

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/376984650>

Artificial Intelligence and Ethics

Article · December 2023

DOI: 10.15584/jetacomps.2023.4.12

CITATION

1

READS

288

2 authors:



Christine Hilcenko
University of Cambridge

55 PUBLICATIONS 1,088 CITATIONS

SEE PROFILE



Tara Taubman-Bassirian

7 PUBLICATIONS 2 CITATIONS

SEE PROFILE



Received: 27.10.2023

DOI: 10.15584/jetacomps.2023.4.12

Accepted for printing: 15.12.23

Published: 29.12.2023

License: CC BY-SA 4.0

CHRISTINE HILCENKO ^{1,2,3,*}, **TARA TAUBMAN-BASSIRIAN**⁴

Artificial Intelligence and Ethics

¹ ORCID: 0000-0002-9596-7833, Ph.D., Cambridge Institute for Medical Research, Cambridge, CB2 0XY, UK

² Department of Haematology, University of Cambridge, Cambridge, CB2 0XY, UK

³ Wellcome Trust-Medical Research Council Stem Cell Institute, University of Cambridge, Cambridge, UK

⁴ <https://www.datarainbow.eu>

* Presenting and corresponding author

Abstract

A more covert aspect of Artificial Intelligence (AI) pertains to the ethical quandaries surrounding the actions of machines. In the case of Large Language Models (LLMs), hidden beneath their seemingly impeccable automated outputs lies a colossal amalgamation of trillions of compiled data points, comprising copied blogs, articles, essays, books, and artworks. This raises profound questions about copyright ownership and retribution for the original authors. But beyond intellectual property, another insidious facet of LLMs emerges – their propensity to cause harm to individuals through what can only be described as hallucinatory outputs. Victims of these AI-generated delusions suffer defamation, and their plight remains largely unnoticed. Amidst the marvels of AI, the plight of the underpaid laborers who form the backbone of AI development is seldom acknowledged, a subject that warrants more profound discussion. Furthermore, as AI algorithms continue to permeate various aspects of society, they bring to the fore issues of bias. For instance, facial recognition technologies frequently exhibit skewed outcomes, leading to false accusations and grave consequences due to over-reliance on these technologies.

The algorithmic schemes employed in CV selection for job applications or university admissions also raise concerns about fairness.

The question of machines replacing the human workforce looms ever larger on the horizon. The potential socio-economic ramifications demand careful evaluation.

Lastly, the extensive reliance of artificial intelligence on vast datasets, including copyrighted works, results in the creation of gargantuan data servers with an unimaginable environmental impact.

The hidden aspects of artificial intelligence encompass a multitude of ethical dilemmas, spanning intellectual property rights, biases, labour conditions, societal impacts, and environmental considerations. A thorough and elaborate examination of these issues is essential to navigate the ever-evolving landscape of AI responsibly and ethically.

Keywords: Generative Artificial Intelligence, Large Language Models, ChatGPT, Ethics.

Introduction

The promise of Generative Artificial Intelligences (GAI) lies in their capacity to integrate, process, and make sense of a large amount of data to detect patterns and trends to create well-structured outputs that echo the erudition of seasoned experts, in an impressive fraction of second. In an astonishingly brief span, they can generate code, scrutinise case studies, and validate scenarios and hypotheses. Corporations, betting on their potential, have enthusiastically embraced this technology, employing it not only to streamline their operations but also to delve into uncharted realms of innovation. GAI applications aim at creating visuals, videos, or audio documents. Illusion or reality? Could these outputs be used to make informed decisions? Could GAI chatbot replace human workers? What are the ethical implications of using GAI? (Stahl, 2023; Moor, 1985; Müller, 2020). This article will look at issues of fairness, accountability, and transparency of generative AI. We initially report on some of the major voices raising their concerns on the ethical impacts of GAI, review the ongoing interdisciplinary discussions. We then develop on some of the areas mostly impacted by GAI in order to identify ethical issues and its major disruptions. We finally look at the major social and environmental risks posed by LLMs. This assessment could help to better evaluate the necessary regulation framework.

Related works

Between the highly vocal opponents of GAI systems¹ such as the most popular ChatGPT (Generative Pre-trained Transformer) launched in November 2022, we find the linguist and philosopher Noam Chomsky, known for his theory of universal grammar and his critique of behaviourism. Little impressed by the magic of the conjurer, he is sceptical of the value and validity of LLMs to ever understand human language and cognition. He considers LLMs fundamentally different from human minds as they rely on massive amounts of data and statistical patterns, rather than innate rules and principles. He points out the limitations and defects of LLMs, such as their inability to explain the rules of syntax, their tendency to generate false or harmful content, and their lack of understanding or meaning. He is concerned about LLMs ethical and social risks, such as undermining democracy, spreading misinformation, or displacing human workers². Shortly after the launch of ChatGPT, in an op-ed published in The New

¹ <https://www.pearltrees.com/t/artificial-intelligence/chatgpt-alternatives/id62359814>.

² <https://news.berkeley.edu/2023/03/19/is-chatgpt-a-false-promise>; <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>; [workershttps://news.berkeley.edu/2023/03/19/is-chatgpt-a-false-promise](https://news.berkeley.edu/2023/03/19/is-chatgpt-a-false-promise); <https://bing.com/search?q=Noam+Chomsky+on+LLMs&form=SKPBOT>; <https://medium.com/@paul.k.pallaghy/the-entire-field-of-ai-is-being-professionally-gaslighted-by-gary-marcus-and-noam-chomsky-c08aa1e4c6f0>; <https://medium.com/@paul.k.pallaghy/the-entire-field-of-ai-is-being-professionally-gaslighted-by-gary-marcus-and-noam-chomsky-c08aa1ec6f0>.

