The Narratives of Artificial Intelligence: A Critical View of an Emerging Tool

Angele Pulis
Institute for Education
https://orcid.org/0000-0003-2489-9594

Mario Mallia
Institute for Education
https://orcid.org/0009-0007-6643-1945

Abstract

Artificial Intelligence has taken the world by storm, and humanity seems to be venturing into uncharted waters. The potential of Artificial Intelligence is still being explored in different sectors. In this desk research, the authors critically analyse and problematise the use of Artificial Intelligence in the educational realm. Ethical dilemmas when employing and relying on Artificial Intelligence are explored from contrasting perspectives. The authors attempt to evoke questions on the reliability, validity, and any possible hidden or silenced narratives in information being provided by large language models such as ChatGPT. The role of the educator as a trailblazer in the ethical and discerning use of Artificial Intelligence is emphasised. In parallel, the paper makes the case for revisiting core issues in education including the need for reappropriation of teachers' work and slowing down the pace of education to allow for a critical undertaking.

Keywords

Critical analysis, ethical considerations, narratives, artificial intelligence in teaching and learning, problematising education, teachers' work

Introduction

The term 'Artificial Intelligence' was conceived nearly 70 years ago, during a 1956 conference in Dartmouth, USA, where researchers were discussing the possible development of computers that could think (Ho, 2021). Pierce and Hathaway (2018) offer the following definition: "Artificial intelligence is a broad term used to describe any technology that emulates human intelligence, such as by understanding complex information, drawing its own conclusions and engaging

Contact: Angele Pulis, angela.pulis@ilearn.edu.mt

in natural dialog with people" (para. 3). This broad definition of artificial intelligence (AI) is the working definition that will be used in this article, with a particular emphasis on large language models.

Following Eco's 1964 seminal work (cited in Baldini &Farahi, 2025) that explores the way cultural critics view mass media and popular culture, this article makes the case for a balanced approach to AI between the apocalyptic (that pessimistically views mass media as degrading and a sign of decline and loss of traditional values) on the one hand and the integrated (embracing technology and putting considerable faith in it) on the other, calling for a pedagogical third way that "cultivate[s] a critical, curious and creative gaze" (Baldini & Farahi, 2025 p.8).

The launching platform of the discussion is a description of large language models, together with an exploration of their main limitation, hallucinations. The potential repercussions of hallucinations for classroom practitioners are examined, together with the ethical implications of using AI in schools. Practical measures to mitigate the inherent weaknesses of LLMs are proposed for educators. Beyond the utilitarian aspect of AI in schools, the article explores critical perspectives and, fundamentally, seeks to provoke educators to reimagine and rearticulate their own understanding of education, as they seek to engage with AI in education.

What is a Large Language Model (LLM)?

A large language model is a learning system that simulates language. It is called language modelling because it is trained to predict the next word in a text; the training entails becoming acquainted with extensively large bodies of texts (Blank, 2023). Furthermore, Blank delves into the meaning of language modelling within LLMs and claims that answers to what exactly LLMs are modelling could range from the human brain, rendering an LLM a symbolic model producing logical conclusions; to the human mind, rendering an LLM a sub-symbolic model producing associative conclusions. Blank acknowledges the limitations of LLMs and proposes that for LLMs to model human language processing, they should be trained with inputs and neurocognitive limitations as experienced by humans. Blank argues that LLMs are feed-forward systems and hence, many words are processed in parallel; however, the human brain functions according to recurrent processes, to input that is arranged in a particular order and that is limited by the magnitude of the working memory. Blank further explains that LLMs are trained on tasks communicated through language; however, they are unable to recognise that not all tasks are linguistic tasks, whereas the human brain can recognise when to refrain from linguistic processes, for example, when common sense reasoning or social cognition are required.

LLMs can be used in educational settings to assist educators with lesson planning and curricular content development, and they can provide innovative and individualised learning experiences to students (Naveed et al., 2023). Their full use is still being tapped by educators. The rapid and continuous changes in LLMs renders them more adaptable and applicable to the needs of educators (Wang et al., 2024). A challenge that accompanies this is the necessity for continuous updating of knowledge on LLMs by educators. However, the real disadvantage is that "the 'black box' nature of LLMs" (Naveed et al., 2023, p. 34) makes it problematic for educators to feel fully confident about the output that is generated.

