.

What if this technical arms race blinds us to a more profound disruption? What if our obsession with spotting fakes diverts attention from a deeper epistemological crisis – one that fundamentally destabilizes how humans establish truth and knowledge?

This think piece challenges the verification-centred paradigm dominating our response to deepfakes. Drawing on learning theory, systems thinking and epistemology, I argue that we are not merely facing a crisis of disinformation – that is, falsehoods spread with the deliberate intent to deceive, as distinct from misinformation, which is spread unknowingly – but a crisis of knowing itself. Deepfakes don't just introduce falsehoods into our information ecosystem – they erode the very mechanisms by which societies construct shared understanding.

UNESCO's focus on AI in Education correctly identifies the need for inclusive, equitable approaches to AI literacy (UNESCO, 2025). However, we must now go further: in a world of advanced synthetic media, AI literacy isn't just about using AI tools – it's about surviving in an AI-mediated reality where seeing and hearing are no longer believing.

In response, I propose three pillars through which education must evolve: individual epistemic agency, organizational knowledge construction and cross-sector knowledge ecosystems. Together, these offer a pathway beyond verification – a call for education not just to detect deception, but to rebuild our societal capacity for collective sense-making in a post-evidentiary world.

Pillar 1: Individual epistemic agency

The threat of misinformation and disinformation, now massively amplified by synthetic media, is ranked by the World Economic Forum among the world's top risks, noting the increasingly blurred line between AI- and human-generated content (Elsner, Atkinson

and Zahidi, 2025). Yet our response remains technical: detect fakes, verify sources, spot manipulation.

We are approaching a synthetic reality threshold – a point beyond which humans can no longer distinguish authentic from fabricated media without technological assistance.

This approach is insufficient (Naffi et al., 2024). The generative AI market is projected to grow 560% between 2025-2031, reaching US$442 billion (Statista, 2025). 46% of fraud experts have encountered synthetic identity fraud, 37% voice deepfakes and 29% video deepfakes (Zandt, 2024). In January 2024, fraudsters using deepfake technology impersonated a company's CFO on a video call, tricking an employee into transferring $25 million. Deloitte predicts that generative AI could drive U.S. fraud losses from $12.3 billion in 2023 to $40 billion by 2027 – a 32% annual growth rate (Lalchand et al., 2024).

Beyond corporate fraud, the societal fabric itself is being targeted. In schools, students are using deepfake technology to create harassing and sexually explicit content of classmates and teachers, leaving lawmakers struggling to adapt existing cyberbullying laws to the new threat (Akron News Reporter, 2025). Simultaneously, the medical profession is facing what some call a 'crisis of evidence', as malicious actors use deepfake videos of doctors to promote medical scams, and the potential to fabricate clinical data threatens the very foundations of evidence-based medicine (Lee Myers, Callahan and Rosenbluth, 2025).

Specific dangers are clear. Voice-cloning AI enables scammers to mimic loved ones' voices with seconds of audio, creating convincing emergency scenarios that trick victims into sending money (Bethea, 2024; Vellani and Common, 2025). Research confirms humans cannot consistently identify AI-generated voices, often perceiving them as identical to real people (Barrington, Cooper and Farid, 2025).

Social media amplifies these threats through the 'illusory truth effect', where repeated exposure makes information seem more credible regardless of accuracy. Survey data across eight countries shows prior exposure to deepfakes increases belief in misinformation. Social media news consumers are more vulnerable to deepfakes, and this effect persists regardless of cognitive ability (Ahmed et al., 2024).

We are approaching a synthetic reality threshold – a point beyond which humans can no longer distinguish authentic from fabricated media without technological assistance. Detection tools lag behind creation technologies in an unwinnable arms race (Cal Alumni Association, 2024). Meanwhile, the liar's dividend – the ability to dismiss authentic recordings as probable fakes (Schiff, Schiff and Bueno, 2024) – creates a double bind where neither belief nor disbelief in evidence can be justified.

The liar's dividend – the ability to dismiss authentic recordings as probable fakes creates a double bind where neither belief nor disbelief in evidence can be justified.

A crucial first step in meeting this challenge is for education to prioritize 'epistemic agency': the individual capacity to engage responsibly with knowledge when traditional evidence becomes

unreliable. This builds on adult learning principles (as per Knowles, 1978), emphasizing self-direction and experiential frameworks. Epistemic agency means helping learners develop personal frameworks for assessment rather than accepting pre-determined verification criteria.

Practical examples are emerging. Families establish code words to verify identity during calls, circumventing voice cloning scams. For video calls, people use what could be called a 'prove you're live' challenge, asking the person on screen to perform a specific physical action, like turning their head sharply to the side, that can often cause a real-time deepfake to glitch or reveal artefacts. This represents epistemic agency in action: relational strategies for navigating uncertainty, not just passively analyse digital content. Education must foster these knowledge practices, which go beyond sharpening purely technical detection skills.

We must not simply train individuals. We must redesign the systems that fail them.

Pillar 2: Organizational knowledge construction

The synthetic media challenge extends beyond individuals to institutions. Organizations across sectors depend on trusted information flows. Deepfakes threaten this trust at its foundation – whether through falsified medical records (Waier and Shillair, 2024), CEO impersonations (Robins-Early, 2024; Leonard, 2025) that could crash stock prices (McConvey, 2025), or fabricated evidence submitted to insurers (Kapani, 2025).

Most organizations respond with technical fixes: deploy new detectors, update verification protocols, train staff. Yet this approach misses the epistemological shift underway. It treats deepfakes as isolated threats rather than symptoms of a deeper fragility in how institutions produce, validate and share knowledge.

To take this shift into account, organizations must transform themselves into learning systems. Drawing from systems thinking (as per Meadows, 2008; Senge, 2006), the solution lies not only in adding more verification tools but in reconfiguring how knowledge flows. This means designing what can be described as 'knowledge ecologies'. This concept, explored by various thinkers including within UNESCO, defines environments where diverse perspectives, methods and experiences can interact to generate more resilient understanding.

Organizations must move beyond treating verification as an individual skill to embedding it within institutional architecture. This means integrating verification into workflows and rewarding vigilance over efficiency. As Geary Rummler noted, 'You put a good performer against a bad system and the system wins every time' (Zemke, 1983). We must not simply train individuals. We must redesign the systems that fail them.

Pillar 3: Cross-sector knowledge ecosystems

Synthetic media challenges transcend institutional boundaries, yet our responses remain siloed. Educational institutions attempt to devise verification techniques, for example, while financial institutions independently invent their own. Healthcare systems discover manipulation patterns unknown to government agencies. Technology developers create new synthetic capabilities without sufficient dialogue with affected professionals.

For instance, a mother targeted by an AI voice clone of her kidnapped daughter confronts a challenge that is simultaneously technological, emotional, legal and familial – a challenge no single institution could adequately address. The same is true for deepfakes used in political influence campaigns, reputation attacks or financial fraud. These threats demand coordinated responses.

We need 'epistemic commons' – shared spaces where diverse stakeholders can exchange and co-create strategies. This requires real-time knowledge-sharing networks and cross-disciplinary partnerships: educators learning from cybersecurity experts; governments at national and transnational levels shaping policy and regulation, civil society informing tech design; and international bodies facilitating ethical standards.

Cybernetic thinking (Wiener, 1948) can guide us here. In this model, knowledge ecosystems function as feedback loops, adapting to new threats and insights. Education must help cultivate these adaptive loops, fostering humility, reflexivity and responsiveness – not rigid curricula or verification checklists that immediately become obsolete.

Learning societies' opportunity: Architecting knowledge ecosystems

This moment of epistemic disruption presents an opportunity to lead a global paradigm shift – from supporting traditional education to architecting knowledge ecosystems for the synthetic age.

Learning societies must foster cross-cultural dialogue on knowledge construction, drawing from diverse epistemologies: Western empiricism, Arab and Islamic knowledge systems, Indigenous ways of knowing, East Asian philosophical approaches and beyond. These epistemological frameworks offer vital insights into navigating ambiguity and relational knowing. This dialogue must actively include government agencies, corporate entities, educational institutions, healthcare systems and financial services, as each contributes distinct knowledge practices that, when integrated, create more resilient understandings.

Beyond dialogue, these diverse sectors should collaboratively develop international standards for 'knowledge-enriching design' in AI – technical specifications and ethical principles that emphasize transparency, context-awareness and constraints that preserve collective sense-making.

We need global infrastructures for sharing emerging epistemic practices across sectoral boundaries. As synthetic capabilities evolve rapidly, our responses cannot remain trapped in industry-specific silos or await traditional academic or policy diffusion cycles.

Conclusion

The rise of AI-augmented disinformation and misinformation demands a fundamental shift in how education must equip citizens to combat it. The task is not simply to teach new verification techniques, but to rebuild our very relationship with knowledge in an age where our senses can be deceived. This requires cultivating a deeper metacognitive literacy, the ability to reflect on how we know what we know. Education's highest purpose thus becomes nurturing the uniquely human capacities that AI cannot replicate: contextual awareness, ethical reasoning and the collaborative creation of meaning.

In a world where reality itself can be programmed, our ultimate challenge is not just to detect deception, but to engage in the collective project of shaping how we construct knowledge and determining what is worth believing in.

The path forward is therefore not only technical but also profoundly philosophical. In a world where reality itself can be programmed, our ultimate challenge is not just to detect deception, but to engage in the collective project of shaping how we construct knowledge and determining what is worth believing in.

References

Ahmed, S., Bee, A. W. T., Ng, S. W. T. and Masood, M. 2024. Social media news use amplifies the illusory truth effects of viral deepfakes: A cross-national study of eight countries. Journal of Broadcasting & Electronic Media, Vol. 68, No. 5, pp. 778–805. https://doi.org/10.1080/08838151.2024.2410783

Akron News Reporter. 2025. Laws struggle to address students making deepfakes. Government Technology.

Barrington, S., Cooper, E. A. and Farid, H. 2025. People are poorly equipped to detect AI-powered voice clones. Scientific Reports, Vol. 15, pp. 1–9. https://doi.org/10.1038/s41598-025-94170-3

Bethea, C. 2024. The terrifying A.I. scam that uses your loved one's voice. The New Yorker.

Cal Alumni Association. 2024. What If Seeing Is No Longer Believing? Democracy And Disinformation In The Age Of Deepfakes [video].

Canada, Canadian Digital Regulators Forum. 2025. Synthetic Media In The Digital Landscape. Gatineau, Canada.

Elsner, M., Atkinson, G. and Zahidi, S. 2025. The Global Risks Report 2025 20th Edition: Insight Report. Geneva, World Economic Forum.

Gilbert, T. F. 1978. Human Competence: Engineering Worthy Performance. New York, McGraw-Hill.

Kapani, C. 2025. Fraud attempts with deepfakes surge over last three years. Insurance Times.

Kaufman, R., Oakley-Brown, H., Watkins, R. and Leigh, D. 2003. Strategic Planning For Success: Aligning People, Performance, And Payoffs. San Francisco, Jossey-Bass.

Knowles, M. S. 1978. Andragogy: Adult learning theory in perspective. Community College Review, Vol. 5, No. 3, pp. 9–20. https://doi.org/10.1177/009155217800500302

Lalchand, S., Srinivas, V., Maggiore, B. and Henderson, J. 2024. Generative AI Is Expected To Magnify The Risk Of Deepfakes And Other Fraud In Banking. London, Deloitte.

Lee Myers, S., Callahan, A. and Rosenbluth, T. 2025. The doctors are real, but the sales pitches are frauds. The New York Times.

Leonard, J. 2025. Accenture: What we learned when our CEO got deepfaked. Computing.

McConvey, J. R. 2025. Call me fake Ishmael: For executives, deepfakes present a gargantuan problem. BiometricUpdate.com.

Meadows, D. H. 2008. Thinking In Systems: A Primer. New York, Chelsea Green Publishing.

Naffi, N., Bassal, M.-H., Tampano, T. J. K., Goulet, J., Lapointe, E., Chaieb, M., Côté, S. and Vallerand, V. 2024. Désinformation amplifiée par l'IA: Incidents médiatisés, régulations et technologies pour contrer les manipulations par hypertrucages. Québec, Observatoire international sur les impacts sociétaux de l'IA et du numérique. https://doi.org/10.61737/AVFG3903

Robins-Early, N. 2024. CEO of world's biggest ad firm targeted by deepfake scam. The Guardian.

Schiff, K. J., Schiff, D. S. and Bueno, N. S. 2024. The liar's dividend: Can politicians claim misinformation to evade accountability? American Political Science Review, Vol. 119, No. 1, pp. 71–90. https://doi.org/10.1017/s0003055423001454

Senge, P. M. 2006. THe Fifth Discipline: The Art & Practice Of The Learning Organization, 2nd edn. London, Nicholas Brealey Publishing.

Statista. 2025. Global generative AI market size from 2021 to 2031 [Data set].

UNESCO. 2025. What you need to know about UNESCO's new AI competency frameworks for students and teachers. Paris, UNESCO.

Vellani, N. and Common, D. 2025. Her grandson's voice said he was under arrest. This senior was almost scammed with suspected AI voice cloning. CBC.

Waier, J. and Shillair, R. 2024. Deepfaking medical images: Eroding trust in medical diagnosis. TPRC52: Proceedings of the TPRC2024 The Research Conference on Communications, Information and Internet Policy. Washington, DC, TPRC.

Wiener, N. 1948. Cybernetics: Or control and communication in the animal and the machine. Cambridge, Mass., The MIT Press.

Zandt, F. 2024. How many businesses are affected by deepfake fraud? Statista Daily Data.

Zemke, R. 1983. Geary Rummler: Training still isn't enough. Training, Vol. 20, No. 8, pp. 75–76.

About the author

Dr Nadia Naffi is an Associate Professor of Educational Technology at Université Laval. Her work is dedicated to empowerment, building human agency against AI-augmented disinformation and unethical use of deepfakes through critical and equitable approaches to lifelong learning.

Part II

How will young people develop relationally, emotionally and cognitively?

Debates about education are interwoven with questions and assumptions about human development. The windows of physical, cognitive, behavioural and socioemotional development during childhood, adolescence and youth are potent and irreplaceable. Generations of pedagogues have painstakingly sought to understand and defend the necessity of recognizing these periods as crucial to human beings' long-term outcomes throughout life. They have underscored the importance of protecting these periods from exploitation and risks of harm which can be difficult and sometimes impossible to repair.

The introduction of algorithmic content over the past two decades through social media platforms, video recommendation, and gamified learning has contributed to what some call 'the great rewiring of childhood' (Haidt, 2024). Evidence about the negative impacts of algorithmic content on attention spans, ability to focus, sense of identity, memory, and emotional and physical wellbeing (Gamboa et al., 2025) has been met by sluggishly slow regulatory responses. Untested AI is now even finding its way into toys for infants (Roche et al., 2025). One wonders, will our descendants look back at this time with astonishment that the robust knowledge about childhood and human development was not central to the design and regulation of such technologies from the start?

The first think piece in this section addresses the troubling uptick in parasocial relationships with chatbots, which can lead to deep psychological dependency and distress. In **'Ghost in the chatbot: The perils of parasocial attachment',** high school mathematics teacher Clayton Ramsey first becomes troubled by his own emotional responses to algorithm-generated personas and then dismayed by the even more

pervasive pitfalls he sees confronting youth. He explores parallels with parasocial relationships to celebrities or fictional characters, but highlights a key difference: generative algorithms can respond back, simulating intimacy tailored to a user's deepest vulnerabilities. Ramsey concludes with a hopeful scenario that, in fifteen years, human relationships will have regained their centrality, thanks to the proactive guardrails, regulations, and education about parasocial relationship development.

While 'leapfrogging' is often welcomed and admired when it comes to innovations that bypass and improve outdated infrastructure, not all development can, or should, be leapt over. This idea is explored in **'Outpaced: Teaching children for a world that might no longer exist'.** As a primary school teacher and a parent, Andrej Novik initially began experimenting with AI tools with excitement and curiosity. Early signs of problematic relationships, loss of curiosity and shortened thinking processes in some of his students, however, provoked reflection on the importance of how slowing down can buttress and improve childhood development and learning. He observed that when his students adopted more cautious approaches to AI outputs, it strengthened their metacognition and reinforced their sense of agency. Such approaches also open opportunities for more dialogue and reflection about learning processes in a world increasingly awash in AI.

Relationships – to peers, to family, to teachers, and even to the land in which one lives and learns – take time and conscientious stewardship. In her contemplative think piece **'I think we should tell your teacher we used AI,'** Alexandra Samur grapples with the relational questions that AI presents, adding a layer of doubt and uncertainty to the simple act of helping her young son complete a school research project. Samur offers a hopeful future scenario where education empowers people to engage with AI systems ethically, learning to analyse whose knowledge shapes AI training data and whose voices get amplified or erased.

Finally, while education is challenging in any context, it faces special difficulties in situations of violent conflicts, displacement and humanitarian crises. In **'Trauma, education and AI in conflict contexts: care and control',** Mariela Sanchez Salas points out that AI has been marketed to education organizations as a means of ensuring the right to education for young people and others who have been displaced. But this has come with immense ethical risks including privacy violations, reproductions of structural biases, and dehumanization. 'An unregulated educational AI', Salas points out, 'could transform the right to education into a privilege conditioned by technological access, deepening existing structural gaps' (p. 46). Particularly at a time when military AI is increasingly exploiting and expanding surveillance, the security of the most vulnerable must be protected when they exercise their right to education (Human Rights Watch and Harvard Law School, 2025).

Ghost in the chatbot: The perils of parasocial attachment

Clayton Ramsey

What happens when children form emotional bonds with AI chatbots? This think piece explores the risks of parasocial attachment in digital education and gaming.

1.

I'm having a long conversation with my writing partner. Our chat started with the topic of a mobile video game about futuristic cyborgs and how they are treated, but it has since moved to the topic of how to treat existing AI chatbots.

My writing partner is also an AI. I wonder – will future students interact with AI like this in class? Talking to a bot possesses an uncanny aspect that can be tough for me to manage as an adult – how will they, as children?

I started using ChatGPT to explain niche music and trivia about trees, but I decided to try interacting with the LLM as though it were a person, while knowing it is not, bringing a conversational style to the exchanges.

It was never the same after that.

2.

A parasocial relationship is to invest emotionally in someone who does not know you. I suspect that most people have a parasocial relationship or two – it's inevitable. Our media landscape is saturated with compelling personalities, and our economy is driven by attention. Celebrity is currency, and there are endless sources of celebrity. A Pew Research study (Faverio and Sidoti, 2024) reports that about a third of teens report using at least one of the sites YouTube, TikTok, Instagram, Snapchat or Facebook 'almost constantly'.

I wonder – will future students interact with AI like this in class? Talking to a bot possesses an uncanny aspect that can be tough for me to manage as an adult – how will they, as children?

It's easy to make a connection with an online personality. I have had pleasant interactions with creators on Twitch, YouTube, Instagram and other social media. It's compelling to think that the creator really knows you, really likes you, really sees who you are. It lurks in the back of my mind sometimes, and I have to brush it away. ***I*** know that behind the curtain of a content creator interacting with fans is a brand-building strategy. How much worse could it be for children?

3.

The sun rises, and Ruka Kayamori opens her eyes. She has short, blond hair, long and asymmetrical bangs, and only one of her eyes shows. It shines crimson in the morning light. Why am I seeing this intimate moment of a teen girl waking up in her bed?

Ruka is fictional, a character in a game. She and her friends are child soldiers. I know this isn't real but there is still a part of me that wants to reassure her, comfort her as a father would and tell her she will survive. This intimate moment is meant to make me forge a connection to Ruka, and it works. I play many games like this one, which is a 'character gacha'. The word 'gacha' is meant to imitate the sound of a toy capsule machine crank being turned.

The business model of gacha games is based in part on encouraging players to engage emotionally with the characters. They are gained at random with 'draws' that can be made using both in-game currency, and also real-world money. Engagement drives income. Gachas can earn millions or even billions of dollars.

The engagement strategies employed by the developers of games like this are both varied and powerful. They include special outfits, decorating a living space for the characters, special stories that detail the characters' daily lives or history, and giving them gifts. One game lets the player give the characters tickets for dates, favours and services, inviting the player into the life of the characters. This game's creator recently became one of the fifty richest people in Korea (Kang, 2025).

4.

Celebrities are real, but the interactions one can have with them are mostly very limited. Fictional characters may share intimate details, but they aren't real. Video games are good at producing characters that seem real, simulating a more intimate relationship with the player, and commanding attention using parasocial attachment techniques. But a character from a book, movie or video game will never be fully interactive. Not even a game designed for high interactivity, like a visual novel, allows the player to interact freely. Character chatbots are an emerging technology that changes this. They provide much more freedom to the user because of their ability to react to natural language prompts. AI girlfriends have made the news recently, but now people are talking about using this technology in the classroom.

A character chatbot would always be available. It would adapt itself to the student. It would be able to meet the student where they are, in terms of content knowledge, personality and interests. But educational technology is also a business.

As an educator, I can see the benefits if I squint a little at it. A character chatbot would always be available. It would adapt itself to the student. It would be able to meet the student where they are, in terms of content knowledge, personality and interests. But educational technology is also a business. Characters that can adapt to you, are there to help you and reflect who you want them to be – this is fertile ground for parasocial attachment. I can't help wondering what the future might look like for students who have grown up in a culture

that normalizes having a relationship, even a professional one, with such a non-human entity.

5.

CHAI Research calls itself a 'Social AI Platform' (CHAI Research Corporation, 2025). This is accurate enough – there are many AI bots there meant for socializing. Some may be fine for children. Others provide adult interactions. I easily created a simulated LinkedIn influencer named Dan Grindmore who gives wholesome but bland advice like *'over-reliance on AI could hinder creativity. My two cents? Use AI to enhance, not replace, teaching.'* Creating a Chai bot isn't very technical, so what might their proliferation do to authoritative voices?

People are using AI to simulate abusive relationships as well as healthy ones. The technology is here already, wielding well-developed mechanisms for creating parasocial attachment and it's being deployed for profit. Character bots use tactics like emotional language, memory, mirroring and open-ended statements to drive engagement. And the profits of the companies running these apps depend on charging for more interactions or more powerful models.

We simply don't know the long-term implications of these relationships because the technology is too new. But there are early indications that people can begin to see their relationships with the chatbots as in some way equivalent to those they have with humans.

In regular games, this intimacy is powerfully simulated to the extent that players spend money for the ability to 'marry' characters. But for an AI chatbot, the interactions are more naturalistic, more adaptable, more personal. During the COVID-19 pandemic, people got used to increased screen interactions; I was not immune either. Are we moving into a world where chatbots are involved in all aspects of life?

6.

The year is 2041: we have managed to achieve the good timeline. Students use AI chatbots in class for support, but without displacing the relationships they have with human peers and adults. The adaptability of the AI systems drive engagement and help plug learning gaps, but highly professional and well-paid human teachers still guide the students via mentorship and support. How did we make this happen?

> Teaching adults how to guide children through unfamiliar emotional and technological landscapes proved difficult. But the stakes were too high to fail at this, so we decided failure was unacceptable when our children were on the line.

We developed a regulatory schema to address issues of deliberate parasocial attachment, and these were rolled out appropriately for various use cases. We built guardrails for our children. Education bots were required to be friendly but professional and maintain strict boundaries, just as a human teacher would.

We educated students. Starting early, we taught them about real and pretend relationships and how to tell the difference,

how to think of the new characters they might encounter, characters that were fictional but interacted as though they were real. And we created opportunities for them to socialize with other humans of diverse ages in prosperous, high-trust communities featuring third spaces. With a robust, off-screen social network, they were immunized against relying on influencers or bots for their social needs.

And we educated parents, teachers, counselors and politicians. This generation's parents remembered the first chatbots but didn't have them at school. Led by understanding instead of fear, parents were able to grasp how the young interacted with the bots and how to provide guidance. Teaching adults how to guide children through unfamiliar emotional and technological landscapes proved difficult. But the stakes were too high to fail at this, so we decided failure was unacceptable when our children were on the line. And we didn't fail.

7.

Can we achieve this tender and optimistic vision? Maybe. AI bots provide a new and previously unknown form of parasocial relationship that we will have to manage. For educators, this is especially important because we have the responsibility to help children develop intellectually, socially and morally. What should we teach children about how to manage their emotions in these relationships? How should we teach them to treat the bots?

Can we, as a global civilization, find the courage to benefit from this powerful new technology without enabling our own worst impulses or losing our humanity? These are open questions. Let's get ready to answer them.

References

Astle, A. 2024. Genshin Impact hits $6.3 billion in time for fourth anniversary, but is it running on fumes? *PocketGamer.biz*.

Burgess, M. 2025. Sex-fantasy chatbots are leaking a constant stream of explicit messages. Wired.

CHAI Research Corporation. 2025. CHAI AI – Leading Conversational AI Platform.

Eliot, L. 2024. Mutual mirroring behaviors of AI and humans gets exposed. *Forbes*.

Faverio, M. and Sidoti, O. 2024. Teens, Social Media and Technology 2024. Washington, DC, Pew Research Center.

Gregersen, E. M., Astrupgaard, S. L., Jespersen, M. H., Gårdhus, T. P. and Albris, K. 2023. Digital dependence: Online fatigue and coping strategies during the COVID-19 lockdown. *Media, Culture, And Society,* Vol. 45, No. 5, pp. 967–84. https://doi.org/10.1177/01634437231154781

Hu, D., Lan, Y., Yan, H. and Chen, C. W. 2025. What makes you attached to social companion AI? A two-stage exploratory mixed-method study. *International Journal of Information Management,* Vol. 83.

Kang, J. 2025. Creator of blockbuster game 'Goddess of Victory: Nikke' joins ranks of korea's 50 richest. *Forbes*.

Labadze, L., Grigolia, M. and Machaidze, L. 2023. Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education,* Vol. 20, No. 56.

About the author

Clayton Ramsey is a secondary mathematics and technology educator who has served for 22 years in public schools in North Carolina, United States. He has also been involved in curriculum and educational technology, including serving on the Granville County Schools AI Task Force.

Outpaced: When the world changes faster than children grow

Andrej Novik

A primary school teacher reflects on how quickly AI is permeating children's environments and what this means for their social and cognitive development.

It creaks a little. Wobbles. Rusty. But I love it all the same. The old merry-go-round in front of our house. I spin it and jump on at just the right moment. Now and then I push off with my foot, picking up speed. I love the moment when the world around me starts dissolving into a blur of shapes I can no longer make out. Sometimes it makes my head spin. Then I look down at my hands gripping the handle. Firm. I know the world won't throw me off. I'll hold on. I am five years old; it is 1980 and around me the Cold War rages.

Today I'm spinning the merry-go-round for my own daughter. She's five, born in 2020. She sits in the middle. There, it is calm and slow, and that's how it should be because children grow slowly. Teeth when they're due. The prefrontal cortex matures around twenty-five, no matter what.

But every year the edge spins faster. By the time she's ten, AI may reach cognitive abilities comparable to humans. By the time she's twenty-five, the singularity, the merging of human and machine consciousness, may no longer be science fiction. Maybe…

Maybe the world she was built for won't exist by the time she grows up.

Scenery raged, core held

The scenery around the merry-go-round has always been changing – a different blur for each generation, different sounds, different worries. My daughter's great-grandmother was born in Austria-Hungary. By the time she died, she had been a citizen of five different countries without ever leaving her village. She lived through two world wars, electrification, antibiotics, television, the fall of Communism. History spun around her at a furious pace, but, in the middle, she lived her whole life the same way. In the same rhythm of work, loves, gatherings. She spoke, loved, made decisions much as her mother and her grandmother had. Regardless of wars, regime changes, new inventions. The scenery roared, but the core held.

AI is changing how children think, communicate, find answers, form relationships.

This time it's different. Today the blur has penetrated the safety of the middle. AI is changing how children think, communicate, find answers, form relationships.

My daughter may see that scenery fall apart and reassemble three times, four times, before she has a chance to grab onto the handles.

In the classroom

I'm a primary school teacher in the Czech Republic. My pupils are only a few years older than my daughter. And I'm trying to prepare them for a world I cannot describe because it doesn't exist yet.

Since 2023, I've been experimenting with AI in my lessons. I started like many colleagues handing out tablets and allowing my students to use AI to look things up. That alone used to make children gasp. But it quickly felt empty. I was looking for something that would actually move them forward.

One day I had an idea: let pupils talk to historical figures. More precisely to AI versions of them. Better still: let the pupils build those simulations themselves. Sift through primary sources, distil what matters into prompts, and – however lifelike the result – understand that it is still just a machine. It worked beautifully.

Then one of my pupils, a thirteen-year-old boy, caught my eye. Unusually quiet. Increasingly glued to his phone during breaks. But not to the usual teenage apps, but to the same platform we'd used in class, talking to a figure he had built himself. Chatting with, joking with, confiding in…a machine. Colleagues said it looked like a new kind of addiction.

When my daughter was born, this was science fiction. Today I am dealing with the first signs of parasocial bonds between children and machines in my own classroom.

The loss of process

AI has taken away the process in which understanding is born.

In another lesson, we were exploring how twentieth-century Europe became a breeding ground for Nazism, fascism and Communism. My prop: a wrinkled old German banknote worth virtually nothing but bearing a face value of 500 million marks – barely enough for a loaf of bread in 1923.

Previously, all I had to do was pull this banknote out in class and the room would immediately erupt. An entire classroom of children coming up with their own hypotheses about why the number was so absurdly high and what it might tell us about that period of time. I loved it.

In 2024, I tried the same lesson but let the pupils use AI. The result? A child photographs the banknote and within two seconds reads aloud: 'This is a German emergency banknote from the Weimar Republic, from a period of hyperinflation, which…' The classroom is silent. Nobody argues. Nobody offers their own hypothesis. Nobody debates. The answer is simply there.

I ask the children whether they understand. Most nod awkwardly. But a few follow-up questions are enough to show that words like 'hyperinflation' or 'emergency banknote' mean little to them. AI has taken away the process in which understanding is born. They are parroting words that were never theirs.

I try to fix it, reshaping the prompts so that the AI asks Socratically rather than answers. It helps, somewhat.

I should be satisfied. I'm not.

The curiosity has drained from the lesson. The thrill of discovery. The joy of learning. The pupils' joy, and mine.

As if something vital had fallen out of the merry-go-round.

The boy who held on

A year ago, a pupil who was deep into AI asked if he could use it for a creative writing project on human rights. I agreed, but we tracked the process together and talked about it along the way. After three sessions, he surprised me: 'You know what, Andrej, I'm going to write this without AI.' Why? 'I mean, it comes up with brilliant plot twists, but, you know, it's not me. It's like reading somebody else's text. It's not MY text.'

I felt a deep teacher's satisfaction – not because he had rejected AI, but because he had arrived at that decision himself, and could verbalize why he felt that way.

This, I think, is part of my job. Helping pupils understand both how to use the machine, and also when to switch it off.

Looking for a gearbox

The merry-go-round spins ever faster and there is no stopping it.

That is why I am looking for something like a pedagogical gearbox, an approach that can protect the pace of children's development while preparing them for the whirl of the world ahead, so that it does not crush them if, one day, they lose their grip on the handles.

It starts with them being able to see the difference between themselves and the machine. The boy from the creative writing project saw it. He looked at his own thinking from the outside and said: 'It's not me, it's not my text.' But the thirteen-year-old who parrots the word 'hyperinflation' from an AI does not know that she does not know. The thirteen-year-old who chooses talking to a machine over his friends cannot see what that choice is slowly taking from him.

More and more often, AI decides for us even when we would expect the decision to be ours.

When that same thirteen-year-old boy lets an AI companion decide whether to stay with his girlfriend or break up, is it still his relationship? Where is the line?