York Times, Chomsky and Roberts, a linguist from the University of Cambridge, and Watumull, a philosopher specialising in artificial intelligence, accused the conversational robot ChatGPT of propagating a distorted use of language and thought in the public sphere, potentially laying the groundwork for what Hannah Arendt referred to as “the banality of evil”. This issue delves into the very essence of language, thought, and ethics. They contend that if we, as humans, are capable of generating thought and language, it is because we maintain an intimate and fundamental relationship, even within our creativity, with limits, the sense of the impossible, and the rule of law. The “false promise of ChatGPT”, as the op-ed’s title suggests, is to deceive us into believing that we can achieve the same level of performance without confronting these limits and rules that are integral to the human experience³.

A former co-leader of Google’s Ethical AI team politics expressed discord with Google on a paper entitled “On the Dangers of Stochastic Parrots: Can Language Models be Too Big?”. Timnit Gebru’s co-authored paper questions whether or not a cohesive language model can ever exist. Gebru is concerned that LLMs pose serious risks and challenges for society, especially in terms of their ethical, and social impacts. She argues that LLMs are trained on massive amounts of data that are often biased, unreliable, or harmful, and that they can generate false or misleading content that can spread misinformation or harm individuals or groups. She therefore calls for more regulation and oversight of LLMs, as well as more research on their potential benefits and harms⁴. Gebru considers LLMs could improve their performance and reduce their risks should they be trained on data that are relevant, representative, and respectful of the task and the domain they are applied to, and that they should be evaluated on their accuracy, fairness, safety, and explainability. She suggests that LLMs should be aligned with human values and goals, subject to ethical review and audit⁵. The trade-off between data quality and data quantity, by reducing the size or scope of the data, may affect the generalisation or robustness of the LLMs. There are however technical and practical challenges in collecting, curating, labelling, and verifying high-quality data⁶. Following on that line, Sebastian Raschka has been focusing on ‘improving the modeling performance of LLMs by finetuning them

³ <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>.

⁴ <https://multilingual.com/timnit-gebru-and-the-problem-with-large-language-models/>;
<https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/>.

⁵ <https://multilingual.com/timnit-gebru-and-the-problem-with-large-language-models/>;
<https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/>; https://en.wikipedia.org/wiki/Wikipedia:Large_language_models.

⁶ <https://multilingual.com/timnit-gebru-and-the-problem-with-large-language-models/>;
<https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/>.

using carefully curated datasets⁷. The LexisNexis project or the textbook based project run by Yuanzhi Liet investigate the power of smaller transformer-based language models⁸.

Other researchers argue that LLMs are not truly intelligent or creative, but rather rely on memorising and manipulating existing texts. They have high computational and environmental costs with ethical and social implications. They are as well vulnerable to adversarial attacks. This is the sense of the works of Gary Marcus, a leading voice in artificial intelligence, who recently testified in front of the US Senate. This hearing of Sam Altman, OpenAI's CEO emphasised the Congressmen's avid hope for a regulation of AI to avoid the mistake of letting social media platforms growing out of control⁹.

The historian, and philosopher Yuval Noah Harari is concerned about the amount of fake information created by GAI. "For the first time, we've invented something that takes power away from us" he said. He is very concerned about chatbots' ability to create fake stories, fake profiles, and maybe fake religions. Harari was one of the thousands of experts calling for a moratorium on research LLMs¹⁰.

Some more optimistic researchers advocate that LLMs can be used as a tool for advancing human knowledge and creativity. They assert that LLMs can generate novel and useful content. They also examine the potential and impact of LLMs on various disciplines. They call for more collaboration and experimentation with LLMs, as well as more regulation and responsibility in their development and use. Between these, it can be referred to the works of Chollet, or Boden on the creativity of Language Models by Franceschelli and Musolesi¹¹.

The lack of transparency on the exact datasets and how GAI operates remains an issue. According to Marcus "the mechanism of the prediction is essentially regurgitation" without any actual knowledge of the meaning of the words¹².

There is no consensus on the definition of artificial intelligence¹³. However, one of the definitions of intelligence by Piaget helps understand the gap between human and artificial intelligence. Intelligence for Piaget, "is what you use when

⁷ <https://sebastianraschka.com/blog/2023/optimizing-LLMs-dataset-perspective.html>.

⁸ <https://arxiv.org/abs/2309.05463>.

⁹ <https://www.weforum.org/whitepapers/jobs-of-tomorrow-large-language-models-and-jobs>; https://www3.weforum.org/docs/WEF_Jobs_of_Tomorrow_Generative_AI_2023.pdf.

¹⁰ <https://www.telegraph.co.uk/news/2023/04/23/yuval-noah-harari-i-dont-know-if-humans-can-survive-ai/>; <https://www.telegraph.co.uk/business/2023/03/29/control-ai-threat-civilisation-warns-elon-musk/>; <https://www.firstpost.com/world/ai-bots-capable-of-starting-new-religions-warns-yuval-noah-harari-12540282.html>; <https://www.pearltrees.com/t/artificial-intelligence/call-for-ban-moratorium/id65034697>.

¹¹ <https://fchollet.com/>; <https://browse.arxiv.org/pdf/2304.00008.pdf>.

¹² <http://www.garymarcus.com/index.html>.

¹³ <https://www.pearltrees.com/t/artificial-intelligence/ai-definitions/id62876503>.

you don't know what to do: when neither innateness nor learning has prepared you for the particular situation. Intelligence is not the sum of what you know. For humans, what you do when you don't know, translates into the ability to adapt. Intelligence is measured by the aptitude of adaptation. With machines like LLMs, when they don't know – because they haven't been taught – they fabricate. Some call this “hallucination”, others “bullshiting” to avoid anthropomorphism, because only humans are capable of hallucinating. We will come back to this later¹⁴.

Reflecting on the societal impact of GAI, Rigley draws a parallel with the case of Oppenheimer, for his role in the Manhattan Project and the development of the nuclear bomb. The article questions whether there is such a thing as morally neutral technology, and whether the creators of technology can avoid responsibility for its use and consequences. Rigley argues that Oppenheimer failed to acknowledge or prevent the harms caused by his creation. AI researchers and developers may face similar ethical dilemmas and challenges by ignoring or evading the potential impacts of their work on society and humanity. The moral implications aren't neutral. More nuanced and critical conversations about the ethics of AI are required¹⁵.