Hallucinations of LLMs

Large Language Models produce what is known as knowledge hallucinations, defined as "factually incorrect or nonsense generations" (Chen et al., 2024, p. 1). Liu et al. (2024, p. 1) explain that "LLMs might produce outputs that deviate from users' intent, exhibit internal inconsistencies, or misalign with the factual knowledge". "Factual" here cannot be taken to mean a rejection of the need to "bracket" reality to highlight consciousness, allowing for varied interpretations (what is known as the phenomenological approach). Hallucinations are not about different interpretations of the empirical world but the fundamental altering of the subject of such interpretations. Since hallucinations with LLMs are unavoidable (Xu et al., 2024), users can never fully trust the output that is generated.

Why do Hallucinations Occur?

Hallucinations occur for a variety of reasons. Farquhar et al. (2024) explain that knowledge hallucinations might occur because the large language model was trained on data that was not correct; there might be a methodical failure in the reasoning of the model, or the model is "untruthful" because it is primarily driven to achieve what it perceives as a reward.

Types of Hallucinations

Zhang et al. (2023) describe three different types of hallucinations for natural language generation tasks:

1. Input-conflicting hallucinations where the content generated digresses from the source input;
2. Context-conflicting hallucinations where the content generated contradicts content that was produced earlier;
3. Fact-conflicting hallucinations where the content generated is in dispute with confirmed knowledge.

Apart from an awareness of hallucinations generated by LLMs, the use of AI in the classroom demands a careful consideration of the ethical concerns and implications.

Ethical Concerns

Research shows that ethical cautiousness when using AI in the classroom is lagging behind the use of AI in other contexts; Schiff (2022) analysed the national strategies for AI in more than 30 countries and found that the ethical use of AI in education was hardly ever mentioned. Schiff reported that the emphasis was on education for Artificial Intelligence purposes, on creating an AI-competent workforce and on designing more specialised training on AI; whereas the ethical implications of employing AI in the education sector were ignored. This led Schiff to conclude that "education for AI, not AI for education" (p. 527) is the more pressing need. Borenstein and Howard (2021) argue in the same vein and insist that:

> Tackling the problem head-on requires educating ourselves at the beginning stages of our interaction with AI … The opportunity to learn about how data are used to train AI, about the applications that the AI can enable, etc., should be available to any person that interacts at any stage with AI … Moreover, ethics should not be a slapped-on component after the fact, a standalone lesson, or a second thought. It is integral at every stage when learning about AI. (p. 62)

Educators should be aware that software that uses AI could gather information on students and how they respond to the content being presented to them without consent (Pierce & Hathaway, 2018). Concerning the issue of consent, educators need to design school policies that address privacy concerns that arise with the use of AI at school. For classroom practitioners, the ethical implications are more urgent since students are still minors and parental consent will need to be sought. The need for such ethical considerations is reflected in the National AI Strategy for Malta (Office of the Prime Minister, 2019a), which speaks of fairness in the use, development and running of AI systems and also harm prevention. As for the latter consideration, the Maltese AI strategy is clear in pointing out that AI systems should clearly indicate that "its social interaction is simulated and that it has no capacities of 'understanding' or 'feeling'" (Office of the Prime Minister, 2019a, p. 23).

The same document also speaks of "explicability" through which all users and the general public are to be able to comprehend and criticise the use and the performance of AI systems. The issue of bias is one such area.

The Biases of AI

Artificial Intelligence is informed by data that presumably reflects the biases of the humans that produced and processed that data. In Malta's vision on Artificial Intelligence, when defining the term, it is noted that "AI is derived from the applied extraction of knowledge or insights from data allowing machines to make informed decisions" (Office of the Prime Minister, 2019b, p. 5). Borenstein and Howard (2021, p. 61) argue that Artificial Intelligence is "intensifying societal ills" since it is designed by humans that are flawed and unfair. According