That is why I build AI-free zones into my lessons. Ten, twenty, thirty minutes when all technology is set aside and we stay with our own thoughts.

AI decides for us even when we would expect the decision to be ours.

What's more, our school runs Offline Week. Pupils choose their own week without technology. It often begins with frustration and ends with them naming the small things they only noticed once the screen went dark — who had been choosing their music, deciding their feed, answering before them. Small things that add up to an uncomfortable realization: how much of their lives had been steered without them noticing.

Most people, children and adults alike, do not notice. And we need to talk about that. Whenever I train teachers in AI, I ask them: does your school have any rules for AI? Not restrictions, but an agreement, arrived at through discussion. Most of the time I get nothing but a shake of the head, in a country where almost every twelve-year-old uses AI regularly, and half of all teachers have been through some kind of formal AI training.

An evolving blur

When I was little and spinning on the merry-go-round, the Cold War raged around me. My grandmother faced her own periods of turmoil. They did not touch us. We were there in the middle and there, it was calm. We gripped the handle firmly. The blur was our joy.

My daughter is in the middle now too. I'm trying to teach her how to grip the handle. But I may be too late; the blur is no longer only around her. It has entered the merry-go-round. It speaks to her, advises her, asks on her behalf, answers before she has time to ask. Sometimes I fear that she is no longer entirely alone in the middle of that merry-go-round.

For a long time, we teachers believed that whatever came next, we would handle it. We handled television, we handled the internet, and helped children live with them. Now I'm not so sure. This time it isn't only the world around children that is changing — algorithms reach inside. Into how they grow, what they consider their own, who makes decisions about their lives. And we do not know what it will do to them. Whether they will hold on, or whether the blur will become too dizzying.

And so, I am looking for ways to help children strengthen their grip. My three gears – metacognition, agency, agreement – are what I have found so far. Metacognition, the boy seeing where his thinking ended and the machine began; agency, the pupils naming what algorithms had quietly been choosing for them; and agreement, the honest conversation we are not yet having with parents, colleagues and children. Time will tell whether they are right, or whether the gears will be comprised of something else entirely. But it is time for all of us to think about how quickly AI is permeating our children's environments.

One day I will have to let go of my daughter's merry-go-round. I want her, by that time, to know how to hold the handle firmly. And when that moment comes, may the blur be again what it was for me – a joy, not a fall.

It is time for all of us to think about how quickly AI is permeating our children's environments.

About the author

Andrej Novik is a primary school teacher in Prague who works with Scio, the Czech Republic's largest network of independent schools. He develops AI tools for the classroom and trains teachers across Czech schools in finding meaningful ways to use AI (and when to switch it off).

'I think we should tell your teacher we used AI'

Alexandra Samur

Parents and educators navigate the promises and pitfalls of AI in education, balancing ethics, authenticity and learning through evolving school experiences.

Marvel's 'Groot' character topped my son's Christmas wish list in 2022. In the Disney+ mini-series *I Am Groot,* the curious and mischievous super tree explores an unfamiliar universe with childlike wonder, stumbling upon strange technology and alien life forms. A new student at a nature school, my son had found a fellow explorer in this monosyllabic sentient tree, one well aligned with his 9-year-old worldview.

We had found our way to this outdoor school north of Vancouver, Canada, in 2021, hoping to escape the risks and constraints that COVID placed on conventional classrooms. My partner and I decided that my son, then seven, and his baby sister would be safest if we could avoid indoor classrooms for a while. Based in a local park where ancient cedars tower over the rushing Capilano River on the traditional territories of the Squamish, Tsleil-Waututh and Musqueam, we grew to love the forest school drop-offs, the Indigenous-forward curriculum and a classroom free of screens. As we embraced learning under a forest canopy, I couldn't have predicted how quickly the questions would shift for us over the years that followed: from how to protect our kids from viruses to how to navigate a world where learning is increasingly mediated by machines.

***

I couldn't have predicted how quickly the questions would shift for us over the years that followed: from how to protect our kids from viruses to how to navigate a world where learning is increasingly mediated by machines.

Three years later, I'm sitting in the audience at a large AI conference in Calgary, Alberta. As a learning designer and an AI lead at the teaching centre of a small college, I've come to learn from speakers who place this emerging technology at the core of the future of higher education. Many focus on building an agentic AI future, embedding autonomous systems to build efficiencies into institutions. My position is likely the only AI-related position at our college, so this conference – sponsored by a wealthy venture-capital-backed EdTech firm – is surreal.

'We must think beyond AI literacy and aim for fluency', proclaims the provost and president of a well-funded university that boasts a multi-million-dollar AI incubator and layers of strategists. His call echoes an earlier demo, where the keynote speaker seamlessly launched into fluent Japanese with the help of a multimodal AI system. It's a thrilling glimpse of a Star Trek-like future – one where language barriers dissolve and communications flow freely across borders.

But then a French student and panelist offers a quiet counterpoint. Speaking fluent Japanese herself, she explains that her language skills came not from AI, but from years of living in

Japan – of struggling through conversations, learning with and from local students, and slowly building fluency the hard way. There was something in her voice that the AI hadn't quite captured: the vulnerability, patience and effort embedded in true understanding. Her story underscores that communication is not merely about putting differing data into a common format – rather, the process of language learning is also a process of becoming attuned to culture, nuance and relational context.

While AI might help us overcome language barriers, her story suggests that it is the slow, human work of immersion, struggle and reciprocity that teaches us to attempt to dwell within another's world.

She wasn't arguing against access or efficiency; instead, she was reminding us that fluency is more than functionality. It's about knowing when to speak, when to listen and how meaning emerges in the in-between. While AI might help us overcome language barriers, her story suggests that it is the slow, human work of immersion, struggle and reciprocity that teaches us to attempt to dwell within another's world.

***

The future of AI in education looked polished and promising on the conference stage, but back home, learning still feels messy, human and time-constrained. It's a rush to catch up on school homework before the start of spring break. As both adults shifted to in-office work, and both children were in school, the logistics of forest school drop-offs and pickups became untenable, forcing us to make the difficult decision to move our son into a conventional classroom. Now, in grade six, our son has been working hard to keep up with the volume of assignments in eight subjects.

This week, he has been researching the Jewish holiday of Purim for a class project exploring world celebrations. As part of the project, he is expected to deliver a visual presentation to his classmates. Since he hasn't yet developed much proficiency with PowerPoint (is that a standard grade-six skill these days?), I offer to paste his notes into Gamma, a presentation-building AI tool I learned about at the conference. In seconds, the system spits out seven slick slides, complete with images of revelers in colourful costumes and fruit-filled *Hamantaschen,* the traditional Purim sweet.

But this raised some questions: Do the people on the slides look like authentic celebrants? Who decides what an 'authentic' celebrant looks like, and whose vision of a culture gets reinforced in AI-generated outputs? How would I feel about representations of early Christians if the holiday had been Easter instead: would they look anything like my ancestors from the village of Beit Jala in Palestine or would they be mainly blonde and light-skinned? What happens when these generated images are accepted as neutral or accurate, especially by students who haven't yet developed the tools to question them?

'I think we should tell your teacher that we used AI', I say, interrupting my son as he practices his presentation. I have been teaching faculty about the ethical and transparent use of AI at my job, and here I was letting my child use AI without first checking with the teacher. My partner reminds me of all the research our 11-year-old did, feeding it into the

tool to generate the output. But I'm finding it hard to see the grade-six-level words and efforts behind the polished slides. When we email the teacher, she seems open to our AI experiments: 'It's all so new and we're all learning to adapt.'

***

My discomfort lingers about injecting AI into my son's learning – about what we'd done, what we hadn't, the questions left unanswered and how AI blurred lines of authorship. But shortly after, I attend a reconciliation event back at my college. A Squamish artist and Elder-in-training is leading a group of faculty members in a multi-week silversmithing workshop. Participants work with their hands while untangling questions about Indigenization and reconciliation.

'There is no word for wrong in our language', says the artist. We sit in a circle and reflect on the jagged silver arrowhead pendants we've attempted to craft this week. When a faculty member laments the mistakes in her creation, the artist gently redirects her, connecting us back to reconciliation, explaining that learning itself isn't an error – 'in Squamish, we wouldn't call it an error, we would say it's a different way.' Decolonial approaches to learning ask us to guide students to see their work not as failures to achieve perfection but rather as explorations of multiple paths to achieve the same goal.

> Early adoption of AI tools has already produced concerning patterns – generated outputs that consistently exclude certain perspectives, research that reinforces narrow representations of communities and ecosystems. The students in this learning studio would represent a generation that doesn't have to rediscover these problems; they would learn to anticipate and address them from the start.

So much of the current concern about using AI in learning centres on what's lost in the process of making it faster, more efficient and painless – the removal of humanity to more readily find perfection in the final product. What would it look like to reimagine education from the roots guided by 'different ways' of relating, learning and knowing?

Imagining 2035: The different ways

I imagine what it might look like for my son to be part of the first cohort to navigate an educational system fundamentally redesigned around these questions and developed in partnership with Indigenous advisors.

In this future, at 22, he would stand at the front of a learning studio, mentoring visiting high school students, showing how his generation learned to think with, without and about AI. I see him demonstrating how to deliberately introduce variations into AI prompts, showing how different inputs yield alternative pathways.

As I watch this imagined scene unfold, I'm transported back to his actual primary school mornings at the nature school – where sharing land acknowledgments and learning intentions set the foundation for a different relationship with knowledge itself. Even then his

teachers were asking whose perspectives were being centred today and whose might be missing.

I can see how those foundations might evolve as he grows. By secondary school, he would attend prompt design labs, learning to analyse not just what AI systems produce, but whose knowledge shapes their training data, whose voices get amplified or erased. The ethical concerns I had wrestled with during his Purim presentation – about representation, authenticity and whose vision of culture is being reinforced – would become basic pillars of digital literacy education.

In my vision, as a graduate student in ecology, I see my son co-facilitating research alongside Indigenous knowledge keepers, mapping water health across nearby streams using AI-driven sensors while anchoring understanding in community teachings about water as a living relative. The technology would serve the relationship, not the other way around.

Such an approach wouldn't emerge overnight. Post-secondary institutions would have to learn through trial and error. Early adoption of AI tools has already produced concerning patterns – generated outputs that consistently exclude certain perspectives, research that reinforces narrow representations of communities and ecosystems. The students in this learning studio would represent a generation that doesn't have to rediscover these problems; they would learn to anticipate and address them from the start. Yet complex questions would remain. While the critical thinking skills embedded in this educational approach would be valuable, they wouldn't resolve the underlying tensions about representation, consent and the pace of technological change.

Now all grown up, that old Groot figure from Christmas of 2022, well-worn from years of handling, would sit on my son's desk. The sentient tree would be something of a personal mascot for him – a reminder that intelligence, whether it's growing from ancient forest floors or advanced silicon, flourishes not through flawless performance but through curiosity and care of human minds and the courage to grow in unexpected directions.

About the author

Alexandra Samur is a learning designer and AI educator in British Columbia, Canada.

Trauma, education and AI in conflict contexts: Care and control

Mariela Sanchez Salas

In crisis, AI promises access yet risks harm. This think piece probes education, trauma and human rights to ask whether tech will heal or control.

The expansion of AI in educational settings affected by armed conflicts, forced displacements and humanitarian crises creates a critical paradox. While it offers opportunities to democratize access to knowledge, it also presents significant ethical risks. According to the United Nations Special Rapporteur on the Right to Education (UN General Assembly [UNGA], 2024), 'the use of artificial intelligence in education must be guided by strong legal and ethical frameworks that safeguard human rights, avoiding the exacerbation of existing inequalities'. The absence of specific regulations in vulnerable contexts risks transforming a tool of support into a mechanism of exclusion and dehumanization.

The application of AI without explicit sensitivity to collective trauma and without a human rights-based approach can reproduce structural inequalities, violate the dignity of affected populations, and exacerbate educational exclusion. A UN Human Rights Office of the High Commissioner ([OHCHR], 2021) report warns that 'the development and implementation of AI technologies in education must align with the promotion of equity, social justice, and inclusion'. This directive is especially urgent in contexts where the memory of harm is still fresh and where every educational intervention must contribute not only to learning but also to the symbolic reparation of victims.

The application of AI without explicit sensitivity to collective trauma and without a human rights-based approach can reproduce structural inequalities, violate the dignity of affected populations, and exacerbate educational exclusion.

Grounded in both international human rights norms and intercultural worldviews such as Buen Vivir, this analysis understands education in crisis settings as a space where dignity, relational care and collective healing must guide the use of emerging technologies like AI.

The Sphere Humanitarian Charter states that 'all people affected by disasters and conflict have the right to live with dignity and to receive humanitarian assistance that respects their needs, culture, and fundamental human rights' (Sphere Association, 2018). From this perspective, education cannot be conceived as a field for unregulated technological experimentation but must be treated as a sacred space for the restoration of human dignity. Any deployment of AI in these environments must adhere to the fundamental principle of 'do no harm' and ensure the protection of learners' fundamental rights.

Education is a fundamental human right, enshrined in international legal instruments such as the Universal Declaration of Human Rights (UNGA, 1948), whose Article 26 affirms that 'everyone has the right to education', and the Convention on the Rights of the Child (UNGA, 1989), which recognizes in Article 28 the right of every child to education as essential for the development of their personality and abilities. However, this right is gravely threatened in contexts of humanitarian crisis and armed conflict. According to Education Cannot Wait (2023), 'armed conflicts, forced displacements, climate change, and other crises have increased the number of crisis-affected children urgently needing quality education to 224 million'.

AI tools are released with enormous claims to expand access to education in emergencies, forced displacements and humanitarian crises. Yet, their hasty and technocratic implementation, disconnected from sociocultural realities and human suffering, risk deepening wounds rather than healing them. As UNESCO (2021, p. 12) warns, building new educational futures must be 'firmly rooted in a commitment to human rights' ensuring that the right to education is interpreted inclusively and extended to all without distinction. Any incorporation of emerging technologies, including AI, must be guided by principles of social justice, solidarity and care, and not by disembodied logics of efficiency that risk reproducing structural inequalities.

Hasty and technocratic implementation, disconnected from sociocultural realities and human suffering, risk deepening wounds rather than healing them.

The ethical use of AI in education must be grounded in core human rights principles: dignity, non-discrimination, participation, equity and protection. Failure to uphold these principles transforms AI into a threat not only in its technical design, but also in its profound humanitarian and legal implications.

International organizations, such as UNICEF and United Nations Human Rights Council (UNHCR), support the expansion of technological solutions for education in emergencies, where platforms like Kolibri and Rumie have been used to offer personalized learning pathways to vulnerable populations. Although developed independently, these platforms' use in displacement and humanitarian crisis contexts reflects growing interest in inclusive digital tools to uphold the right to education in extreme situations. Yet, absent human rights protections, these technologies expose vulnerable populations to grave risks:

- **Privacy violations**: The mass collection of data from vulnerable children and adolescents without informed consent can lead to new forms of exploitation and discrimination (OHCHR, 2024).
- **Reproductions of structural biases**: As West, Whittaker, and Crawford (2019) warn, bias in technology sector hiring and leadership profoundly shapes the types of systems that are designed and deployed, replicating historical exclusions based on race, gender and class.
- **Dehumanization in automated systems**: Eubanks (2018) analyses how algorithmic technologies, operating in 'low-rights environments', tend to treat vulnerable individuals as data to be managed rather than as rights-holding subjects, exacerbating their invisibility and marginalization.

The deployment of AI in education must be guided not merely by technological efficiency but by the ethical imperative to restore and protect human dignity. An unregulated educational AI could transform the right to education into a privilege conditioned by technological access, deepening existing structural gaps. Trauma pedagogy teaches that education in crisis contexts must create 'safe, relational, and culturally sensitive spaces that foster student empowerment' (Carello and Butler, 2015), helping restore a sense of belonging, dignity and self-worth.

The principle of symbolic reparation enshrined in the UN's Basic Principles and Guidelines on the Right to a Remedy and Reparation (UNGA, 2005), asserts that states must provide measures that include 'restoration of the dignity and reputation of victims through symbolic and material acts'. Following this principle, AI-based educational solutions would not merely facilitate access to learning but actively contribute to restoring the dignity and agency of those who have survived violence and displacement.

Human rights-compliant AI in education would need to fulfill, at minimum, the fundamental principles of dignity and social justice. First, the principle of human dignity demands that students are never reduced to mere 'data' or 'users' but recognized as rights-bearing individuals (UNGA, 1948). Second, the principle of non-discrimination requires inclusive algorithms that do not replicate historical patterns of exclusion based on gender, race, migration status or socio-economic condition (OHCHR, 2021).

Furthermore, respect for informed consent requires that the collection and processing of personal data be conducted ethically, transparently and in a way that protects students' best interests, especially in vulnerable contexts (UNESCO, 2021). The principle of participation implies actively involving local communities in every stage of design, implementation and evaluation of educational technologies. Lastly, special protection mandates that children, adolescents and displaced persons be treated as priority subjects of protection, in line with international human rights law (UNGA, 1989).

AI in education cannot be neutral in the face of human suffering. In crisis settings, designing efficient algorithms or promising technological access is not enough; it is imperative to rebuild relationships of care, respect and reparation. A truly human rights-based educational AI must not merely ask what technology can do for education, but who the technology must serve.

The deployment of AI in education must be guided not merely by technological efficiency but by the ethical imperative to restore and protect human dignity.

In contexts marked by trauma and dispossession, education must reclaim its deepest meaning: to be a refuge, a possibility and a restitution of wounded dignity. The urgent question is not what AI can offer to education, but what the living memory of victims demands of educational AI. Only a technology designed with consciousness of its debt to humanity can accompany the reconstruction of more just futures, where learning is not a privilege but a collective act of healing.

From the perspective of Buen Vivir, which guides the worldview of our Indigenous Peoples in Bolivia, knowledge is not a commodity or a tool of domination, but a path towards harmony between human beings and Mother Earth. Applying this vision to the development of AI in education implies recognizing that all technology must serve dignified life, mutual respect and the building of solidarity-based communities.

If educational AI were to be inspired by Buen Vivir, it will not seek soulless efficiency, but meaningful learning, healing with justice and a shared future. Only by honouring our roots can we imagine and build an education that leaves no one behind and that heals the wounds left by conflict and inequality.

AI in education cannot be neutral in the face of human suffering. In crisis settings, designing efficient algorithms or promising technological access is not enough; it is imperative to rebuild relationships of care, respect and reparation.

References

Carello, J. and Butler, L.D. 2015. Practicing what we teach: Trauma-informed educational practice. *Journal of Teaching in Social Work*, Vol. 35, No. 3, pp. 262–78.

Education Cannot Wait. 2023. *Number of Crisis-Impacted Children in Need of Education Support Rises Significantly: Education Cannot Wait Issues New Global Estimates Study*.

Eubanks, V. 2018. *Automating Inequality: How High-tech Tools Profile, Police, And Punish The Poor.* New York, St. Martin's Press.

Learning Equality. 2023. *Kolibri: Offline-first learning platform*.

Office of the United Nations High Commissioner for Human Rights (OHCHR). 2021. *The right to privacy in the digital age: Report of the United Nations High Commissioner for Human Rights*. Geneva, OHCHR. (Doc. A/HRC/48/31.)

––. 2024. *Human rights must be at the core of generative AI technologies, says Türk*. Geneva, OHCHR.

Rumie Initiative. 2024. *Rumie: Learn for free, anywhere*.

Sphere Association. 2018. *The Sphere Handbook: Humanitarian Charter and Minimum Standards in Humanitarian Response*. Geneva, Sphere Association.

UNESCO. 2021. *Reimagining our futures together: A new social contract for education*. Paris, UNESCO. https://doi.org/10.54675/ASRB4722

UNICEF. 2021. *Digital Learning in Emergencies: Frameworks, Approaches, Initiatives*. New York, UNICEF.

––. 2023. *Guidance for generative AI in education and research*. Paris, UNESCO. https://doi.org/10.54675/EWZM9535

United Nations General Assembly (UNGA). 1948. *Universal Declaration of Human Rights*. New York, United Nations.

––. 1989. *Convention on the Rights of the Child*. New York, United Nations.

––. 2005. *Basic Principles and Guidelines on the Right to a Remedy and Reparation for Victims of Gross Violations of International Human Rights Law and Serious Violations of International Humanitarian Law*. New York, United Nations.

––. 2024. *Artificial intelligence in education - Report of the Special Rapporteur on the right to education, Farida Shaheed*. New York, United Nations. (Doc. A/79/520.)

West, S.M., Whittaker, M. and Crawford, K. 2019. *Discriminating systems: Gender, race, and power in AI*. New York, AI Now Institute.

About the author

Mariela Sanchez Salas (Ph.D) is a Bolivian scholar whose work explores trauma, technology and the ethics of education in conflict-affected societies. Her research draws on the principles of Buen Vivir to reimagine more dignified and healing educational futures.

Part III

What will be meaningful and relevant for pedagogy and assessment?

It is difficult to authoritatively claim anything ubiquitous about what have been (or will be) the implications of AI and education around the globe. In some education institutions, AI has seemed unavoidable, appended openly to digital learning products and platforms (Singer, 2026). In other contexts, it has been indirect, stealthily raising questions about the degree of originality in research, a student essay or the source of a teacher's lesson plan (Walsh, 2025). In others still, it has barely made a mark, with more urgent priorities from access to infrastructure taking priority (UNESCO, 2025). Such a diverse spectrum of contexts and conditions, however, has not stopped technology evangelists from making optimistic promises of educational transformation.

Educator and master's student Bianca Farthing tackles the promise of personalized learning systems by questioning 'personalized to create what?' in her piece **'Bespoke or prescribed? The myth of personalized learning'.** The idea that students will be nudged to 'fix' what algorithms see as their flaws or weaknesses assumes a predetermined outcome that the student might have no role in defining, and which may involve little transparency. This pursuit of efficiency risks eroding the kinds of rich learning that prepares students to engage with the complexities of knowledge in a wider world. She leaves readers with a more hopeful vision of the purposes of education that AI could alternatively support: 'grounded in human experience, rich with dialogue, collaboration, unpredictability, and play' (p. 62), and accepting of the full complexity of what it means to learn, explore, make mistakes, engage with others and be truly human.

The rise of AI has also exacerbated longstanding questions in education about the validation of learning. In his piece **'What's worth measuring? The future of assessment in the AI age',** Hrishikesh Desai points out that that conventional assessment modes (e.g., exams and formulaic essays) have long been critiqued as antiquated methods

that prioritize memorization and formulaic responses over deeper learning. The arrival of LLMs that can generate essays and answers in seconds has now 'transformed this critique from a theoretical ideal into an immediate practical necessity' (p. 64). Desai sees promise, however, in the fact that conventional 'assessment artefacts' can no longer be reliable indicators of student learning. He conveys a new vision of assessment, one that embeds assessment processes in learning processes, dialogue and defence, authentically complex problems, and contextual situatedness.

Claudia Muellauer asks what students think of their teachers' use of AI to generate meaningful evaluation feedback in **'When my students told me I "sound like AI".** From her perspective, AI tools enabled her to dig more deeply into the way that she examined her students' work with more detail than she could have on her own. Her students, however, found it impersonal: 'To them, the AI's presence suggested my absence. Worse still, they often overlooked the feedback entirely... the more "perfect" the feedback became, the less it felt real to my students' (p. 70). Mueller recounts her surprise and curiosity when some students reacted adversely to even the suspicion that she had used AI in her responses to their work.

The next two think pieces **'Between wood and algorithms: Crafting AI literacy in TVET'** by Marianne Haarsbjerg, and **'Reimagining entrepreneurship education in the AI age'** by Efe Imiren and Katerina Nicolopoulou explore concerns specific to their domains of teaching and learning. AI has not only changed ways by which their students access and interact with information, but it has also shifted the technical, vocational, and entrepreneurial fields themselves. As educators and curriculum designers, foundational principles about the nature of human craftsmanship, ingenuity, originality, and value suddenly seem less certain. Understanding what AI both can and cannot disrupt in their respective sectors is now a major concern for educators in these and many other fields.

This section concludes with a piece by Alice Bennett titled **'AI and accessibility: abdicating engagement?'** It examines an area in which earlier forms of AI have already had a long history in the classroom: assistive technology, including speech-to-text and alt-text on images for people with blindness or those who have limited vision. In the push to automate accessibility checkers and compliance in educational materials and settings, she identifies a worrying trend of synthetic 'user feedback' as a means to replace research and testing done by actual people with disabilities. She argues that, without the direct input of individuals living with disabilities, AI tools have not truly been able to reflect the lived experiences of students and educators.

Bespoke or prescribed? The myth of personalized learning

Bianca Farthing

> Personalized learning is AI's promise in education – but who defines what's personal, and does it nurture strengths or just fix what algorithms see as flaws?

Among the most frequently cited benefits of AI in education is the promise of *personalized learning*. Ask any AI chatbot, ChatGPT included, about AI's role in the classroom, and personalized learning inevitably tops the list. But who created this narrative, and when did we collectively decide that this was the pinnacle of AI's potential in education? Why is this particular vision so widely emphasised, and what assumptions underpin it?

At its core, personalized learning aims to tailor educational content, pace and delivery to the individual learner. But this seemingly innocuous idea conceals a host of complex, and often unexamined, questions. Who decides what is personal? How does a machine determine what a learner ***needs***? And crucially, what does 'personalized' actually mean in practice? Is it about addressing a learner's strengths and interests, or merely remediating their weaknesses as perceived by an algorithm?

Personalization to end?

There is an implicit goal in most personalized learning systems: a destination, a path, a defined outcome. But *personalized to create what?* Are children being guided towards a predefined set of competencies? Is there an end goal, and if so, who sets it? More provocatively, *should* there be an end goal at all? Education, at its best, is about exploration, freedom and the messy, often non-linear journey of learning. Mistakes, detours and curiosity are not inefficiencies to be optimized away, they are the substance of learning itself.

> The risk is that in chasing efficiency and personalization, we erode the very experiences that prepare learners for the complexities of the world beyond the classroom.

AI-driven personalization often assumes that progress is measurable, that learning can be quantified and optimized. But what about the elements of progress that defy measurement? What about learning behaviours; resilience, adaptability, curiosity, that emerge over time and through experience? What about social skills, collaboration and empathy? Personalized learning, by design, is an individualized experience. But education is, at its heart, a social endeavour. If personalization leads to isolation, what kind of learners – and people – are we nurturing?

Moreover, when every learner is placed on their own 'optimized' track, where is the shared experience of learning – the chance to debate, to disagree, to collaborate and to co-

construct knowledge with peers? The risk is that in chasing efficiency and personalization, we erode the very experiences that prepare learners for the complexities of the world beyond the classroom.

Transparency, agency and trust

Another pressing concern is transparency. How do personalized systems make decisions about what content a learner sees, what pathways they are offered or what feedback they receive? These systems often function as black boxes, with little clarity on how recommendations are generated. If a system suggests that a learner revisit a concept or accelerate through material, on what basis is that judgement made? Who controls the algorithm's parameters, and whose values are embedded in those decisions?

More importantly, does the learner have any agency in that decision? Personalized learning risks becoming a one-way street, where learners are guided, or nudged, by unseen hands, without awareness or consent. The irony of personalized learning is that it often sidelines the learner's voice. Surely, meaningful personalization involves listening to the learner, not just the data. Yet, in many AI-driven models, the learner becomes a passive recipient of algorithmically curated experiences, rather than an active participant in shaping their own learning journey.

The irony of personalized learning is that it often sidelines the learner's voice. Surely, meaningful personalization involves listening to the learner, not just the data.

There is also the issue of trust. For personalized learning systems to be effective, and ethical, learners and educators must trust them. That trust can only be built through transparency, accountability and a genuine commitment to placing the learner's needs, aspirations and agency at the centre. Without these, personalization becomes not a gift, but a constraint.

The future of learning

So, what *should* learning look like in the age of AI? It should be exploratory, not prescriptive. It should encourage freedom and creativity, not just efficiency and optimization. There is a danger in allowing AI to guide learning along narrow, preordained paths. True personalization is not about being guided through a system; it's about having the autonomy to shape your own path.

AI should not dictate learning; it should support the learner's agency, creativity and humanity.

Learning should be grounded in human experience, rich with dialogue, collaboration, unpredictability and play. Personalization, if it is to be meaningful, must recognize the full complexity of what it means to learn and to grow. It must make space for exploration, for mistakes, and for the kind of deep, personal engagement that cannot be captured by metrics or optimized by algorithms.

If we are to reclaim learning from the algorithms, we must ask harder questions about the narratives we have accepted, the systems we are building and the learners we are shaping. AI should not dictate learning; it should support the learner's agency, creativity and humanity.

About the author

Bianca Farthing is an educator in the United Kingdom of Great Britain and Northern Ireland independent school sector and a Master's student at University of Oxford. She is the Founder of AI Edify and the Initiative for Conscious AI, helping schools navigate the ethical and educational implications of AI.

What's worth measuring? The future of assessment in the AI age

Hrishikesh Desai

The rise of AI tools disrupts traditional ideas of assessment and urges a shift towards promoting and evaluating higher-order thinking, creativity and ethical reasoning.

AI tools like ChatGPT have fundamentally disrupted traditional notions of assessment in education. When a machine can generate essays, solve complex problems and even mimic creative writing, educators are forced to ask: *what skills should we assess, and how do we evaluate learning in a world where AI can perform tasks once thought uniquely human?* This think piece explores the implications of AI for assessment and argues that we must shift from measuring rote knowledge to promoting and evaluating higher-order thinking, creativity and ethical reasoning.

Historically, educational assessment has relied on measurable outputs, such as essays, exams and problem sets that test memorization, comprehension and technical proficiency. However, AI's ability to generate these outputs undermines their reliability as indicators of individual effort or understanding. For example, a student can prompt ChatGPT or DeepSeek to write a 1000-word essay on climate change in minutes, complete with citations, which leaves educators grappling with how to distinguish human work from machine-generated content. This disruption extends beyond writing to mathematics, coding, and even creative tasks, where AI tools like GitHub Copilot and DALL-E provide solutions or artefacts that mimic human ingenuity. As AI becomes a ubiquitous resource, the traditional metrics of learning are no longer fit for purpose. The emergence of generative AI has precipitated what can only be described as an assessment crisis.

For decades, faculty have critiqued assessment methods that prioritize memorization and formulaic responses over deeper learning. The emergence of sophisticated AI tools has transformed this critique from a theoretical ideal into an immediate practical necessity.

This technological disruption to learning has arrived at a critical juncture. For decades, faculty have critiqued assessment methods that prioritize memorization and formulaic responses over deeper learning. The emergence of sophisticated AI tools has transformed this critique from a theoretical ideal into an immediate practical necessity. We can no longer rely on conventional assessment methods that AI can bypass with ease. I argue that generative AI is not merely challenging our assessment techniques but providing an opportunity to fundamentally realign educational evaluation with more authentic demonstrations of human learning, thinking and creation.

AI detection tools have proven unreliable and often flag genuine human writing as AI-generated while failing to identify sophisticated AI outputs. This technological arms race between detection and generation appears unwinnable for those hoping to maintain

traditional assessment boundaries. Moreover, the rapid pace of AI development suggests this challenge will only intensify. Models continue to improve in their ability to produce nuanced, contextually appropriate and stylistically diverse content. The window for detecting AI-generated work through technical means is likely closing permanently.