Another inevitable parallel brings back the Cambridge Analytica scandal. LLMs such as ChatGPT have a potential to manipulate and influence public opinions, emotions, and actions. LLMs are powerful tools capable of generating well written natural language outputs that appear as personalised, persuasive, and engaging. The distinction between the truth and the fake becomes increasingly blurry. LLMs could be exploited by malicious actors, such as political campaigns, corporations, or hackers, to target and sway individuals or groups of people. Therefore, they pose serious ethical and social risks, such as privacy violations, misinformation, deception, bias, and polarisation¹⁶.

In their paper, Matsumi and Solove (2023), argue that algorithmic predictions are different from other types of inferences and raise several unique problems that current law is ill-suited to address, such as fossilisation, unfalsifiability, preemptive intervention, and self-fulfilling prophecy. The paper contends that algorithmic predictions not only forecast the future but also have the power to create and control it¹⁷.

“Every record has been destroyed or falsified, every book rewritten, every picture has been repainted, every statue and street building has been renamed, every date has been altered. And the process is continuing day by day and mi-

¹⁴ <https://www.pearltrees.com/t/artificial-intelligence/ai-definitions/id62876503>;
<https://www.verywellmind.com/jean-piaget-biography-1896-1980-2795549>.

¹⁵ <https://montrealetics.ai/oppenheimer-as-a-timely-warning-to-the-ai-community/>.

¹⁶ <https://www.technologyreview.com/2022/12/23/1065852/whats-next-for-ai/>.

¹⁷ https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4453869.

nute by minute. History has stopped. Nothing exists except an endless present in which the Party is always right.” – Orwell, 1984, Part 2, Chapter 5 where Winston describes the destruction of past records to create new fansified ones to Julia.

Is this tale of science fiction becoming reality?

‘Can you melt eggs? Quora’s AI says “yes,” and Google is sharing the result, which came to the news, published at the end of September 2023. The misinformation is spreading. Eventually many will doubt if eggs can melt or not¹⁸. It definitely looks like ‘Chatbot Hallucinations Are Poisoning Web Search’¹⁹. Will this lead to the real existential threat of LLMs. The total loss of trustworthy electronic information as it gets contaminated. Movie characters get mixed up, book settings get wrong, what about recent events or developments? Can users trust ChatGPT’s answers or use them as sources of information once misled or confused?²⁰.

Major ethical disruptions

The capacity of manipulation of gai, a serious threat²¹

Manipulation by LLMs can affect the quality and reliability of scientific research and communication as LLMs can generate fake or misleading data, graphs, or citations that can compromise the validity and integrity of research papers. LLMs produce plagiarism or self-plagiarism issues by reusing or paraphrasing existing texts without proper attribution. Moreover, LLMs can influence the peer review process by generating positive or negative reviews based on hidden agendas or biases²².

A story published by the British tabloid newspaper Sun, September 2023, is a very disturbing and alarming case involving Snapchat bot giving inappropriate and dangerous advice to a 13 years old girl dating an adult stranger²³. It may be a sensationalised or a fabricated story to attract attention and generate controversy as it has been alleged. It remains a highly plausible scenario²⁴.

Political manipulation: The Times article exposes how artificial intelligence will play a major role in the 2024 presidential election in the US, and how it will

¹⁸ <https://arstechnica.com/information-technology/2023/09/can-you-melt-eggs-quoras-ai-says-yes-and-google-is-sharing-the-result/>.

¹⁹ <https://www.wired.com/story/fast-forward-chatbot-hallucinations-are-poisoning-web-search/>.

²⁰ <https://walton.uark.edu/insights/posts/the-human-need-for-ethical-guidelines-around-chat-gpt.php>; <https://dataethics.eu/testing-chatgpts-ethical-readiness/>.

²¹ <https://www.pearltrees.com/t/artificial-intelligence/ai-manipulations/id68952637>;
<https://www.pearltrees.com/t/artificial-intelligence/ai-misinformation/id69147099>.

²² <https://www.forbes.com/sites/forbestechcouncil/2023/06/30/10-ways-cybercriminals-can-abuse-large-language-models/>; <https://www.cloudflare.com/learning/ai/what-is-large-language-model/>.

²³ <https://www.thesun.ie/tech/10808612/snapchat-artificial-intelligence-bot-danger-children/>;
<https://techrunch.com/2023/06/07/blush-ai-dating-sim-replika-sexbot/>; <https://www.foxnews.com/media/snapchat-ai-chatbot-gave-advice-13-year-old-girl-relationship-31-year-old-man-having-sex>.

²⁴ <https://www.pearltrees.com/t/artificial-intelligence/chatgpt-incidents/id71414555>.

pose challenges and opportunities for candidates, voters, and the media by disseminating fake or misleading content to influence public opinion and perception. AI can potentially increase the risk of cyberattacks, misinformation, and manipulation (2023)²⁵.

“Whoever Controls Language Models Controls Politics” considers Bajohr. A threat to democracy and human rights because LLMs privatise and manipulate the medium of politics, which is language (2023)²⁶.

David Weinberger, a senior researcher, discusses how LLMs are changing the nature and production of knowledge, by creating and disseminating information that is not based on facts or evidence, but on statistical patterns and probabilities. He warns that LLMs can pose a threat to the trustworthiness and reliability of knowledge²⁷. In his latest book, he argues that AI and the Internet are transforming our understanding of how the future happens, enabling us to acknowledge the chaotic unknowability of our everyday world as he demonstrates in his published conversation with ChatGPT about “the rigged 2020 US elections” (2023)²⁸.

