Read the text below and write an essay (7200–9000 characters with spaces) in response to the following questions: How do the arguments presented in the article align with or challenge current debates in Estonia or the European Union regarding education policy, particularly in relation to digital literacy and AI in education? Based on the discussion in the article, imagine a future where digital literacy is a core component of English Studies. What should digital literacy look like in this field, and how should it be taught? Consider both the risks and opportunities AI presents for linguistic and literary scholarship. Provide specific examples.

Your essay should engage with the ideas and claims in the article, have a clear thesis and structured logical argumentation, be structured following the rules of academic English and be written in correct and idiomatic English. The text should be your original work. All additional sources and tools used should be appropriately referenced.

Deconstructing the AI Myth: Fallacies and Harms of Algorithmification

Dagmar Monett[1] and Bogdan Grigorescu[2]
[1]Computer Science Dept., Berlin School of Economics and Law, Germany
[2]Technical Architecture, eBay, London, UK

1. Introduction

With the latest advancements in the field of Artificial Intelligence (AI), especially those concerning the growing use of "machine learning"-based techniques in many domains since 2010, it has become of utmost importance to understand and critically assess the role and impact that "intelligent"[1] algorithms have on our lives. Such a critical appraisal is incomplete without knowing their particularities (e.g. how they work), which are no longer unique to academic labs, as well as their limitations (e.g. what they cannot and likely will not ever be able to do). Knowing and critically evaluating the current and potential consequences of using them, their fundamental characteristics, and how they might affect minorities, entire cultures and the environment, are more essential than ever for a well-functioning society. Yet, misconceptions about algorithms' capabilities, the narratives used to refer to them, a new high in hype inflating narrow successes, and the unregulated and/or bad uses of AI technologies are not only affecting disempowered populations but also damaging the AI field: they distort the public's perception of AI and destroy trust. They might have far-reaching consequences that, if not addressed, generate pushback and favour lawsuits becoming the new normal (GW Law, 2024).

Where critical AI literacy[2] is inadequate or non-existent, the risks that people use AI to damage others, make wrong decisions, or complain about people addressing the issues are high. Choosing the right narratives about AI is fundamental for dealing with and understanding AI's implications whatever their type. This is also why it is insufficient to encourage its use in educational settings just because it is new or appealing: pressing issues and the actual harms, technological limitations, potential dangers, and inflated narratives about AI must also be part of educating people.

[1] They are not intelligent, for several reasons discussed in (Bishop, 2021; Larson, 2021; McDermott, 1976).
[2] I.e. questioning what AI is used for, knowing about how their techniques work, not falling for the fallacy of AI functionality (Raji et al., 2022), and taking a critical stance about its negative impact and limitations.

Digital literacy, especially AI literacy, is about the how and when (how to use the latest technology in a given context), about the what and how (what it is and how it functions). It is also about why not use or even develop it, and about the power structures behind AI, its limitations, and negative implications. Most of the latter are avoided in the AI discourse, are superficially discussed, or are simply not present in certain academic settings, e.g. where AI tools are promoted as solvers of major systemic problems, disregarding how the harms outweigh the benefits and overlooking how the trends of economic inequalities are exacerbated (Acemoglu and Johnson, 2023).

This paper aims to help understand why deconstructing common myths around AI is essential. The methodology used follows an analytical approach and builds upon extensive related work. The paper explores the origins and determining factors of past and present AI, critically abounds in the actual processes behind its developments and limitations, and contributes to demystifying the AI hype by showing the hidden purposes and causes that hinder understanding.

2. What AI?

Much of the confusion around what AI stems from the lack of consensus on its definition (Monett et al. 2019). It has been a very controversial and contextual concept over decades, part of a "conceptual borrowing" of AI as a discipline (Floridi and Nobre, 2024). AI is software; no magic is behind it (Landgrebe and Smith, 2023; Larson, 2021). However, the term has been misused and abused, and its meaning rebranded depending on the concrete application at hand, particularly for corporate or individual interests or for capturing the media's attention, VC's support, or attracting potential users.

In the recently approved AI Act of the European Parliament, for instance, an AI system is defined (AI Act, 2024, Article 3) as "a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments." For regulatory purposes, this one might be a general enough definition. Other authors believe, however, that there are no truly "intelligent" artefacts (Crawford, 2021) and there probably never will be, which they explain is mathematically, computationally, evolutionarily, and neurobiologically demonstrable (Landgrebe and Smith, 2023).