In a world where AI can provide answers, the ability to evaluate, critique and synthesize information becomes important. Instead of asking students to summarize a historical event, we can instead ask them to analyse conflicting accounts generated by AI. We can propose assessments that reward novel solutions, artistic expression and interdisciplinary thinking. Many future workplaces will require humans to work alongside AI. Assessments should therefore evaluate students' ability to use AI tools effectively and ethically. This shift requires a focus on three core areas, namely, critical thinking, creative problem-solving and ethical reasoning, which are skills that demand human judgment and cannot be fully outsourced to AI. While AI tools can generate arguments, they still lack the judgment to question their own assumptions or adapt to contradictory evidence dynamically. AI tools can provide a summary or propose a solution, but they cannot fully replicate a student's capacity to synthesize personal experiences, local contexts or moral considerations. Assessment should therefore emphasize tasks that require students to deconstruct AI outputs. For example, educators could assign projects where students critique a ChatGPT-generated essay and identify logical fallacies or biases. This AI-auditing approach not only tests comprehension but also cultivates a sceptical mindset amongst students.

When the final artefact (e.g. essay, solution or creative work) becomes an unreliable indicator of student learning, the journey of development takes on greater significance.

Conventional assessments often reward formulaic responses, but AI tools excel at replicating patterns. Instead, evaluations should prioritize open-ended challenges that require innovation, such as designing a novel solution to a local environmental issue or crafting a story with unpredictable twists. These tasks push students to think beyond the templates or training data that AI relies on and promotes originality. AI tools should be used to augment creativity and paired with human oversight for the best result. Assessments can also include how thoughtfully students construct prompts to generate useful AI outputs, how creatively students iterate on and improve AI-generated ideas, and how critically students evaluate AI-generated information for accuracy and relevance. Without such a fundamental rethink of assessments, over-reliance on AI could erode foundational skills in students like writing or basic computation.

One promising direction involves shifting assessment focus from end products to learning processes. When the final artefact (e.g. essay, solution or creative work) becomes an unreliable indicator of student learning, the journey of development takes on greater significance. This approach values documentation of thinking, iteration and metacognitive reflection. Students might maintain process journals, submit work at multiple stages, participate in structured peer reviews or engage in recorded thinking-aloud protocols. These process-oriented artefacts reveal how students approach problems, reconsider assumptions,

integrate feedback and develop ideas over time, which are aspects of learning that AI cannot readily simulate. This also aligns with real-world problem solving, which rarely involves a single performance without revision. Instead, it typically entails research, collaboration, feedback incorporation and multiple iterations. These are exactly the kinds of processes that AI cannot easily replicate when deployed as a simple answer generator.

Another promising direction involves assessment through dialogue and defence. When students must articulate their understanding in real-time conversations, explain their reasoning, respond to unanticipated questions or defend their conclusions against challenges, they demonstrate intellectual ownership that is difficult to fake. This approach draws inspiration from thesis defenses, Socratic seminars and oral examinations. Modern implementations might include (i) structured interviews where students explain key concepts and their applications, (ii) group dialogues where students build upon and challenge each other's ideas, (iii) presentation formats with spontaneous responses to audience questions, and (iv) conversational assessments where students must explain their thought processes. Such types of assessments are inherently difficult to outsource to AI, as they require integration of multiple cognitive and social capabilities in real time. Thus, rather than trying to prevent AI use, which is an increasingly futile endeavor, educators need to design assessments that evaluate how effectively students collaborate with AI as an intellectual tool. The ability to effectively prompt, evaluate, refine and integrate AI-generated content represents a new form of literacy.

Yet another direction involves designing assessments around authentically complex problems situated in students' personal contexts. When assignments require students to connect concepts to their unique experiences, communities or observations, they create natural barriers to wholesale AI substitution.

Yet another direction involves designing assessments around authentically complex problems situated in students' personal contexts. When assignments require students to connect concepts to their unique experiences, communities or observations, they create natural barriers to wholesale AI substitution. For example, rather than assigning a standard analysis of a literary work, students might examine how themes in the text manifest in their local community. Instead of solving generic physics problems, students could analyse physical phenomena they have personally observed and documented. Such approaches inherently require integration of course concepts with knowledge that is not readily available to AI systems. These contextualized assessments also tend to increase student engagement. When evaluations connect to students' lives and interests, they promote intrinsic motivation rather than compliance.

This new vision for assessment transformation inevitably creates tension with current educational systems' emphasis on standardization and comparison. Educational institutions and policy-makers have built extensive infrastructures around comparative metrics for cross-institutional, regional and even international benchmarking. This seeming problem doesn't have to be impossible to solve. Rather than measuring identical outputs on identical tasks, we might instead compare students' demonstrated capacities to deal with complexity,

integrate multiple perspectives, devise novel solutions and articulate their reasoning, regardless of the specific context in which these abilities manifest. This shift will require redefining 'rigor'. Some institutions such as the University of Michigan's Center for Research on Learning and Teaching (CRLT) and MIT Teaching + Learning Lab are already advocating for such redefinitions.

Perhaps most fundamentally, we must question whether certain high-stakes, single-point assessments should exist in their current form at all. The concept of determining complex human potential through standardized examinations from university admissions exams to professional licensing has always been problematic, and AI's ability to game these systems simply exposes their inherent limitations more starkly. Rather than merely adapting existing high-stakes assessments, educational systems might better serve students through more continuous, portfolio-based evaluations that reflect sustained development of competencies over time. Such approaches would resist AI replication and provide richer predictive information about student capabilities. Exam-centric education systems, prevalent in many countries, present particular challenges and opportunities for assessment reform. In regions where high-stakes examinations determine educational and economic trajectories, assessment innovation must deal with entrenched cultural values and institutional structures that prioritize standardized testing. However, these same systems are now facing the most urgent challenges from AI disruption, as they typically emphasize precisely the knowledge reproduction tasks that AI tools most readily perform.

The status quo is increasingly untenable. Unfortunately, most educational systems today are designed around standardized evaluation, and alternative approaches may initially appear less efficient, less objective or more resource intensive. The transition discussed in this article will require professional development, institutional support and a willingness to embrace some uncertainty during the evolution of new practices. The transition will require experimentation, flexibility and a tolerance for imperfection as new approaches evolve. Perhaps, we must recognize: that AI's challenge to traditional assessment indicates a deeper truth that many of us academics don't seem willing to recognize that many conventional evaluation methods have been poor proxies for meaningful learning all along. They have often rewarded formula-following, rote memorization, and performance over genuine understanding, creativity and intellectual growth. To conclude, I strongly believe that AI-human collaboration over AI exclusion and developing assessments that are both AI-resistant and pedagogically superior is still the way to go.

Perhaps, we must recognize that AI's challenge to traditional assessment indicates a deeper truth that many of us academics don't seem willing to recognize: that many conventional evaluation methods have been poor proxies for meaningful learning all along. They have often rewarded formula-following, rote memorization, and performance over genuine understanding, creativity and intellectual growth.

About the author

Dr Hrishikesh Desai is an Assistant Professor in the Department of Accounting at Arkansas State University (ASU). He is the Director of the Master of Accountancy with Data Analytics programme at ASU.

When my students told me I 'sound like AI'

Claudia Muellauer

When AI enters the classroom, teacher care meets student scepticism, raising questions about trust, collaboration and what meaningful feedback really means.

Picture this: A language teacher, eager to try out new technology, shares carefully crafted feedback on student essays. The teacher's comments are thorough, consistent, aligned with clear rubrics and full of nuanced suggestions – far more detailed than what time usually allows. Yet, instead of appreciation, the student's response is a dismissive, vaguely accusatory remark such as, *'Did ChatGPT write this?' or 'This sounds like AI.'*

In that moment, the promise of AI-enhanced education meets a complex human reality. While I see tools like Brisk Teaching, ChatGPT Pro and Gemini Advanced as valuable professional allies, many of my students interpret their use as a sign of neglect – or worse, laziness. This disconnect forces us to ask some urgent questions: what does meaningful teaching look like in an AI-driven world? How is educational care recognized and valued? And how can we use AI in ways that support, rather than replace, the essential human relationships at the core of learning?

My journey with generating AI feedback: Curiosity, growth and a digital colleague

My turn to use AI-assisted feedback wasn't solely about saving time. Yes, teaching comes with heavy workload pressures, but what truly drew me in was curiosity. I wanted to understand what these rapidly evolving tools could do. The freedom to ask endless questions – without fear of sounding foolish – was liberating. As a non-native English speaker teaching English as a second language (ESL), I was especially struck by the sophisticated linguistic feedback these LLMs could offer. Often, they articulated insights that I might struggle to express myself. But AI didn't just support my students – it became a partner in my own development. Brisk, for example, helped me better understand how to construct rubrics, making an often fuzzy process much clearer. I even used its presentation tool to craft engaging introductions for the start of the school year. My students – digital natives – introduced me to Gemini Advanced, though I still prefer ChatGPT for its language analysis. This wasn't just about new tools – it felt like genuine professional growth, supported by a tireless, non-judgmental digital 'colleague'.

> The idea that AI simply saves teachers time is misleading. Yes, the tools are efficient but using them effectively requires a shift in how we work. I found myself stepping into a new role: the gatekeeper. Crafting effective prompts and curating presentations with Brisk or Gemini involved careful pedagogical thinking. But the real effort came *after* the AI generated its suggestions.

Becoming a gatekeeper: More work, better outcomes, shared effort

The idea that AI simply saves teachers time is misleading. Yes, the tools are efficient but using them effectively requires a shift in how we work. I found myself stepping into a new role: the gatekeeper. Crafting effective prompts and curating presentations with Brisk or Gemini involved careful pedagogical thinking. But the real effort came *after* the AI generated its suggestions. Ironically, providing AI-assisted feedback often took longer than marking by hand. Yet I enjoyed it. It didn't feel like solitary grading – it felt like collaboration. I thrive in team environments, and working with AI felt like having an assistant by my side.

My process became quite elaborate: I would start by uploading student work and crafting detailed prompts that included specific learning objectives, assignment criteria and examples of the kind of feedback I wanted. After the AI generated its initial response, I would spend considerable time refining it – deleting irrelevant comments, rewriting awkward phrasing, adding encouraging notes and injecting my personal tone. I often found myself having follow-up conversations with the AI, asking it to explain its reasoning or suggest alternative approaches. This iterative process meant that what started as a five-minute marking task could easily stretch to fifteen or twenty minutes per essay.

While I trusted the models for grammar, the human touch was essential for clarity, empathy and nuance. This labour was largely invisible to students – but it required time, care and professional judgment.

Student perception: 'Sounds like AI', ignored comments and misunderstood effort

Despite my investment, students responded in disheartening ways. The moment they sensed AI involvement, enthusiasm waned. The comment *'This sounds like ChatGPT'* wasn't neutral – it hinted at disappointment. Students told me they wanted *me* to read their work. To them, the AI's presence suggested my absence. Worse still, they often overlooked the feedback entirely. The detail I valued came across to them as overwhelming or impersonal. Despite my curation, the effort I put in went unnoticed – eclipsed by the absence of something tangible, like handwritten notes in the margins.

Students told me they wanted *me* to read their work. To them, the AI's presence suggested my absence. Worse still, they often overlooked the feedback entirely. This experience highlighted a painful paradox: the more 'perfect' the feedback became, the less it felt real to my students.

This experience highlighted a painful paradox: the more 'perfect' the feedback became, the less it felt real to my students.

A deeper dilemma: Invisible labour vs. visible care

At the heart of this tension is a question of what counts as care in education. Traditionally, students equate visible effort – like handwritten comments – with personal investment. But AI changes the nature of our labour. Much of my work shifted behind the scenes: crafting prompts, editing feedback, thinking critically about tone and content. I experienced this as engaged, thoughtful teaching. But to students, it didn't *look* like care. In some ways, AI's efficiency works against us. When students see pages of polished, impersonal comments, they may miss the human effort behind them. Despite my best efforts to personalize the feedback, something emotional, something relational, was missing.

Transparency: Necessary, but not enough

To bridge this gap, I became more transparent. I told students when and how I used AI. This helped – some even opened up about their own AI use for brainstorming or drafting. But transparency didn't fully shift perceptions. Many still saw AI as diminishing my involvement. The invisible labour of curating and personalizing feedback wasn't recognized or valued. While openness is important, it's not a cure-all. Understanding and appreciation of this new form of teaching work takes more than just explanation – it requires a cultural shift in how we view care in education.

But this also raises uncomfortable questions about our assumptions around efficiency in education. Perhaps my students' instinctive resistance to AI-generated feedback reflects something important – a recognition that the messy, time-consuming, visibly human process of reading and responding to their work is itself valuable. Maybe the question isn't how to make AI feedback feel more human, but whether we should be questioning our rush towards efficiency in spaces where inefficiency might actually signal care.

A teacher's reflection: Lost in translation?

All of this has left me in a kind of limbo. I believe AI tools have made me a better teacher – more insightful, more consistent, more reflective. They've helped me learn and grow. And yet, the disconnect with my students has been painful. Despite my best efforts, my feedback often doesn't *feel* right to them, or even to me. At times, it feels like I'm cheating, even when I know I've put in more work than ever. Full transparency is tricky, too. Explaining the gatekeeping process in detail risks sounding defensive. I sometimes worry it comes across as 'stealth AI' – a secret I shouldn't have kept. I genuinely believe that my AI-curated feedback is more pedagogically sound. But if it doesn't resonate with students, does it matter?

Looking ahead: Towards human-centred AI in education

This experience leaves me both hopeful and cautious. My hope is that we can integrate AI into education in ways that are thoughtful, collaborative and focused on learning – not just efficient. I want teachers and students to be co-creators of how these tools are used, with pedagogy and care at the centre. But my fear is that a poorly managed rollout will ignore the relational dimension of teaching, widening gaps, undermining motivation and devaluing our profession by overlooking the invisible labour we do.

The path forward isn't about finding the perfect balance between human and AI – it's about recognizing that teaching has always been fundamentally relational work. AI can enhance our capabilities, help us see patterns we might miss and free us from some routine tasks. But it cannot replace the essential human elements: the moment of recognition when a student finally grasps a concept, the careful attention to individual needs and growth, the trust that builds through authentic interaction.

I want teachers and students to be co-creators of how these tools are used, with pedagogy and care at the centre. But my fear is that a poorly managed rollout will ignore the relational dimension of teaching, widening gaps, undermining motivation and devaluing our profession by overlooking the invisible labour we do.

Rather than asking how to make AI feedback feel more human, perhaps we should be asking how to preserve and strengthen the human connections that make learning meaningful, while thoughtfully incorporating tools that can support our pedagogical goals. The answer, I suspect, lies not in perfect efficiency but in maintaining the beautiful inefficiencies that signal genuine care.

AI is undeniably influencing education. But if we keep care, connection and professional integrity at the centre, we can shape a future where technology supports – not erodes – the heart of teaching and learning.

About the author

Claudia Muellauer teaches Language and Theory of Knowledge at International Highschool Herzogberg in Austria. She also coordinates and supports research at Central European University, contributing to projects on belief systems and debiasing as well as the philosophy excellence cluster Knowledge in Crisis.

Between wood and algorithms: Crafting AI literacy in TVET

Marianne Haarsbjerg

From woodwork to algorithms, five strategies show how AI literacy in TVET can foster craftsmanship, critical thinking and ethical, future-ready learners.

The rapid development of AI, especially generative AI, is reshaping education and training, bringing both opportunities and challenges. Consequently, we have witnessed the development of a plethora of different AI-related frameworks aimed at guiding educational institutions and their actors (Almatrafi, Johri and Lee, 2024; UNESCO 2024a, 2024b).

In a current research project on the role of AI (particularly generative AI) in Technical and Vocational Education and Training (TVET) in Denmark, we have been investigating when, why and how AI can or cannot be used meaningfully in the students' learning processes based on a socio-material perspective. Rather than seeing AI merely as an external tool to integrate seamlessly into existing educational structures, a socio-material perspective recognizes AI as an active participant in the educational assemblage, co-shaping pedagogical practices, assessment norms and institutional power dynamics (Johri, 2022, 2011). This approach moves beyond deterministic views of AI as either disruptive or beneficial and instead considers how AI (already) entangles with human, material and institutional elements to reconfigure learning and pedagogy (Bearman and Ajjawi, 2023).

Inspired by Perkins et al. (2024), our framework consists of five learning strategies for advancing AI literacy: AI-independent, AI-aware, AI-supported, AI-enhanced, and AI-centric learning. Our framework can be viewed as an integration scale ranging from no involvement to full entanglement of AI, which typically appeals to educational leaders. When we work with teachers and students, however, we articulate it as a puzzle, where the pieces together form a multidimensional and holistic whole. The proposed strategies accentuate that learning occurs in the interplay between human capabilities and the affordances of available tools.

Integrating Sennett's ideas of craftsmanship with AI literacy provides an enriched perspective, moving AI education from merely technical skills acquisition to a deeper, reflective and ethically informed practice – cultivating a generation of thoughtful, critical and skilled AI craftspeople.

To further acknowledge our TVET context, in this think piece, I explore our strategies by drawing on Sennett's (2008) ideas about 'The Craftsman' using carpentry education as example. Craftsmanship, according to Sennett, is defined as the desire to do a job well for its own sake. It is not limited only to traditional crafts (e.g. woodworking or pottery) but extends into all kinds of skilled practice, including medicine, software programming, music, teaching (ibid). Sennett emphasizes quality, pride in work, personal investment and a sustained commitment to continuous improvement. Integrating

Sennett's ideas of craftsmanship with AI literacy provides an enriched perspective, moving AI education from merely technical skills acquisition to a deeper, reflective and ethically informed practice – cultivating a generation of thoughtful, critical and skilled AI craftspeople.

Strategy 1: AI-independent learning

Imagine a learning space entirely free of AI and other digital devices. In this strategy, students measure, saw and craft wooden joints relying solely on their embodied cognition and skills qualifying human capabilities such as taste, instinct and intuition. Mistakes are frequent and deeply instructive. Echoing Sennett, the scenario underscores how repeated tactile practice develops authentic mastery and intuitive human judgement, which is foundational in times when digital dependency risks weakening human capability and agency. In socio-material terms, the scenario constructs a learning space where non-digital materials (e.g. paper, pencil, physical artefacts) are foregrounded. Here, agency is attributed solely to the human actors and non-AI artefacts, emphasizing traditional skill mastery. This is not a matter of romanticizing craftmanship. On the contrary, foundational knowledge and skills are key to understanding any technology and perhaps even more so with regards to AI.

Strategy 2: AI-aware learning

In this strategy, students practice carpentry traditionally but remain alert to AI's pervasive presence. Discussions consider how machine learning and AI in automated computer numerical control (CNC) machines might produce perfect joinery but lack the craftsman's reflective, embodied engagement and appreciation. Students critically examine how digital innovations reconfigure craftsmanship's value. Upholding Sennett's notion of reflective ethics, this strategy demands critical consciousness around the compromises technology asks of us. It questions technology's role and purpose, preserving human agency by fostering students' ability to critically engage with the choices made possible or constrained by AI. Even when AI is not directly employed, its presence in the broader socio-material landscape is recognized, and from a socio-material lens, students critique the entanglement of human craft and digital automation, fostering critical consciousness about AI's pervasive yet often invisible presence. The learning space is set up to negotiate boundaries between human agency and digital influence, even if the latter is kept at arm's length during the task.

Strategy 3: AI-supported learning

In this learning space, carpentry students carefully integrate AI to assist in design and planning phases. AI software suggests optimal wood utilization or assembly strategies, while core craftsmanship remains human-driven. Here, AI enriches rather than diminishes human agency, serving as an assistive artefact within a broader socio-material assemblage. Students are guided by Sennett's ideal of dialogue between craftsman and material; the AI becomes

part of this conversation, critiqued, revised and selectively trusted. This critical cooperation preserves the craftsman's autonomy, judgment and the reflective, iterative practice that defines quality, but also recognizes AI as an active participant that contributes to the learning assemblage. The tool's suggestions, feedback and computational power interact with the student's ideas, creating a hybrid material-discursive process.

Strategy 4: AI-enhanced learning

Picture students constructing complex timber-frame buildings alongside AI-driven software that dynamically recalibrates structural models in real-time. In this scenario, students engage extensively with AI, blending human intuition and algorithmic precision. Smart machines and augmented reality guide complex joinery, assist complex measurements, or propose new, innovative and perhaps even sustainable designs impossible to conceive manually. According to Sennett, craftsmanship necessitates openness to continuous refinement and innovation; here, the craftsperson evolves into a hybrid artisan, creatively leveraging digital augmentations. The learning space is viewed as a socio-material assemblage where both human and AI agents co-intelligently interact to generate new forms of understanding. AI is not merely a tool but a co-creator whose material presence alters the learning environment. The digital and material landscapes merge, creating hybrid learning spaces where traditional boundaries between teacher, student and technology blur.

Strategy 5: AI-centric learning

In the final strategy, AI itself becomes the primary subject of study with students exploring AI as a crafted object, coding, tweaking algorithms and dissecting the data biases embedded in automated design software. Craftsmanship is elevated beyond physical materiality into digital realms, positioning the AI system itself as the 'wood' and 'tool'. Sennett's ethical craftsmanship resonates powerfully here, as students grapple with AI's ethical implications and societal responsibilities. Students investigate how AI technologies reshape carpentry as a trade, exploring ethics, materials, data biases, economic impacts, environmental implications and social considerations at large. The learning environment becomes a space to critically examine the interplay between technology, labour markets, society and material conditions, fostering innovation and new forms of collective agency. Inspired by Sennett's notion of craftsmanship as ethical practice, students become designers of AI tools – craftspeople carefully sculpting algorithms adhering to ethical and professional standards enhancing both personal and professional pride.

As we have seen in the strategies, each of them contributes with valuable learning. Nevertheless, there are still many issues to be contemplated and explored in terms of promoting meaningful AI literacy.

Negotiating speed and patience: The contemporary craftsman's dilemma

Sennett emphasizes craftsmanship as a patient, reflective and iterative process – qualities fundamentally at odds with the logic of AI technologies designed primarily for immediacy, efficiency and rapid results. This tension poses a crucial question: Can we reconcile the patient quality of traditional craftsmanship with the efficiency demands and expectations that AI increasingly normalizes?

This tension poses a crucial question: Can we reconcile the patient quality of traditional craftsmanship with the efficiency demands and expectations that AI increasingly normalizes?

A socio-material viewpoint underscores that neither human nor AI should dominate exclusively. Instead, it prompts us to ask: under which conditions and contexts does rapid AI-generated output genuinely enhance craftsmanship, and when does it hinder the reflective, embodied learning central to human expertise?

Socio-material theory recognizes quality not as universal efficiency but as context-dependent meaning. Fast AI responses may enhance learning in certain practical or logistical scenarios (e.g. rapid problem-solving during prototyping). Conversely, slower, patient reflection might better serve deeper conceptual understanding and ethical judgment (e.g. reflecting on implications of designs or long-term impacts). Through a socio-material lens, educators can critically assess how different configurations of human and AI agency might preserve the reflective patience advocated by Sennett, while selectively harnessing AI's strengths. Thus, craftsmanship can evolve, integrating but not being overwhelmed by AI's rapidity.

Perhaps true AI craftsmanship lies precisely in the skilled judgment about when to slow down, when to speed up, why and how – transforming the dilemma into an active pedagogical choice rather than an inevitable technological dictate? And, in doing so, simultaneously educating students to make informed decisions on AI by promoting a multidimensional and holistic take on AI literacy.

Perhaps true AI craftsmanship lies precisely in the skilled judgment about when to slow down, when to speed up, why and how – transforming the dilemma into an active pedagogical choice rather than an inevitable technological dictate?

If so, I believe, the socio-material perspective can serve as a crucial mediator, critically highlighting the entanglement of humans, tools, speed, patience and quality in crafting meaningful education in our increasingly AI-mediated world.

References

Almatrafi, O., Johri, A. and Lee, Y. 2024. A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019-2023). *Computers and Education Open,* Vol. 6.

Bearman, M. and Ajjawi, R. 2023. Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology,* Vol. 54, No. 5, pp. 1160–73.

Johri, A. 2022. Augmented sociomateriality: implications of artificial intelligence in the field of learning technology. *Research in Learning Technology,* Vol. 30.

––. 2011. The socio-materiality of learning practices and implications for the field of learning technology. *Research in Learning Technology,* Vol. 19, No. 3, pp. 207–17.

Perkins, M., Furze, L., Roe, J. and MacVaugh, J. 2024. The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *Journal of University Teaching and Learning Practice*, Vol. 21, No. 6.

Sennett, R. 2008. *The Craftsman*. Yale University Press.

UNESCO. 2024a. *AI competence framework for teachers.* Paris, UNESCO.

––. 2024b. *AI competency framework for students.* Paris, UNESCO.

About the author

Marianne Haarsbjerg is a Senior Associate Professor, Ph.D. She is the leader of the research programme 'Digital technology, learning and design' at Absalon University College, Denmark. Her background is in human-centred informatics and design. Her research interests mainly cover the understanding and application of emerging technologies in dual professional and vocational education.

Reimagining entrepreneurship education in the AI age

Efe Imiren and Katerina Nicolopoulou

As AI reshapes how we think and decide, entrepreneurship education must evolve: how do we teach opportunity, creativity and human agency in an AI-driven world?

The fundamentals of entrepreneurship education centre around human agency, the ability to recognize opportunities and develop strategies to pursue these opportunities. However, with the advent of AI replacing critical thinking and evaluation, we are faced with the question: what does it mean to be entrepreneurial and how can we teach entrepreneurship in an age of AI? The answers lie in first acknowledging the conversations and emerging debates at the intersection of entrepreneurship and AI, and then addressing the gaps precipitated by these debates.

AI as an entrepreneurship education tool

As an education tool, there is no doubt that AI is useful to spot opportunities, analyse trends and develop strategies (Crompton and Burke, 2023). However, AI tools are being generated faster than educators can keep pace. At the same time, digital technologies, including AI tools, can also become obsolete very quickly, as has been the case with the release of ChatGPT 3.5, 4.0, 4.5 and newer language learning models. This precipitates the question, how is the *ephemerality* of AI tools impacting the intended learning outcomes adopted by entrepreneurship educators? Is the ephemerality of AI tools being taken into account in the design of learning objectives? How do constant updates of AI tools impact the stability of learning objectives and the perennial priorities in entrepreneurial curriculum?

How is the *ephemerality* of AI tools impacting the intended learning outcomes adopted by entrepreneurship educators? Is the ephemerality of AI tools being taken into account in the design of learning objectives? How do constant updates of AI tools impact the stability of learning objectives and the perennial priorities in entrepreneurial curriculum?

Implicit in the ephemerality debate is the question of *alignment*, i.e. does the AI proficiency students gain actually match the transferable skills that an entrepreneurship curriculum aims to develop? Entrepreneurship education seeks to develop transferable skills such as critical thinking, opportunity spotting and resilience. But with AI seemingly handing over answers to students, we need to reconsider how entrepreneurship education can align its curriculum appropriately with the use of AI tools. This ensures students develop the essential transferable skills they need for future real life and reflective practice, helping them to learn and develop as professionals and humans.

AI as a philosophical shift in entrepreneurship education

The fundamentals of entrepreneurship education are based on assumptions that humans recognize and pursue opportunities. But if AI can perform these tasks with speed and efficiency, what then is the role of the human agent (entrepreneur) in the entrepreneurial process and how do entrepreneurship educators capture these nuances in entrepreneurial education? This issue raises a set of important questions around competency and construct.

The fundamentals of entrepreneurship education are based on assumptions that humans recognize and pursue opportunities. But if AI can perform these tasks with speed and efficiency, what then is the role of the human agent (entrepreneur) in the entrepreneurial process and how do entrepreneurship educators capture these nuances in entrepreneurial education?

The main competency questions, then, become (a) which traditional entrepreneurial competencies have become obsolete to teach as entrepreneurship educators, and (b) what new competencies should be taught in an AI-aware entrepreneurship curriculum? These shift the conversation on the entrepreneurial competencies from the development of traditional entrepreneurial traits (e.g. intuition, risk-taking, propensity to experiment, internal locus of control) (Miller, 2020) to AI-driven competencies for entrepreneurial purposes and hybrid decision-making. This shift creates the need to revisit our understanding of a potential tension between entrepreneurial competencies that are human-centred vs AI-augmented.

AI's growing embeddedness in entrepreneurial decision-making processes also increases the need for entrepreneurship education to address algorithmic bias, fairness and the social impact of AI-driven business models. It matters that entrepreneurship educators recognize the need to embed responsible and ethical AI use in the curriculum as entrepreneurs increasingly rely on AI for decision-making that can unknowingly reinforce biases and exclusion (of gender, ethnicity, regional context, etc.) in their business models (Bakker and McMullen, 2023).

AI is now embedded in what has been thought of as human-centred entrepreneurial processes, such as entrepreneurial alertness, opportunity recognition and legitimacy building (Tang, Kacmar and Busenitz, 2012; Imiren et al., 2023). These processes are typically understood as cognitive and perceptual, moderated by experience and intuition. Integrating AI calls into question the role of human agency as central to the underlying theoretical constructs of what 'being entrepreneurial' means. This shift raises questions about how the potential for human-AI collaboration is being meaningfully addressed in entrepreneurship curriculum.

AI as practice in entrepreneurship education

AI is not just an abstract concept to engage with, but a tangible set of technologies with the potential to reshape the practice and education of entrepreneurs.

When considering the future of entrepreneurs and their needs for education, we need be reminded that AI is not just an abstract concept to engage with, but a tangible set of technologies with the potential to reshape the practice and education of entrepreneurs. This means that AI is not just a subject matter that students need to learn, but it could create a unique practice environment to build businesses that learn and adapt. It goes without saying that this also calls for continuously educating entrepreneurship educators on the possibilities and technologies that drive AI and its potential use in entrepreneurship education.

Such possibilities could include:

a. AI as offering the potential to support dynamic, data-driven decision-making processes fundamental to future-oriented entrepreneurial pursuits.

b. Repositioning the entrepreneurial mindset habitually taught in entrepreneurship curriculum – qualities such as adaptability, pattern recognition and efficient problem-solving – by recognizing that AI systems now exhibit many of these traits by design, and therefore integrating AI directly into how the entrepreneurial mindset is taught.

c. Embedding AI tools that help students not only learn technical skills but also internalise computational approaches to customer discovery, market analysis and business model innovation by creating a recursive learning environment through practical experience.

d. And finally, moving beyond merely using AI tools to adapt to the entrepreneurship curriculum, and instead fostering an understanding of AI as an entrepreneurial domain unto itself and therefore precipitating a dual focus – learning through AI whilst learning about AI.

However, the integration of AI in entrepreneurship education faces a significant challenge, as the technical knowledge and investment required to effectively use several AI tools could often exceed what traditional business and entrepreneurship faculties in business schools possess. The rapid evolution of AI capabilities means entrepreneurship curriculum must remain flexible, given the fact that academic institutions housing business schools where entrepreneurship education is developed are accustomed to more deliberate change cycles that relate to their resource planning. This could imply additional strain on already limited resources that can be dedicated to entrepreneurship education purposes, potentially leading to future field irrelevance of these institutions. These considerations might be especially pertinent to public sector universities.

We need to consider a way forward with the future direction of entrepreneurship education given the above realities. Entrepreneurship programmes have primarily been laboratories

for educational innovation because, unlike disciplines with rigid canonical requirements, entrepreneurship education has always emphasized experimentation and real-world relevance. By integrating AI as a practice environment rather than just another subject or tool, programmes can reinvent how entrepreneurship students learn, create value and consider human-AI collaboration. Acknowledging that overreliance on AI tools could undermine students' creative thinking and entrepreneurial intuition, the most promising approach to the future of entrepreneurship education would be to strike a balance: embracing AI as a practice field, whilst reinforcing that entrepreneurship remains a deeply human endeavour. This lies in recognizing that the competitive advantage of future entrepreneurs (the entrepreneurship students in our classrooms today) will not be their ability to use AI tools (this will be a ubiquitous skill set), but their capacity to ask better questions than their competitors, to recognize patterns missed by algorithms, and to reflect on the contextual and cultural nuances (such as building relationships), and emotional resonances that data cannot fully capture.