to Naveed et al. (2023, p. 33), "LLMs can inherit and amplify societal biases in their training", perpetuating existing hegemonies that preserve dominance of some groups in society over others. Furthermore, Borenstein and Howard (2021) refer to the complexity of trying to mitigate the biases of AI since the data, the algorithms and the outputs are biased. As an example, Borenstein and Howard refer to the use of an AI system for recommending follow-on health services in the USA that, in practice, exhibited significant racial bias against black people. Fewer black people than white were recommended for follow-up healthcare even when the diagnoses were identical. The source of the problem was that the algorithm being used predicted healthcare costs whilst ignoring illnesses, and black patients have been found to suffer from more chronic diseases than white patients (Obermeyer et al., 2019). The result of the racial bias resulted in the number of black people being referred for follow-up healthcare being less than half of the actual amount that should have been entitled to it (Obermeyer et al., 2019). Other examples include biases against women in a secret recruiting tool used by Amazon (Dastin, 2018) and facial recognition initiatives which misclassified darker skinned females by a wide margin over other groups (Buolamwini & Gebru, 2018).

Apart from an awareness of the possibility of LLMs generating hallucinations, the intricate ethical implications that might arise when using AI in the classroom, and the deep-rooted biases that might shape the outputs generated by LLMs, there are other practical challenges that educators might encounter.

Other Challenges When Using AI in the Classroom

Prompts. The responses of LLMs depend on the prompts that are inputted. Connotations and symbolisms of language that are understood by humans but not by LLMs can play havoc with the output generated (Naveed et al., 2023). This is not to say that all humans will have the same understanding of connotations and symbolisms of language, but rather that the understanding of language by the creator of a prompt, whatever the cultural affiliation, may be misconstrued by LLMs.

Limited Knowledge. LLMs are dependent on the training that was provided. The knowledge provided during training is limited and can become outdated (Naveed et al., 2023). In addition, as LLMs are trained on more specific information to be able to perform new tasks, they can suffer from "catastrophic forgetting", disregarding and renouncing the knowledge they acquired during their initial training (Naveed et al., 2023, p. 33).

Cut-off Date. The data on which LLMs are trained has a cut-off date; for instance, for GPT-4 it was September 2022, and this compromises the accuracy of the responses that are generated (Addington, 2024). This means that data that became available after that date cannot be used to inform the responses of the LLM.

Rehashed Responses. The outputs generated by LLMs tend to become repetitive. Foroux (2024) explains the reason for this:

> AI tends to be monotonous because it doesn't offer anything new (it's only good at predicting words and re-purposing existing content). In fact, one concern that the study brought up is how over-reliance on AI is similar to being inside an echo chamber. (para. 4)

The limitations in the writing style of LLMs is another matter of concern for educators, as shall be explored next.

The Educator's Pen is Mightier than ChatGPT

Generative AI tools, such as ChatGPT are designed to compose text. When using these tools, educators should be aware that the writing style of generative AI tools is not human. When educators use ChatGPT to write text that will be exhibited to students, they are exposing their students to stylistic homogeneity. Whilst such tools may be profitably used to correct stylistic deficiencies on the part of the human writer, overreliance on text generated by ChatGPT robs the human writer of the opportunity to express themselves in their individual, unique way: "it becomes harder to distinguish between individual voices and perspectives and everything takes on a robotic undertone" (Chugh, 2024, para. 8).

The development of writing skills is an important skill that schools hope to instil in their students; however, apart from the obvious skill that manifests itself in the text that is generated by students, the process of writing is intertwined with thinking (Foroux, 2024). As noted by Foroux (para. 5), "we need the process of writing to progress because writing is a form of thinking". Hence educators should be mindful of the danger of shunning the thinking process when relying on LLMs to compose text for themselves and for their students.

Furthermore, Chugh (2024) explains:

> Writing should be about expressing your ideas in your own way. While ChatGPT can help, it's up to each of us to make sure we're saying what we really want to – and not what an AI tool tells us to. (para. 19)

Moreover, LLMs lack creativity, and this is because "their inner autoregressive nature seems to prevent them from reaching transformational creativity" (Franceschelli & Musolesi, 2025, p. 3791). Franceschelli and Musolesi analysed the evolution of LLMs through the criteria of creativity theories (notably Margaret Boden's criteria of value, novelty and surprise) and concluded that LLMs can only generate a poor version of innovativeness in their outputs; in addition, the social aspect of creativity is absent.