A digital artefact (e.g. program, algorithm, application) has different levels of representation. Most people hyping or idealising digital artefacts or their capabilities, like those based on AI techniques or those advertised as having AI inside but have none, might lack a basic understanding of at least one of their levels of representation. Winograd and Flores (1986, 87ff) grouped those levels of representation into five categories, from the physical machine (its circuits and hardware components) to high-level language representation schemes for (symbolic) structures and how to operate with them. They state (p. 89) that these levels of representation and their complex interdependencies change "as new kinds of hardware are designed and as new programming concepts evolve", and suggest that "people who have not programmed computers have not generally had experiences that provide similar intuitions about [computer] systems" like those that programmers usually have and use (Winograd and Flores, 1986, p. 90). Such people normally lack a basic understanding of hardware or software breakdowns that could happen at a given point in time, how resources (e.g. memory) were or could be used, which and how higher-level operations were coded, which interdependencies between the representation levels (and

even external dependencies) exist, how the problem domain was represented, and what are the implicit biases of their programmers when coding and working with it, as bias is unavoidable and history and culture-dependent.

3. Factors Conditioning the AI Momentum

Several factors have triggered the AI momentum or AI "spring" of recent years, in particular the one from the 2010s and early 2020s. Advances in the field of computer science have been the least contributing to it, despite the Janus-faced hype, doom-loaded with utopian and dystopian idioms (Goodlad, 2023). The reasons for the supposed AI renaissance have more of a technical, economic, and even socio-political character.

The following is a non-exhaustive list of conditioning factors for the AI "spring" we are in:

- Computing power and hardware, i.e. chips and micro- and multiprocessors, especially AI chips developed for parallel computations required by AI algorithms; also technology other than computer chips, like cameras, sensors, wearables, etc. that enable data collection and processing at scale;
- Data, amassed and controlled by tech corporations, often extracted without consent from the data providers or producers. Huge amounts of data in the form of text, images, video, and audio, but also derived from users' behaviour (clicks, likes, purchases, hovers, time spent, location);
- Algorithms and models, the former mostly around for a longer while, their basic characteristics dating back to the origins of neural networks and machine learning (from the 1950s to the 1980s), now revamped, tested, and developed further; the latter consolidating theoretical insights from the fields of statistics and linguistics, now powered with massive computing hardware and fed with huge amounts of data;
- Exploitation—as Crawford (2021, p. 64) writes, "The technical AI research community relies on cheap, crowd-sourced labour for many tasks that can't be done by machines." There is a massive amount of miserably paid people annotating data for the algorithms to work and moderating other data for the companies developing such algorithms to continue extracting value from data (Castaldo, 2023; Hao, 2023; Williams et al., 2022; Zuboff, 2019);
- Investments, without it wouldn't be possible to buy technology, build huge data centres and research facilities, do research, as well as procure resources and technical infrastructure with which AI-based technology and applications can be developed;
- Courses and study programs (and also blogs, podcasts, etc.) e.g. with instructional purposes devoted to lowering the entrance level to the AI domain;
- Practitioners and self-declared instructors, i.e. people with or without professional scholarship in the AI field, also sharing, tweaking, and using algorithms for new kinds of applications and problems than ever before;
- Conferences and (open-access) publications, with a major increase in the number of papers and AI-related events (Maslej et al., 2024), written or organised by almost all types of actors, from laymen to AI insiders, and lately also including AI-generated content, directly affecting trust and review processes;
- Development platforms and tools, where (open-source) code is collaboratively produced, shared, maintained, and made accessible to novices and experts alike. These are enablers of mass consumption and of the development of new algorithms that fuel the algorithmification of society;

- Marketing and media that reverberate about the possibilities of AI algorithms and models, mostly hyping them.