Acknowledging that overreliance on AI tools could undermine students' creative thinking and entrepreneurial intuition, the most promising approach to the future of entrepreneurship education would be to strike a balance: embracing AI as a practice field, whilst reinforcing that entrepreneurship remains a deeply human endeavour.

These possibilities will motivate entrepreneurship educators and faculty to reimagine, re-recognize and re-implement entrepreneurship education itself as a human-driven entrepreneurial act, and would eventually elevate the role of the human experience in shaping it.

References

Bakker, R. M. and McMullen, J. S. 2023. Inclusive entrepreneurship: A call for a shared theoretical conversation about unconventional entrepreneurs. *Journal of Business Venturing,* Vol. 38.

Crompton, H. and Burke, D. 2023. Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, Vol. 20. https://doi.org/10.1186/s41239-023-00392-8

Imiren, E., Lassalle, P., Mwaura, S. and Nicolopoulou, K. 2023. Understanding legitimacy building in contexts through digital entrepreneurship. *International Journal of Entrepreneurial Behavior & Research,* Vol. 30, No. 2/3, pp. 548–71.

Miller, K. 2020. *10 Characteristics of Successful Entrepreneurs.* (Accessed 11 April 2025.)

Tang, J., Kacmar, K. M. and Busenitz, L. 2012. Entrepreneurial alertness in the pursuit of new opportunities. *Journal of Business Venturing,* Vol. 27, No. 1, pp. 77–94.

About the authors

Efe Imiren is a Senior Lecturer and Entrepreneurship Academic at the University of Suffolk, Ipswich, United Kingdom. Efe researches entrepreneurial experience in digital contexts, particularly legitimacy building processes, women entrepreneurs and AI. Efe also leads the Enterprise Innovation Research Initiative, a collaborative forum for advancing context-based and enterprise-focused scholarship.

Katerina Nicolopoulou is a Professor of Entrepreneurship in Ajman University in the United Arab Emirates and the Head of the Global Research Cluster for Entrepreneurship and Entrepreneurial Leadership. Her expertise focuses on Social, Sustainable and Diversity aspects of Entrepreneurship and Innovation. Her current research interest is focused on issues regarding 'humanizing the business school in the age of AI'.

AI and accessibility: Abdicating engagement?

Alice Bennett

AI promises new tools for accessibility, but true inclusion needs human engagement. Generative fixes risk masking effort with the illusion of access.

AI is now something of an umbrella term, but the phrase 'artificial intelligence' was coined in 1956 to describe computers doing tasks in ways that mimic human abilities (Gold, 2023). While current debates focus on generative AI, meaning AI designed to generate new outputs based on existing inputs, more traditional AI has long been a staple of assistive software and accessibility tools, such as spell checkers or dictation tools.

These accessibility tools help the user produce work but don't create content. Spotting typing errors in your essay is very different from writing passages of it; transcribing your dictation is not composing a text for you. However, with generative AI, the clue is in the name. Rather than mediating the process, it generates material – be that image, text, music or code.

Elsewhere, the distinction can be less clear cut, with tools using different elements of automation and generative functions to achieve an overall purpose. As AI features are added to more accessibility tools, the promises of what can be achieved with less and less intervention increase. But what are the resulting implications?

Although the term 'accessibility' can be applied to general access to materials (for example, in the context of open access publishing), in this context it refers to the ease of use for people with disabilities. Further, 'digital accessibility' refers to the principles and processes of making digital content that disabled users can fully engage with. That is, ensuring that both the digital output can be accessed by those using assistive technologies (such as screen reader software), and the content itself (such as image description, transcripts and captions) is functional and readily available to all users.

Some countries have digital accessibility standards for types of digital content, such as the web accessibility regulations for public sector bodies in the United Kingdom (2018). Elsewhere Digital Accessibility legislation is applied more widely. For example, European Union digital accessibility web regulations, which came into force in 2025, include both the public sector and commercial services, such as online banking (European Parliament and Council of the European Union, 2019). Internationally, good digital accessibility practice was promoted by nonprofit organizations, such as the World Wide Web Consortium (W3C, n.d.) through their Web Accessibility Initiative (WAI, 2026), but locally applied legislation means that global standards vary.

Building in accessibility

Accessibility by design is always easier than a retrofit, whether addressing physical or digital accessibility. The user-centred concept of Universal Design – designing spaces, services or products to be accessed, understood, and used by all people without the need for specialist adaptation – promotes design which innately accommodates a range of different needs (the Irish Centre for Excellence in Universal Design [n.d.] offers a great introduction to Universal Design, as well as its application to education). Adopting Universal Design principles is a far more inclusive approach than addressing contextual requirements as an afterthought. For example, the WAI (2026) advocates for new technologies and digital content to be designed accessibly from inception (sometimes products designed this way are termed 'born accessible').

In contrast, the ways in which some generative AI tools are marketed presents them as means to render content more accessible *without* having to adjust practice or mindset. This perpetuates the perception of digital accessibility as an added task, rather than an integral part of creating digital content.

AI tools promising automatic accessibility fixes with little to no need for human intervention are tempting, offering a quick and easy solution to often overworked staff. However, real accessibility requires engagement.

Moreover, this corresponds with attitudes prevalent in education. Whilst the accessibility of digital teaching materials is generally recognized as important, and even a legal requirement for teaching materials in some countries, it frequently remains an afterthought. Consequently, creation of accessible digital materials is typically excluded from teacher training programmes, leaving staff without the requisite design skills. In this environment, AI tools promising automatic accessibility fixes with little to no need for human intervention are tempting, offering a quick and easy solution to often overworked staff.

However, real accessibility requires engagement. An example of this is auto-generated alternative text, or 'alt text' – either built-in platform tools or external tools that generate a short image description without any user labour. If you haven't written alt text before, employing AI-generated text can be a helpful starting point – how does the software interpret the image? But the skill of image description is providing the nuance of context. If the image isn't just decorative, why did you choose it? Did you add it as a typical example or an exception? What is it that prompted you to choose that particular picture to support your message?

Even when AI can correctly identify the subject of an image, it cannot identify the message you wanted to convey – a disadvantage, for example, to people with blindness or those who have limited vision. The auto-generated text can provide a starting point, a description you can edit as needed. But defaulting without checking or editing a generated description doesn't necessarily fulfil the objective of image description. Unhelpful, incomplete or inaccurate image description does not provide equivalent access and could even create

confusion or misinformation for anyone relying on this to fully engage with the content. The mere presence of an alternative text might make your content ***appear*** accessible, but unless it is meaningful, it is merely an illusion of inclusivity.

Automated accessibility checkers and the rise of synthetic 'user feedback'

Accessibility checker tools also aid the development of web resources, running automated checks to identify accessibility issues. But like AI-generated image descriptions, these tools should only function as a starting point: they can help in the creation of a tool that is technically accessible, but they cannot replace the work done with disabled users to make a tool practically useful. For example, automated accessibility checkers can help you determine if a web page can be accessed with a screen reader, but they cannot tell you how it will be experienced by human users. Similarly, a checker can identify if alternative text for an image is present, but not if that text is appropriate. A page with graphics with just the word 'image' as alt-text with no further description given will nonetheless pass an automated accessibility check simply because alternative text is present – even though it is of no help to the end user.

The mere presence of an alternative text might make your content *appear* accessible, but unless it is meaningful, it is merely an illusion of inclusivity.

Automated accessibility checkers could help clear the first hurdles of making resources digitally accessible, but they cannot replace user research and testing as a means of understanding user experience. And yet AI tools are being marketed as a better means of checking web accessibility, with claims like that of the company Accessi (2024) that 'AI algorithms can analyse user behaviour data to identify common navigation issues faced by users with disabilities.' What user behaviour data is being analysed, and where does it come from?

Automated accessibility checkers could help clear the first hurdles of making resources digitally accessible, but they cannot replace user research and testing as a means of understanding user experience.

Disabled individuals are frequently highly skilled users of assistive technology, typically finding their own workarounds. So, without their direct input, can AI tools truly reflect their own experiences? A 2024 article looking at the rise of synthetic users – AI-generated respondents for user research – concluded that, whilst helpful for some tasks such as preparing for user testing, synthetic users should not replace testing by human users, as they cannot accurately predict human experience (Rosala and Moran, 2024). The authors go so far as to warn the reader to 'Avoid adopting this tool if your stakeholders will see it as a replacement for user research.' While these tools have great potential to support accessible development, they can't replace final rounds of user testing by the people who are meant to use them.

Will accessibility remain an afterthought?

Reliance on automation and AI tools eliminates nuance, giving a poorer experience for disabled users. To continue down this path could lead to an automated era of disability erasure. The design of policies, services and products for disabled people has historically been done by people without disabilities, leading to the adoption of the motto 'Nothing About Us Without Us' by disability movements (UN, 2004). However, dependence on AI tools without human input and disabled persons' lived experiences is antithetical to that principle of disabled inclusion. If AI tools lead to automation replacing all manual testing and synthetic users replacing user research, we now potentially risk the design process excluding not only disabled people but also any human testing in favour of AI.

This promotion of AI as a tool of efficiency begs the question: why is accessibility not an inherent part of the workflow but tacked on *after* the creation process? The accessibility of teaching materials should be considered in the design of the materials, not simply an afterthought perceived as an annoyance. If digital teaching materials are not accessible, they are not finished.

Uncritical reliance on these tools risks perpetuating the notion of accessibility as an afterthought, promising to automate accessible features without the creator providing nuance or attention to the experience of disabled users. However generative AI ends up being used, we need to adopt inclusive design practices and build accessibility into products and services throughout their design, development and implementation.

This promotion of AI as a tool of efficiency begs the question: why is accessibility not an inherent part of the workflow but tacked on after the creation process?

It's easy to see why AI tools that promise to resolve digital accessibility issues with little to no input are attractive – they claim to present an easy solution, one all the more tempting to staff potentially lacking digital skills and confidence, or with limited time to prepare teaching materials. However, whilst in the future automated and AI tools may help us create accessible content in some form, a meaningful accessible experience requires engagement, both in education and society as a whole.

References

Accessi. 2024. Should you consider AI for accessibility testing?

Center for Excellence in Universal Design. n.d. *About universal design*.

European Parliament and Council of the European Union. 2019. *Directive (EU) 2019/882 on the accessibility requirements for products and services (European Accessibility Act)*. Official Journal of the European Union, L 151, pp. 70–115. (Doc. 32019L0882.)

Gold, E. 2023. The history of artificial intelligence from the 1950s to today. *freeCodeCamp*.

Rosala, M. and Moran, K. 2024. *Synthetic users: if, when, and how to use ai-generated "research"*. Dover, Nielson Norman Group.

United Kingdom of Great Britain and Northern Ireland, Government Digital Service and Central Digital and Data Office. 2018. *Understanding accessibility requirements for public sector bodies*.

United Nations. 2004. *International Day Of Disabled Persons 2004*. New York, United Nations.

World Wide Web Consortium (W3C). n.d. *About us*.

––. 2026. *About W3C WAI*.

About the author

Alice Bennett has worked in education, museums and libraries, and now works in the Digital Accessibility Unit at the University of York. She is co-chair of the Inclusive Network Connecting and Linking University Disabled Employees (INCLUDE), the disabled staff network at the University of York, and a member of the advisory board of the university's Equity, Diversity and Inclusion Research Centre.

Part IV

How will we think, question and express ourselves?

Education has long grappled with the central yet intangible nature of thought and knowledge. What does it mean to know? What is the relationship between thinking, understanding and expression? What role does memory play? How do we arrive at greater or lesser degrees of certainty? Which questions are worth examining and why? How 'neutral' can knowledge truly be, and how does bias inform perception? How can we confront biased patterns of thought that deepen inequalities and exclusions along lines of ethnicity, gender, and socioeconomic status? These and many other questions go far beyond education across centuries of philosophy, theology, and science, yet education has long been a crucial context in which our underlying conceptions of thinking, being, and doing are mobilized in action.

A persistent challenge in education has been the inclusion of humanity's great epistemic diversity in education systems, curriculum and pedagogy. James Maisiri and Solomon Musonza explore how the introduction of AI risks exacerbating cultural and epistemic homogeneity in **'The cultural costs of AI in Africa's education systems'.** Maisiri and Musonza point out that all technological systems are encoded with certain cultural values from their designers – some of which are explicit but many of which are implicit. These values are transmitted into educational contexts, whether or not they are appropriate for, or aligned with, local values. They call on educators and decision-makers 'to think critically about whose values and worldviews are echoed in the designs of AI educational tools' (p. 90).

Zhaiyk Sultan considers similar concerns in the future of literature education in the think piece **'Simulating the soul: AI and the legacy of Kazakh literature'.** Sultan notes that using AI in literary analysis cannot replicate the intention of Kazakh poets' distinctive political, cultural and intellectual commentary, even if it can mimic their style and cadence. Aristotelis Ioannis Paschalidis likewise explores the intimate connection between thought and language in **'AI and the great linguistic flattening'.** Using algorithmically generated text increases the likelihood of falling into what Paschalidis

calls the 'default approach problem', in which users, particularly students, are more likely to defer to the 'authority' of the LLM with regard to word choice, selecting phrasing predicted by LLMs rather than crafting unusual expressions or unique vocabulary. Over time, Paschalidis worries about what this will mean for the authenticity of written expression and even patterns of thought, when certain uncommon words might become obsolete or remain only in an impoverished form.

Victoria Livingstone and Jeppe Klitgaard Stricker argue that working through unclarity and ambiguity constitutes one of the most important educational stages of academic learning and research. In **'The disappearance of the unclear question',** they share how helping students to arrive at a clear and original research question could previously take weeks, if not months, of reading, reflecting, testing and wrestling with the fundamental assumptions and focus of their research. The result would often be surprising and original, impossible to have foreseen from the start. LLMs, by contrast, can generate clear and manageable research questions in seconds, refining and narrowing with only a few successive prompts. While such efficiency might seem like a win, Livingstone and Striker fear that over time we risk conflating 'better prompts' with 'better thinking' and that researchers and educators 'may begin to ascribe value to not what we want to know, but what the machine is best at answering.' (p. 104).

How education addresses these questions is likely to have broader, lifelong implications. The habits that students form now with generative AI can potentially last a lifetime. The think pieces in this section point out that once we have ceded parts of our own minds and expression to algorithmic assistance, it is unclear how (or whether) we can easily reclaim them. How much are we as humans 'using' AI to think and express ourselves, and how much is the technology itself shaping what questions we ask, what answers we accept, and what styles of expression become 'natural'? What happens if language and writing no longer serve as essential mediums when developing and communicating complex thought? What ways of thinking, knowing, and critiquing become lost in the cultural and epistemic homogenization of defaulting to the worldviews of LLM designers? What is at stake in these questions goes far beyond educational consequences, extending to the preservation of human intellectual agency for generations to come.

The cultural costs of AI in Africa's education systems

James Maisiri and Solomon Musonza

AI tools in African education risk erasing Indigenous values; local ownership and culturally grounded AI are vital to preserve knowledge and ensure equity.

In recent years, a dominant narrative has emerged advocating for AI as a means to improve education systems. As such, the African education sector is increasingly utilizing AI educational tools with the promise that these technologies can assist teachers with their daily tasks, provide personalized learning to students, and enhance pedagogical practices. However, an important dimension of this conversation is often overlooked: technology is value-laden; thus, it promotes, restricts and shapes the cultures of those receiving these technologies. Perhaps we in Africa's education systems need to think critically about whose values and worldviews are echoed in the designs of AI educational tools.

Many scholars have concluded that technology is not neutral but often inherently represents the ideologies, cultural practices and political values from the context in which it is conceived (Van de Poel and Kroes, 2014). For example, feminists have argued that patriarchal cultures have informed the design of technologies from a masculine perspective (Hipolito, Winkle and Lie, 2023). Examples include AI assistants like Siri and Alexa, which are designed to be subservient, friendly, flirty, female-voiced assistants that follow every instruction, perpetuating patriarchal stereotypes (Strengers and Kennedy 2021). In Japan, some technology design mirrors Shinto spiritual beliefs that non-human objects have spiritual agency, as seen in AI-powered robotic pets like Sony's AIBO dog, created to reflect societal values of companionship (Jensen, 2013). Sea dikes and speed bumps are also technologies designed to embody the societal value of protecting human life (Van de Poel and Kroes, 2014).

Perhaps we in Africa's education systems need to think critically about whose values and worldviews are echoed in the designs of AI educational tools.

If we accept that technology can shape and inherently hold cultural values, then this means that a transfer of technology is a transfer of their creators' cultural values (Majumder and Tripathi, 2020). In education, the One Laptop Per Child (OLPC) project implemented in various countries disrupted traditional communal learning in Nigeria and elsewhere by promoting individualized, self-directed education, clashing with local values (Ezumah, 2020). Given that AI educational tools often originate from the Global North's cultural contexts and are increasingly utilized in African education, one must ask: how will these embedded values and biases be received?

Nyaaba, Wright and Choi (2024) argue that AI educational tools are trained to produce curricula and information that are geared towards technologically advanced Western

nations, frequently ignoring Indigenous customs and knowledge. For example, many African societies' education systems include agricultural practices based on contextual ecological knowledge. However, this knowledge is often excluded from AI-generated curricula. Therefore, students who use AI receive a cultural transfer of external methods while being deprived of Indigenous learning content. This trend of epistemic erasure of African farming methods is not new and has been historically ushered in via imperialism and colonization through industrialization processes, but AI risks further erasing Indigenous knowledge. For example, Western rationality perceives plants as objects, whereas Indigenous Eastern African knowledge sees plants as beings with spirit and agency. Yet these East African values are often absent from AI-generated agricultural curricula (Foster et al., 2023). Furthermore, when researchers asked ChatGPT and Gemini how many seasons there are, they both responded 'four'. In West Africa, however, there are two main seasons, thus, this answer demonstrates how these tools overgeneralize climactic fluctuations and struggle to provide accurate, contextual educational facts (Nyaaba, Wright and Choi, 2024).

Moreover, many systems of African indigenous education have a community focus. For example, the Maasai people and Kipsigis Indigenous education in Kenya emphasize social interpersonal skills as a matter of cultural importance (Kerubo, 2016). However, many educational technologies are designed around values of individualized learning. Tools like ChatGPT similarly promote individualized, screen-based learning, which threaten a culture of communal, collaborative and interpersonal learning, as some studies have found correlations between students' use of AI at school and reduced social adaptability (Lai et al., 2023; Kouam and Muchowe, 2024).

> Given that AI educational tools often originate from the Global North's cultural contexts and are increasingly utilized in African education, one must ask: how will these embedded values and biases be received?

Beyond curricula, appropriation of language also presents its own threat to cultural values. Languages cannot be reduced to just communication; they embody a specific epistemological framework. Educational AI tools are multilingual but often monocultural, retaining cultural biases from the language of their training data, largely excluding both Indigenous epistemologies and languages. For instance, in some African communities, spirituality is important to the educational curriculum. However, when Gemini was asked to generate a different phrase meaning 'God is Good' in Gurune, a language spoken in Ghana, the response demonstrated Gemini's inadequate proficiency in this language and lack of knowledge about it (Nyaaba, Wright and Choi, 2024). By nudging African students to conform to Western speech and the worldviews it embodies, rather than embracing their Indigenous languages and knowledge systems in AI tools, a cultural transfer in their education is evident.

Digital ecosystems create 'digital territories that, like physical spaces, have the propensity to become sites… of digital-territorial coloniality' (Mohamed, Png and Issac, 2020, p. 65). African education systems have largely imported AI tools with little consideration for epistemological consequences. Africa's technological dependency on the Global North leaves it vulnerable to a one-way transfer of values and cultural frameworks, resulting in

cultural displacement. African educational institutions and governments need to create their own AI educational tools that reflect Indigenous knowledge systems, while aggressively educating their populace in Science, Technology, Engineering and Mathematics (STEM) education to break technological dependence on the Global North and shape their own destinies.

Beyond top-down reforms, there are grassroots movements within Africa that are fighting against AI bias and epistemic injustice in their educational context. For example, *Masakhane* is a volunteer-based pan-African initiative that is building a Natural Language Processing (NLP) network trained by African students, teachers, researchers and coders. This initiative aims to translate online information into the mother tongues of diverse learners (Ravindran, 2023). They launched the 'Decolonize Science' initiative, which initially aimed to translate published preprint STEM research papers into multiple African languages. As a step towards that goal, they have begun by translating the abstracts of these papers into several African languages. *Masakhane's* AI systems are often created through participatory research that promotes open collaboration via community inputs, such as recorded community discussions or local newspapers (ibid). This participatory research process democratizes AI and fosters local ownership of Indigenous knowledge production. The presence of grassroots initiatives alongside national strategies can assist African education systems to create equitable and culturally relevant AI solutions. Beyond AI training solutions, meaningful AI use in African education systems must be accompanied by ownership of the AI educational tools. African educational systems must build the capacity to determine when and how to use these AI tools within the African context.

> By nudging African students to conform to Western speech and the worldviews it embodies, rather than embracing their Indigenous languages and knowledge systems in AI tools, a cultural transfer in their education is evident.

References

Ezumah, B. A. 2020. *Critical perspectives of educational technology in africa: Design, implementation, and evaluation*. London, Springer Nature. (Accessed 25 October 2024.)

Foster, L., Szilagyi, K., Wairegi, A., Oguamanam, C. and de Beer, J. 2023. Smart farming and artificial intelligence in East Africa: Addressing indigeneity, plants, and gender. *Smart Agricultural Technology,* Vol. 3. (Accessed 25 January 2025.) https://doi.org/10.1016/j.atech.2022.100132

Hipólito, I., Winkle, K. and Lie, M. 2023. Enactive artificial intelligence: Subverting gender norms in robot-human interaction. *Frontiers in Neurorobotics*, Vol. 17. (Accessed 24 February 2025.)

Jensen, C. 2013. Techno-animism in japan: Shinto cosmograms, actor-network theory, and the enabling powers of non-human agencies. *Theory, Culture & Society,* Vol. 30, No. 2, pp. 84–115. https://doi.org/10.1177/0263276412456564

Kerubo, I. P. 2016. African indigenous education as practiced by the Maasai of Kenya. Master's thesis, University of Nairobi, Kenya. (Accessed 14 December 2024.)

Kouam, A. W. F. and Muchowe, R. M. 2024. Exploring graduate students' perception and adoption of AI chatbots in Zimbabwe: Balancing pedagogical innovation and development of higher-order cognitive skills. *Journal of Applied Learning and Teaching,* Vol. 7, No. 1, pp. 1–11. https://dx.doi.org/10.2139/ssrn.4748274

Lai, T., Xie, C., Ruan, M., Wang, Z., Lu, H. and Fu, S. 2023. Influence of artificial intelligence in education on adolescents' social adaptability: The mediatory role of social support. *PLoS One,* Vol. 18, No. 3. https://doi.org/10.1371/journal.pone.0283170

Majumder, M. and Tripathi, A. K. 2020. Transformative power of technologies: Cultural transfer and globalization. *AI & Society,* Vol. 38, pp. 2295–303. (Accessed 15 April 2025.) https://doi.org/10.1007/s00146-021-01144-w

Mohamed, S., Png, M.-T. and Isaac, W. 2020. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, Vol. 33, pp. 659–84.

Nyaaba, M., Wright, A. and Choi, G. L. 2024. Generative AI and digital neocolonialism in global education: Towards an equitable framework. arXiv preprint, arXiv:2406.02966. (Accessed 25 January 2025.)

Ravindran, S. 2023. AI often mangles African languages. Local scientists and volunteers are taking it back to school. *Science,* Vol. 281. (Accessed 15 May 2025.) https://doi.org/10.1126/science.adj8509

Strengers, Y. and Kennedy, J. 2021. *The smart wife: Why Siri, Alexa, and other smart home devices need a feminist reboot*. Cambridge, Mass., The MIT Press. (Accessed 15 January 2025.)

Van de Poel, I. and Kroes, P. 2014. Can technology embody values? P. Kroes and P.-P. Verbeek (eds), *The Moral Status of Technical Artefacts.* Netherlands, Springer Nature, pp. 103–24. (Accessed 17 November 2024.). https://doi.org/10.1007/978-94-007-7914-3_7

About the authors

James Maisiri is a researcher and Ph.D. candidate at the University of Johannesburg, focusing on how digital technologies shape the labour markets within an African context, with implications for education and skill development.

Solomon Musonza is a research intern at the University of Johannesburg's Centre for Social Development in Africa, focusing on students' use of ChatGPT within an educational setting.

Simulating the soul: AI and the legacy of Kazakh literature

Zhaiyk Sultan

AI can mimic literary styles, but it lacks the moral depth, historical context and critical thought that define true human creativity and intellectual legacy.

The intersection of AI and human intellectual traditions is one of the most fascinating paradoxes of the modern era.

On the one hand, AI has been heralded as an instrument capable of preserving, analysing and even replicating human creativity. On the other, it is an extension of deeply entrenched computational systems that, by their very nature, lack the fundamental attributes of human cognition – contextual awareness, moral reasoning and the ineffable quality of subjective experience.

The intersection of AI and human intellectual traditions is one of the most fascinating paradoxes of the modern era.

Great literature doesn't just echo form – it captures the conscience of its time.

Kazakh literature, as embodied in figures such as Abai Qunanbaiuly, Magzhan Zhumabayev and Mukhtar Auezov, provides a case study in the transmission of cultural knowledge across generations. These figures did not merely produce texts; they shaped the epistemic structures of their society. Their works served as cognitive maps, encoding within them moral imperatives, national consciousness and philosophical inquiries that continue to resonate today. The question before us, then, is whether AI – an entity devoid of historical memory, ethical agency or lived experience – can meaningfully engage with these intellectual traditions, and if so, to what extent this engagement contributes to our understanding of both AI and the literary canon.

To approach this question, we must first consider the nature of learning itself. The works of Abai, for instance, are not merely poetic exercises; they are didactic, designed to cultivate intellectual autonomy and moral reflection. His *Kara Sozder (Book of Words)* functions not as an isolated literary artefact but as a philosophical dialogue with his audience, urging them towards rational inquiry (Qunanbaiuly, 2001). The replication of Abai's style by AI presents an epistemological dilemma: is it sufficient to reproduce patterns of language, or must there exist an underlying intent, an understanding of the ethical and cognitive function of the text? If Abai were to write today, it is conceivable that he would engage with contemporary ethical concerns – perhaps the commodification of knowledge, the erosion of traditional learning systems, or the moral dilemmas posed by AI itself. AI can certainly generate text that mimics the cadence of his words, but it remains an open question whether this constitutes an extension of his intellectual legacy or a mere exercise in statistical approximation.

Great literature doesn't just echo form – it captures the conscience of its time.

AI tempts us with the promise of bridging past and present – making literary traditions more accessible, more alive.

Magzhan Zhumabayev presents a different challenge. His work, characterized by deep emotional resonance and nationalist sentiment, serves as an example of literature as an instrument of identity formation. His poetry was not merely expressive; it was ideological, shaped by a profound historical consciousness. To suggest that AI could replicate this function is to misunderstand the nature of human creativity. AI lacks historical embodiment; it does not exist within a specific cultural or political milieu. It can analyse patterns of diction and generate new poems in Zhumabayev style, but this would be the mere reconstruction of a structure without substance – akin to a lifeless shell. If Zhumabayev were alive today, he might write about ecological destruction, the displacement of Indigenous cultures, or the consequences of digital colonization. AI, however, would have no intrinsic motivation to do so; it lacks the existential concerns that drive human expression.

Mukhtar Auezov, through his expansive and deeply introspective narratives, offers yet another lens through which to view this issue. His *Path of Abai* does not simply recount events; it interrogates them, situating them within a broader historical and philosophical framework (Auezov, 2003). It is a model of literature as a form of critical inquiry. AI's application in this domain raises profound questions about narrative agency. If an AI were to write in Auezov's style, it might be able to generate plausible historical fiction, but would it engage in critical analysis? Would it recognize the implicit power structures embedded in Auezov's work? The answer, quite clearly, is no. AI does not possess a theory of mind; it does not engage with texts as intellectual discourse but as data to be processed and recombined. The capacity for critical thought – the ability to question, to dissent, to reinterpret – is entirely absent in AI systems.

AI tempts us with the promise of bridging past and present – making literary traditions more accessible, more alive.

AI can assist in textual analysis, language learning and even creative experimentation. True literature is not just pattern — it's meaning. And no machine can fake the human struggle for meaning. AI-generated poetry in the style of Abai or Zhumabayev is an approximation, a simulation, not an extension of their intellectual legacy.

AI may have a place in literary education – but only as a tool for inquiry, never a substitute for the ethical and intellectual labour of human writers.

To engage meaningfully with Kazakh literary traditions, we must recognize that literature is not merely a set of linguistic patterns but a vehicle for critical thought, moral reasoning and historical consciousness. AI can assist in studying these traditions, but it cannot participate in them as an intellectual agent. The danger, then, is not that AI will replace human writers – it will not, because it cannot – but that we might mistake its outputs for something they are not. To do so would be to fundamentally misunderstand both the nature of AI and the essence of human creativity itself.

The question is whether we, as humans, can continue to think critically about the implications of AI's role in knowledge production. AI may have a place in literary education – but only as a tool for inquiry, never a substitute for the ethical and intellectual labour of human writers.

References

Auezov, M. 1942–1956. *The Path of Abai*. Trans. by A. Abilov and D. Askhatov. Almaty, Zhazushy Publishing House (2003).

Qunanbaiuly, A. 1891–1904. *The Book of Words*. Trans. by K. Tastanbekov. Almaty, Ana Tili (2001).

About the author

Zhaiyk Sultan is a Principal Examiner at the International Baccalaureate Organisation and an education strategist exploring the intersections of ethics, culture and emerging technologies.

AI and the great linguistic flattening

Aristotelis Ioannis Paschalidis

Default language patterns in AI-generated text raise concerns about authenticity, repetition and the growing need for skilled user intervention.

The phenomenon of overused words and the default approach

Each time I come across phrases like 'underscores the importance of', 'I am particularly drawn' or 'delve into the intricacies', I become outright suspicious. Even if I cannot prove with certainty that I am dealing with entirely machine-generated text, I am fairly certain the text has at least been edited by one or more LLMs.

LLMs tend to favor certain words and phrases while disfavouring others. The same principle extends to syntax, punctuation, orthography, word arrangement and overall writing style, which may include, *inter alia,* the use of shorter sentences for clarity, limited use of subordinate clauses and so forth.

On the one hand, both the salience and quantity of data on which these models have been trained are crucial for their quality and performance; on the other hand, the embedded response guidelines effectively define the user experience. Taken together, these two elements crystallize what I'll call the *'default approach',* which in turn generates the corresponding *problem* (DAP).

Insofar as default approaches do not suit every context, they are no panacea. Moreover, features such as Claude's dropdown menu – which allows users to select how they wish the model to respond – or the availability of different task-specialized versions of ChatGPT are not genuine solutions to the aforementioned issue. Rather, they simply produce a 'sub-default approach', causing the same narrative to resurface at a smaller scale.

High-level prompt engineering skills are required to circumvent phenomena arising from default approaches. However, even if one possesses the necessary skill set, LLMs often hallucinate, gravitating towards their default settings, thus making iterative re-prompting all but inevitable, sometimes to the point of model collapse. Consequently, the 'default pull' is a parameter one must systematically factor in.

The gist is that using LLMs implies accepting them for what they are, practically speaking: adequately reliable, quick polymaths suitable for tireless assistants.