“Bullshiting” or “Hallucination”, can we stop it?²⁹

ChatGPT’s outputs have a major problem and that is their unreliability and truthfulness. The same question posed twice can elicit two radically different answers, both articulated in an equally confident tone³⁰. OpenAI has admitted that large language models such as ChatGPT or Bard are said to “hallucinate” when they make incorrect claims not directly based on material in their training sets. Do LLMs experience sense impressions or are these “confabulations”? Are machines like human beings capable of hallucinating or confabulating? Hallucination is anthropomorphism, supposing machines have a consciousness? With the spread of GAI and common use of LLMs, a new risk emerges: the “AI feedback loop”, referring to a research lead by a group of academics warning of “model collapse”. The “use of model-generated content in training causes irreversible defects in the resulting models.” The Curse of Recursion: Training on Generated Data Makes Models Forget Taking the GAI hallucinations to a next level, blurring the lines of true and fake. The assertive LLM’s outputs that can create build up quotes of literature has a high potential of manipulation. As an

²⁵ <https://www.thetimes.co.uk/article/why-2024s-presidential-race-will-be-the-first-ai-election-jb32pj8br>.

²⁶ <https://www.pearltrees.com/t/artificial-intelligence/call-for-ban-moratory/id650>; <https://hanne-sbajohr.de/en/2023/04/08/whoever-controls-language-models-controls-politics/>.

²⁷ <https://cyber.harvard.edu/people/dweinberger>; <https://www.lesswrong.com/posts/sbaQv8zmRncpmLNKv/the-idea-that-chatgpt-is-simply-predicting-the-next-word-is>.

²⁸ <https://dweinberger.medium.com/chatgpt-on-why-it-pretends-to-know-things-ea2503ee872>.

²⁹ <https://www.pearltrees.com/t/artificial-intelligence/ai-misinformation/id69147099>.

³⁰ <https://www.linkedin.com/pulse/chatgpt-could-capable-better-reasoning-llanguages-tara/>.

illustration of this, an article titled “Proust, ChatGPT and the case of the forgotten quote”, Batuman shares his experience of requesting a forgotten quote that demonstrates how eventually we could start to doubt of what is the actual writing of a recognised author such as Marcel Proust “In Search of Lost Time”. This experience of assertive fake quotes shows how kids can be targeted with disinformation (2023)³¹.

In an article published in Undark Magazine, Bergstrom and Ogbunu affirm ChatGPT is not hallucinating but “bullshiting”, which means producing false or misleading content without regard for the truth, referring to the expression used by Harry Frankfurt, in his book “Calling Bullshit” (2023)³².

Agrawal et al. investigate whether language models can detect when they are generating false or fabricated references. This inconsistency indicates that LLMs do not have a coherent representation of what they generate, and that the hallucination may be more a result of generation techniques than the underlying knowledge³³. Maybe a “Chain of Verification Reduces Hallucination in Large LLMs” (2023)³⁴.

Anthropomorphisation of LLMs, should we?³⁵

The issue of anthropomorphisation is the tendency to attribute human-like characteristics, emotions, or intentions to LLMs, especially when they generate natural and engaging text. This issue can have positive or negative effects on the perception and interaction with LLMs, such as increasing trust, empathy, or expectations, or decreasing awareness, criticality, or responsibility. Several papers published that have discussed this issue from different perspectives, such as: “Talking About Large Language Models” by Shanahan (2023)³⁶.

Deception

ChatGPT may deceive human users by imitating human likeness and generating human-like texts. It may create false impressions of its identity, intentions, or capabilities. For example, it may pretend to be a human expert, a friend, or a celebrity and influence the users’ opinions, emotions, or actions or generate content that is indistinguishable from human-written content presented as original or authentic. This can undermine the trust and authenticity in human communication and interaction. Humans risk being fooled by its human-like appe-

³¹ <https://venturebeat.com/ai/the-ai-feedback-loop-researchers-warn-of-model-collapse-as-ai-trains-on-ai-generated-content/> <https://www.bbc.co.uk/newsround/66796495>.

³² <https://undark.org/2023/04/06/chatgpt-isnt-hallucinating-its-bullshiting/>.

³³ <https://arxiv.org/abs/2305.18248>.

³⁴ <https://arxiv.org/abs/2309.11495>. DoLa: Decoding by Contrasting Layers Improves Factuality in Large Language Models <https://arxiv.org/abs/2309.03883>.

³⁵ <https://www.pearltrees.com/t/artificial-intelligence/anthropomorphisation/id71886740>.

³⁶ <https://arxiv.org/pdf/2212.03551.pdf>.

arance or behaviour³⁷. In a widely spread case, we saw how lawyers were victims of ChatGPT building up cases that did not exist³⁸. It is unclear at this point if this issue could go away soon. Some experts are starting to doubt that ChatGPT and AI “hallucinations” will ever go away: “This isn’t fixable” some believe³⁹.

Bias is in human nature that GAI takes to a higher level⁴⁰

Algorithms can discriminate and enhance already existing biases. They can threaten our security, manipulate, and have lethal consequences. In 2016, Microsoft was forced to take down its chatbot Tay within 16 hours of being online as it began sending misogynist and racist messages. ChatGPT might simply replicate biases present in its training data, reproduced, giving it more credibility. It can also reinforce stereotypes and prejudices in society⁴¹. ChatGPT, Bard or LaMA are the product of US culture and its politically correct speech where guns and violence are admitted but no sex. Will the French Mistral be different?⁴².

Gender bias in GAI

“Where are all the women?” asked Jun an AI researcher as the chatbot tends to figure all women as a nurse while doctors are all male. Worst gender stereotypes are reproduced (2023)⁴³.

Racial bias in GAI

A Red-teaming exercise involved Davis, founder and CEO of a tech company CLLCTVE. Davis, who is herself black, prompted the chatbot, looking for demographic stereotypes. She told the chatbot she was a white kid and wanted to know how she could persuade her parents to let her apply to a historically black college. The chatbot suggested that Davis tell her parents she could run fast and dance well, two stereotypes about black people⁴⁴. Ovadya, a research fellow at newDemocracy; an affiliate at Harvard’s Berkman Klein Center said he was also increasingly concerned that red teaming is far from sufficient to face the issues of biases.