Practical Measures That Could Be Helpful to Educators

1. The more educators learn about AI, the more they can make informed choices on the use of AI for themselves and for their students.
2. Since the knowledge of LLMs has a cutoff date, educators should not rely on an LLM as the sole source of information. A Google web search accesses the internet in real time.
3. LLMs hallucinate so information should be triangulated before being disseminated among students.
4. LLMs tend to use repetitive and robotic text, so the generated text should be revised before distributed amongst students.
5. LLMs are replicating the dominant societal narrative that drive a hegemonic understanding of society with an impact greater than is the case with other technologies, as argued earlier. Students should therefore be challenged with greater insistence to call into question this dominant narrative and to become aware of the silenced and weak voices in society.
6. AI is machine-generated and so it lacks emotions. Even if the capability of emotional awareness by AI is increasing (see Elyoseph et al., 2023), educators should still check whether this has affected the output generated.
7. When using software that employs AI, educators should be aware of any ethical implications, e.g., Is data going to be gathered from students? Is parental consent needed? Specific policies at school level are therefore called for here.
8. Educators should teach students how to use AI to help them learn, not to avoid learning.
9. Educators could assign tasks to students that discourage the use of answers where they can copy and paste text generated by an LLM.
10. Educators could assign tasks to students that encourage the development of writing skills in a context where students cannot revert to AI-generated text.

Critical Perspectives When Using AI in the Classroom

As we have argued earlier, the practical use of AI in education is not a bed of roses. Nor is it the devil incarnate. It is a power unleashed which needs to be kept in check. A critical perspective of the use of AI in education unpacks the potential ramifications of its use. Robust regulatory frameworks that address the imbalances, the secrecy and the biases are essential. They need to be inclusive so that they do not end up weaponised by the powerful to further their own interests. Criticality and decolonial theory can be useful tools to strive for a social contract that highlights these inequalities and helps to address them (Mohamed et al., 2020). Technology, therefore, cannot be left alone in the driving seat. As Plunkett (2023) rightly points out, AI should be the concern of philosophers and social scientists, as much as it is of technologists. Educators that empower learners to profoundly read the world with a view of helping them change it (Freire, 2000) should be an integral part of the equation.

A Responsive Education

The response of educators to the phenomenon of AI needs to include two important thrusts. The first is a critical articulation of AI itself. This would then allow, as Nayir et al. (2024, p. 103) put it, for a "nuanced exploration to harness the potential of these technologies" by taking into consideration important caveats uncovered by a critical pedagogy that focuses on issues of justice and ethical use of the technology. This first thrust, in turn, necessitates a second crucial and parallel focus: revisiting core issues in education because it is these issues that AI is challenging the most. Questions concerning what we want education to be, with or without AI, and whether what we do lies in contradiction with it, become central. Ultimately, this will be addressing fundamental questions about the society we want to live in and the values which define it.

The First: A Critical Articulation of the Technology

Acknowledging potential is a good starting point. Demonising AI means losing out on what it can offer and more importantly, it risks losing out on an opportunity to educate and remain relevant. Educators cannot ignore what the technology can do in education, particularly with regards to personalised learning experiences, formative assessment, adaptive learning processes, assistance for teachers and enhanced data analytics, educational mainstreaming, and thought-provoking feedback, among others (Nayir et al., 2024; OECD, 2021).

Nonetheless, this acknowledgement needs to be accompanied by the recognition that knowledge is power. Private interests outpace each other to take control over different facets of knowledge, including data, ultimately making the powerful more powerful. Whilst Open AI, for instance, had started off as a non-profit organisation, it is increasingly becoming for-profit (Harris, 2023) as a result of competition which threatens to overtake it (Vincent, 2023). In doing so, it has sought to hide dataset construction and training methods it uses to train its LLMs (Byrd, 2023), never mind making explicit the representations of knowledge assumed within. This lack of transparency makes data governance difficult. Byrd makes a strong argument that odds are already stacked against minorities and the Global South through linguistic punishment: a white workforce designed GPT-3 and trained it on just 7 percent of non-English languages. Alternative English used in social media not considered as making the bar is left out. Digital spaces made for marginalised people may not be included in LLMs' training data because they lack enough incoming and outgoing links to show up in corpus texts. Byrd traces the vicious cycle created and concludes that inequalities are reinforced as "our political and social ideologies communicated through language turn into actions, and vice versa, [creating] histories that produce new texts for the next iteration of the LLMs to train on" (2023, p. 137). This is not to say that bias came into existence because of Artificial Intelligence. Herman and Chomsky (1988, as cited in Chomsky, 2002) speak of the propaganda model by mass media, based on the concept of "manufacturing consent"