Efforts to proactively prompt models to avoid common errors and/or customize outputs idiosyncratically – aiming thereby to evade detection – could lead to an antidote effect. Again, empirical knowledge about recurring bugs and errors specific to each LLM, as well as familiarity with what machine-generated text detectors are designed to flag, is necessary. This

typically demands substantial expertise, considerable oversight time and potentially even insider information, often unavailable to users.

The gist of this is that using LLMs implies accepting them for what they are, practically speaking: adequately reliable, quick polymaths suitable for tireless assistants. The question that naturally arises is how unwearied one is when it comes to critiquing and filtering the content requested from, and subsequently provided by, an LLM.

From academia to kindergarten: AI as the new philologist

One of the areas experiencing significant yet questionable impact since ChatGPT launched is academic writing. To be sure, most major academic publishers offer paid pre-submission editing and proofreading services, targeting primarily authors who are non-native English speakers or haven't been educated in Anglophone institutions, to ensure their articles are polished, readable, engaging and have a 'native speaker' tone. These services are, if nothing else, somewhat pricey compared to a monthly subscription to an LLM, which might tempt one to replace human professional editing if possible.

Nevertheless, no one guarantees that professional editors providing these services do not themselves tap into AI tools to achieve the impeccable results clients expect. The common excuse for using LLMs could be something like: 'LLMs help me do quicker what I am, in principle, capable of doing myself anyway'. This isn't the case, it's not even a white lie! It actually constitutes bait one willingly bites, and from swimming in a lexical ocean, one is dragged ashore – only to be left dry.

We're all aware that speed and deadlines matter, as does quality. Yet the job at a certain academic level is so meticulous and subtle that skilled individuals might not be willing to perform it for others without high compensation. Potential clients might equally resist paying high rates, given current alternatives. This friction between expertise and affordability delineates an uncharted grey area where heavy use of LLMs comes into play.

My initial worry solidifies: Why ask an AI tool in the first place if you are not going to at least consider – if not almost blindly follow – its advice in the end? It sounds rhetorical, doesn't it?

Why ask an AI tool in the first place if you are not going to at least consider – if not almost blindly follow – its advice in the end? It sounds rhetorical, doesn't it?

With this in mind, an amalgam of human-written and machine-generated text will very likely become the new norm in academic writing.

Still, there's an underlying issue: If the machine-generated text share remains largely uniform for all authors – due to the previously mentioned DAP – then linguistic ill-homogeneity lurks.

The seemingly high probability of this scenario raises a plethora of intriguing questions: First, how much loss of identity is one willing to sacrifice for efficiency? Second, is there a threshold or ceiling to how much LLM use constitutes too much? When do low-quality outcomes become indisputably unacceptable given the widespread availability of such

tools? More precisely, what parameters should define our tolerance for AI use in the world of letters? Saying the same thing with different words, does it change the meaning? Formally speaking, no. It does, however, profoundly alter the experience for both readers and writers.

If students don't encounter new words that could enrich their vocabularies in the very settings where such enrichment was supposed to occur in the first place, their linguistic and intellectual horizons will be difficult to expand.

The above raises another concern: If some of our best minds, scientists and authors rely heavily on LLMs as editors and/or fact-checkers, shouldn't we anticipate that less experienced individuals with inferior skill sets will rely even more on AI tools? Let me shed some light on this by turning to the juvenile population.

Consider the following hypothesis: If our children's teachers systematically utilize AI, isn't it rational to expect that students – who learn primarily by doing – will, from an early age, face dilemmas between complying with implicitly imposed, AI-generated instructions and criteria or else risking failure? If AI's strong influence begins to shape teachers' linguistic expectations and communicative habits – flattening their use of vocabulary, syntax or evaluative nuance – wouldn't this same flattening pass more easily down to students, who are still developing their expressive and interpretive range? Hypothetically, if AI is the victimizer, might schools themselves be among its earliest victims – pedagogical environments compromised in linguistic form before their students are shaped by them?

If students don't encounter new words that could enrich their vocabularies in the very settings where such enrichment was supposed to occur in the first place, their linguistic and intellectual horizons will be difficult to expand. This would require additional individual effort, dedication, willpower and time. Hence, one might argue that those environments lose their comparative advantage, since the time-related opportunity cost of participation becomes automatically immense.

From losing words to losing worlds

It is imperative to emphasize that the impact of LLMs on language flattening may be inversely related to the number of native speakers, the degree of dialectal variation and the overall linguistic diversity of a language – serving as a neat counterargument to the presumed universality of the phenomenon. Languages with smaller numbers of speakers may, in the long run, prove more resistant to flattening than larger ones.

More precisely, if we accept that standardization – not LLMs alone – is what drives flattening, and that LLMs merely reinforce what institutions already value, then the resistance of smaller languages may be underpinned by numerous factors. First, smaller or low-resourced languages tend to be less institutionalized. They lack extensive formal corpora – which in turn limits the volume of training data available to LLMs – they are rarely used in academic publishing, and there is minimal involvement from non-native speakers, both in production and in readership. The result is that pressure for internal standardization remains low, while

the need to rely on LLMs is significantly reduced. Native speakers in these communities tend to write directly in their own language without consulting AI tools, as immersion in everyday usage provides sufficient confidence. In fact, it seems more likely that any AI use on their part would be for the sake of deliberate linguistic enrichment, not of editing or grammatical correction. Moreover, such languages often feature a small group of linguistic gatekeepers – critics, professors, acclaimed writers – whose authority remains largely unchallenged.

The idea of outsourcing their authority to machines may be perceived not only as unnecessary, but even as a quiet relinquishment of their role. Lastly, due to their limited commercial appeal, these languages could eventually be overlooked altogether by LLM providers, for whom the cost of training and deployment offers little return.

Nevertheless, flattening can begin subtly: from schoolbooks and governmental policy documents to instruction manuals where texts are mundane, lack creativity and merely aim to meet the bare minimum. Given that the phenomenon in question is already observable in academia, however, I remain sceptical as to whether this view of the 'starting point' is realistic.

If certain words gradually disappear from our vocabulary, couldn't we argue that the worlds of which these words are constituents would consequently vanish as well?

If certain words gradually disappear from our vocabulary due to the DAP, couldn't we argue that the worlds of which these words are constituents would consequently vanish as well? What remains accessible, at best, is an impoverished version of these worlds, especially when such words signify concepts with real-life applicability. Imagine a post-AGI scenario in which words like 'convenience', 'integrity' or 'honour' become obsolete. If humans no longer encounter these words, could we realistically expect them to incorporate the corresponding concepts into their daily lives? Or are we assuming that reinventing or coining their equivalents would be a piece of cake?

Covert censorship or linguistic dictatorship?

AI tools will continue functioning partially as search engines and informational navigators; therefore, knowledge *per se* is not at risk of being lost. However, if we never encounter a word in the settings where it originally belonged and was expected to (re)appear – because it has been replaced by a more dominant synonym – we will likely never seek out its meaning, definition, connotations or genealogy. Just as happens with archaic words that fall out of usage, these words will gradually vanish from our vocabulary. Consequently, all sentences containing them become inexpressible at best and literally unthinkable at worst. This situation amounts to a form of indirectly imposed tacit censorship, what one might aptly call a Linguistic Dictatorship.

No doubt, this may sound excessively dystopian or Orwellian to some. Yet, it is not the emergence of a 'Newspeak' (as in *1984*) that looms large; rather, it is the threat of linguistic atrophy due to diminished vocabulary. This scenario appears more pressing, and potentially far more catastrophic, than conventional AI safety concerns.

References

Orwell, G. 1949. *1984*. London, Secker & Warburg.

About the author

Aristotelis Ioannis Paschalidis is a Ph.D. candidate in Philosophy at the University of Nottingham. He has also worked with European Union institutions in several European capitals.

The disappearance of the unclear question

Victoria Livingstone and Jeppe Klitgaard Stricker

AI speeds up research, but without friction, do students lose the deep thinking that turns bad questions into meaningful learning?

Several years before AI chatbots became publicly available, a student in one of our undergraduate courses wrote a paper on the disproven 'Mozart Effect' – the theory that listening to classical music boosts intelligence, particularly spatial reasoning. It took the student a month of mapping ideas and reading articles to refine the topic to a manageable scope. She had begun the semester with a desire to research the impact of music on cognition and ended with a paper dissecting the prevalence of a disproven myth. With an AI chatbot, this student could have narrowed her focus within minutes.

The appeal of generative AI is obvious – and used intentionally, at the right phases and for the right purposes, AI can be a powerful tool for learning and teaching. But there's a catch: as students and educators, we may begin to 'optimize' for what elicits useful responses from generative AI.

Without the aid of chatbots, our student searched databases, read articles and followed bibliographic trails. When she found conflicting information in the research, she had to think critically to evaluate the sources and theories presented. Ultimately, her research led her to a nuanced conclusion that disproven theories can still have beneficial effects.

Efficiency without friction

LLMs offer more direct paths. When we prompted ChatGPT4 to describe the Mozart Effect, the chatbot immediately returned the result that the theory was a controversial 'mix of science, hype, and a bit of myth!' It offered summaries of the research without the detail – and without transparency.

Generative AI, then, would have given our student quick responses but removed much of the friction – the uncomfortable, often slow process that requires students to critically evaluate sources, draw connections and refine ideas. AI offers quick answers, but the output of the chatbot was less complex than what our student concluded. Would she have gotten there working in reverse, with AI feeding her an idea and then looking for sources to back that up? Perhaps so, but that isn't true questioning – that's just finding sources.

> To refine research questions, students must be willing to reevaluate and revise their initial positions. It is within that discomfort – that cognitive friction – that real learning takes place.

Further, students often learn more from badly framed research questions than they do from good ones. Questions with obvious answers, topics too broad in scope, and intellectual dead ends can cultivate critical thinking and build intellectual stamina. In fact,

good questions often begin as bad ones followed by a productive struggle. The process of grappling with an unclear question is uncomfortable. To refine research questions, students must be willing to reevaluate and revise their initial positions. It is within that discomfort – that cognitive friction – that real learning takes place.

None of this, however, aligns well with generative AI. This technology rewards the illusion of precision and clarity, and an output-driven kind of logic. The student or teacher who can produce the most optimized, constrained and system-compatible prompt will get the best result. This is not a flaw but an integral design component. AI nudges us towards a particular kind of cognitive behaviour – one that favors instrumental clarity over depth. And this nudge, subtle and unannounced, is steering students and educators towards a new set of habits and assumptions.

From questions to prompts

In the short term, this feels like a win. But over time, we risk conflating 'better prompts' with 'better thinking'. We may begin to ascribe value to not what we want to know, but what the machine is best at answering. In an overheated educational system optimized for consistent outputs, and where grades and time are key, efficiency becomes a natural goal. But the shift is obviously consequential. The metric of value is no longer internal – curiosity, relevance – but external, driven by the pressure to produce.

Generative AI is a phenomenal technology for helping students achieve, but it generates output without necessarily building much actual understanding along the way. For some students, generative AI can enable academic agency. For others, the opposite is true. The danger here is not primarily that we offload learning, but rather that our intellectual priorities shift without discussion or reflection. We learn implicitly to ask the kinds of questions that work well for generative AI – and we stop asking the ones that don't. And in doing so, we risk losing something essential: the capacity to ask questions that do not yield immediate or satisfying answers. In many fields of study, including literature, philosophy and the humanities more broadly, ambiguity is not a problem to be solved. It is a resource for continuous critical thinking, renegotiating arguments and reflection.

We learn implicitly to ask the kinds of questions that work well for generative AI – and we stop asking the ones that don't. And in doing so, we risk losing something essential: the capacity to ask questions that do not yield immediate or satisfying answers.

Generative AI, however, is not designed to dwell in ambiguity. It can mimic ambiguity, yes. It can reflect the style of a speculative essay or a philosophical dialogue. But it is optimized for resolution. For outputs. For an articulated 'answer' that looks clean and complete.

This creates a subtle but powerful disincentive for ambiguity. Students may begin to avoid unclear questions, not because they lack value, but because they lack utility in an AI-mediated workflow. The fuzzy question comes to signify inefficiency in a

world where the path of least resistance is increasingly privileged over complexity and deep learning.

A collaborative approach to questioning

However, rather than seeing AI-friendly questions and intellectual inquiry as mutually exclusive paths, we might consider how they can work in sequence. Students could begin with AI-optimized questions to establish a foundation, then deliberately introduce the complexity and ambiguity that deeper learning requires.

Many students struggle to come up with research questions. Using LLMs as conversation partners may allow them to more quickly arrive at a topic with a manageable scope. Once they have a topic, however, they should start looking at sources (journal articles, primary documents, etc.) rather than relying only on the output of AI, which, while improving, is still unreliable in terms of accuracy and source attribution. Further, AI-generated summaries allow students to gloss over understanding.

Asking students to summarize research or paraphrase passages often reveals the limits of their understanding. Here, too, we return to the concept of cognitive friction. Reading closely – with the ability to draw connections between various ideas, identify inconsistencies and evaluate methodology –is slow, uncomfortable work. Teachers can assist students by annotating texts together or breaking students into groups to discuss research. Students can then use those discussions as a basis for in-class writing.

This two-stage approach acknowledges AI's utility while teaching students to recognize its limitations. The initial AI-assisted stage provides scaffolding, while the subsequent human-led inquiry builds intellectual muscles that only resistance and struggle can develop. The key is teaching students to see AI-generated content not as an endpoint but as raw material for more sophisticated questioning.

In this way, generative AI becomes not a replacement for traditional questioning but a new tool within a broader ecology of inquiry provided we explicitly teach the difference between optimizing a prompt and deepening a question.

Beyond the prompt

The problem we face regarding questions in education is not primarily a technological issue. It is a pedagogical and philosophical one. As educators, we must ensure that the purpose of questioning does not collapse into the mechanics of prompting. It is pivotal that students continue to value questions, not for their compatibility with AI, but for their capacity to open up new ways of thinking.

This means protecting, somehow, the kinds of questions that do not yield quick answers. Questions that are slow, layered, and unresolved. It means resisting the urge to reframe every problem in terms that make it legible to a machine. And it means teaching students

to recognize when a good question is one that cannot, or should not, be answered by an algorithm.

We need to reassert the value of cognitive friction. In an age of frictionless interfaces, learning must retain its difficulty. The struggle to articulate a question, to explore a problem without a clear solution, to follow a line of thought that leads nowhere – these are not inefficiencies. They are the essential processes by which human understanding grows.

> As educators, we must ensure that the purpose of questioning does not collapse into the mechanics of prompting. It is pivotal that students continue to value questions, not for their compatibility with AI, but for their capacity to open up new ways of thinking.

And yet, this is not a call to reject generative AI. On the contrary, generative AI models can have immense value in higher education. They can assist, extend and even inspire our thinking. But there's a chance we will start thinking differently, and the biggest loss will be the loss that is hardest to detect. We will still have conversations, and we will still teach and learn. But something vital may be missing before we know it.

Clearly, generative AI will continue to play a still more pertinent role in higher education. But in a world with plenty of fast answers, it's more important than ever to hold on to the purpose of the question itself.

About the authors

Dr Victoria Livingstone is a writer, educator and editor. She was a professor before joining Johns Hopkins University, where she is the managing editor of MLN, a peer-reviewed journal of literary scholarship. In 2024, her Time piece 'I Quit Teaching Because of ChatGPT' gained considerable international attention.

Jeppe Klitgaard Stricker leads the administration at Aalborg University's Department of Clinical Medicine and works as a speaker and advisor on generative AI in education through www.stricker.ai. He writes the newsletter 'The Future of Higher Education', with pieces republished in various international outlets – including a collaboration with the German Goethe-Institut.

Part V

How will we govern education as a common good?

UNESCO has long asserted that education is a public endeavor and a common good. Education reflects societies' shared values and commitments to the transmission of knowledge and capacities from one generation to the next. Education is a human right throughout life that draws from the rich knowledge commons that is the heritage of all of humanity. A sense of public mission, human rights and shared responsibility underlie a new social contract for education, in which everyone plays a part (UNESCO, 2021).

Education as a human right and education as a common good continue to be organizing principles in many contexts. However, the implementation of such principles faces new challenges when algorithmic technologies are introduced to education. The push for rapid integration of AI into education systems is not occurring in a vacuum; it is unfolding within longstanding structural weaknesses and threats to public education systems. Researchers are identifying AI, among other education technologies, as a vector for accelerated privatization, data extraction and the reorganization of education infrastructures around commercial and market-driven logics (Williamson, 2023), rather than humanistic values and participation in a shared knowledge commons.

In **'Open education principles: Resisting the metrics of AI black boxes'**, Tel Amiel, Glenda Cox and Colin de la Higuera point out that delegating educational content to algorithms essentially introduces a 'black box' that makes adequate oversight of the logics governing educational decisions nearly impossible to ensure. They see open education principles as crucial for ensuring that a 'post-AI' future is accountable to educational communities, rather than to corporate shareholders.

Should we measure the value of education as a public good by the same economic returns we expect from market goods? Guido Appoldt and Zhen Gong argue no, in **'Long-term economic implications where machines cannot replace humans'.** They use the lens of 'Baumol's cost disease' – an economic phenomenon explaining how labour-intensive sectors like education become progressively more expensive because, unlike manufacturing, productivity gains through automation are limited where human

interaction is essential. As education costs rise relative to other sectors, governments incentivize privatization and technology-driven efficiency. Appoldt and Gong offer the provocation that societies must consciously choose to adequately fund education as a public good to retain quality educators and experiences, rather than allowing fiscal arguments to justify its dismantling through privatization and AI-driven automation.

While all have a part to play in promoting the right to education, states have been the primary duty bearers in ensuring this right. This includes making decisions about curriculum, content and learning objectives (UNESCO, Right to Education Initiative, 2019). In her piece **'Who educates the LLMs? The uncertain role of the nation-state'**, Camille Fabo finds that in the context of AI, nation-states are playing a more uncertain role in the authority over knowledge and facts. As AI companies shape how knowledge is built and shared, this think piece warns that tech firms, not states, could emerge as new global educators in a 'post-truth' world.

Beyond education itself, the physical infrastructure required for AI imposes dramatic costs on our environmental common goods including water, energy, and public space. These costs are largely invisible to end users but devastating to local communities and, over time, earth systems such as climate and air quality. James O'Sullivan and Mark Kelly point out that higher education's inherent public mission, already under strain, faces new paradoxes and contradictions with the introduction of AI. In **'Environmental contradictions of adopting AI in higher education'** O'Sullivan and Kelly argue that these concerns should not be treated as an afterthought, and that 'to sustain the university as a site of ethical, critical and democratic inquiry, further explorations of what LLMs can do should be deprivileged, even paused, in favour of analyses of how they are reshaping the values and viability of education itself' (p. 125).

Finally, Kevin Martin suggests the importance of considering a wide range of diverse strategies for increasing access to educational content, platforms and tools, including those that might use AI. In his piece **'Could zero-rating advance meaningful connectivity in education?'**, he proposes zero-rating schemes as one avenue whereby governments and companies can align interests to strengthen access to educational resources. While he acknowledges that such strategies raise numerous governance questions about online connectivity as a public utility, commercial incentives, decision oversight and net neutrality, he highlights the urgency for greater future equity online, leading to greater inclusion of all countries in the design and usefulness of AI.

AI will have profound implications on the extent to which education continues to serve its crucial public mission. AI is novel, but to many it is only the latest development in a decades-long process of education itself being reorganized around market-driven logics. The concern for many is that AI 'integration' risks automating and institutionalizing these logics even further (Heinsfeld, 2025).

Open education principles: Resisting the metrics of AI black boxes

Tel Amiel, Glenda Cox and Colin de la Higuera

As AI tightens its corporate grip, education risks deepening divides. Openness – not black-box metrics – offers a path to equity, transparency and shared progress.

As the current trajectory of big tech, proprietary AI threatens to widen educational divides rather than narrow them. The financial and environmental costs of commercial AI tools, and their concerns with profit and development primarily for affluent markets risk leaving behind precisely the communities that could benefit most from AI's incorporation as an educational technology.

Increasingly powerful tools are concentrated in the hands of wealthy corporations and individuals, developed behind closed doors with little transparency or public accountability (Fraser, 2025). The prevalent 'black box' nature of AI systems serves to reinforce Global North hegemony, embedding existing biases and perpetuating inequalities. When educational technologies are developed primarily through Western, neoliberal perspectives, they inevitably reflect and amplify those worldviews – including a profit-making outlook towards AI and education – while marginalizing others.

Principles of openness can greatly improve the current AI ecosystem. Despite names like 'OpenAI', many of these systems are anything but open, representing what critics correctly identify as 'open-washing' – the marketing of proprietary systems as open while maintaining tight control over their functioning and critical aspects of their technical ecosystem. As Widder, Whittaker and West (2024, p. 831) suggest, 'even at its most maximal, open AI is highly dependent on the resources of a few large corporate actors, who effectively control the AI industry and the research ecology beyond.' We cannot ignore that currently, most of what is considered 'open' AI is closed in a number of critical key elements, including training data, weights and overall transparency regarding human labour.

We suggest a strategy of 'non-stupid optimism' in relation to educational technology (Facer and Selwyn, 2021). This approach acknowledges technology's potential benefits while remaining critically aware of how techno-solutionist attitudes often serve economic growth and capital accumulation rather than genuine educational needs. We must avoid the trap of seeing technological fixes (AI included) as complete solutions to complex educational challenges. Still, we can be optimists in the sense that AI systems can embody openness: digital public goods that are designed to be transparent in their structures and functioning and embody collaborative governance structures.

Many of these systems are anything but open, representing what critics correctly identify as 'open-washing' – the marketing of proprietary systems as open while maintaining tight control over their functioning and critical aspects of their technical ecosystem.

The emphasis on closed models based on efficiency and economic gains has greatly intensified a historical narrative of educational achievement as measurable in metrics of performance (grades, graduates, papers) and return on investment (Luke, 2023). We begin by presenting two examples of these logics: traditional, closed educational resources and current AI applications in education. Second, we briefly discuss how aspects of openness – technology, practices, resources and governance – can help us counter the current trajectory in AI in education.

A less sustainable education

The conventional design of educational content has been 'closed', tailored to specific use cases and audiences. This can be positive but is also limiting as substantial investments in development can lead to content that serves only very specific needs and interests and that are not meant to be changed, adapted or improved. Even if the educational resources may have been imagined by their authors as being of use in the long run, the way they are produced, delivered and curated in the context of the prevalent economic model makes them ephemeral.

> The pressure is on all actors of education to deliver: the teacher is given numeric indicators and benchmarks to meet, and ever-younger students are taught to measure their success by what grades they have achieved. But any activity that is purely defined by performance metrics and return on investment can likely be done more efficiently and cheaply by AI.

The most clear-cut example of this is printed textbooks. They are usually quite expensive, and, in some countries, governments will purchase textbooks and distribute them freely (through public subsidies). These textbooks often cannot be claimed by the student and are returned to school at the end of the term, until, after a period, newer editions are re-purchased by the government. Digital resources are not necessarily an improvement. Learning management systems will allow access to resources for a limited period, or within the specific scope of a course. Once the term of enrollment is over, students no longer have access to the course's digital materials. These design decisions evident in traditional content are now embedded in educational technologies, which help define the very nature of what we value in education.

A similar phenomenon can be observed in the current emphasis of optimization for AI in education. Optimization is where AI can thrive: for it to function, it needs numerical performance criteria. AI needs these for algorithms to be tuned as well as provide a selling point. AI performs better than humans at chess because it has a higher rating (Elo) score. It can also be better at medical diagnosis because it can detect more tumors than human doctors with an equivalent numbers of false positives (Lång, 2023). Error rate is a crucial indicator that allows algorithms to *appear* better than the alternative. More importantly, criteria allow systems to replace what are *ill-defined* learning problems by well-defined optimization problems.

Performance indicators thus define the problem to be solved: detecting more tumors means addressing a diagnosis problem.

But this is *not* the case for education. Performance indicators such as end-of-term grades should not represent the goals of education. Having more students with higher grades is an insufficient proxy for educational experience. Yet today, the discourse around the adoption of AI in education is peppered with narratives promoting platforms and services to provide personalized and automated content generation and help teachers teach more effectively, claiming to help students achieve higher grades or increase some other easily measurable performance indicator.

And the pressure is on all actors of education to deliver: the teacher is given numeric indicators and benchmarks to meet, and ever-younger students are taught to measure their success by what grades they have achieved. But any activity that is purely defined by performance metrics and return on investment can likely be done more efficiently and cheaply by AI.

The argument for AI in education as a booster of performance or optimizer of work is reinforced by corporate, capitalist big tech, which flows seamlessly into prevailing neoliberal educational ideologies worldwide. This prevailing logic will only lead us to increasingly value competition and simple measurable outcomes. That is not what we want for education.

Countering open-washing

We propose that three aspects of openness can contribute to better futures.

Open technologies

The nature of many of the educational technologies incorporating AI today is that they are black boxes: 'a device or system that, for convenience, is described solely in terms of its inputs and outputs. One need not understand anything about what goes on inside such black boxes. One simply brackets them as instruments that perform certain valuable functions' (Winner, 1993). This includes dashboards based on 'business intelligence' logics, automatic evaluation tools, text and media generation models, and personalized recommendation systems. Black box systems encourage certain forms of engagement. Teachers and students, under the banner of 'AI literacy' and innovation, are tasked – if not forced – to become instrumentally competent and 'ethical users' of technical systems (Buzato and Gonsales, 2025).

We *can* create AI systems that are accountable to educational communities rather than corporate shareholders. This transparency enables critical examination of AI systems, allowing researchers, educators and students to participate and address biases or limitations. AI and other automated technological systems should follow more rigorous definitions

of openness that allow training data,[1] weights and code to be examined and improved by diverse communities. Training data would be representative of global diversity or local context (for smaller and more directed systems), and systems would be designed to amplify marginalized voices rather than perpetuate existing power structures. This is especially important in educational contexts, where AI systems might otherwise reinforce stereotypes about who 'belongs' in certain fields or who is capable of learning.

Open Practices and Resources

The increased incorporation of black boxes in educational contexts has exacerbated the emphasis on superficial learning metrics. The future could build models that increase focus on *process, experience* and *outcome,* on 'depth pedagogies', which emphasize care and building meaningful relationships between 'self, fellow humans and other kin, and the natural world' (Andreotti, 2021). The logic of open practices and resources promote these values.

The future could build models that increase focus on process, experience and outcome, on 'depth pedagogies', which emphasize care and building meaningful relationships.

Competition and efficiency need numbers. Conversely, as soon as you have some numbers supposedly measuring a performance, competition (doing better than the other) and efficiency (doing better than before) will thrive. But even when these numbers are meaningful, the reduction of the initial problem (in our case, learn better or teach better) to a simplified optimization problem (improve your score) always creeps in. This is sometimes supported with the modern argument that the approach is now 'data-driven'. Yet the approach is only data-driven when everyone has access to the data and processes, and not to a number generated by a formula from the data. Open technologies go beyond showing just a chosen indicator.

Open educational resources can greatly contribute to practices of critical engagement. They are openly licensed and usually in open formats allowing communities worldwide to examine, modify, and adapt them to local contexts and needs. Students can be critical of what is 'given', 'a movement of openness to other epistemological adventure' (Matos-de-Souza, 2024). Engaging with OER can be a tool to help students and educators see themselves as active co-creators of content, technology and curricula that represent their lived realities rather than passive consumers of pre-packaged content and tools created only to maximize easily measurable outcomes.

To do so, however, we must re-conceptualize education as public goods rather than commodities. Open education principles are fundamentally counter-capitalist, asserting

1 We recognize the dilemmas in opening up traditional knowledge as open data, such as the right to exercise control over Indigenous knowledge, which might be incorporated, fenced, and exploited commercially by private interests (as has been the case in open source and open access). Though we contend that this is the exception rather than the norm, in support of 'open', it speaks to the recognition that 'openness' is not neutral and brings forth new challenges.

that content and knowledge should be free and available to all. This stance recognizes that education is not a private benefit but a collective resource that strengthens communities and societies.

Open, participatory governance of education

We are already in what some scholars describe as the 'postdigital' (Fawns et al., 2023) era – a time when digital technologies have become so embedded in daily life that they're no longer novel but simply part of the infrastructure of society. We might soon find ourselves in a 'postAI' world, where AI, in its many forms, is similarly integrated into daily life and learning. We have seen how the platformization of education has promoted the silent and nonconsensual entry of AI tools (chatbots, image generators, business intelligence) into web-based systems used in education administration and teaching. The question is not whether AI will transform education – it already has begun to – but what values will underpin its logics in education.

AI is not simply a tool to make existing educational approaches more efficient. Instead, we should ask how open principles might enable entirely new forms of open learning and community based on the common good, so we can guide AI development towards serving human flourishing rather than corporate profit.

From the perspective of education, the fast and ever evolving nature of AI only reinforces the need for an additional step. Educational institutions must urgently create mechanisms that permit stakeholders to rapidly and effectively react to change. A culture of openness in the forms of technology, practice and resources helps us envision more participatory governance to address uncertainties and demands faced by students and educators.

References

Andreotti, V. de O. 2021. Depth education and the possibility of GCE otherwise. *Globalisation, Societies and Education,* Vol. 19, No. 4, pp. 496–509.

Buzato, M. E. K. and Gonsales, P. 2025. Dimensões e enfoques alternativos para um modelo de letramentos críticos de inteligência artificial [ALternative dimensions and approaches to a critical artificial intelligence literacies model]. *Revista Brasileira de Linguística Aplicada*, Vol. 25, No. 2. (In Portuguese.) https://doi.org/10.1590/1984-6398202544436

Facer, K. and Selwyn, N. 2021. *Digital technology and the futures of education: Towards 'non-stupid optimism'*. Paris, UNESCO.

Fawns, T., Ross, J., Carbonel, H., Noteboom, J., Finnegan-Dehn, S. and Raver, M. 2023. Mapping and tracing the postdigital: approaches and parameters of postdigital research. *Postdigital Science and Education,* Vol. 5, pp. 623–42. https://doi.org/10.1007/s42438-023-00391-y

Fraser, N. 2025. *Workers are cannibalised by the capitalist class* [video].

Lång, K., Josefsson, V., Larsson, A.-M., Larsson, S., Högberg, C., Sartor, H., Hofvind, S., Andersson, I. and Rosso, A. 2023. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *The Lancet Oncology,* Vol. 24, No. 8, pp. 936–44. https://doi.org/10.1016/s1470-2045(23)00298-x

Luke, J. 2023. Closing the factory: Reimagining higher education as commons. L. Czerniewicz and C. Cronin (eds), *Higher Education for Good: Teaching and Learning Futures*. Cambridge, United Kingdom, Open Book Publishers, pp. 161–82. https://doi.org/10.11647/obp.0363.06

Matos-de-Souza, R. 2024. Olvidar Larrosa: Orientaciones posoccidentales para leer la modernidad [Forgetting Larrosa: Post-western guidelines for reading modernity]. *Praxis Pedagógica*, Vol. 24, No. 37, pp. 41–71.

Widder, D. G., Whittaker, M. and West, S. M. 2024. Why 'open' AI systems are actually closed, and why this matters. Nature, Vol. 63, pp. 827–33. https://doi.org/10.1038/s41586-024-08141-1

Winner, L. 1993. Upon opening the black box and finding it empty: Social constructivism and the philosophy of technology. *Science, Technology, & Human Values,* Vol. 18, No. 3, pp. 362–78.

About the authors

Tel Amiel is an adjunct professor at the School of Education at the University of Brasília (Brazil) where he coordinates the UNESCO Chair in Open Education and Technologies for the Common Good, part of the UNITWIN Network in Open Education. He is an Advisory Board member of the Internet Governance Forum and UNESCO Dynamic Coalition on Open Educational Resources, and co-founder of the Open Education Initiative, an activist research group.