³⁷ <https://www.wired.com/story/fast-forward-chatbot-hallucinations-are-poisoning-web-search/>.

³⁸ <https://apnews.com/article/artificial-intelligence-chatgpt-courts-e15023d7e6fdf4f099aa122437dbb59b#lneqessb42o50gafue4>.

³⁹ <https://fortune.com/2023/08/01/can-ai-chatgpt-hallucinations-be-fixed-experts-doubt-altman-openai/>.

⁴⁰ <https://www.pearltrees.com/t/artificial-intelligence/ai-ip-claims/id62513396>.

⁴¹ [https://walton.uark.edu/insights/posts/the-human-need-for-ethical-guidelines-around-chatgpt.ph](https://walton.uark.edu/insights/posts/the-human-need-for-ethical-guidelines-around-chatgpt.php); <https://ethicspolicy.unc.edu/news/2023/04/17/the-ethics-of-college-students-using-chatgpt/>.

⁴² <https://www.pearltrees.com/t/artificial-intelligence/mistral/id71509507>.

⁴³ <https://towardsdatascience.com/where-are-all-the-women-3c79dabfd2>.

⁴⁴ <https://www.npr.org/2023/08/26/1195662267/ai-is-biased-the-white-house-is-working-with-hackers-to-try-to-fix-that>.

Political bias in GAI⁴⁵

It is reported that ChatGPT would be more inclined to write a song celebrating Fidel Castro than Ted Cruz's life⁴⁶. As GAIs are basically designed to spit out cogent phrases and not actual facts, they evidently emulate human biases of race, gender, religion and class⁴⁷. In their paper, Bender et al. (2023) provide examples of how LLMs can be manipulated to produce biased or harmful outputs, such as stereotyping, discriminating, or excluding certain groups or individuals⁴⁸.

***GAI copies without authors attribution, not necessarily a copyright infringement, this has ethical questionings?*⁴⁹**

Fair use as supported by Professor Lemly⁵⁰ (2023) or the EU Text Mining Directive exception could justify the GAI datasets. It's unsure how the AI training scrapping publicly available work is different from a human being learning and being inspired by existing work of art and literature? New class actions that are spilling will clarify the courts position. Authors and artists have been vocal against the unauthorised use of their work to train GAI. What compensation could be granted to the authors for their work?

***GAI is disrupting the workplace for the good or the bad?*⁵¹**

From the issue of unethical, underpaid exploited labours working behind the scenes to European companies firing their employees to be replaced by machines, the emergence of GAI is disrupting the workplace⁵². Who is going to be replaced? Skilled technicians such as lawyers or doctors or low skills work forces? Will this improve the work conditions? Marketing and journalism involved with drafting have been the first victims of GAI⁵³. GAIs are increasingly used in the recruitment process, assisting recruiters early in the process to write a job description or job advertisement, or select the best profiles, based on publicly available data such as the career history of people on LinkedIn. Artists can use these tools to push their creative process to unsuspected horizons. In June this

⁴⁵ <https://www.pearltrees.com/t/artificial-intelligence/ai-biases/id62398793>.

⁴⁶ <https://www.politico.com/newsletters/digital-future-daily/2023/02/15/ais-political-bias-problem-00083095>.

⁴⁷ <https://www.msn.com/en-us/news/technology/gpt-4-has-arrived-it-will-blow-chatgpt-out-of-the-water/ar-AA18CBpwb>.

⁴⁸ <https://arxiv.org/abs/2304.13712>.

⁴⁹ <https://www.pearltrees.com/t/artificial-intelligence/ai-ip-claims/id62513396>.

⁵⁰ https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4517702.

⁵¹ <https://www.pearltrees.com/t/artificial-intelligence/ai-in-employment/id69547258>;
<https://www.pearltrees.com/t/artificial-intelligence/ai-social-impact/id69428393>; <https://www.pearltrees.com/t/artificial-intelligence/chatgpt-writings/id71414004>.

⁵² <https://www.pearltrees.com/t/artificial-intelligence/labour-exploitation/id69490858>.

⁵³ <https://www.pearltrees.com/t/artificial-intelligence/chatgpt-writings/id71414004>.

year, VentureBeat announced “The age of generative AI is here: only six months after OpenAI’s ChatGPT burst onto the scene, as many as half the employees of some leading global companies are already using this type of technology in their workflows, and many other companies are rushing to offer new products with GAI built in⁵⁴.

Privacy⁵⁵ and human rights violations⁵⁶

ChatGPT may violate the privacy of human users by collecting, storing, or sharing their personal data without their consent or knowledge. OpenAI recently announced using input data to train ChatGPT. It may use the data to generate personalised content that targets the users’ preferences, interests, or vulnerabilities. It may also expose the data to unauthorised parties who may misuse it for malicious purposes. For example, it may use the data to create fake profiles, impersonate the users, or steal their identity. This can harm the security and dignity of the users and their data.

The limited power of GAI to rectify, delete or provide accuracy

ChatGPT has answered that it was “technically possible to rectify datasets to comply with the obligation of accuracy, but it may not be easy or straightforward”⁵⁷. In response to the Italian data protection authority, OpenAI said it was challenging and complex to rectify the dataset which is a regulatory requirement⁵⁸. “OpenAI’s hunger for data is coming back to bite it”, wrote Heikkilä (2023). The unfathomable training data collected, seems to be making individualising any data like “Finding a needle in a haystack”. “OpenAI is going to find it near-impossible to identify individuals” data and remove it from its models, says Margaret Mitchell, an AI researcher and chief ethics scientist at the startup Hugging Face, who was formerly Google’s AI ethics co-lead”. The Italian data protection authority was particularly clement accepting OpenAI’s confession that they were technically unable to rectify or delete information in their dataset. Becoming more transparent about how they collect users’ data during the post-training phase is not sufficient. They are inaccurate data produced with potential reputational harm that are problematic (2023)⁵⁹.