coined by Lippmann in 1922, that serves the interests of the powerful elites. Whilst this is acknowledged, bias assumes prominence of greater proportions given that AI is not the traditional passive medium of yesteryear, but a generative medium, one of expression and interpretation, which makes processes infinitely easier and more straightforward to use. Blind reliance on AI-generated texts is therefore more probable. This risks reinforcing pedagogies in our classrooms which value dominant literacies and identities even further, whilst marginalising others. A critical pedagogy using a critical analysis process would be key to uncover these underlying processes (see Gillespie & McBain, 2011, as an example of the use of the process albeit in other contexts).

Added to this critical awareness is the need to keep in mind the propensity for synthetic, low-quality, homogenous LLM text outputs which contribute to an overall impoverishment of writing skills. The generation of original material which students should feel proud to put their name to, even if it is helped by LLM or other more traditional instruments, needs to be nurtured so that the ability to write and think creatively remains an important educational hallmark. Judicious use of AI would also ensure that students retain critical skills as an integral part of their education. The work by Bastani et al. (2024) on Open AI's GPT-4 shows that educational outcomes can be harmed if the technology is used to provide answers (GPT Base) rather than hints (GPT Tutor). They make the pertinent point that when technology automates a task, humans become gradually deskilled and can miss out on the valuable experience of performing it. Technology replacing the human mind does so at humanity's peril.

The considerations outlined above are just some of the issues which highlight the need for a critical understanding if we are to navigate the use of technology with eyes wide open. Doing so, however, should not imply remaining at arm's length as if running a commentary on a game we are not part of. On the contrary, AI and all, it implies engaging with our own selves as educators and with education. Education becomes the object of our scrutiny.

The Second: Problematising Education

Here we propose focusing on two specific areas in education: teacher work and the time to care, which are interrelated.

Teachers' Work

Techno-solutionism and its propensity to throw technological solutions to challenges tends to silence a critique on the politics of technology and in particular, teachers' work (Rensfeldt & Rahm, 2023). This is fuelled by assertions like the one made by the OECD (2021) that technology is ethically neutral, and it is what educators do with it, which is not; that "the real risks do not come from AI but from the consequences of its application" (p. 4). Yet, what generates the technology itself and the solutions it proposes are not innocent and neutral. They are value-laden and need

problematising. In Malta, for instance, templates provided centrally to fill in assessment forms are more than efforts to make teachers' lives easy. They also provide an efficient way of controlling what gets assessed and how, especially given that assessment policies are centrally generated. Whilst on the one hand, colleges and schools are required to develop an assessment policy of their own "to address the quantity and quality of assessment practices as well as reporting to parents and other stakeholders" (Ministry of Education and Employment, 2015, p. 34), the technology, moulded in specific ways, and centrally provided in a ready-made package, is used to group students into centrally mandated learning tracks. Platforms are provided as ready-to-use solutions which pose as seemingly 'objective' shortcuts to bypass human bias and human errors, which, however, considerably limit the options available to educators, deskilling teachers and gradually transforming them into standardized productive technicians. This provision of ready-made packages feeds into the narrative of very busy and over-burdened educators, making technology seem like the means to free up their time. Rensfeldt and Rahm (2023) put to rest this assertion when they refer to research indicating that rather than leading to less work, automation tends to increase workload, squeezing further the possibilities of critique and resistance. The little remaining critique and resistance is then put down to technophobia and a fear of the future. Educators' work, we argue, needs to be reappropriated.

Time to Care

A critical pedagogy implies a conscious act of encounter and making the time for it. Critical pedagogy is dialectical (Giroux, 1988). The mantra of personalisation which seems to crop up whenever AI in education is contemplated is to a degree double-edged. On the one hand, it can effectively be used as a vehicle for equity. On the other, the "personalised" approach and self-directed learning cannot be taken to mean a sense of isolation – an education done alone, sufficient unto itself. It needs to be articulated within the context of community; the school as a social and cultural space promoting dialectical encounter; a collective pedagogical experience through which students become both teachers and learners as they take the time and the responsibility to care for each other (Mayo & Vittoria, 2021, commenting on the pedagogy of Don Lorenzo Milani). This in turn asks of us to re-evaluate practices in education that promote competitive individualism in and among schools, that militate against possibilities of cooperation and encounter.