Often, the creators and corporate players driving those factors redesign human activity and behaviour in the social, historical and cultural variance of its practices, intending to make AI work but not designing AI to work for humans; rather, for humans working or adapting their behaviour so that AI can work. The outcomes are models that only work in narrow scenarios. Others are rather useful for corporate profits that are silent about the AI risks and the people they harm (O'Neil, 2016), i.e. ignoring or diminishing the potential economic, political, social (Acemoglu, 2021), or environmental and planetary costs (Crawford, 2021) of the technology. Yet, those same factors conditioning the AI momentum, once abused, misused, and hyped, give rise to the so-called AI "winters," which are periods of high disappointment with major cuts in funding and field reorientation. Several works address the AI hype from different perspectives, as was the case before other AI "winters" (Floridi, 2020). Unfortunately, in the contemporary AI field local AI "winters" are still happening: in all those other AI subfields not in the media's attention, nor within venture capitalists' favours, and therefore devoid of the advantages of AI "spring"-conditioning growth. This is the case of almost all AI subfields other than machine learning, where major developments have occurred recently. For more on AI hype, AI "winters," AI pitfalls and stepping stones for progress, we refer the reader to the works of (Birhane, 2021; Floridi, 2020; Markelius et al., 2024; McQuillan, 2022; Mitchell, 2021; Strickland, 2021).

The optimisation promise and the machinisation of our lives

The algorithmic fabric that conditions the current development of AI has been part of a continued multi- objective optimisation process for the sake of innovation and technological progress. This is a fertile ground for hype and the optimisation fallacy, i.e. thinking that optimising complex processes and societies through their simplification and fragmentation is the best option for understanding and dealing with them. Those who control the data and the technology, a few unelected players, are increasingly optimising and affecting society's functioning through AI (Whittaker, 2021; Zuboff, 2019). The vast complexity of social interactions and life itself are reduced to abstract, limited mathematical formulations that cannot cope with novelty, are translated to machinistic (i.e. machine-dependent) and mechanistic expressions, and result in restricted and narrow mathematics-based algorithms that operate in utilitarian ways. Those algorithms are then used "to find, expand, fine-tune, and experiment with [human] processes [and activities] that unfold in time" (Hilbert, 2022, p. 4), thereby allowing for the algorithmification of humans and their behaviours and directly affecting them and their future.

Digital machine-accelerated optimisation is not neutral at any level. It is run on a social and human framework that is affected by different levels of power, where the core subjects of the optimisation, the humans producing the data, are the least served, directly affected, or forgotten (Buolamwini and Gebru, 2018; Raji et al., 2020; Zuboff, 2019). The extracted pieces of what we signalled online in the past, disregarding how they actually model or represent reality, are used for predicting future human behaviours that have still not occurred and are inherently incomplete, cannot be finalised, are biased, and are dynamic in a world that is open, uncertain, and non-determinable (Birhane, 2021; Birhane et al., 2023).

The AI optimisation promise is essentially a substrate on which societal and political decisions are technologically manipulated to serve a few and concentrate power. It is the gear of what Zuboff (2019) calls a surveillance economy and parasitic economic logic for

expropriating critical human rights and people sovereignty, for deskilling and corroding human agency, and for exacerbating dependency on technology. It is the product of a covert Silicolonisation dressing of a digital soft-totalitarianism which Sadin (2020) defines as "the conjunction between territorial expansion and the advent of a 'life-industry' [...] organizing the world according to self- interests, while making people believe that we have never known a 'cooler,' more 'collaborative,' and 'creative' period in history." It is an optimisation that runs ad infinitum, where, increasingly, algorithmic breakthroughs, continual technical progress, and greater corporate convenience are what matter most, usually without considering the computational costs, the ethical implications, or the impacts they have on people's lives, society, and the planet (Bender et al., 2021; Bode and Goodlad, 2023; Birhane, 2021; Buolamwini and Gebru, 2018; Crawford, 2021; Sadin, 2020; Zuboff, 2019).

4. The Machine Intelligence Fallacy: Why Changing the Narrative is Imperative

There is an urgent need for a global optimisation boost, but one that introspects and looks inward to question fundamental problems across the AI subfields and the narratives around what machine intelligence means. The machine intelligence concept is fundamentally wishful and has turned fallacious, being prone to misunderstandings about what the technology can truly achieve. In its current state, it still resembles science fiction and wishful thinking (McDermott, 1976; Mitchell, 2021). Hyping narratives and techno-solutionism still monopolise airtime and have spread out to other sectors of society, e.g. education (Bender, 2024). For example, the fear of missing out (FOMO) has become a serious problem in education. It has been undermining some educators' behaviours (Bender, 2024) who see in AI opportunities for revamping their (possibly stuck) research objectives and academic goals (Watermeyer et al., 2023). In an exercise of abundant gullibility, education and academia depend on more corporate greed than ever. With generative AI-based chatbots, for instance, the low- hanging fruits of attractive publications or visibility in an AI era rippled with hype, function like a step higher in the already deteriorated academic ladder (Lin et al., 2023; Watermeyer et al., 2023).