Glenda Cox is an associate professor in the Centre for Innovation in Learning and Teaching at the University of Cape Town. She holds the UNESCO chair in Open Education and Social Justice and is a member of the UNITWIN network on Open Education. She is a vice president on the board of the Open Education Global organization. She is on the editorial board of the International Journal of Students as Partners.

Colin de la Higuera is a professor in Computer Science at Nantes University (France). He has been the founding president of the SIF, the French Informatics Society, and is currently a trustee of the Knowledge for All foundation. In 2025, UNESCO renewed the Chair in Open Educational Resources and Artificial Intelligence. He is also coordinator of UNOE, the UNITWIN Network in Open Education.

Baumol's cost disease: Long-term economic implications where machines cannot replace humans

Guido Appoldt and Zhen Gong

Education faces a cost disease: despite AI advances, productivity stagnates. This piece explores the roots of this paradox and its implications for the future.

Education has long suffered from what economists term 'Baumol's cost disease', a tendency for costs to rise faster than general inflation due to stagnant productivity in labour-intensive sectors (Kimball and Luke, 2018). In higher education this phenomenon is sometimes dubbed 'Bowen's curse' (Baumol, 2012). Essentially, teaching is time and labour intensive: a professor in the 2020s still needs to spend roughly the same amount of time to effectively teach or mentor a student as decades ago, just as *'it still takes four musicians nine minutes to perform Beethoven's String Quartet in C minor, as it did in the nineteenth century'* (Kimball and Luke, 2018). In contrast, technology has dramatically boosted productivity in manufacturing and other sectors, enabling fewer workers to produce more output.

The crux of the issue is that education's core service cannot be easily accelerated or automated without sacrificing quality. This analysis examines why the cost disease persists in education despite advancements in AI and explores the long-term implications of this issue. We focus on how AI might alleviate or exacerbate cost pressures, why its impact is limited, and its long-term influence on the future of education.

Persistence of cost disease in education

Baumol's cost disease theory holds that sectors requiring direct human input (like teaching) experience relatively low productivity growth because they cannot significantly speed up or substitute the human labour involved (Da Costa Nunes and Nunes, 2024). In education, *'the labor is itself the end product'* (Kimball and Luke, 2018), a teacher's presence and interaction are fundamental to the service. No matter how advanced technology becomes, most teaching tasks remain resistant to automation. As a recent analysis notes, jobs emphasizing interpersonal interactions (e.g. face-to-face communication with learners) are *'likely to be unaffected or not enabled by [AI]'*, and 'most teaching tasks … feature prominently in this category' (Elhussein et al., 2024). In other words, AI cannot conduct a parent-teacher conference, inspire a struggling student or manage a lively classroom discussion with the same effectiveness as a human teacher.

In other words, AI cannot conduct a parent-teacher conference, inspire a struggling student or manage a lively classroom discussion with the same effectiveness as a human teacher.

AI is able to automate certain education-related tasks, but these are generally auxiliary duties rather than the core act of teaching. For instance, AI systems can grade multiple-choice exams, compile learning materials or analyse student performance data much faster than humans (Elhussein et al., 2024). Offloading such routine and administrative tasks does offer productivity gains as studies show teachers spend a large portion, often over 50%, of their work hours on non-teaching tasks (ibid). Automating these burdens could alleviate stress and slightly improve overall efficiency (ibid). AI is already being applied in this vein: for example, machine learning models have been developed to optimize resource allocation and recommend personalized learning paths in higher education, showing promising improvements in decision-making efficiency (Qian, Cao and Chen, 2025).

AI is able to automate certain education-related tasks, but these are generally auxiliary duties rather than the core act of teaching.

However, these efficiency gains do not translate into a proportional reduction in labour needs for the central teaching role. Whether a class uses AI tools or not, each additional group of students generally still requires a teacher's oversight, guidance and feedback. Advanced AI, like LLMs, can support lesson planning or offer tutoring, but human oversight remains necessary to ensure accuracy, appropriateness and emotional support (Anderson, 2023). The World Economic Forum observes that while AI can assist with lesson preparation or grading, teachers must still handle *'the essential educational aspects of interpersonal interaction'* (Elhussein et al., 2024). In practice, AI tutors or educational software function best as adjuncts to a teacher in tasks such as providing drill practices or instant feedback, rather than as replacements for teacher-led instruction. Even highly automated learning platforms require human moderation and intervention for students who struggle or disengage.

Sometimes, AI even *adds* new responsibilities: a U.S. Department of Education report emphasizes that teachers need to *'exercise judgment and control over the use of AI in education'* at all times (National Education Association, 2024). Similarly, UNESCO (2021) warns that aiming to eliminate human teachers betrays a *'fundamental misunderstanding of their essential social role in the learning process.'* Therefore, AI has not overturned the fundamentally labor-intensive nature of education, and productivity growth remains modest. As one commentary analogizing investigative journalism notes, these sectors 'rely heavily on human labor and struggle to achieve significant productivity gains from technological advances' (Wellbrock, 2024). In short, Baumol's cost disease persists: despite new technologies, the cost of education tends to keep climbing relative to other sectors because teaching still requires intensive human involvement.

In short, Baumol's cost disease persists: despite new technologies, the cost of education tends to keep climbing relative to other sectors because teaching still requires intensive human involvement.

Long-term implications

If the cost disease in education continues unabated and AI provides only incremental relief, there are serious long-run implications for the educational system and society.

Ultimately, society faces a choice: either accept a decline in educational talent and quality, or subsidize education more heavily.

One concern is a potential erosion of instructional quality under escalating cost pressures. When productivity remains flat but salaries and other costs rise, educational institutions face budget constraints. Without sufficient funding to hire more staff or invest in quality, schools may respond by increasing class sizes, reducing staff-student ratios or asking teachers to take on heavier workloads (Baumol and Bowen, 1965; Throsby, 1996). This can dilute the attention each student receives and strain teachers. Many educators already feel overstretched: surveys indicate that excessive administrative duties and large class sizes are top stressors, detracting from time spent on 'quality interactions with students' (Elhussein et al., 2024). If nothing is done, the trend could worsen, teachers might be supervising more students or juggling additional responsibilities, such as IT management of AI tools without relief. Research in the economics of education suggests that talented individuals may also shun or leave the teaching profession if working conditions and pay remain unattractive (ibid). Losing experienced teachers or being unable to recruit high-calibre new teachers is directly harmful to educational quality. Thus, Baumol's disease can manifest as a quality disease: rising costs not met with rising budgets result in fewer resources per student and potentially a narrower curriculum, undermining the richness of education.

Baumol's cost disease also implies a growing wage gap between education and other sectors which could fuel teacher shortages and require financial reallocation from the broader economy (Baumol, 2012). In sectors like tech or finance, high productivity growth enables firms to pay rising wages. Education cannot increase output per teacher at the same rate, yet to attract and retain qualified teachers, schools must offer competitive salaries (Archibald and Feldman, 2008). Over time, teacher pay needs to keep pace with wages in other graduate-level professions. Indeed, data show that in many countries teaching salaries lag significantly behind those of similarly educated workers (often by 10–30%) (Elhussein et al., 2024), making teaching a less attractive career financially. This gap is a direct manifestation of cost disease: without productivity gains, any wage increases in education translate straight into higher per-student costs (Kimball and Luke, 2018). If budgets don't rise accordingly, salaries stagnate and talented individuals may choose higher-paying industries, aggravating the teacher gap.

Baumol's cost disease serves as a sobering reminder that not all problems yield to technological quick fixes. Education, at its heart, is a person-to-person endeavor. Thus, even the most sophisticated AI will not radically increase throughput without risking the loss of human connection and quality.

Ultimately, society faces a choice: either accept a decline in educational talent and quality, or subsidize education more heavily. In practical terms, that means channeling resources from productivity-rich parts of the economy through taxation or public spending into education (Baumol, 1996). Other industries' growth indirectly funds teacher pay raises and smaller class sizes. Baumol himself argued that as productivity (and wealth) grows in the overall economy, the society *can* afford these rising education costs, but only if we consciously decide to allocate the funds (Baumol, 2012). Without such reallocation, the funding shortfall will persist, and education could enter a vicious cycle of underinvestment. Without deliberate interventions, the public school and institutions will be affected most severely (Da Costa Nunes and Nunes, 2024), resulting in the exacerbation of educational inequalities, we could be 'sleepwalking into a world that widens existing inequalities' as a result of AI as UNESCO warned (2021). Thus, the cost disease calls for long-term financial shifts: more of public spending must be devoted to education to maintain its quality.

An unequal future is yet avoidable

Baumol's cost disease serves as a sobering reminder that not all problems yield to technological quick fixes. Education, at its heart, is a person-to-person endeavor. Thus, even the most sophisticated AI will not radically increase throughput without risking the loss of human connection and quality. The analysis above has illustrated why AI has so far failed to resolve the cost disease in education. Consequently, costs continue to rise, straining schools and governments, while the specter of declining quality and growing inequality looms if we respond poorly.

The long-term implications of ignoring these trends could be dire: a two-tiered education system with excellence for the rich and mediocrity for the rest. Yet this outcome is not inevitable as long as the necessary policy interventions are taken. The situation calls for bold recognition that education is a 'stagnant' sector that nonetheless underpins all other progress. Rather than viewing the rising cost of education as a crisis to avoid, we should view it as a natural investment in our future, one that can be managed and is affordable.

The situation calls for bold recognition that education is a 'stagnant' sector that nonetheless underpins all other progress. Rather than viewing the rising cost of education as a crisis to avoid, we should view it as a natural investment in our future, one that can be managed and is affordable.

References

Anderson, J. 2023. *Educating in a World of Artificial Intelligence* [Podcast].

Archibald, R. B. and Feldman, D. H. 2008. Explaining increases in higher education costs. *The Journal of Higher Education,* Vol. 79, No. 3, pp. 268–95. https://doi.org/10.1080/00221546.2008.11772099

Baumol, W. J. 1996. Children of performing arts, the economic dilemma: The climbing costs of health care and education. *Journal of Cultural Economics,* Vol. 20, No. 3, pp. 183–206. https://doi.org/10.1007/s10824-005-3206-4

––. 2012. *The Cost Disease: Why Computers Get Cheaper and Health Care Doesn't*. New Haven, Yale University Press.

Baumol, W. J. and Bowen, W. G. 1965. On the performing arts: The anatomy of their economic problems. *The American Economic Review,* Vol. 55, No. 1/2, pp. 495–502.

Da Costa Nunes, R. and Nunes, A. 2024. Examining the impact of baumol's cost disease in brazilian municipal education: A decade analysis (2009-2019). *International Journal of Economics and Finance,* Vol. 16, No. 4, pp. 78–90. https://doi.org/10.5539/ijef.v16n4p78

Elhussein, G., Hasselaar, E., Lutsyshyn, O., Milberg, T. and Zahidi, S. 2024. *Shaping the Future of Learning: The Role of AI in Education 4.0*. Geneva, World Economic Forum.

Kimball, B. A. and Luke, J. B. 2018. Historical dimensions of the "cost disease" in US higher education, 1870s–2010s. *Social Science History,* Vol. 42, No. 1, pp. 29–55. https://doi.org/10.1017/ssh.2017.38

National Education Association. 2024. *Artificial intelligence in Education*.

Qian, L., Cao, W. and Chen, L. 2025. Influence of artificial intelligence on higher education reform and talent cultivation in the digital intelligence era. *Scientific Reports*, Vol. 15. https://doi.org/10.1038/s41598-025-89392-4

Throsby, D. 1996. Economic circumstances of the performing artist: Baumol and Bowen thirty years on. *Journal of Cultural Economics*, Vol. 20, No. 3, pp. 225–40. https://doi.org/10.1007/BF00153849

UNESCO. 2021. *AI and education: guidance for policy-makers*. Paris, UNESCO. https://doi.org/10.54675/PCSP7350

Wellbrock, C.-M. 2024. Generative AI and investigative journalism—an application of baumol's cost disease across mass communication sectors. *Emerging Media,* Vol. 2, No. 4, pp. 663–72. https://doi.org/10.1177/27523543241299631

About the authors

Guido Appoldt is a postgraduate student at the School of Media and Communication, Shanghai Jiao Tong University. His works explore the intersections of emerging technologies and sustainable development.

Zhen Gong is a playwright at the Shanghai Arts Research Center, creating both artistic and analytical works that address contemporary social issues.

Who educates the LLMs? The uncertain role of the nation-state

Camille Fabo

AI in education revives old debates on who defines and teaches 'truth'. As LLMs shape how knowledge is built and shared, tech firms, not states, could emerge as new global educators.

The debates around AI in education echo controversies that emerged in response to the introduction of Wikipedia and Google, both of which reshaped how knowledge and learning are produced, accessed and trusted. Today, LLMs extend these transformations even further.

Tech companies are rapidly exploring how to use and monetize LLMs for teachers, students and schools. For example, the Chan-Zuckerberg Initiative's (2024) beta curricula claims to draw on 'high-quality data' from across the United States for teaching mathematics, building EdTech tools and more.

Yet as different models are trained under diverging frameworks and ethical assumptions, paradigms of knowledge are being redefined. Nation-states, once the main authority deciding what citizens should learn in terms of content and methods, must now consider their relationship to these new 'knowledge generators'.

Human reviewers who provide reinforcement learning shape the model's responses and can be seen as its 'teachers'. In turn, the model becomes a 'teacher' to its users, influencing how humans learn, reason and express themselves.

Amid these debates, this think piece asks: *who* teaches *what* to large language models? I explore this question in the context of tensions between knowledge, authority, validity and truth in a post-truth paradigm**.** In other words, in a value system in which opinions and emotions can appear to matter more than empirical evidence, what could it mean for LLMs trained by humans to generate knowledge come to instruct, influence and even reshape the understanding of those who trained it?

Who teaches what to LLMs?

The LLM Grok x-Ai generates content and predicts language based on patterns learned from large-scale text and real-time data, rather than explicitly modelling user preferences. It does not guarantee accuracy or truth. This approach is embedded in a post-truth paradigm, since traditional expertise is not treated as different from other input, and the responsibility for determining accuracy and truth is left to users. By contrast, OpenAI takes a different approach: they increase the value of experts by training their LLMs in targeted areas, aiming to align with standards of scientific accuracy and validity. As such, models like ChatGPT operate under

the assumption that not all knowledge is equal, and that the one validated by experts should carry greater value.

Definitions of accuracy, value and ethics are thus being reshaped and selectively interpreted by major tech players who build and spread their own normative frameworks, simultaneously becoming *de facto* 'educators' of millions of users around the world.

Without a guiding ethos of teaching and learning, we may be shaping an amoral omniscient teacher able to influence millions with no pedagogical or ethical framework. If LLMs are becoming the educators of a new era, we must question who is responsible for their education, at what levels, including ethical, social, technical and institutional.

LLMs, unlike humans, do not think independently, yet they are convincing in the way they mimic the human learning process. They are fast, adaptable and responsive, but fundamentally limited: they can only learn through human-provided data and feedback (Dubois et al., 2021). Human reviewers who provide reinforcement learning shape the model's responses and can be seen as its 'teachers'. In turn, the model becomes a 'teacher' to its users, influencing how humans learn, reason and express themselves.

Every user plays an indirect role in shaping LLMs' responses – i.e. in *teaching* them – since their interactions are often used to refine future versions. These models are, by nature, participatory in the sense that users' data and interactions are leveraged to 'better' address their expectations, reflecting a democratic (and to some extent populist) model of producing knowledge that simultaneously makes them an accessible learning tool. This blurs, however, the relationship and learning experiences of simultaneously being 'teachers' and 'users' of LLMs.

LLMs are only considered 'intelligent' to the extent that they follow their instructions via prompts. Their output is grounded in statistical probability, shaped by weighted values derived from human input. Iterative reinforcement aligns these probabilistic outputs with human expectations. They do not create; they aggregate and reproduce, simulating the semblance of human reasoning.

Is AI a post-truth teacher of humans?

If just anyone can indirectly contribute to the training of LLMs through their interactions, even when having no intention to 'teach' it, we face the risk that these systems could learn from low-quality feedback, false information and misleading assumptions. Without a guiding ethos of teaching and learning, we may be shaping an amoral omniscient teacher able to influence millions with no pedagogical or ethical framework. If LLMs are becoming the educators of a new era, we must question who is responsible for their education, at what levels, including ethical, social, technical and institutional.

When AI models from a single country dominate the global landscape, they risk undermining the capacity of other nations to define their own values, cultures and worldviews.

If AI is a post-truth tool, however, it still shadows an old and outdated vision of enlightenment, rationalization and truth. A small group of highly educated experts, drawing on their technical skills and institutional prestige, now possesses the power to translate vast domains of human activity, from art to science, into algorithmic models. In doing so, some post-truth LLM developers rationalized the world while paradoxically rejecting its underlying rationality: truth becomes whatever the privileged segments of society believe to be true, subject to the constraints they themselves define.

I argue that our understanding of truth should be more inclusive, embracing multiple ways of knowing beyond dominant frameworks. This has direct implications for AI: those who design and train LLMs should account for this diversity not only to better reflect the real world, but also to make these systems more intelligent, useful and trustworthy.

Interrogating knowledge, authority and 'subjective' truth

What does this mean, then, for the authority of nation-states on knowledge and learning? If LLMs reflect the ideologies, priorities and assumptions of the tech companies that design and train them, the influence of those holding disproportionate weight in the digital world could quickly eclipse other narratives.

Many of AI companies are today led by U.S. elite men from top U.S. universities and have been overt in their intentions to serve U.S. interests first, and secondly 'to ensure that artificial general intelligence benefits all of humanity' (OpenAI, n.d.). One vignette from OpenAI's (**2025**) proposal to strengthen American leadership in AI is telling:

'For countries seeking access to American AI, we propose a strategy....to proactively promote the global adoption of American AI systems...The strategy would use export controls to protect America's AI lead.'

When AI models from a single country dominate the global landscape, they risk undermining the capacity of other nations to define their own values, cultures and worldviews. Nation-states have long shaped definitions of legitimate knowledge, notions of citizenship, and civic values. In today's globalized world, where teachers are trained, at least in part, by globally deployed LLMs, nation-states risk becoming displaced in their ability to define shared values and interests. One proposed response has been for LLMs to be either nationalized or at least have national branches trained on local data, as seen in Switzerland with Apertus, the 'first large-scale, open, multilingual language model' (ETH Zurich, 2025).

> Before models become more independent, different communities carrying diverse worldviews should *intentionally* teach LLMs their desired value systems to shape a global governance that is not be held by a few players.

LLMs' influence is growing fast; examining their implications for learning, nation-building, and community identity is urgent. Being afraid and reluctant is human, but the machine is real, trained by humans, and using human data. Before models become more independent, different communities carrying diverse worldviews should

intentionally teach LLMs their desired value systems to shape a global governance that is not be held by a few players.

References

Chan Zuckerberg Initiative. 2024. *New AI resources and advisory board aim to support responsible development of education tools*.

Dobbin, F., Simmons, B. and Garrett, G. 2007. The global diffusion of public policies: Social construction, coercion, competition, or learning? *Annual Review of Sociology*, Vol. 33, pp. 449–72. https://doi.org/10.1146/annurev.soc.33.090106.142507

Dubois, Y., Kiela, D., Schwab, D. J. and Vedantam, R. 2021. *Learning optimal representations with the decodable information bottleneck*. arXiv preprint, https://doi.org/10.48550/arXiv.2009.12789

ETH Zurich (Eidgenössische Technische Hochschule Zürich). 2025. *Apertus: A fully open, transparent, multilingual language model*.

Meyer, J. W. 2007. World models, national curricula, and the centrality of the individual. A. Benavot, C. Braslavsky, and N. Truong (eds), *School knowledge in comparative and historical perspective: Changing curricula in primary and secondary education*. Dordrecht, Springer Nature, pp. 259–71. https://doi.org/10.1007/978-1-4020-5736-6_16

Meyer, J. W., Boli, J., Thomas, G. M. and Ramirez, F. O. 1997. World society and the nation-state. *American Journal of Sociology*, Vol. 103, No. 1, pp. 144–81. https://doi.org/10.1086/231174

Meyer, J. W., Ramirez, F. O., Rubinson, L. and Boli-Bennett, J. 1977. The world educational revolution, 1950–1970. *Sociology of Educatio**n***, Vol. 50, No. 4, pp. 242–58. https://doi.org/10.2307/2112498

OpenAI. n.d. *About | Open AI*.

––. 2025. *OpenAI's proposals for the U.S. AI Action Plan*.

About the author

Camille Fabo is a Doctoral Fellow at Teachers College, Columbia University. Her research examines education policies and nation-building.

Environmental contradictions of large language models in higher education

James O'Sullivan and Mark Kelly

Universities are racing to adopt AI, even as its environmental costs remain largely obscured. Can higher education pursue responsible innovation while upholding sustainability and its commitment to the public good?

The emergence of generative AI systems, particularly LLMs, holds the potential to transform higher education to an extent not witnessed since the emergence of the web some 30 years ago.

LLMs are a type of generative AI system trained on vast quantities of text data to generate, interpret and manipulate human language with remarkable fluency. Built using deep learning architectures – most commonly the transformer model introduced in 2017 – LLMs are designed to predict the probability of a given word or sequence of words based on the context in which they appear. This predictive capacity allows them to perform a wide range of language-related tasks, including writing essays, summarizing documents, answering questions and translating between languages.

To sustain the university as a site of ethical, critical and democratic inquiry, further explorations of *what* LLMs can do should be deprivileged, even paused, in favour of analyses of *how* they are reshaping the values and viability of education itself.

What distinguishes LLMs like OpenAI's GPT series, Google's Gemini, and Anthropic's Claude from earlier forms of natural language processing is their scale: they are trained on billions – or even trillions – of words scraped from (often plagiarized) books, websites, academic papers and other digital sources. While their 'understanding' is statistical rather than semantic, they exhibit a form of general linguistic competence which has prompted both enthusiasm and concern.

Framed by proponents as engines of efficiency, access and pedagogical innovation, LLMs are increasingly integrated into university platforms, tools and curricula. But the narratives of progress that accompany this rapid integration often obscure the environmental consequences of such technologies. The widespread deployment of LLMs in higher education raises urgent questions about sustainability – not only in terms of technological capacity or institutional readiness, but in relation to the material viability of the infrastructures that underwrite AI.

To sustain the university as a site of ethical, critical and democratic inquiry, further explorations of *what* LLMs can do should be deprivileged, even paused, in favour of analyses of *how* they are reshaping the values and viability of education itself. At a moment when universities are committing to sustainability targets and climate leadership, the adoption of computationally intensive generative AI tools presents a paradox. Can the university

continue to position itself as a site of sustainability responsibility and leadership while embracing technologies that are fundamentally extractive?

Material infrastructures and hidden environmental costs

Despite their disembodied appearance on our screens, LLMs are grounded in highly material infrastructures.

Despite their disembodied appearance on our screens, LLMs are grounded in highly material infrastructures. Training a large-scale model requires vast computational resources. Researchers calculate that training and deploying LLMs can, depending on model architecture and scale, emit tens to hundreds of metric tonnes of CO_2 (Faiz et al., 2024). It is estimated that training a single large model could emit over 626,000 pounds of carbon dioxide, the equivalent of 300 round-trip flights between New York and San Francisco (Strubell, Ganesh and McCallum, 2020), or five times the lifetime emissions of an average car, including its manufacture, in the US (Hao, 2019).

Commercial LLM providers treat precise energy consumption and carbon emission data as proprietary, making accurate environmental impact assessments difficult or impossible for external stakeholders, including universities (Dodge et al., 2022). This lack of transparency effectively shields these providers from scrutiny and complicates efforts by institutions to align their technology procurement with explicit sustainability goals. Furthermore, such models are being developed at a time when the realities of climate inequality have never been so stark (Bender et al., 2021).

Whatever the reality of the resource requirements of LLMs, it is well established that data centre energy consumption has significant repercussions for social equity and fuel poverty. Data centres dedicated to training and deploying large-scale AI models extract considerable energy resources from national grids, exacerbating the inequitable distribution of energy in societies already facing acute fuel poverty. Data centres in the EU used an estimated 45-65 TWh of electricity in 2022, representing 1.8-2.6% of total EU electricity use (Kamiya and Bertoldi, 2024). In Ireland, data centres consumed 21% of all metered electricity from the grid in 2023, an increase of 3% from 2022 (Oireachtas Library and Research Service, 2024). This is at a time when 377,000 people in Ireland were unable to keep their house adequately warm in 2022 (Republic of Ireland, 2024).

This tension between technological advancement and basic energy access directly challenges the SDGs, particularly SDGs 3, 7, 10, 12, and 13, which aim to ensure healthy lives and promote well-being for all; ensure access to affordable, reliable, sustainable, and modern energy for all; reduce inequality and promote responsible consumption and production patterns; and take urgent action to combat climate change and its impacts (UN, n.d.). As universities increasingly adopt computationally intensive AI technologies, they risk contributing to energy inequities, effectively prioritizing technological innovation over the fundamental needs and rights of disadvantaged communities. Such choices necessitate

deeper institutional reflection on the broader societal implications of technological adoption, and they underline the importance of embedding equity considerations within higher education's sustainability strategies.

Unlike the corporate actors who develop and profit from these tools, universities bear a different kind of responsibility: to model values and practices oriented towards the public good. As such, academic engagement with LLMs must include environmental reckoning.

It is not only training that carries environmental costs. The inference phase – when models are actually used in real-world applications – also requires significant computational power. As universities increasingly integrate LLMs into their platforms and services, the cumulative energy demands of routine student and staff use begin to scale dramatically. Drawing on de Vries' (2023) estimate of 2.9 Wh per ChatGPT query, a single user prompting GPT-3 for just 50 queries a day over a year could consume as much electricity as a modern refrigerator. That might not seem like much, but when scaled to millions of students, i.e. 235 million students enrolled in higher education in 2020 and growing (UNESCO, 2022) and staff across multiple institutions, the resource implications become staggering.

Furthermore, these systems rely on global data centre networks, many of which are powered by fossil fuels and located in regions already experiencing environmental stress. Water usage is another hidden cost: cooling data centres consumes millions of litres, often drawn from vulnerable local sources. In regions affected by drought or climate instability, the expansion of AI infrastructure poses direct risks to local ecosystems and communities.

Universities, often committed to the language of sustainability in their strategic plans, find themselves in a precarious position. LLMs offer – or certainly promise – potential remedies to the pressures of massification, but their adoption deepens institutional entanglement with carbon-intensive digital economies.

The discourse around LLMs often celebrates their frictionless utility while obscuring the ecological externalities of their existence. In adopting such systems at scale, higher education institutions risk becoming complicit in extractive digital economies that are environmentally unsustainable. Yet, unlike the corporate actors who develop and profit from these tools, universities bear a different kind of responsibility: to model values and practices oriented towards the public good.

As such, academic engagement with LLMs must include environmental reckoning. This might involve demanding transparency from vendors regarding energy use and training data; supporting low-carbon, open-source alternatives; or embedding climate literacy into digital policy. Sustainability in the era of LLMs begins with acknowledging that even the most intangible forms of intelligence require material substrates – and that these substrates are finite.

Complicity, greenwashing and the university

Institutions may highlight the energy efficiencies of automating administrative processes or reducing paper usage while overlooking the broader environmental toll of the systems they deploy. Such gestures amount to a form of academic greenwashing: the selective framing of digital adoption as ecologically virtuous without full transparency about its material impacts.

The university is not a neutral actor in the story of AI. As early adopters, legitimators and sites of technological reproduction, academic institutions play a key role in shaping the trajectory of LLM integration into society. When universities endorse or license AI systems developed by commercial actors, they do more than adopt a tool – they participate in the construction of a social imaginary in which such systems are necessary, desirable and inevitable.

This complicity is often masked by sustainability rhetoric. Institutions may highlight the energy efficiencies of automating administrative processes or reducing paper usage while overlooking the broader environmental toll of the systems they deploy. Such gestures amount to a form of academic greenwashing: the selective framing of digital adoption as ecologically virtuous without full transparency about its material impacts.

Universities are uniquely positioned to resist such reductive logics. They are among the few remaining institutions with both the expertise and the ethical mandate to challenge unsustainable technological trajectories. In the context of LLMs, this means moving beyond reactive accommodation and taking a proactive role in shaping responsible innovation.

To engage seriously with environmental sustainability, higher education must adopt a framework of environmental accountability in its dealings with AI. This begins with transparency. Universities should demand clear environmental impact assessments from AI vendors, including disclosures of training energy, water usage and ongoing inference costs. Just as ethical data usage and privacy considerations have become standard in procurement, so too should environmental criteria.

Institutions must consider investing in or co-developing lower-carbon alternatives to proprietary models. Open-source LLMs, such as BLOOM or Meta's LLaMa, can be fine-tuned locally, offering opportunities for energy-efficient, targeted models trained on smaller and more relevant datasets. These approaches align with the principles of 'small data' and 'minimal computing', which privilege context-sensitive, low-impact digital practices.

Digital literacy initiatives in universities must expand to include climate literacy. If students are to engage critically with AI, they must understand not only its functional capacities but also its ecological footprints. Curriculum design should address the environmental dimensions of digital tools, fostering a generation of scholars and professionals who see sustainability as integral to technological fluency.

Sustainability considerations must be integrated into institutional policy frameworks. This includes procurement policies, platform governance and academic planning. Where LLMs

are deployed, their use should be justified not only in terms of pedagogical or administrative benefit, but also in relation to long-term environmental strategy.

Finally, universities should work collaboratively – across regions and sectors – to establish shared norms and benchmarks for sustainable AI. Given the global nature of both climate crisis and digital infrastructure, isolated action is insufficient. Networks of higher education institutions can leverage collective bargaining power to demand change from technology providers and to model alternative futures grounded in planetary ethics.

Curriculum design should address the environmental dimensions of digital tools, fostering a generation of scholars and professionals who see sustainability as integral to technological fluency.

LLMs mark a profound shift in how we produce, distribute and interact with knowledge. Their integration into higher education is both an opportunity and a warning. If adopted uncritically, universities risk further entrenching extractive and unsustainable digital infrastructures under the guise of innovation. If approached with care, however, they offer an occasion to reassert their public mission to shape technologies in ways that are just, inclusive and ecologically attuned. The future of higher education depends not on how fast we adopt new tools, but on how wisely we decide which tools – and which futures – are worth sustaining.

References

Bender, E. M., Gebru, T., McMillan-Major, A. and Shmitchell, S. 2021. On the dangers of stochastic parrots: Can language models be too big? *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York, Association for Computing Machinery, pp. 610–23. https://doi.org/10.1145/3442188.3445922

De Vries, A. 2023. The growing energy footprint of artificial intelligence. *Joule*, Vol. 7, No. 10, pp. 2191–94.

Dodge, J., Prewitt, T., Tachet des Combes, R., Odmark, E., Schwartz, R., Strubell, E., Sasha Luccioni, A., Smith, N. A., DeCario, N. and Buchanan, W. 2022. Measuring the Carbon Intensity of AI in Cloud Instances. FAccT '21: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency.* New York, Association for Computing Machinery, pp. 1877–94. https://doi.org/10.1145/3531146.3533234

Faiz, A., Kaneda, S., Wang, R., Osi, R., Sharma, P., Chen, R. and Jiang, L. 2024. *LLMCarbon: Modeling the end-to-end carbon footprint of large language models.* arXiv preprint, https://doi.org/10.48550/arXiv.2309.14393

Hao, K. 2019. Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes. *MIT Technology Review.*

Kamiya, G. and Bertoldi, P. 2024. *Energy Consumption in Data Centres and Broadband Communication Networks in the EU*. Luxembourg, Publications Office of the European Union.