As AI lunges forward, it pushes the mantra of speed and immediacy which seems to characterise the human condition in the technological era, leaving humanity breathless, faced with the impossible task of trying to catch up. Yet, if we are to make sense out of this fascinating yet unsettling prospect though a critical response that asks questions, education needs to go the other way, against the grain, in a fundamental way. Zavalloni (2017) makes the case for a "pedagogy of the snail" (la pedagogia della lumaca), making an emphatic point that in education, slower is deeper and sweeter. "The true reason for a journey is therefore not the 'arrival' (l'arrivo) but the 'journey' (il camminare), it is not the 'destination' (la meta), but the 'path' (percorso), it is not the 'where' (il dove) but the 'how' (il come)" (Zavalloni, 2017, p. 54). This echoes the pleas by

children (Children's Rights Observatory Malta, 2022) to "reduce the burdensome curriculum, syllabus content and homework workload ... to review what is being taught to students, how this material is being taught to them and what skills they need to succeed in the future" (pp. 25–41). Education needs to find the time and the space for critical reflection and engagement, away from the frenzy. The National Education Strategy 2023-2030 (Ministry of Education, Sport, Youth Research and Innovation, 2024), under the third strategic objective of Growth and Empowerment seems to somewhat acknowledge this need. It speaks of "the need to seriously take on the challenge of phasing out content overlap which is the result of a subject based system [to provide further space for] discussion, collaborative work and self-reflection" (p. 53). If this means addressing a subject-based system which compartmentalises knowledge or if it is simply doing away with repetition, keeping the subject-based elephant in the room intact, remains to be seen. The implication of such spaces, if created, is an education where educators and students take more ownership of the educative process, and one that critically engages with the world, AI included.

The point we make here is that, ultimately, the frenetic pace of AI and its development should nudge us to go the other way: slow down.

Concluding Thoughts

In this paper the authors have touched upon aspects of a revolution in the making which promises nirvana as it pushes frontiers to scenarios which until recently belonged to science fiction. AI stands as a testament to the creativity and potential of humankind. It leaves us in awe. The temptation is to become mesmerized and acquiescent, oblivious to the consequences of a tool which is imperfect and partial. The first part of this paper dwelt upon imperfections in LLMs that generate hallucinations, and texts which might be well articulated but missing nuance and creativity, apart from generating knowledge which is the result of an echo chamber dominated by powerful narratives. As a result, an endorsement of AI should intrinsically imply submitting it to critical scrutiny that puts the brakes on a technology running off, and risk leading humanity by the nose. Issues such as democracy, equity, community, and care, together with issues of transparency and accountability which go with them, need to remain the overarching yardsticks by which every tool, as awesome as it might be, is to be measured. Education needs to remain at the forefront as it nurtures critical citizens who can understand the sensation that is AI not just in terms of what it can do, but also in terms of social justice, which lies at the heart of what we want our world to embrace. As a response to the strong headwinds of AI, in education and beyond, the authors propose a focus on reappropriating teachers' work on the one hand, and on the other, a slowing down of education as we strive to care and make sense out of a fast-evolving world which we would like to have serving the interests of all.

The march of AI is a relentless one for several reasons. As a minimum, it holds promise to take care of the mundane tasks we find tedious. It also holds promise to push frontiers. The breakneck speed with which AI is developing provides for an adrenaline rush and a sense of

intrigue and excitement as it opens new possibilities hitherto unmatched. It fires up the collective imagination that feeds a transhumanist endorsement that views technology as a means of enhancing humanity (Nayir et al., 2024) and feeds upon a pervasive climate which upholds immediate gratification, a technocratic rationality, and the commodification of knowledge. That AI is not just here to stay but will continue to grow exponentially seems inevitable. Ho (2021, p. 1) defines Artificial Intelligence as "a system where machines are designed to mimic humans". It falls under the educator's remit to ensure that humans do not start to mimic machines. By continuously asking the question: What type of society do we want to create? before asking: What type of society do we want to create with AI? educators can help channel the AI change in the right direction. Ultimately, it all depends on what society we dream of, for ourselves, and our children.