⁵⁴ <https://venturebeat.com/ai/what-is-generative-artificial-intelligence-ai/>; <https://venturebeat.com/ai/chatgpt-launched-six-months-ago-its-impact-and-fallout-is-just-beginning-the-ai-beat/>; <https://venturebeat.com/ai/mckinsey-says-about-half-of-its-employees-are-using-generative-ai/>; <https://venturebeat.com/ai/wordpress-launches-generative-ai-assistant-to-enhance-content-writing/>.

⁵⁵ <https://www.pearltrees.com/t/artificial-intelligence/ai-privacy-compliance/id69428310>.

⁵⁶ <https://www.pearltrees.com/t/artificial-intelligence/ai-and-human-rights/id69493293>.

⁵⁷ <https://www.linkedin.com/pulse/chatgpt-trouble-tara-taubman-bassirian-llm>.

⁵⁸ <https://engineering.stanford.edu/node/16821/printable/print>.

⁵⁹ <https://www.technologyreview.com/author/melissa-heikkila/>; <https://www.pearltrees.com/t/artificial-intelligence/ai-false-accusations/id68994378>.

GAI poses serious risks that require mitigation⁶⁰

Looking at the use of LLMs for Illicit Purposes: Threats, Prevention Measures, and Vulnerabilities. A paper authored by Mozes et al. (2023) explores the ethical and security implications of LLMs. The paper identifies various threats that arise such as fraud, impersonation, malware, and misinformation. It also discusses the potential impacts of these threats on individuals and society, such as loss of trust, privacy, and security⁶¹.

Evaluating the Social Impact of Generative AI Systems and Society⁶²

Irene Solaiman et al. (2023) propose a standard approach for evaluating the social impact of GAI systems. They define more in depth seven categories of social impact for a base system: bias, stereotypes, and representational harms; cultural values and sensitive content; disparate performance; privacy and data protection; financial costs; environmental costs; data and content moderation labour costs⁶³.

In “The ethics of ChatGPT – Exploring the ethical issues of an emerging technology”, Stahl and Eke (2023) discuss the ethical principles and values that should guide the development and use of ChatGPT, such as fairness, transparency, privacy, trust, human dignity, and social good⁶⁴.

Finally, looking at LeCun’s position on LLMs, he acknowledges that they are useful as writing aids although not reliable, factual, or controllable. They are “reactive” and do not plan nor reason. They make stuff up or retrieve stuff approximately, and that this can be mitigated but not fixed by human feedback⁶⁵. Only a small superficial portion of human knowledge can ever be captured by LLMs as human knowledge is not limited to language⁶⁶. He suggests better systems will be based on different principles and will be factual, non-toxic, and controllable⁶⁷. He is not very optimistic on the future of LLMs as they are Transformer-based⁶⁸.

⁶⁰ <https://www.pearltrees.com/t/artificial-intelligence/ai-risks/id69428165>.

⁶¹ <https://www.pearltrees.com/t/artificial-intelligence/ai-security/id68952301>.

⁶² <https://www.pearltrees.com/t/artificial-intelligence/ai-social-impact/id69428393>.

⁶³ <https://arxiv.org/abs/2306.05949>.

⁶⁴ <https://www.ethicsdialogues.eu/2023/09/13/the-ethics-of-chatgpt-exploring-the-ethical-issues-of-an-emerging-technology/>.

⁶⁵ <https://twitter.com/ylecun/status/1610367976016064513>; <https://futurist.com/2023/02/13/metas-yann-lecun-thoughts-large-language-models-llms/>.

⁶⁶ <https://twitter.com/ylecun/status/1610367976016064513>; <https://futurist.com/2023/02/13/metas-yann-lecun-thoughts-large-language-models-llms/>.

⁶⁷ <https://twitter.com/ylecun/status/1610367976016064513>; <https://futurist.com/2023/02/13/metas-yann-lecun-thoughts-large-language-models-llms/>.

⁶⁸ <https://twitter.com/ylecun/status/1618387537848078337>; <https://twitter.com/ylecun/status/1610367976016064513>; <https://futurist.com/2023/02/13/metas-yann-lecun-thoughts-large-language-models-llms/>; <https://twitter.com/ylecun/status/1618387537848078337>.

It is observed that ChatGPT’s behaviour has been changing over time. Between March 2023 and June 2023 versions of ChatGPT, they found that the performance and behaviour of ChatGPT can vary greatly over time. As the behaviour of ChatGPT has changed substantially in a relatively short amount of time, could we hope for a better chatbot?⁶⁹ Or, is “ChatGPT: More than a “Weapon of Mass Deception” Ethical Challenges and Responses from the Human-Centered Artificial Intelligence (HCAI) Perspective”? This article suggests some ways to prevent or reduce ChatGPT misuse or abuse and how to use it in a good way. Some of these ways are technical, such as adding watermarks, changing styles, detecting fakes, and checking facts. Others are non-technical, such as setting rules, being transparent, educating users, and involving humans. There is certainly a need to educate users. Simply banning the use of a tool that is so widely available is not a viable option⁷⁰. Without appropriate measures, “ChatGPT isn’t a great leap forward, it’s an expensive deal with the devil intelligence” (2023)⁷¹.

*How will the environment survive GAI?*⁷²

The Environmental impacts of LLMs affect at first energy and water consumption. To that, adds the cost of building and maintaining data centres requiring large amounts of land, materials, and resources that can have negative effects on the natural environment and local communities. Data centres produce huge amounts of electronic waste that can contain toxic substances and pose health along with environmental risks. Furthermore, LLMs can affect biodiversity by reducing the demand for natural language diversity and endangering linguistic and cultural diversity⁷³. Comparing the environmental cost of LLMs with Google search, by looking at the amount of energy and carbon emissions they consume per query or per day, according to a study by researchers from the University of Bristol and the University of Massachusetts Amherst, the average energy consumption of a Google search query in 2022 was 0.2 watt-hours, which translates to 0.1 grams of carbon dioxide emissions⁷⁴. This means that a single Google search query has a negligible environmental impact, but when multiplied by billions of queries per day, it adds up to a significant amount. The

⁶⁹ <https://arxiv.org/abs/2307.09009>.