Similar problems, ethical dilemmas, and discussions originated in the past, actually since the beginnings of the AI field itself. Seminal works on "past" AI, its limitations, and pressing ethical implications and consequences remain seldom read or discussed (Dreyfus, 1992; Lighthill,[3] 1973; Weizenbaum, 1976; Weizenbaum and Wendt, 2015; Wiener, 1950, 1988)). For more on those and other related matters see also (Larson, 2021; Lipton and Steinhardt, 2019; McQuillan, 2022; Mitchell, 2019; Wooldridge, 2020). AI and its techniques, whatever their mathematical brightness, colourful pixelated appearance, or media and corporate furore, won't ever solve academia's structural problems. Those problems needed in the past, still need, and will remain needing in the future, human touch and agency. Avoiding or algorithmificating those problems distracts from the actual negative consequences they pose to learners, educators, entire educational institutions, and society.

If it were *only* education! Sadly, there is much more to AI and how a flawed narrative erodes or destroys trust in it. Take for example the irrational exuberance with which "new" algorithms or applications are announced ("the future is promising," "for the benefit of humanity") and how they are debunked shortly after (Mitchell, 2021; Schaeffer, Miranda and Koyejo, 2023); or the serious problems with deep fakes, mis-, and disinformation massively

[3] The "Lighthill report" was commissioned by the UK's Science Research Council and summarised the limitations of its techniques and how their actual development was far away from what was promised. It led to major cuts in research funding and to the first AI "winter" in the late 1970s.

originated *thanks* to generative AI, fuelling data contamination and leaking, as well as malpractices like reproducibility issues (Balloccu et al., 2024); or the many intellectual property infringements and ongoing lawsuits (GW Law, 2024) curbing value creation and knowledge share by artists; or the severe privacy issues not just through data collection but also through inferences from the data collected (Bartneck et al., 2021); or the many security issues of AI models with major unsolved vulnerabilities allowing for attack vectors emerging faster than any time in history before (Vijayan, 2024); or the unsustainable and disproportionate waste of resources (rare metals, electricity, water, etc. (Crawford, 2021; Luccioni, Jernite, and Strubell, 2023)) for revamping the algorithmification of our lives and what we do to even higher levels, touched with unscientific claims becoming the norm, like machines "understanding," "reasoning," "comprehending," "thinking," "having emotions," "being sentient," "acting" and "thinking" like humans (the wish!) (McDermott, 1976), i.e. already living the grand dream of AI which is still very different from reality (Mitchell, 2019, 2021; Wooldridge, 2020). None of the eventually well-intended AI applications counterbalances the dangers they pose to disempowered populations and political, societal, and planetary ecosystems (Crawford, 2021; McQuillan, 2022; Zuboff, 2019). The AI field is slowly imploding before everyone's eyes; trust has been eroded and the discourse around AI politicised. The next debacle or AI "winter" might be around the corner. It is time to put an end to all that nonsense once and for all.

5. Recommendations for a new Type of AI Literacy

The narratives we need about AI should focus on factual information that demystifies AI. In this regard, AI literacy based only on its functionality (i.e. what it is and how it works) is not enough. The critical components must come prominently into the AI equation. We must demand responsibility proactively when dealing with AI and through regulation. We must contest malpractices and seek accountability. We need to be sceptical, to question why, for whom, for which purposes, who is profiting from our data and why, which values and rights are eroded, and why they must be at the centre of a well-functioning society, among many other questions.