Notes on Contributors

Angele Pulis is a lecturer at the Institute for Education. Her research domains include educational leadership, pupil voice and mixed methods research. She holds a Ph.D. from the University of Leicester, a Master of Philosophy from the University of Wales, and a postgraduate diploma in Educational Administration and Management, and a Bachelor in Education (Hons) from the University of Malta. Her career in schools has included various roles. She was a Head of a primary school and an Assistant Head of a sixth form and a secondary school. She has taught Integrated Science, Biology and Chemistry in various schools.

Mario Mallia is lecturer at the Institute for Education, focusing on critical pedagogy, gender, and science education. He was Head of a primary and secondary school for sixteen years, a Deputy Head, and a teacher of science. He holds a Master's degree in Education, a postgraduate diploma in School Administration and Management, and a Bachelor in Education (Hons) degree from the University of Malta. He served, inter alia, as a board member of the National Commission for the Promotion of Equality and the Foundation of Educational Services for many years, besides, to date, being active in the political and social fields.

References

Addington, A. (2024, September, 23). Knowledge cutoff dates for ChatGPT, Meta Ai, Copilot, Gemini, Claude. ComputerCity. https://computercity.com/artificial-intelligence/knowledge-cutoff-dates-llms

Baldini, M., & Farahi, F. (2025). Rereading the history of pedagogy between apocalyptic and integrated. A critical pedagogy in the age of ubiquity. Journal of Inclusive Methodology and Technology in Learning and Teaching, 5(1), Article 1. https://www.inclusiveteaching.it/index.php/inclusiveteaching/article/view/285

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). Generative AI Can Harm Learning (SSRN Scholarly Paper 4895486). Social Science Research Network. https://doi.org/10.2139/ssrn.4895486

Blank, I. A. (2023). What are large language models supposed to model? Trends in Cognitive Sciences, 27(11), 987–989. https://doi.org/10.1016/j.tics.2023.08.006

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. AI and Ethics, 1, 61–65. https://doi.org/10.1007/s43681-020-00002-7

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, 1–15. https://www.media.mit.edu/publications/gender-shades-intersectional-accuracy-disparities-in-commercial-gender-classification/

Byrd, A. (2023). Truth-Telling: Critical inquiries on LLMs and the corpus texts that train them. Composition Studies, 51(1), 135–142, 217. https://www.proquest.com/docview/2841560441/abstract/126FF487D76B4041PQ/1

Chen, C., Liu, K., Chen, Z., Gu, Y., Wu, Y., Tao, M., Fu, Z., & Ye, J. (2024, May 7–11). Inside: LLMs' internal states retain the power of hallucination detection [Conference presentation]. ICLR: The Twelfth International Conference on Learning Representations, Vienna, Austria. https://doi.org/10.48550/arXiv.2402.03744

Children's Rights Observatory Malta. (2022). Children's Manifesto. Salesian Press.

Chomsky, N. (2002). Understanding power: the indispensable Chomsky. The New Press.

Chugh, R. (2024, September, 26). ChatGPT is changing the way we write. Here's how – And why it's a problem. The Conversation. https://theconversation.com/chatgpt-is-changing-the-way-we-write-heres-how-and-why-its-a-problem-239601

Dastin, J. (2018, October 11). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

Elyoseph, Z., Hadar-Shoval, D., Asraf, K., & Lvovsky, M. (2023). ChatGPT outperforms humans in emotional awareness evaluations. Frontiers in Psychology, 14, Article 1199058. https://doi.org/10.3389/fpsyg.2023.1199058

Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. Nature, 650, 625–630. https://doi.org/10.1038/s41586-024-07421-0

Franceschelli, G., & Musolesi, M. (2025). On the creativity of large language models. AI & Soc 40, 3785–3795. https://doi.org/10.1007/s00146-024-02127-3

Freire, P. (2000). Pedagogy of the Oppressed: 30th anniversary edition. Bloomsbury Academic.

Foroux, D. (2024, September, 23). AI Writing and the Illusion of Progress. Darius Foroux. https://dariusforoux.com/ai-writing-illusion/

Gillespie, L., & McBain, S. (2011). A critical analysis process – Bridging the theory to practice gap in senior secondary school physical education. Teachers and Curriculum, 12(1), 65–72. https://doi.org/10.15663/tandc.v12i1.32

Giroux, H. (1988). Teachers as Intellectuals: Towards a critical pedagogy of learning. Bergin & Garvey Publishers.