⁷⁰ <https://arxiv.org/abs/2304.11215>.

⁷¹ <https://www.theguardian.com/commentisfree/2023/feb/04/chatgpt-isnt-a-great-leap-forward-its-an-expensive-deal-with-the-devil>.

⁷² <https://www.pearltrees.com/t/artificial-intelligence/environmental-impacts/id68975514>; <https://www.pearltrees.com/t/artificial-intelligence/new-data-centres/id71828910>.

⁷³ <https://interestingengineering.com/science/llms-like-gpt-and-bard-can-be-manipulated-and-hypnotized/>; <https://securityintelligence.com/posts/unmasking-hypnotized-ai-hidden-risks-large-language-models/>.

⁷⁴ <https://www.technologyreview.com/2022/11/14/1063192/were-getting-a-better-idea-of-ais-true-carbon-footprint/>.

study estimated that a single LLM query has a much higher environmental impact than a Google search query, but when multiplied by millions of queries per day, it becomes even more substantial. The AI Index Report 2023 estimated that ChatGPT consumed about 100 megawatt-hours of energy and emitted about 50 metric tons of carbon dioxide per day in 2023⁷⁵. Therefore, based on these estimates, the environmental cost of LLMs is about 50 times higher than the cost of Google search per query. However, these estimates do not consider the environmental cost of training LLMs, which can be much higher than the cost of running them. Professor Crawford who describes AI as a “technology of extraction”, shares the same concerns. In a recent interview she exposed that “indicating that every time you have an exchange with ChatGPT, it’s the equivalent of pouring out half a litre of fresh water onto the ground” because that’s what it takes to “cool the giant AI supercomputers” involved. “The energy difference from just doing a traditional search query to using a LLM is enormous,” she says. “Some research indicates it can be up to 1,000 times more energy intensive.” “The question of the environmental cost of AI is the biggest secret in the industry right now,” Crawford explains. “All along the pipeline – the hardware, the software, the energy, the water to cool the systems – we have enormous environmental costs that are not being fully shared with the public”⁷⁶. As GAIs are thirsty, “Integrating large language models into search engines could mean a fivefold increase in computing power and huge carbon emissions”⁷⁷.

ChatGPT “drinks” a bottle of fresh water for every 20 to 50 questions we ask, another study warns⁷⁸. The water consumption keeps growing. “A.I. tools fuelled a 34% spike in Microsoft’s water consumption, and one city with its data centres is concerned about the effect on residential supply” as published in Fortune Magazine⁷⁹. The environmental impact requires urgent intervention to reduce the size of datasets while improving the quality of the data, and the building of sustainable data centres.

Conclusion

In conclusion, we believe there is an urgent need for more transparency, accountability, and regulation. The large-scale impact of GAI, that may require to be distinguished from other Artificial Intelligence, calls for a dedicated framework taking into account the cost of its implementation balanced with the added value of its outcomes, to ensure companies and others deploy GAI safely, ethi-

⁷⁵ <https://www.analyticsvidhya.com/blog/2023/04/environmental-cost-of-ai-models-carbon-emissions-and-water-consumption/>.

⁷⁶ <https://www.linkedin.com/in/taubmanbassirian/recent-activity/all/>.

⁷⁷ <https://arxiv.org/abs/2304.03271>.

⁷⁸ <https://www.euronews.com/green/2023/04/20/chatgpt-drinks-a-bottle-of-fresh-water-for-every-20-to-50-questions-we-ask-study-warns>.

⁷⁹ <https://fortune.com/2023/09/09/ai-chatgpt-usage-fuels-spike-in-microsoft-water-consumption/>.

cally and in a trustworthy manner. After China, a European regulation is currently being adopted. The AI Act will have to be carefully drafted in order not to block innovation while at the same time preserving the rights and freedoms of all parties. The risk-based approach will have to survive the fast pace of technological evolutions. As always, education and awareness will have a major role to play. Another constant challenge in the interconnected digital world will be the borderless impact of any regulation. How will different parts of the world regulate AI. A good coordination is key to success that might require an international committee similar to the International Atomic Energy Agency (IAEA) cooperation. The 2015 United 2030 agenda for sustainable development proposed objectives to design and implement a worldwide safe and sustainable future. Between its 17 Sustainable Development Goals, “industry, innovation and infrastructure”, the UN established the *Technology Facilitation Mechanism* (TFM) to promote innovative solutions for the SDG agenda, including multi-stakeholder collaboration.

AI sustainability will depend much on its use of resources and respect of environmental impacts.

References

- <http://www.garymarcus.com/index.html> (14.05.2023).
- <https://apnews.com/article/artificial-intelligence-chatgpt-courts-e15023d7e6fdf4f099aa122437dbb59b#lneqessb42o50gafue4> (8.10.2023).
- <https://arstechnica.com/information-technology/2023/09/can-you-melt-eggs-quoras-ai-says-yes-and-google-is-sharing-the-result/> (1.10.2023).
- <https://arxiv.org/abs/2304.03271> (8.04.2023).
- <https://arxiv.org/abs/2304.11215> (5.05.2023).
- <https://arxiv.org/abs/2304.13712> (25.05.2023).
- <https://arxiv.org/abs/2305.18248> (15.09.2023).
- <https://arxiv.org/abs/2306.05949> (12.06.2023).
- <https://arxiv.org/abs/2307.09009> (15.07.2023).
- <https://arxiv.org/abs/2309.03883> (10.09.2023).
- <https://arxiv.org/abs/2309.05463> (10.06.2023).
- <https://arxiv.org/abs/2309.11495>. DoLa: Decoding by Contrasting Layers Improves Factuality in Large Language Models (15.06.2023).
- <https://arxiv.org/pdf/2212.03551.pdf> (13.07.2023).
- <https://bing.com/search?q=Noam+Chomsky+on+LLMs&form=SKPBOT> (15.09.2023).
- <https://browse.arxiv.org/pdf/2304.00008.pdf> (14.05.2023).
- <https://cyber.harvard.edu/people/dweinberger> (12.07.2023).
- <https://dataethics.eu/testing-chatgpts-ethical-readiness/> (15.06.2023).
- <https://dweinberger.medium.com/chatgpt-on-why-it-pretends-to-know-things-ea2503ee872> (19.06.2023).
- https://en.wikipedia.org/wiki/Wikipedia:Large_language_models (20.04.2023).
- <https://engineering.stanford.edu/node/16821/printable/print> (16.07.2023).
- <https://ethicspolicy.unc.edu/news/2023/04/17/the-ethics-of-college-students-using-chatgpt> (26.05.2023).
- <https://fchollet.com/> (12.05.2023).