Our concrete recommendations for a much-needed new type of AI literacy are:

- Technological developments, especially digital ones, depend on much more than just AI-based algorithms (see Section 3 for details). Emphasise those other factors; demystify with factual information the magic thinking about what AI could do but still does not.
- Reclame autonomy back and make informed decisions related to technology. "You have to choose between [emphasis added] freedom or servitude to powers and systems that decide for us the course of things" (Sadin, 2020).
- Put human ethical and moral judgement into focus. If individuals or whatever the system is not ethical, we cannot have ethical outcomes. They should be more than just soft, vacuous wording to industry, academia, government, and society.
- Prioritise human rights and values. By surrendering human agency to algorithmification one bit at a time, we gradually give up our rights and values. Vulnerable populations, children, and disempowered people are the most affected. AI systems cannot be truly compatible with societal standards and will not align well with human values if part of its members are left behind.
- Learn about and try out new products, tools, and frameworks. They are skills that may be important in several fields or for using certain applications. Learn also that, other

times, however, some of them do not deserve the attention they are given and are only enablers of deskilling and mediocrity. Is it the future you want to live in? Question the what, how, why, by whom, for what, etc. of AI and digital technologies. This helps to understand or even avoid the AI fog.

- Collaborate, break the siloed thinking: multidisciplinarity is key to today's AI. E.g. involve people affected directly by AI in all phases of its lifecycle; that is the best way to design, develop, test, deploy, and use, i.e. deal with AI.
- Strive for informed decisions. Replace being told what to do or what to think concerning AI, with critical AI literacy. This needs critical thinking and grounded knowledge about AI. Verify, then trust!

Critical AI-literate people with a healthy dose of scepticism, if in doubt, should stop blindly accepting uncertain decision making and start asking the hard questions. Educators play a distinctive role thereby; they should shift from FOMO and corporate will to contesting both technology and technologists; they should stop educating generations of learners on wishful futures and techno-idealised resources-burning algorithmification. Teaching and learning spaces must shift from digital swamps and human automatons to active and critical shapers of society's future. The rule of law plays an essential role in guaranteeing that technologists, VCs, and anyone deploying software systems that might be life impacting are accountable. The reader is referred to the many policies, safeguards, and regulatory frameworks that are already into force or proposed to be introduced. For example, the EU AI Act (EU, 2024), the US Blueprint for an AI Bill of Rights (US, 2022), the UK AI Regulation Bill (UK, 2023), or the preliminary draft of China's proposed AI Law (China, 2024), among others.[4] Neither AI nor the software industry are exempt from accountability. Opponents have been voicing concerns about regulation stifling innovation; however, regulation is not de facto a barrier to innovation. Correct regulation is actually an enabler of innovation and a guarantor of a safety baseline. Other industries and their products are strongly regulated and they continue to innovate at scale.

6. Conclusions

Society is being algorithmificated, but not just through AI. The danger is a shift to the "just automate" mindset and over-reliance on machines leading to surrendering human agency, which, in the end, only empowers those who will make decisions for us instead of us. Digital literacy, especially critical AI literacy must be part of national education curricula at all levels, from school to higher education and to postgraduate, vocational, and corporate education. Those are core skills for life that, left unattended, will reinforce the power and serve the purposes of only a few, mostly not democratically elected players. The true purpose and essence of "modern" automation are hidden behind the algorithmification of society and encompass practices and processes that are not always obvious or easy to detect or contest (McQuillan, 2022). Only when people are affected, e.g. disadvantaged populations or those already marginalised (which are actual majorities), is the true impact of automation known or visible (these are mostly the problems that make it into press news). It is already too late to tackle them or develop countermeasures by then. At this time, the lives and future of uncountable people, especially the disempowered ones, might have been affected

[4] The OECD.AI Policy Observatory is also a good referent for tracking national AI policies and strategies from around the world: https://oecd.ai/en/dashboards/overview.

irreversibly. Educating ourselves and others, being aware, and understanding that the answer should not be to rush or surrender to the algorithmification of our lives, are the critical skills of the digital era that need urgent attention.

References

Acemoglu, D. (2021) Harms of AI. NBER Working Paper 29247, National Bureau of Economic Research. SSRN Electronic Journal. https://ssrn.com/abstract=3922521.

Acemoglu, D. and Johnson, S. (2023) Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity. UK: Basic Books.

AI Act (2024) "Artificial Intelligence Act," European Parliament. https://www.europarl.europa.eu/doceo/document/TA-9- 2024-0138_EN.pdf.

Balloccu, S., Schmidtová, P., Lango, M., and Dusek, O. (2024) "Leak, Cheat, Repeat: Data Contamination and Evaluation Malpractices in Closed-Source LLMs," in Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics, Vol. 1, pp. 67–93.

Bartneck, C., Lütge, C., Wagner, A., and Welsh, S. (2021) "Privacy Issues of AI," in An Introduction to Ethics in Robotics and AI, SpringerBriefs in Ethics. Springer, Cham.