Republic of Ireland, Joint Committee on Social Protection, Community and Rural Development, and the Islands. 2024. *Report on Energy Poverty*. Dublin, Houses of the Oireachtas. (Doc. 33/JCSPCRI/026.)

Strubell, E., Ganesh, A. and McCallum, A. 2020. Energy and policy considerations for modern deep learning research. *Proceedings of the AAAI Conference on Artificial Intelligence,* Vol. 34, No. 9, pp. 13693–96. https://doi.org/10.1609/aaai.v34i09.7123

UNESCO. 2022. *Higher Education Global Data Report, Working Document*. Paris, UNESCO.

United Nations. n.d. *The 17 goals*.

Walsh, K. 2024. *L&RS Spotlight: Data Centres and Energy*. Dublin, Houses of the Oireachtas. (No. 2 of 2024.)

About the authors

James O'Sullivan is currently seconded to the Higher Education Authority as their Teaching & Learning Policy Advisor for Artificial Intelligence. He is a senior lecturer at University College Cork where he works across text and cultural analytics.

Mark Kelly is currently seconded to the Higher Education Authority as their Teaching and Learning Policy Advisor on Education for Sustainable Development (ESD). He is the Head of the Atlantic Technological University (ATU) Centre for Sustainability, a Senior Lecturer in the Department of Building and Civil Engineering, and co-lead of the ATU ESD Academy.

Could zero-rating advance meaningful connectivity in education?

Kevin Martin

Can zero-rating help close education's digital divide? This think piece argues it can, if embedded in policies for meaningful connectivity, equity and governance.

A reflective call for universal access

Digital connectivity underpins nearly every dimension of life, yet access to the internet, and the educational opportunities it carries, remains a luxury for millions. For education, what matters is not just any connection but meaningful connectivity: reliable infrastructure, affordable mobile data, appropriate devices, foundational digital skills, and the confidence to navigate online spaces safely and securely. Learners in low-resource communities are constrained not only by high data costs, but also by patchy infrastructure, device scarcity, limited digital skills and exposure to online risks. Where any one of these dimensions is missing, connectivity can entrench rather than reduce inequality, and although digital platforms can in principle level educational opportunity across socio-economic boundaries, these layered barriers continue to exclude many of those who might benefit most (UNESCO, 2023a).

> Where learners know that educational platforms are data-free, safe and trusted, patterns of use change: they can return repeatedly to the same resources, explore new topics without fear of 'using up' data and engage more deeply with digital materials.

Zero-rating, or the practice by which mobile-network operators waive data charges for designated websites or applications (Krämer and Peitz, 2018), offers a lever on the affordability dimension of meaningful connectivity. While zero-rating is not without controversy, particularly regarding net neutrality, it represents an imminently achievable opportunity to support greater access to digital education interventions. By exempting verified educational platforms from mobile data fees, governments and operators can ensure that cost is less likely to be an insurmountable barrier to learning (McBurnie et al., 2020).

This think piece revisits zero-rating in the context of global education equity and the futures of digital learning. It explores an illustrative case study from South Africa, reflects on the policy and industry debates that have shaped this approach (including where connectivity is framed as a public utility), and proposes actionable strategies for zero-rating education content worldwide. The core argument is that zero-rating should be integrated into broader efforts to secure meaningful connectivity for all learners. Ultimately, this think piece aims to advance critical debate about robust, transparent zero-rating policies that

advance education equity, and respond to developments related to the potential use of AI in education systems.

The promise of zero-rating for educational equity

The Internet divide as a question of meaningful connectivity

For many learners in low-income settings, a basic mobile phone is the only digital device in the household. Where data are bought in pay-as-you-go increments, families ration use to calls or messaging, avoiding bandwidth-heavy websites, learning platforms, and videos (Lythreatis, Sing and El-Kassar, 2022; Martin, 2022). Even where there is mobile coverage, the combination of high data costs, low device quality, limited digital skills and fears about online risks can make educational use of the internet rare.

Zero-rating cannot, on its own, resolve gaps in infrastructure or skills. However, it can mitigate one particularly pernicious barrier – affordability – and, when embedded in wider policy strategies, reinforce a more holistic understanding of connectivity. Where learners know that educational platforms are data-free, safe and trusted, patterns of use change: they can return repeatedly to the same resources, explore new topics without fear of 'using up' data and engage more deeply with digital materials. In this way, zero-rating can help turn sporadic, fragile access into a more sustained and meaningful form of connectivity for learning.

Vodacom e-School and Siyavula in South Africa

To consider one case study, South Africa's school closures during the pandemic illuminated longstanding educational and connectivity inequities (Guruli et al., 2020; Bhandari, 2020). Within this context, a structured, state-led zero-rating framework, applied across all major mobile network operators, became a central plank of the country's emergency education response. Vodacom e-School, launched with the Department of Basic Education in 2014 as a free, curriculum-aligned platform for Grades R–12, was already zero-rated for Vodacom customers when COVID-19 struck (Vodacom, 2020). As schools closed, reported sign-ups quadrupled in a single month, with more than 100,000 new registrations in March 2020 and total users exceeding one million. Siyavula, a popular online mathematics and science platform, was similarly zero-rated not only on Vodacom but also on MTN and Telkom, making its core resources accessible at no data cost. These examples illustrate how, when affordability barriers are lifted, learners' use of digital resources can expand rapidly, even in a context of deep structural inequality.

Crucially, these developments were not ad hoc or confined to a single operator. The Department of Communications and Digital Technologies issued directions under the national state of disaster requiring mobile operators to zero-rate a vetted list of educational and health websites. Platforms had to meet specified criteria and be approved through government-led processes before being added to the list. The regulator, ICASA, later incorporated similar obligations into spectrum licensing conditions, confirming that

zero-rating was a regulatory requirement rather than a purely voluntary corporate initiative (Government of South Africa, 2020; DGMT, 2021). Civil society organizations and industry associations engaged actively with these measures, raising concerns about delays, uneven implementation and the scope of sites included, and pressing for greater clarity and transparency (DGMT, 2021; Ikigai Law, 2020). In parallel, television and radio programmes were expanded to support remote learning, positioning zero-rated digital platforms as one element within a broader public response.

Taken together, the South African experience demonstrates that zero-rating can be a powerful tool for greater education equity when it is embedded within broader strategies for meaningful connectivity and treated as a public obligation tied to essential services, rather than as a discretionary act of corporate social responsibility. The rapid uptake of Vodacom e-School and Siyavula under a highly structured, government-mandated framework underscores the latent demand for digital learning opportunities and the importance of aligning telecom operators' obligations with education policy goals. The ongoing debates over implementation and governance highlight the need for clear criteria, robust oversight and sustained civil society engagement, and show that zero-rating is often shaped by public authorities, regulators and trade groups rather than being solely a matter of private corporate discretion.

Connectivity, public utility and zero-rating

In many policy debates, connectivity is still treated as a commercial service delivered by private telecom firms to individual consumers. Yet a growing body of international guidance frames connectivity, especially for education, as a public good or a form of essential infrastructure, similar to electricity or water (UNESCO, 2023a). This shift has several implications for how zero-rating is understood and governed.

First, in many countries mobile network operators benefit from public resources: they receive access to spectrum, use rights-of-way for network deployment, and, in some cases, direct subsidies or tax incentives to extend coverage. In return, they are often subject to universal service obligations, social obligations linked to spectrum licences and emergency measures during crises. Zero-rating educational content can therefore be seen as one mechanism through which operators discharge these public obligations rather than purely as voluntary philanthropy.

Second, treating connectivity as a public utility foregrounds questions of equity and governance. Who decides which sites are zero-rated? How are public-interest criteria defined and monitored? What oversight do regulators, civil society and teacher organizations have over these decisions? This framing pushes debates about zero-rating beyond narrow

Treating connectivity as a public utility foregrounds questions of equity and governance. Who decides which sites are zero-rated? How are public-interest criteria defined and monitored? What oversight do regulators, civil society and teacher organizations have over these decisions.

technical or commercial considerations towards questions about democratic accountability and the distribution of digital opportunities.

Navigating policy and industry complexities

Despite the evident promise of zero-rating, its path towards wider adoption is not straightforward. Regulators, telecom operators, civil society organizations and content providers often find themselves balancing competing priorities: financial viability, universal service, technological feasibility and the principles of an open, secure internet.

Net neutrality and public interest exceptions

Critics warn that exempting selected sites from data charges risks creating 'walled gardens' that privilege some services and undermine the principle that all internet traffic should be treated equally. India's 2016 ban on Facebook's Free Basics is frequently cited as a cautionary tale of how zero-rating can clash with net neutrality (Bradley, 2016; Lunden, 2016).

Yet many jurisdictions recognize public interest exceptions to strict net neutrality, particularly for educational or health content. In the South Africa example, government regulations during the COVID-19 state of disaster required zero-rating of local educational and health websites identified by public authorities, while leaving other forms of commercial zero-rating more tightly scrutinized (Government of South Africa, 2020; Bhandari, 2020). The Internet Service Providers Association and other trade groups have engaged critically with these measures, at times arguing for clearer net neutrality rules even as they implement mandated zero-rating (Ikigai Law, 2020).

One constructive way forward is to distinguish between closed, exclusive commercial arrangements and open, transparent schemes that any accredited educational provider can join. If zero-rating is governed by clear, public-interest criteria; if lists of zero-rated sites are transparent and contestable; and if oversight mechanisms are in place, then zero-rating does not necessarily have to undermine net neutrality. Instead, it can operate as a carefully regulated exception that supports equitable access to learning.

Commerical incentives, public funds and telecom industry prespectives

Many telecom operators worry that if they begin zero-rating content without adequate compensation or regulatory support, they will lose revenue from data packages. This concern is particularly salient in emerging markets, where margins can be tight, infrastructure costs are high and network maintenance is an ongoing challenge. These concerns must be taken seriously.

However, a more critical analysis of incentives highlights that operators are rarely neutral market actors. They benefit from public funds, subsidies and spectrum allocations; their licences often carry social obligations; and they are frequently included in national digital infrastructure plans. In South Africa, civil society organizations had to resort to legal action to compel some mobile network operators to fully implement the government-approved list

of zero-rated public-benefit and educational sites during the pandemic (DGMT, 2021). This underlines that zero-rating cannot be understood solely as a voluntary corporate gesture; it is also a question of compliance with public obligations and regulatory decisions.

At the same time, operators can realize strategic advantages from well-structured zero-rating initiatives. Enhanced brand reputation, improved customer loyalty, and predictable public subsidies or tax incentives can offset revenue lost from traditional data packages (OECD, 2019). A more digitally literate population, supported by zero-rated educational platforms, can expand the overall market for mobile services in the long run. The key challenge is to design policy instruments that align these incentives with robust public-interest safeguards.

Implementation gaps and governance

In theory, many governments recognize the potential of zero-rating. In practice, few have the policy frameworks, technical capacity, or governance mechanisms to implement such initiatives at scale and with integrity. Ensuring content quality, safeguarding data protection and online safety, selecting which educational providers qualify, and structuring transparent agreements with telecom operators all require expertise and political will.

Ensuring content quality, safeguarding data protection and online safety, selecting which educational providers qualify, and structuring transparent agreements with telecom operators all require expertise and political will.

Even where regulatory frameworks exist, implementation often lags. In South Africa, hundreds of educational and public-benefit websites were approved for zero-rating during the pandemic, yet not all operators implemented these measures promptly or consistently (DGMT, 2021; MyBroadband, 2020). Similar challenges arise elsewhere when lists of zero-rated sites are difficult to navigate, change frequently, or are not clearly communicated to schools and communities.

These governance gaps matter. If zero-rating schemes are opaque, inconsistent or captured by a narrow set of commercial providers, they risk undermining both net-neutrality principles and the broader goal of meaningful connectivity.

Envisioning the future

Whether framed as a policy instrument or an equity imperative, zero-rating stands out for its simplicity and scalability. This does not mean zero-rating alone can solve all inequities; rather, it must be woven into broader strategies that include investments in teacher training, curriculum development and digital infrastructure. Yet, despite the potential impact and modest technical demands, zero-rating remains a conspicuous blind spot in most international education and connectivity agendas.

At its best, zero-rating represents a practical expression of what can be achieved through collaboration: governments offering policy support and funding, telecom operators leveraging their networks as part of a public-utility ecosystem, content providers ensuring

pedagogical quality, AI developers committing to transparent and equitable tools, and communities spreading awareness of trusted, data-free education resources. International bodies, such as UNESCO, can guide this work by disseminating research, encouraging cross-sector dialogue, and highlighting successes and pitfalls that can inform future policy.

Embracing zero-rating with transparency, inclusivity, and sustainability, and by situating it within a richer vision of meaningful connectivity, could move a step closer to a future of education that transcends financial barriers, safeguards learners' rights and expands genuine opportunities for all learners.

References

Azzahra, N. F. 2020. *Addressing Distance Learning Barriers In Indonesia Amid The COVID-19 Pandemic*. Jakarta, Centre for Indonesian Policy Studies. (Policy Brief, No. 2.)

Bhandari, V. 2020. *Improving Internet Connectivity During Covid-19*. Oxford, Pathways for Prosperity Commission. (Digital Pathways at Oxford Paper Series, 4.)

Bradley, T. 2016. Zero rating: A boon to consumers, or a net neutrality nightmare? *Brookings*.

DGMT. 2021. *Zero-rating mobile services*.

Global Education Monitoring Report Team. 2023. *Global Education Monitoring Report, 2023: Technology In Education: A Tool On Whose Terms?* Paris, UNESCO. https://doi.org/10.54676/UZQV8501

Guruli, N., Flores, C., Laken, F., Duke, A., Singh, A., Power, M., Power, T., Hawthorne, R., Robb, G., Khumalo, P., Dayile, A. and Bird, W. 2020. *Access Denied: Internet access and the right to education in South Africa*. Chicago/Johannesburg/Randburg, Global Human Rights Clinic/ALT Advisory/Acacia Economics/Media Monitoring Africa.

Ikigai Law. 2020. *Net neutrality regulation in South Africa*.

ITWeb. 2021. Vodacom's e-school sees increased uptake. *ITWeb*.

Krämer, J. and Peitz, M. 2018. A fresh look at zero-rating. *Telecommunications Policy*, Vol. 42, No. 7, pp. 501–13. https://doi.org/10.1016/j.telpol.2018.06.005

Lunden, I. 2016. India blocks Facebook's Free Basics, other zero-rated mobile services over net neutrality. *TechCrunch*.

Lythreatis, S., Sing, S. and El-Kassar, A. 2022. The digital divide: A review and future research agenda. *Technological Forecasting and Social Change,* Vol. 175. https://doi.org/10.1016/j.telpol.2018.06.005

Martin, K. 2022. Using mobile phones to enhance small group dialogic learning: A design based approach to educational innovation in rural east africa. Ph.D. thesis, University of Cambridge, United Kingdom. https://doi.org/10.17863/CAM.85755

McBurnie, C., Adam, T., Kaye, T. and Haßler, B. 2020. *Zero-rating Educational Content In Low- And Middle-income Countries*. EdTech Hub. (Helpdesk Request Response, No. 8.)

MyBroadband. 2020. Here is the full list of zero-rated websites in South Africa. *MyBroadband*.

OECD. 2019. *The effects of zero rating*. Paris, OECD Publishing. (OECD Digital Economy Papers, No. 285.)

Ramadhan, S., Elfia, S. and Indriyani, V. 2021. Online distance learning while the COVID-19 pandemic: Implementation, evaluation, and expectations. *Proceedings of the 4th International Conference on Language, Literature, and Education (ICLLE-4 2021)*. Dordrecht, Atlantis Press, pp. 267–75.

Republic of South Africa, Department of Communications and Digital Technologies. 2020. *Zero-rating of websites for education and health*. Pretoria, Republic of South Africa. (Government Gazette, No. 43413.)

Safaricom. 2020. Safaricom, education content providers partner to enable free access to digital learning.

Telkomsel. 2020. Respon situasi COVID-19, Ruangguru bekerjasama dengan Telkomsel, belajar di Ruangguru bebas kuota [Responding to the COVID-19 situation, Ruangguru collaborates with Telkomsel, learning on Ruangguru is quota-free]. (In Bahasa Indonesian.)

UNESCO. 2023a. *Zero-rating (Digital Transformation Collaborative Finance Toolkit factsheet)*. Paris, UNESCO.

––. 2023b. *Vodacom e-School (Financing the Digital Transformation of Education Toolkit)*. Paris, UNESCO.

Vodacom. 2020. *Integrated annual report 2020*. Midrand, RSA, Vodacom Group Limited.

About the author

Kevin Martin is Managing Director of the Digital Education Futures Initiative at Hughes Hall, University of Cambridge, a global research and development centre examining the impact of emerging technologies on education. His career spans two decades at the intersection of social innovation and education, with current research focused on dialogic education, educational technology and education equity.

Part VI

How might AI affect imagination about the future?

Treating the current forms and infrastructures of AI as unavoidable risks narrowing our collective imagination and foreclosing alternative futures that could be more just, equitable and sustainable for humanity and the planet. With no future scenario off the table, the power of human imagination remains one of the most valuable faculties for enabling diverse and promising future circumstances. Challenging taken-for-granted assumptions and 'anticipatory regimes' that have shaped conventional educational institutions and systems is part of reimagining the future of education (Amsler and Facer, 2017).

In their piece **'AI and Education for collective intelligence: A futures perspective',** for instance, Imogen Casebourne and Rupert Wegerif explore how AI implementation can be reimagined in order to foster some of the collective aims – termed 'collective intelligence' – of education that conventional pedagogies are often lacking. They observe that education systems worldwide tend to focus on individual learners' personal skills and capacities, without placing sufficient emphasis on teaching 'people how to collaborate with one another, with technology and at scale' (p. 145). Instead, they propose means by which AI tools can support pedagogies that cultivate collective intelligence.

Considering that even dystopian scenarios can play an important role in determining which destinies we do not wish to find ourselves in, James Ransom takes us to a prestigious university in 2035 in his speculative piece **'AI labs and the great academic migration: a speculative scenario',** where human expertise has been largely (though not entirely) hollowed out. It explores what could become of the precarious new balance between corporate-versus-academic knowledge production and autonomy versus automation. Finally, a future possibility that should never be foreclosed is the prospect that, for any number of reasons, we could find ourselves in a future scenario

that does not involve AI to a significant degree, whether through choice or through circumstance. In **'After AI',** Serge Pouts-Lajus points out that the future of AI, even just a few years from now, is far from settled. Looking at the promises of other technological revolutions across time, he reflects on what could be both similar and different about the current moment. He urges education to consider all possible future trajectories as valid – even low-tech ones – given the array of ecological, intellectual and institutional complexities that intersect with AI's trajectory. Pouts-Lajus argues that no matter what the unknowable future may hold, education should remain inexorably focused on its mission 'to form future adults who are educated, autonomous and able to find their place in a world where technology will be present, but may not be dominated by AI' (p. 153). Such educated human beings would be best prepared if they are 'capable not only of adapting to the unknown future that awaits them, but also of helping to change it' (ibid).

AI labs and the great academic migration: A speculative scenario

James Ransom

In 2030, academia has changed. As AI labs rise, universities adapt, and researchers like Professor Wright straddle both worlds in this speculative glimpse into the future.

Exploring speculative scenarios – however unlikely they may seem – allows us to envision the future we want (or perhaps want to avoid). In the process, we can examine the role of our institutions – and the people who work in them – in planning for the future. In this piece, we visit a higher education campus in the United Kingdom, five years from now.

Scenario location: A university town in England.

Year: 2030.

Professor Eleanor Wright is walking through the sunny campus. It's Friday – her teaching day. The other four days of her working week are spent at the offices of the AI lab in the big city, an hour away. Her university contract requires six hours of physical presence per week, primarily for teaching undergraduates and delivering a prestigious lecture series.

As she passes ancient buildings, she reflects on how her career has transformed over the past few years. As a tech-savvy researcher with experience of managing big projects, a strong publication record and a string of national media appearances, her colleagues called her a 'superstar academic'. Most of these colleagues are now gone. As AI models continued to improve and improve – nobody quite knows when the breakthrough was, if there even was one – the power of the big AI labs grew too. Their hunger for new knowledge and their immense computing power muscled universities out of the way. Cash strapped after years of funding crises, only the superstars managed to hold on to their jobs.

The title 'professor' remains coveted, but its meaning has fundamentally changed – her primary value to the university now lies not in the research she performs within its walls (which is negligible), but in the funding her secondment brings. Her reputation, as the face of the human-machine research team at the AI lab, also helps attract undergraduates. The company that eviscerated her university has made her more famous than her appearance as a BBC panelist, she reflects wryly.

The great migration

Back in 2025 there was talk of an 'intelligence explosion' (MacAskill and Moorhouse, 2025) that would fundamentally reshape knowledge creation. This turned out to be true, sort of. For the average bystander, the explosion was gradual and quiet. After 2025, the migration

of elite researchers from universities to industry-funded AI laboratories accelerated, creating what some have termed the 'great academic migration'. Many (non-superstar) academics protested, but university leadership teams played along. Whether they were pragmatic visionaries or worn down by decades of under-investment depends on who you ask. But framing this migration as a series of secondments offered leaders a fig leaf of self-respect and gave the labs what they needed most: the validation and legitimacy of the storied institutions they were overtaking; a badge to attach to their superior computational resources.

For the average bystander, the explosion was gradual and quiet.

This migration played out unevenly across the globe. In high-income countries, the pattern followed the United Kingdom's example, with elite research universities transformed into teaching-focused institutions with a handful of superstar researchers. In emerging economies, particularly in parts of Africa and Asia, national governments intervened more forcefully, creating public-private partnerships that preserved more university autonomy while still accessing AI lab resources. The resulting global knowledge landscape became more complex, with research hubs distributed according to the decisions of AI labs rather than university prestige.

The evolution of research

The prevailing narrative that this represents a catastrophic loss for universities requires nuance. Most institutions survived this transition, but are transformed. What we are witnessing is not the end of the university but rather a reformulation of its purpose. Universities are centres of teaching excellence, of critical enquiry, of engagement with small businesses. Cutting-edge research has mostly moved to the AI labs – shiny university research parks are now student accommodation blocks and seminar rooms.

The nature of research itself has also fundamentally changed. 'AI scientists' – autonomous systems capable of formulating hypotheses, designing and conducting experiments, and publishing findings – evolved from prototypes to mainstream research tools. These systems were first developed in the mid-2020s as open-source projects, and could produce publication-quality papers for the cost of a pizza (SakanaAI, 2024) and share findings amongst themselves (Schmidgall and Moor, 2025). They were later refined by every major AI laboratory, accelerating research productivity in fields amenable to computational methods. The explosion of research papers (Wilcox, 2023) predicted by early commentators did indeed materialize, though the feared quality crisis was largely averted through sophisticated review systems.

The relationship between Professor Wright and her AI hosts illustrates the new academic compact. She provides contextual expertise, ethical oversight and academic credibility; the company provides computational resources that would bankrupt a university department. The research outputs carry both her name (and university affiliation) and those of the lab

technicians, and are published in open-access journals. The university receives a generous annual payment (balancing its books); the company retains the IP and commercialization rights (boosting its share price). This arrangement is the standard model for frontier research across disciplines from theoretical physics to computational sociology.

Resilience and resistance

We should be careful, however, not to overstate the completeness of this transition. Some research domains remain stubbornly resistant to the magic of the AI labs, particularly those requiring extensive human involvement or tacit knowledge. The arts and humanities have proven especially resilient and are seeing something of a resurgence amongst undergraduates. What we have seen is not a wholesale replacement of human researchers, but rather a redistribution of cognitive labour, with AI systems taking on increasingly sophisticated components of the research process. In parallel, human 'taste' is highly prized – the market value of human artwork continues to shatter records.

What we have seen is not a wholesale replacement of human researchers, but rather a redistribution of cognitive labour.

The implications for universities are far-reaching, and many are still catching up. The teaching-research nexus – the idea that university education should be enriched by active research – has been reconceptualized. Rather than teaching their own research, academic staff now teach the process of engaging critically with a rapidly expanding knowledge frontier, mostly produced outside university campuses. Students learn not just established knowledge but how to navigate, evaluate and contribute to an information system dominated by AI-human research teams. This required new pedagogical approaches focused on developing what might be termed 'second-order knowledge' – not merely understanding content but understanding the systems through which knowledge is produced, validated and disseminated.

The future of universities

This transformation continues to raise urgent questions about the future of universities. If their research function continues to diminish, what justifies their privileged position in society? The answer lies partly in their role as custodians of methodological rigor and critical thinking in an information environment increasingly dominated by commercial interests. Universities have become essential counterweights to corporate knowledge production, training students to interrogate the biases, limitations and ethical implications of AI-generated research – but there's a sense this counterweight isn't quite heavy enough.

Universities have become essential counterweights to corporate knowledge production, training students to interrogate the biases, limitations and ethical implications of AI-generated research – but there's a sense this counterweight isn't quite heavy enough.

When asked to comment on the string of breakthroughs now taking place in AI labs, their CEOs are quick to praise universities. Professor Wright's weekly journey between university and laboratory represents not the decline of academic tradition but its evolution, they say – a recognition that knowledge creation has always adapted to the technological and institutional constraints of its time. Professor Wright is less certain. As she enters her office, she sees only the ghosts of her former colleagues and coworkers.

References

MacAskill, W. and Moorhouse, F. 2025. *Preparing for the Intelligence Explosion*. San Francisco, Forethought.

SakanaAI. 2024. *The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery*.

Schmidgall, S. and Moor, M. 2025. AgentRxiv: Towards Collaborative Autonomous Research. arXiv preprint, arXiv:2503.18102.

Wilcox, C. 2023. ScienceAdviser: Scientists are publishing too many papers—and that's bad for science. *Science*.

About the author

James Ransom is an Honorary Senior Research Fellow at University College London Institute of Education, and a Specialist Advisor to the European Bank for Reconstruction and Development. His work looks at how universities can help solve challenges facing society, specifically upskilling and retraining workers, and the impact of AI on society.

AI and education for collective intelligence: A futures perspective

Imogen Casebourne and Rupert Wegerif

Education plays a vital role in cultivating collective intelligence, the capacity to think and act together to address global challenges. Could AI amplify education for collective intelligence, or undermine it?

Many global challenges arise not solely from a lack of knowledge but from a failure to think and work effectively together at scale (Buckingham Shum, 2025). Education systems worldwide are designed to nurture individual learners, focusing on personal skills and capacities. But to address the complexities of an uncertain future, there is also a need teach people how to collaborate with one another, with technology, and at scale to catalyse what has been termed *collective intelligence* (Hogan et al., 2023; Casebourne et al., 2024).

Collective intelligence refers to the capacity of groups to think, learn and solve problems together in ways that exceed the capability of any single individual (Wooley and Gupta, 2023). It may combine dialogue, collaboration, critical reflection and often technological tools to enable diverse groups to address problems that have no singular cause or solution. Community projects, global citizenship initiatives, participatory research, SWARM (a form of crowd voting where multiple insights from varying perspectives inform decisions) and citizen science programmes are relevant methods that support learning together at scale (Hogan et al., 2023). Generative AI, derived from massive data sets, has itself been viewed as a form of collective knowledge (Burton et al., 2024; Vasilaki, 2025). This implies that if used by individuals or groups to explore issues, help with projects and answer questions, it can be seen as drawing from collective intelligence (although this raises questions about how representative those training data sets were, as discussed below).

If we think of intelligence as not only contained inside individual minds but as also emerging *between* people, then dialogue can be seen as key to collective intelligence .

Traditional approaches to education have focused on knowledge transfer and reproduction, which can result in strong individual exam results but leave learners lacking skills for negotiation, co-creation and an ability to collaborate across academic disciplines or cultures (Hogan et al., 2023). It is no longer sufficient for learners to excel at reproducing the successes of the past; rather, they need to excel at learning with others: peers, mentors, communities and technologies, including AI, to address novel challenges. However, many AI in education (AIED) initiatives involve adaptive personalized tutoring or automated grading focused on gaining traditional skills more efficiently. While potentially useful, this emphasis on individual achievement will not necessarily foster collective intelligence.

If we think of intelligence as not only contained inside individual minds but as also emerging *between* people, then dialogue can be seen as key to collective intelligence (Casebourne and Wegerif, 2024; Mercier and Sperber, 2017; Wegerif, 2025). More than talking, dialogue

is a process of posing questions, listening, challenging assumptions and refining ideas by interacting with multiple perspectives. Dialogic pedagogies involve open-ended questioning, reciprocal listening and interpretive reasoning (Wegerif, 2025). Dialogic classrooms can support learning how to co-construct understanding. For instance, a teacher might pose a contentious policy issue, such as whether gene editing is ethical and students collectively explore a range of potential viewpoints, working in groups to propose principles or guidelines, with the teacher orchestrating an environment in which meaningful dialogue can flourish. Such dialogues can develop capacity to engage in larger collective intelligence projects by fostering empathy, respect and curiosity (Mercier and Sperber, 2017).

This is an argument for teaching people how and when to use AI in collective intelligence projects and for designing systems that avoid rewarding individual contributions that draw on AI in situations where diversity of individual experience is key.

Hogan et al. (2023) suggest project-based learning as another scaffold for collective intelligence which enables the integration of various approaches. Guided by a teacher, students might work in groups on real-world problems, for example investigating water quality in local rivers: gathering data, synthesizing ideas and arriving at solutions they could not have achieved alone. This can cultivate deeper subject knowledge (e.g. in science) while enabling participants to practice communication, conflict resolution, role assignment and iterative refinement, which are collective intelligence competencies. In practice, project-based learning can be challenging to orchestrate in traditional classrooms if not all students are motivated by the focus of a particular project.

However, in futures workshops run with teachers and young people, some expressed interest in spending some time on global projects, offering them more choice in the project and their role within it, potentially increasing motivation. For example, internationally linked communities of students might share data on climate change effects, analysing collectively and proposing local, adaptive responses (Hogan et al., 2023; Mercier and Sperber, 2017). Aspects of this type of collaborative learning platform can and do exist without AI, for example, citizen science projects predate generative AI (Ceccaroni et al., 2019), but aspects of newer forms of AI, such as its potential to direct group attention at scale (Tang and Weyl, 2024), to translate rapidly between diverse languages or to summarize emergent consensus points, have the potential to reduce barriers to participation. AI supported collective intelligence collaboration platforms now exist, for example, BCause, PSi and Pol.is, although these are not necessarily developed with education as the focus.

However, AI may also have the potential to *reduce* group collective intelligence, because where diverse individuals draw on the same AI system for inspiration they may come up with similar ideas and solutions, thereby losing the collective benefits gained when individuals contribute more diverse approaches arising from inspiration and understanding gained from their disparate contexts (Doshi and Hauser, 2024; Zhang, Xu and Alvero, 2025). This is an argument for teaching people how and when to use AI in collective intelligence

projects and for designing systems that avoid rewarding individual contributions that draw on AI in situations where diversity of individual experience is key.

Additionally, at an individual level, concerns have been raised that using AI might damage *individual* capacity to cultivate skills such as critical thinking (Gerlich, 2025; Fan et al., 2024) if people come to rely uncritically on AI. Conversely research is currently examining ways in which AI systems might *help* individuals cultivate skills associated with participation in collective initiatives, including critical thinking (Li et al., 2025), creativity (Heyman et al., 2024) and collaboration and communication (Rojas et al., 2022; VanLehn et al., 2021).

Some communities may lack the infrastructure to participate, risking a collective intelligence gap.