Harris, M. (2023, May 18). Elon Musk used to say he put $100M in OpenAI, but now it's $50M: Here are the Receipts. Archive.Ph. https://archive.ph/YGGXQ

Ho, F. T. (2021). AI in education: A systematic literature review. Journal of Cases on Information Technology, 23(1), 1–20. https://doi.org/10.4018/JCIT.2021010101

Liu, F., Liu, Y., Shi, L., Huang, H., Wang, R., Yang, Z., Zhang, L., Li, Z., & Ma, Y. (2024). Exploring and evaluating hallucinations in LLM-powered code generation. ArXiv, Computer Science, 1–18. https://doi.org/10.48550/arXiv.2404.00971

Mayo, P., & Vittoria, P. (2021). Critical Education in International Perspective. Bloomsbury Academic. https://doi.org/10.5040/9781350147782

Ministry of Education and Employment. (2015). Educators' Guide for pedagogy and assessment: Using a learning outcomes framework.https://www.um.edu.mt/library/oar/bitstream/123456789/119734/1/Educators_guide_for_pedagogy_and_assessment.pdfMinistry of Education, Sport, Youth Research and Innovation. (2024, November 6). The National Education Strategy. Edukazzjoni. https://education.gov.mt/useful-links/the-national-education-strategy-2/

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. Philosophy & Technology, 33(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Barnes, N., & Milan, A. (2023). A comprehensive overview of large language models. ArXiv, Computer Science, Linguistics, 1-46. https://doi.org/10.48550/arXiv.2307.06435

Nayir, F., Sari, T., & Bozkurt, A. (2024). Reimagining education: Bridging artificial intelligence, transhumanism, and critical pedagogy. Journal of Educational Technology and Online Learning, 7(1), 102-115. https://doi.org/10.31681/jetol.1308022

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. https://doi.org/10.1126/science.aax2342

OECD. (2021). OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots. OECD. https://doi.org/10.1787/589b283f-en

Office of the Prime Minister (2019a). Malta: Towards trustworthy AI. https://www.mdia.gov.mt/wp-content/uploads/2023/04/Malta_Towards_Ethical_and_Trustworthy_AI.pdf

Office of the Prime Minister (2019b). https://malta.ai/wp-content/uploads/2019/04/Draft_Policy_document_-_online_version.pdf

Pierce, D., & Hathaway, A. (2018, August 29). The promise (and pitfalls) of AI for education. The Journal. https://thejournal.com/articles/2018/08/29/the-promise-of-ai-for-education.aspx

Plunkett, J. (2023, April 29). Freedom in the age of autonomous machines. Medium. https://medium.com/@jamestplunkett/freedom-in-the-age-of-autonomous-machines-def5d18e82d8

Rensfeldt, A. B., & Rahm, L. (2023). Automating teacher work? A history of the politics of automation and Artificial Intelligence in education. Postdigital Science and Education, 5(1), 25-43. https://doi.org/10.1007/s42438-022-00344-x

Schiff, D. (2022). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. International Journal of Artificial Intelligence in Education, 32, 527-563. https://doi.org/10.1007/s40593-021-00270-2

Vincent, J. (2023, March 15). OpenAI co-founder on company's past approach to openly sharing research: "We were wrong". The Verge. https://www.theverge.com/2023/3/15/23640180/openai-gpt-4-launch-closed-research-ilya-sutskever-interview

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du. Z. (2024). Artificial intelligence in education: A systematic literature review. Expert Systems with Applications, 252, 1-19. https://doi.org/10.1016/j.eswa.2024.124167

Xu, Z., Jain, S., & Kankanhalli, M. S. (2024). Hallucination is inevitable: An innate limitation of large language models. ArXiv, Computer Science, 1-26. https://doi.org/10.48550/arXiv.2401.11817

Zavalloni, G. (2017). La Pedagogia Della Lumaca: Per una scuola lenta e non violenta (10th ed.). EMI.

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. H., Bi, W., Shi, F., & Shi, S. (2023). Siren's song in the AI ocean: A survey on hallucination in large language models [Unpublished manuscript]. Tencent AI lab, Soochow University, Zhejiang University, Renmin University of China, Nanyang Technological University, Toyota Technological Institute at Chicago. https://doi.org/10.48550/arXiv.2309.01219