Bender, E. (2024) "Doing their hype for them: Defeatist, second-hand hype goes to college," Mystery AI Hype Theater 3000: The Newsletter. https://buttondown.email/maiht3k/archive/doing-their-hype-for-them/.

Bender, E.M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021) "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT'21, pp. 610–623.

Birhane, A. (2021) "The Impossibility of Automating Ambiguity," Artificial Life, Vol. 27, No., 1, pp. 44–61. Birhane, A., Prabhu, V., Han, S., and Boddeti, V.N. (2023) "On Hate Scaling Laws for Data-Swamps," arXiv [cs.CY]. https://arxiv.org/abs/2306.13141.

Bishop, J.M. (2021) "Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It," Frontiers in Psychology, Vol. 11, 513474.

Bode, K. and Goodlad, L.M.E. (2023) "Data Worlds: An Introduction," in Bode, K. and Goodlad, L.M.E. (eds.), Critical AI, Vol. 1, No. 1–2.

Buolamwini, J. and Gebru, T. (2018) "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Vol. 81, pp. 77–91.

Castaldo, J. (2023, September 16) "Meet the gig workers making AI machines more accurate, capable and powerful," The Globe and Mail. https://www.theglobeandmail.com/business/article-ai-data-gig-workers/.

China (2024). "Artificial Intelligence Law of the People's Republic of China (Draft for Suggestions from Scholars)" [Translation], Center for Security and Emerging Technology, Georgetown University, Washington, D.C. https://cset.georgetown.edu/publication/china-ai-law-draft.

Crawford, K. (2021) Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press. Dreyfus, H.L. (1992) What Computers Still Can't Do: A Critique of Artificial Reason. MIT Press.

EU (2024). "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)," EUR-Lex, European Union. https://eur-lex.europa.eu/legal- content/EN/TXT/?uri=OJ%3AL_202401689.

Floridi, L. (2020) "AI and Its New Winter: from Myths to Realities," Philos. Technol., Vol. 33, pp. 1–3.

Floridi, L. and Nobre, A.C. (2024) "Anthropomorphising machines and computerising minds: the crosswiring of languages between Artificial Intelligence and Brain & Cognitive Sciences," Centre for Digital Ethics (CEDE). http://dx.doi.org/10.2139/ssrn.4738331.

Goodlad, L.M.E. (2023) "Editor's Introduction: Humanities in the Loop," in Bode, K., and Goodlad, L.M.E. (eds.), Critical AI, Vol. 1, No. 1–2.

GW Law (2024) "AI Litigation Database," The George Washington University, Washington DC, USA. https://blogs.gwu.edu/law-eti/ai-litigation-database/.

Hao, K. (2023, July 11) "The Hidden Workforce That Helped Filter Violence and Abuse Out of ChatGPT," [Podcast]. The Journal, WSJ Podcasts. https://www.wsj.com/podcasts/the-journal/the-hidden-workforce-that-helped-filter- violence-and-abuse-out-of-chatgpt/ffc2427f-bdd8-47b7-9a4b-27e7267cf413.

Hilbert, M. (2022) "Societal Algorithmification and its Computational Measurement," in Yasseri, T. (Ed.), Handbook of Computational Social Science, Edward Elgar Publishing.