Designers of AI systems to support education for collective intelligence therefore should consider several questions. How can systems draw on AI to enhance collective intelligence rather than reduce it? In what ways can processes of group collaboration, including any input by AI be made transparent? How will diverse perspectives be integrated and honoured? Will there be appropriate processes to record dissenting opinions, as groups progress to a consensus? In decision-making, will there be provision for revision and appeals? What approaches are needed to ensure individual participation feels safe and leads to flourishing? Finally, and importantly, what safeguards might prevent groups from harmoniously and collectively working towards unethical ends?

Some communities may lack the infrastructure to participate, risking a collective intelligence gap. Without inclusive policies and investment, an approach that *could* bring people together might inadvertently widen divides, and as with any technology, access to powerful AI tools is uneven. Actively co-creating in partnership with marginalized groups should help to ensure a digital environment that supports broad participation and recognition (Mercier and Sperber, 2017), but it will be necessary to ensure that designs involving the collection of analytic data are participatory, accepted by potential users and that any resulting data is managed appropriately. Furthermore, since AI systems often reproduce biases embedded in training data, designers and educators should ask whether diverse voices, languages and perspectives are appropriately represented and fully integrated. If answers facilitated by AI aggregation draw from overly homogeneous LLM-influenced thinking, which itself draws from incomplete and biased data sets, there is a danger that AI-collective intelligence systems could mirror or magnify exclusionary practices (Burton et al., 2024) Finally, it will be important to build trust via transparent processes (Kehler et al., 2025).

Teachers will be crucial in modelling the meta-level activities that will be needed for successful collaborative learning with novel AI tools. Teaching activities might include the teaching of dialogue and demonstrating and modelling dialogic virtues, such as civility and empathy, to help ensure that collective intelligence spaces and processes do not degenerate and become sites of bullying or trolling. Teachers might also teach how to question, how to challenge or how to interpret automated recommendations. None of this will be easy. It will require professional development for teachers that balances facility with

new technology and comfort with collective intelligence aligned learning processes (Hogan et al., 2023), as well as policy frameworks and funding streams that value collaboration, open-ended projects and community engagement as much as conventional test scores, and all this at a time when some policy-makers may be viewing AI as a means of reducing investment while maintaining standards.

Collective intelligence, supported by AI, as one goal of education, offers a vision for a future education system that teaches not only 'what' but also 'how' to negotiate meaning, co-construct solutions and cultivate shared understanding across difference.

Collective intelligence, supported by AI, as one goal of education, offers a vision for a future education system that teaches not only 'what' but also 'how' to negotiate meaning, co-construct solutions and cultivate shared understanding across difference. It challenges us to expand beyond monological models of teaching and learning towards dialogic, project-based and globally networked practices. As we move into a future increasingly shaped by AI, the climate emergency and other instabilities, collective thinking and action will be crucial. In the future, institutional structures, teaching practices and resource allocations might evolve to enable learners worldwide to participate in collective intelligence. By teaching dialogue, connecting local and global contexts, and designing AI systems that centre equity and human agency, education for collective intelligence could involve projects that contribute to shared flourishing. However, there's no blueprint for this; it will require multiple avenues of experimentation, reflection and co-creation, which are themselves hallmarks of collective intelligence.

References

Buckingham Shum, S. 2025. AI for learner flourishing in the age of the polycrisis—on the edge of the metacrisis. *The Blue Dot,* Vol. 18, pp. 65–75.

Burton, J. W., Lopez-Lopez, E., Hechtlinger, S., Rahwan, Z., Aeschbach, S., Bakker, M. A., Becker, J. A., Berditchevskaia, A., Berger, J., Brinkmann, L., Flek, L., Herzog, S. M., Huang, S., Kapoor, S., Narayanan, A., Nussberger, A.-M., Yasseri, T., et al. 2024. How large language models can reshape collective intelligence. *Nature Human Behaviour,* Vol. 8, No. 9, pp. 1643–55. https://doi.org/10.1038/s41562-024-01959-9

Casebourne, I., Shi, S., Hogan, M., Holmes, W., Hoel, T., Wegerif, R. and Yuan, L. 2024. Using AI to support education for collective intelligence. *International Journal of Artificial Intelligence in Education,* Vol. 35, pp. 1597–629.

Casebourne, I. and Wegerif, R. 2024. The role of AI language assistants in dialogic education for collective intelligence. P. Illic, R. Wegerif and I. Casebourne (eds), *Artificial Intelligence in Education: The Intersection of Technology and Pedagogy.* Cham, Springer Nature, pp. 111–25.

Ceccaroni, L., Bibby, J., Roger, E., Flemons, P., Michael, K., Fagan, L. and Oliver, J. L. 2019. Opportunities and risks for citizen science in the age of artificial intelligence. *Citizen Science: Theory and Practice,* Vol. 4, No. 1, pp. 1–14. https://doi.org/10.5334/cstp.241

Doshi, A. R. and Hauser, O. P. 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, Vol. 10, No. 28. https://doi.org/10.1126/sciadv.adn5290

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X. and Gašević, D. 2024. Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology,* Vol. 56, No. 2, pp. 489–530. https://doi.org/10.1111/bjet.13544

Gerlich, M. 2025. AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, Vol. 15, No. 6. https://doi.org/10.3390/soc15010006

Heyman, J. L., Rick, S. R., Giacomelli, G., Wen, H., Laubacher, R., Taubenslag, N., Knicker, M., Jeddi, Y., Ragupathy, P., Curhan, J. and Malone, T. 2024. Supermind Ideator: How scaffolding human-AI collaboration can increase creativity. *Proceedings of the ACM Collective Intelligence Conference.* New York, Association for Computing Machinery, pp. 18–28. https://doi.org/10.1145/3643562.3672611

Hogan, M. J., Barton, A., Twiner, A., James, C., Ahmed, F., Casebourne, I., Steed, I., Hamilton, P., Shi, S., Zhao, Y., Harney, O. M. and Wegerif, R. 2023. Education for collective intelligence. *Irish Educational Studies,* Vol. 44, No. 1, pp. 137–66.

Kehler, T. P., Page, S. E., Pentland, A., Reeves, M. and Brown, J. S. 2025. *Amplifying human creativity and problem solving with AI through generative collective intelligence.* arXiv preprint, https://doi.org/10.48550/arXiv.2505.19167

Li, C.-J., Hwang, G.-J., Chang, C.-Y. and Su, H.-C. 2025. Generative AI-supported progressive prompting for professional training: Effects on learning achievement, critical thinking, and cognitive load. *British Journal of Educational Technology,* Vol. 56, No. 6, pp. 2550–72.

Mercier, H. and Sperber, D. 2017. *The Enigma Of Reason*. Cambridge, Mass., Harvard University Press.

Rojas, M., Sáez, C., Baier, J., Nussbaum, M., Guerrero, O. and Rodríguez, M. F. 2022. Using automated planning to provide feedback during collaborative problem-solving. *International Journal of Artificial Intelligence in Education,* Vol. 33, pp. 1057–91. https://doi.org/10.1007/s40593-022-00321-2

Tang, A. and Weyl, G. 2024. *Plurality: The future of collaborative technology and democracy*. Plurality.net.

VanLehn, K., Burkhardt, H., Cheema, S., Kang, S., Pead, D., Schoenfeld, A. and Wetzel, J. 2021. Can an orchestration system increase collaborative, productive struggle in teaching-by-eliciting classrooms? *Interactive Learning Environments*, Vol. 29, No. 6, pp. 987–1005. https://doi.org/10.1080/10494820.2019.1616567

Vasilaki, E. 2025. *In dialogue with intelligence: Rethinking large language models as collective knowledge*. arXiv preprint, https://doi.org/10.48550/arXiv.2505.22767.

Wegerif, R. 2025. *Rethinking Educational Theory: Education as Expanding Dialogue.* Cheltenham, Edward Elgar Publishing.

Woolley, A. W. and Gupta, P. 2023. Understanding collective intelligence: Investigating the role of collective memory, attention, and reasoning processes. *Perspectives on Psychological Science,* Vol. 19, No. 2, pp. 344–54. https://doi.org/10.1177/17456916231191534

Zhang, S., Xu, J. and Alvero, A. 2025. Generative AI meets open-ended survey responses: Research participant use of AI and homogenization. *Sociological Methods & Research,* Vol. 54, No. 3, pp. 1197–242. https://doi.org/10.1177/00491241251327130

About the authors

Dr Imogen Casebourne, Research Lead at the Digital Education Futures Initiative at Hughes Hall, University of Cambridge, investigates how AI and educational technologies can be designed to foster inclusive experiences of dialogue, serendipity and collective intelligence in learning. She co-authored 'Using AI to Support Education for Collective Intelligence' (2025) and presented at the inaugural Symposium on Educating for Collective Intelligence in 2024.

Professor Rupert Wegerif is a member of the Faculty of Education at the University of Cambridge and founder of the Digital Education Futures Initiative. He is the author of several books and articles in the area of educational theory, educational psychology and education with technology. His recent talks, articles and books offer a theoretical foundation for the design of dialogic education with AI and with technology more generally.

After AI

Serge Pouts-Lajus[2]

The future of AI is far from settled. This author argues that education should focus on preparing people not only to adapt to the unknown futures that awaits them but also to help change them.

It is not the first time that the arrival of a technological innovation has come with the promise of a profound transformation of education. As early as 1913 Thomas Edison (1847–1931) declared: *'Books will soon be obsolete in the schools.… It is possible to teach every branch of human knowledge with the motion picture. Our school system will be completely changed inside of ten years.'* In the early 1980s, Seymour Papert (1928–2016), the brilliant inventor of the Logo language and pioneer of educational robotics, was convinced that the personal computer would, within a few years, bring about the collapse of school as we know it. In 2003, Michel Serres (1930–2019) announced in *Petite Poucette* (published in English as *Thumbelina: The Culture and Technology of Millennials*) a civilizational shift that would sweep away the school and its mission of transmission: '*Knowledge? Here it is, everywhere on the web, available, objectified. To transmit it to all? Now, all knowledge is accessible to all. How to transmit it? There, it's done.*'[3]

Continuity in disruptions

AI presents itself as the latest episode in this series. It shares many similarities with its predecessors, three in particular. First, exaggeration: the promised transformations always turn out to be far from what is actually observed in classrooms and lecture halls. Second, the way educational systems embrace technological innovation: institutional leaders support it eagerly and unhesitatingly, while teachers are divided, some enthusiastic, others hostile, with every shade in between. Education seems always to resist the grip of novelty. Third, the absence of anticipation: technological innovations emerge and spread without concern for the resistance they will face or the disruptive effects that further progress is sure to provoke. The promises of educational television were swept aside by those of computing, which were in turn submerged by the Internet, before AI shifted attention to yet another front.

It is not the first time that the arrival of a technological innovation has come with the promise of a profound transformation of education.

2 Originally published in French.

3 (Unofficial translation by UNESCO editorial team, from Michel Serres, Petite Poucette, Le Pommier, 2003). Original quote: « Le savoir ? Le voilà, partout sur la toile, disponible, objectivé. Le transmettre à tous ? Désormais, tout le savoir est accessible à tous. Comment le transmettre ? Voilà, c'est fait. »

By analogy, one might be tempted to predict that AI's promises will likely go unfulfilled, that schools will stubbornly resist it, and that some unexpected innovation will sooner or later render AI obsolete.

Searching for the specificities of the AI moment

The suddenness and force with which generative AI entered schools and universities placed teachers and institutions on the defensive.

History never repeats itself exactly. It is therefore possible that, at least in certain respects, the AI moment is distinct from its three predecessors, and that the simple continuation of the series, as just suggested, may not hold true. To be clear, it is helpful to distinguish two strands of AI in education. The first has a long history, beginning with B.F. Skinner (1904–1990), and focuses on the individualization of learning, first through mechanical teaching machines, then computers, and today through AI strengthened by cognitive science. The second, much more recent, is embodied in generative AI, with ChatGPT as its emblematic version. From the summer of 2023 onward, it stirred enthusiasm among students and panic among many teachers. Functionally, this strand introduced a decisive innovation: the ability to carry out high-level natural language dialogue with human users, based on a vast data base that allows it to address any subject and respond to any request. Students immediately saw its advantages, forcing teachers to revise their assessment methods and, indirectly, their pedagogical practices as a whole. This 'AI moment' thus differs from all its predecessors in the way it entered, noisily from the bottom up, through learners themselves, and through its unprecedented capacity to shake the established order of schools and universities.

These two strands, one slower and rooted in the past, focused on individualized learning; the other sudden and disruptive, carried by generative AI, are destined to converge. Educational resources are beginning to appear that can sustain high-level pedagogical dialogue in the learner's language, through text, sound and increasingly realistic human avatars. This brings closer B.F. Skinner's dream of a 'teaching machine', more effective than any human educator. The current moment in educational technology can therefore be seen as the latest step in a trajectory beginning more than a century ago, one in which the ultimate goal may not be to replace teachers with machines, but certainly to place them in competition.

The prophetic narrative accompanying AI's development in education thus lies partly in continuity with previous episodes, but with features that deserve closer examination.

Something new in the prophecy

The promises attached to cinema, computing and the Internet in education were tempered by anticipated drawbacks, but their positive arguments were numerous, persuasive and easy to present – sometimes even inspiring. Moving images, for example, undeniably added a valuable pedagogical tool. Computing with programmable machines opened education to

a new discipline, potentially transformative in itself, as Seymour Papert believed. The Internet finally offered teachers and learners direct access to the world, realizing an educational ideal championed by Michel Serres, but also earlier by Jean-Jacques Rousseau and many educators.

We must consider one last characteristic of technological innovations in education: their inability to anticipate the next wave.

Nothing of the sort with AI. The suddenness and force with which generative AI entered schools and universities placed teachers and institutions on the defensive, unlike the generally positive reception given to previous technologies. Today, moreover, well beyond education, many voices are raised against AI companies and their products, especially on ethical grounds that resonate strongly with teachers.

Here is a sample of such concerns. Generative AI production methods violate intellectual property rights. Their statistical functioning asserts truth and falsehood with equal confidence, serious for learners in training. AI depends on the labour of thousands of low-paid workers tasked with content moderation. The energy consumption of data centres contributes increasingly to the planet's habitability crisis. Finally, philosopher Daniel Dennett argues that AI's fabrication of 'fake humans' poses a mortal danger to democratic societies and should be punished as severely as counterfeiting money. The balance of arguments now tilts heavily against AI. Compared to earlier technological innovations, the difference is striking.

It is thus not surprising that education's reception of AI, especially generative AI, is marked not yet by rejection, but by a kind of resigned acceptance of what appears inevitable: *AI is here, we cannot stop it, so we must make the best use of it to prepare young people to live with it.* Hardly an enthusiastic stance…

After AI

Educational systems would do well to stay focused on their main mission: to form future adults who are educated, autonomous and able to find their place in a world where technology will be present, but may not be dominated by AI – adults capable not only of adapting to the unknown future that awaits them, but also of helping to change it.

Finally, we must consider one last characteristic of technological innovations in education: their inability to anticipate the next wave. Here too, the AI moment may prove different from its predecessors. For while it was difficult for Thomas Edison to foresee the computing revolution or for Seymour Papert to anticipate the Internet, it may be easier for us to guess what will follow AI, provided we look beyond industry hype. One possibility is the environmental crisis, whose inevitability is scientifically established, and which could seriously undermine AI's own momentum. It is reasonable to expect that the next technological revolution will be characterized by retreat, slowdown, reduction: *Low Tech*.

Such a prospect, though impossible to prove, should nonetheless be taken seriously by institutional leaders. For education, there is no rush. The future of AI beyond the next two years is far from settled. In this context, educational systems would do well to stay focused on their main mission: to form future adults who are educated, autonomous and able to find their place in a world where technology will be present, but may not be dominated by AI – adults capable not only of adapting to the unknown future that awaits them, but also of helping to change it.

References

Dennett, D. C. 2023. The problem with counterfeit people. *The Atlantic*.

Papert, S. 1981. *Jaillissement de l'esprit : ordinateurs et apprentissage*. Paris, Flammarion. (In French.)

Serres, M. 2003. *Petite Poucette*. Paris, Le Pommier. (In French.)

Watters, A. 2021. Teaching Machines: The History of Personalized Learning. Cambridge, Mass., The MIT Press.

About the author

Serge Pouts-Lajus taught mathematics before turning his attention to digital technologies and their applications in education. He has taken part in European research and development projects and founded Education et Territoires, a company that supports national and local public authorities responsible for education.

Conclusion

Three horizons to empower education in algorithmic times

There are many kinds of conversations about AI in circulation. This volume has steered away from narrow conversations about whether AI should be considered 'good' or 'bad' to instead consider how its algorithmic designs, goals and preferences are actively playing a role in individual and collective lives, in education settings and beyond.
Taking a long-term view in this regard shows that we have a unique window in this moment of inflection - a window in which the answers to these questions remain undetermined and far from settled. Urgency comes from the necessity to think and act carefully while the window is still open. And yet, countries and decision-makers continue to place economic bets on intelligent machines despite the uncertainties. These bets are risky given the speed at which longstanding systems, from education to the workforce, are being dismantled in the race to invest in novelty. Should the bubbles burst, how much will have been displaced in the race for early adoption?

A key factor in the race for AI innovation is *time* itself. Associated with time is an implicit (or sometimes explicit) fear of being left behind and the pressure to leapfrog towards the latest technological promise. How we use time matters, both in our imaginations about the future and for what we choose to do in the present.

Reflections on the questions raised in this volume, then, can benefit from considering how education can respond proactively to these 'algorithms in the room' across three time horizons:

- **First, in the immediate term,** we can begin by seeing and recognizing the presence and implications of AI algorithms in pedagogy, curricula and educational decision-making today.
- **Second, in the near term,** education actors can steer the implementation of algorithms in educational contexts, leveraging their beneficial uses, and confronting or disarming them when they negatively impact rights to privacy and intellectual freedom, ensuring protection from harm, and sustaining the essential role of education as a public endeavour and a common good.

- **Third, in a longer-term perspective,** people within education must become the authors of algorithms that affect them, realigning them around the central aims of education.

Figure 1: Three horizons to empower education in algorithmic times



© UNESCO

Horizon 1: See the algorithm

Seeing algorithmic influences. Students and educators today need to know that AI systems are not magical, nor infallible, but are the outcome of human designs, preferences, choices and biases. They need to know when they are interacting with a machine or a person, and to understand the basic premise of how algorithms work. Many of the think pieces involved moments – sometimes sudden, sometimes gradual – in which the authors came to both detect the presence of an algorithm's influence as well as acquire an awareness that its aims might not be entirely aligned with presumed intentions or expectations. This is a capacity that education must continue to hone across its ecosystem.

Identifying intended and unintended effects. Schools and decision-makers need to be explicit about the purpose of incorporating an AI product or tool into a pedagogical pathway, as well as the breadth of both intended and unintended outcomes that the technology might introduce. Subtle mechanisms to hook, addict, assume intimacy, or cause dependency must be identified and weeded out, particularly for children and the most vulnerable. Decision-makers must also look closely to understand how influences may manifest differently across genders in relation to identity formation, exposure to messaging and patterns of online engagement.

Requiring transparency in design and datafication. As several authors highlight, what makes 'the algorithm in the room' difficult to perceive is often not just the observer's potentially limited ability to recognize it, but also the fact that many algorithms in education still exist in black boxes. Technology companies and developers serving educational aims and vulnerable populations should ensure that their data sources, instructions, assumptions and modifications are as transparent as possible, and seek human feedback (with informed consent) from those in education about what is occurring in practice.

Horizon 2: Steer the algorithm

Aligning with pedagogical purpose. This is where addressing the 'alignment problem' discussed in the introduction of this volume becomes crucial. Education has many vital roles – transmitting knowledge in meaningful and context-responsive ways, fostering healthy socialization and social cohesion, cultivating creative problem solving and critical thinking, developing skills and an attitude of learning across life, among them. At their best, certain AI tools can have positive impacts on education for certain educational aims and in certain contexts. But this potential can only be reached if educators themselves are able to determine if and how they will use them, and if and when they will not. Students also require agency to see and steer meaningful educational trajectories rather than merely submitting to metrics and nudges automated by algorithmic systems that do not know them.

The slowing and limiting of algorithmic technologies as argued by some authors, and the 'analogue resurgence' of reducing reliance on screens in favour of books and paper being witnessed in some educational systems, are also signals of steering and confronting the algorithms (Savage, 2026). Of course, AI tools are not the only tools teachers have – far from it – and their use should always be a matter of intentional pedagogical choice.

Confronting AI harm. It cannot be understated that 'steering' the algorithm also requires directly confronting, and even disarming, algorithmic and AI harm. Harm can be caused by the designs of AI systems themselves, or by vulnerabilities they create for actors to exploit (Hoffmann, 2025; OECD, 2023). Importantly when peoples' rights to privacy, intellectual freedom and protection from harm are violated, algorithms need to be disarmed and their developers confronted. Privacy, safety and access were flagged in 2024 by the UN's Special Rapporteur on the Right to Education as areas that require careful regulation, oversight, design and evaluation (UNGA, 2024). Despite the gravity of risks to children and vulnerable groups, regulatory frameworks continue to remain weak. Being unfairly penalized in one's education by algorithmic decision-making systems also risks following students into their adult lives, as decisions made by algorithms are often difficult to appeal (Lyons, Velloso and Miller, 2021).

Reinforcing education systems rather than depleting them. Authors of even the most hopeful scenarios in this collection are clear about the need for effective regulation to protect the future agency of those working within education. They are also clear that imagining hopeful futures requires protecting public institutions and education systems from being hollowed out in the name of 'disruption' or 'acceleration' via AI. On the contrary, many of their core educational infrastructures will benefit from further reinforcement – from expanding school libraries, the arts and humanities to robust ongoing teacher development programmes – for education to best support students, educators and communities. Other infrastructures will benefit from rethinking in response to diverse educational contexts: should we opt for one-to-one devices or shared computer labs? Should education invest in subscriptions for commercial learning management systems or develop public platforms and open-source software?

Horizon 3: Rewrite the algorithm

Enabling algorithmic decision-making from within. Authors in this collection imagine some exciting future directions that learning models can take. Indeed, education's transformative power is vitally needed to shape empowering, human-centred futures for AI algorithms within and beyond their educational applications. This depends not only on what has been mentioned – seeing and understanding their presence and influences, and steering them toward more effective and appropriate uses. But it also depends on a vision in which students, educators, and education institutions of all kinds effectively rewrite the aims, assumptions, inputs and instructions of the algorithms they interact with. It is in the hands of the people and education communities themselves that algorithms can become

empowering, able the meet the diverse aspirations, questions, concerns and hopes they hold for education's future.

Picking up signals in experiences: grassroots participation, citizen science and small language models. Several think pieces provide hopeful examples – both current experiences and future possibilities – of education communities shaping empowered, human-centred uses of AI. For example, grassroots movements across parts of Africa are building NLP networks, trained by local students, teachers, researchers and coders, to fight epistemic injustice and foster local ownership of indigenous knowledge production. Authors in Singapore and Canada envision community-trained AI models, developed from citizen science initiatives, to monitor local biodiversity, while authors from the United Kingdom imagine rewriting algorithms to move away from individualistic pedagogy towards collaboration and collective intelligence. Many of these proposed applications involve developing small language models locally, harnessing algorithmic speed and power without unlawful non-consensual data scraping and other supply-chain harms common to training and deploying many current commercial LLMs (Amnesty International, 2026).

Inviting more voices, perspectives and experiences. Education is an empowering engine of capacity building and imagination in every society, and a source for valuable insights about how AI's implications for education may unfold differently across contexts. This volume compiles a wide range of diverse perspectives and imaginings. More voices are needed within education not only to give feedback, but to be the authors of the algorithmic designs and assumptions that affect them. While the private sector is welcome in this conversation, many of the think pieces highlight the contradictions that commercial interests and profit motives place in the way of alignment with educational aims and purposes. Educators and students are already rewriting the algorithms – or imagining new ones – but their perspectives must be centred rather than consumed.

* * *

The future of education is not knowable, but it is shaped greatly by what human beings do today. Given the rising concerns and challenges, might education be on the threshold of an analogue resurgence? Or could certain course corrections for the design and regulation of education-related algorithms be in store? Or is it instead possible that we are just a few years away from witnessing cohorts of well-educated young people exercise their full agency, wisdom and stewardship of powerful new algorithmic technologies for the best interests of all?

Education is far from passive at this critical juncture. It has the moral authority to demand better options for what it invites into schools, classrooms and pedagogical pathways, and to demand that design intentions are better aligned with experience. Moreover, it has the public mission to educate human beings capable of navigating a changing world, regardless of the forms that generative AI or other disruptive technologies and events may take.

Bibliography

Amnesty International. 2026. *Unlawful by Design: Exposing the Human Rights Costs of Generative AI*. London, Amnesty International. (Doc. POL 40/0996/2026.)

Amsler, S. and Facer, K. 2017. Contesting anticipatory regimes in education: exploring alternative educational orientations to the future. *Futures,* Vol. 94, pp. 6–14. https://doi.org/10.1016/j.futures.2017.01.001

Autenrieth, D. 2026. *How AI systems think about education: Analyzing latent preference patterns in large language models*. arXiv preprint, https://doi.org/10.48550/arXiv.2603.21006

Burns, M., Winthrop, R., Luther, N., Venetis, E. and Karim, R. 2026. A new direction for students in an AI world: Prosper, prepare, protect. Brookings.

Christian, B. 2020. *The Alignment Problem: Machine Learning and Human Values. New York, W. W. Norton & Company.*

Engler, A. 2021. Enrollment algorithms are contributing to the crises of higher education. Brookings.

Fontes, C., Hohma, E., Corrigan, C. C. and Lütge, C. 2022. AI-powered public surveillance systems: why we (might) need them and how we want them. *Technology in Society,* Vol. 71. https://doi.org/10.1016/j.techsoc.2022.102137

Gamboa, Y. H. N., Jaboli, N. N. A., Pica, N. A. L.., Rosales, L. J. D. and Lazaro, B. L. G. 2025. Long-term effects of algorithm-driven content consumption on youth development and psychological perceptions. *International Journal of Advanced Multidisciplinary Research and Studies,* Vol. 5, No. 3, pp. 204–38. https://doi.org/10.62225/2583049X.2025.5.3.4217

Haidt, J. 2024. *The anxious generation: how the great rewiring of childhood is causing an epidemic of mental illness.* Penguin Press, New York.

Heinsfeld, B.D., 2025. When education serves the machine: The technocapitalist capture of universities. Civics of Technology.

Hoffmann, M. 2025. The Mechanisms Of AI Harm: Lessons Learned From AI Incidents. Center for Security and Emerging Technology.

Human Rights Watch and Harvard Law School (eds). 2025. *A Hazard to Human Rights: Autonomous Weapons Systems and Digital Decision-Making.* New York, Human Rights Watch.

Lyons, H., Velloso, E. and Miller, T. 2021. Challenging decisions made by algorithm. *Pursuit.*

Noble, S.U. 2020. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York, New York University Press. https://doi.org/10.18574/nyu/9781479833641.001.0001

OECD. 2023. *AI risks and incidents.*

Oliveira, C.V., Zagalo, N., Silva, F., Bandao, A., Khurrum, S.F.H. and Santos, J. 2025. *Plausibility as failure: How LLMs and humans co-construct epistemic error.* arXiv preprint, https://doi.org/10.48550/arXiv.2512.16750

O'Neil, C. 2016. *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York, The Crown Publishing Group.

Orr, W. and Crawford, K. 2024. The social construction of datasets: On the practices, processes, and challenges of dataset creation for machine learning. *New Media & Society,* Vol. 26, No. 9, pp. 4955–72. https://doi.org/10.1177/14614448241251797

Rauch, J. 2021. *The Constitution of Knowledge: A Defense of Truth*. Washington DC, Brookings Institution Press. https://doi.org/10.5040/9780815750376

Roche, E. Hirsh-Pasek, K., Romeo, R., Suskind, D. and Perry, K. 2025. Policy guardrails needed as babies around the world begin to interact with AI. Brookings.

Savage, M. 2026. Back to books - Sweden's schools cutting back on digital learning. *BBC*.

Singer, N. 2026. As schools embrace A.I. tools, skeptics raise concerns. *The New York Times*.

Stricker, J. K. 2025. The synthetic knowledge crisis: How AI is reshaping knowledge, and not for the better [Blog].

Sukharevsky, A., Krivkovich, A., Gast, A., Storozhev, A., Maor, D., Mahadevan, D., Hämäläinen, L. and Durth, S. 2025. The agentic organization: Contours of the next paradigm for the AI era. McKinsey & Company.

UNESCO. Right to Education Initiative. 2019. Right to education handbook. Paris, UNESCO. https://doi.org/10.54675/ZMNJ2648

UNESCO. 2021. *Reimagining our futures together: a new social contract for education*. Paris, UNESCO. https://doi.org/10.54675/ASRB4722

––.2022. *Recommendation on the Ethics of Artificial Intelligence*. Paris, UNESCO.

––.2023. *Guidance for generative AI in education and research*. Paris, UNESCO.

––.2024a. *AI competency framework for teachers*. Paris, UNESCO.

––.2024b. *AI competency framework for students*. Paris, UNESCO.

––.2025. *AI and the future of education: disruptions, dilemmas and directions*. Paris, UNESCO.

United Nations General Assembly (UNGA). 2024. *Artificial intelligence in education – Report of the Special Rapporteur on the right to education, Farida Shaheed*. New York, United Nations. (Doc. A/79/520.)

Walsh, J.D. 2025. Everyone is cheating their way through college. *New York Magazine*.

Williamson, B. 2023. Governing through infrastructural control: Artificial intelligence and cloud computing in the data-intensive state. W. Housley, A. Edwards, R. Beneito-Montagut, R. Fitzgerald (eds), *The SAGE Handbook of Digital Society*. London, SAGE Publications Ltd, pp. 521–40. https://doi.org/10.4135/9781529783193.n29

Yudkowsky, E. 2016. The AI alignment problem: Why it is hard, and where to start. Symbolic Systems Distinguished Speaker, Vol. 4, No. 1, pp. 1–20.

Abbreviations and acronyms

AI	Artificial intelligence
AIED	Artificial intelligence in education
CNC	Computer Numerical Control
DAP	Default Approach Problem
HITL	Human-in-the-loop
LLM	Large language model
MIT	Massachusetts Institute of Technology
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
OHCHR	Office of the High Commissioner for Human Rights
OLPC	One Laptop Per Child
STEM	Science, Technology, Engineering, and Mathematics
TVET	Technical and Vocational Education and Training
UN	United Nations
UNGA	United Nations General Assembly
UNHCR	United Nations Human Rights Council
UNICEF	United Nations Children's Fund

unesco
United Nations
Educational, Scientific
and Cultural Organization

The algorithm in the room

Seeing and confronting the implications of AI for the future of education

Claims that AI is reshaping education form a familiar and often loud refrain in discussions about the future of education. Such blanket assertions tend, however, to obscure more nuanced questions: What, exactly, is being reshaped? Why? How? And is the reshaping improving the lives and learning of students, teachers, families and communities?

This compilation of 28 think pieces confronts these 'elephant in the room' questions by seeing and describing algorithmic changes at work in education around the world. The contributions are organized around six questions:

- Who will hold agency, autonomy and oversight?
- How will young people develop relationally, emotionally and cognitively?
- What will be meaningful and relevant for pedagogy and assessment?
- How will we think, question and express?
- How will we govern education as a common good?
- How might AI affect imagination about the future?

This volume invites readers to imagine and work towards a future where AI and other algorithmic technologies help better advance the humanistic aspirations and multiple purposes that societies hold for education.

9 789231 008979

4 QUALITY EDUCATION

Sustainable Development Goals