Landgrebe, J. and Smith, B. (2023) Why Machines Will Never Rule the World: Artificial Intelligence without Fear. New York: Routledge.
Larson, E.J. (2021) The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do. Cambridge, MA: Harvard University Press.
Lighthill, M.J. (1973) "Artificial Intelligence: A General Survey," Artificial Intelligence: A paper symposium. UK: Science Research Council. http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm.
Lin, S.M., Chung, H.H., Chung, F.L., and Lan, Y.J. (2023) "Concerns About Using ChatGPT in Education," in Huang, Y.M., Rocha, T. (eds), Innovative Technologies and Learning. ICITL 2023. Lecture Notes in Computer Science 14099. Springer, Cham.
Lipton, Z.C. and Steinhardt, J. (2019) "Troubling Trends in Machine Learning Scholarship: Some ML papers suffer from flaws that could mislead the public and stymie future research," Queue, Vol. 17, No. 1, pp. 45–77.
Luccioni, A.S., Jernite, Y., and Strubell, E. (2023) "Power Hungry Processing: Watts Driving the Cost of AI Deployment?" arXiv [cs.LG]. https://arxiv.org/abs/2311.16863.
Markelius, A., Wright, C., Kuiper, J., Delille, N., and Kuo, Y.-T. (2024) "The mechanisms of AI hype and its planetary and social costs," AI Ethics. https://doi.org/10.1007/s43681-024-00461-2.
Maslej, N., Fattorini, L., Perrault, R., et al. (2024) "The AI Index 2024 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA.
McDermott, D. (1976) "Artificial Intelligence meets Natural Stupidity," SIGART Newsletter. Vol. 57, pp. 4–9.
McQuillan, D. (2022) Resisting AI: An anti-fascist approach to artificial intelligence. Bristol University Press, UK.
Mitchell, M. (2019) Artificial Intelligence: A Guide for Thinking Humans. UK: Pelican Random House.
Mitchell, M. (2021) "Why AI is Harder Than We Think," arXiv [cs.AI]. https://arxiv.org/abs/2104.12871.
Monett, D., Hoge, L., and Lewis, C.W.P. (2019) "Cognitive Biases Undermine Consensus on Definitions of Intelligence and Limit Understanding," in the Joint Proceedings of the Workshops LaCATODA 2019 & BtG 2019, Vol. 2452, pp. 52–59.
O'Neil, C. (2016) Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. UK: Allen Lane, Penguin Books.
Raji, I.D., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J., and Denton, E. (2020) "Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing," in Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society, AIES '20, pp. 145–151.
Raji, I.D., Kumar, I.E., Horowitz, A., and Selbst, A. (2022) "The Fallacy of AI Functionality," in Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), pp. 959–972.
Sadin, É. (2020) La Silicolonización del Mundo: La irresistible expansión del liberalismo digital (engl.: The Silicolonisation of the World: The irresistible expansion of digital liberalism), Buenos Aires: Caja Negra.
Schaeffer, R., Miranda, B., and Koyejo, S. (2023) "Are Emergent Abilities of Large Language Models a Mirage?" in Oh, A. et al. (eds.), Proceedings of the 37th Annual Conference on Neural Information Processing Systems, NeurIPS 2023, Vol. 36, pp. 55565–55581.
Strickland, E. (2021) "The Turbulent Past and Uncertain Future of Artificial Intelligence: Is there a way out of AI's boom-and-bust cycle?" IEEE Spectrum. https://spectrum.ieee.org/amp/history-of-ai-2655064200. UK (2023). "Artificial Intelligence (Regulation) Bill," House of Lords, UK Parliament. https://bills.parliament.uk/publications/53068/documents/4030.
US (2022). "Blueprint for and AI Bill of Rights," The Office of Science and Technology Policy, The White House, U.S. https://www.whitehouse.gov/ostp/ai-bill-of-rights/.
Vijayan, J. (2024) "Pervasive LLM Hallucinations Expand Code Developer Attack Surface," DarkReading, Informa. https://www.darkreading.com/application-security/pervasive-llm-hallucinations-expand-code-developer-attack-surface.
Watermeyer, R., Phipps, L., Lanclos, D., and Knight, C. (2023) "Generative AI and the Automating of Academia," Postdigit Sci Educ. https://doi.org/10.1007/s42438-023-00440-6.
Weizenbaum, J. (1976) Computer Power and Human Reason: From Judgment to Calculation. W. H. Freeman & Co. Weizenbaum, J. and Wendt, G. (2015) Islands in the Cyberstream: Seeking Havens of Reason in a Programmed Society. Litwin Books.
Whittaker, M. (2021) "The steep cost of capture," Interactions, Vol. 28, No. 6, pp. 50–55.
Wiener, N. (1950, 1988) The Human Use of Human Beings: Cybernetics and Society. Da Capo Press.
Williams, A., Miceli, M., and Gebru, T. (2022, October 13) "The Exploited Labor Behind Artificial Intelligence," Noema. https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/.
Winograd, T.A. and Flores, F. (1986). Understanding Computers and Cognition: A New Foundation for Design. Norwood, NJ: Ablex Publishing Corporation.

Wooldridge, M. (2020) The Road to Conscious Machines: The Story of AI. UK: Pelican Random House.
Zuboff, S. (2019) The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power. USA: Public Affairs, Hachette Book Group.