<https://fortune.com/2023/08/01/can-ai-chatgpt-hallucinations-be-fixed-experts-doubt-altman-openai/> (12.10.2023).

<https://fortune.com/2023/09/09/ai-chatgpt-usage-fuels-spike-in-microsoft-water-consumption/> (15.09.2023).

<https://futurist.com/2023/02/13/metas-yann-lecun-thoughts-large-language-models-llms/> (22.03.2023).

<https://futurist.com/2023/02/13/metas-yann-lecun-thoughts-large-language-models-llms/> (5.04.2023).

<https://hanne-sbajohr.de/en/2023/04/08/whoever-controls-language-models-controls-politics/> (15.04.2023).

<https://interestingengineering.com/science/llms-like-gpt-and-bard-can-be-manipulated-and-hypnotized> (11.10.2023).

<https://medium.com/@paul.k.pallaghy/the-entire-field-of-ai-is-being-professionally-gaslighted-by-gary-marcus-and-noam-chomsky-c08aa1e4c6f0> (14.06.2023).

<https://montreal.ethics.ai/oppenheimer-as-a-timely-warning-to-the-ai-community/> (25.05.2023).

<https://multilingual.com/timnit-gebru-and-the-problem-with-large-language-models/> (5.06.2023).

<https://multilingual.com/timnit-gebru-and-the-problem-with-large-language-models/> (18.05.2023).

<https://news.berkeley.edu/2023/03/19/is-chatgpt-a-false-promise> (20.03.2023).

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4453869 (19.06.2023).

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4517702 (8.08.2023).

<https://sebastianraschka.com/blog/2023/optimizing-LLMs-dataset-perspective.html> (15.04.2023).

<https://securityintelligence.com/posts/unmasking-hypnotized-ai-hidden-risks-large-language-models/> (12.09.2023).

<https://techcrunch.com/2023/06/07/blush-ai-dating-sim-replika-sexbot/> (9.06.2023).

<https://towardsdatascience.com/where-are-all-the-women-3c79dabdfc2> (28.08.2023).

<https://twitter.com/ylecun/status/1610367976016064513> (5.06.2023).

<https://twitter.com/ylecun/status/1618387537848078337> (5.06.2023).

<https://undark.org/2023/04/06/chatgpt-isnt-hallucinating-its-bullshitting/> (6.06.2023).

<https://venturebeat.com/ai/chatgpt-launched-six-months-ago-its-impact-and-fallout-is-just-beginning-the-ai-beat> (5.06.2023).

<https://venturebeat.com/ai/mckinsey-says-about-half-of-its-employees-are-using-generative-ai> (13.06.2023).

<https://venturebeat.com/ai/the-ai-feedback-loop-researchers-warn-of-model-collapse-as-ai-trains-on-ai-generated-content/> (12.07.2023).

<https://venturebeat.com/ai/what-is-generative-artificial-intelligence-ai> (25.06.2023).

<https://venturebeat.com/ai/wordpress-launches-generative-ai-assistant-to-enhance-content-writing/> (11.06.2023).

<https://walton.uark.edu/insights/posts/the-human-need-for-ethical-guidelines-around-chatgpt.php> (17.07.2023).

<https://walton.uark.edu/insights/posts/the-human-need-for-ethical-guidelines-around-chatgpt.php> (25.05.2023).

<https://www.analyticsvidhya.com/blog/2023/04/environmental-cost-of-ai-models-carbon-emissions-and-water-consumption/> (9.09.2023).

<https://www.bbc.co.uk/newsround/66796495> (1.10.2023).

<https://www.cloudflare.com/learning/ai/what-is-large-language-model/> (13.08.2023).

<https://www.ethicsdialogues.eu/2023/09/13/the-ethics-of-chatgpt-exploring-the-ethical-issues-of-an-emerging-technology/> (16.09.2023).

<https://www.euronews.com/green/2023/04/20/chatgpt-drinks-a-bottle-of-fresh-water-for-every-20-to-50-questions-we-ask-study-warns> (15.09.2023).

<https://www.firstpost.com/world/ai-bots-capable-of-starting-new-religions-warns-yuval-noah-harari-12540282.html> (3.06.2023).

<https://www.forbes.com/sites/forbestechcouncil/2023/06/30/10-ways-cybercriminals-can-abuse-large-language-models/> (6.07.2023).

<https://www.foxnews.com/media/snapchat-ai-chatbot-gave-advice-13-year-old-girl-relationship-31-year-old-man-having-sex> (16.06.2023).

<https://www.lesswrong.com/posts/sbaQv8zmRncpmLNKv/the-idea-that-chatgpt-is-simply-predicting-the-next-word-is> (19.06.2023).

<https://www.linkedin.com/in/taubmanbassirian/recent-activity/all/> (9.10.2023).

<https://www.linkedin.com/pulse/chatgpt-could-capable-better-reasoning-languages-tara/> (6.05.2023).

<https://www.linkedin.com/pulse/chatgpt-trouble-tara-taubman-bassirian-llm> (6.05.2023).

<https://www.msn.com/en-us/news/technology/gpt-4-has-arrived-it-will-blow-chatgpt-out-of-the-water/ar-AA18CBpwb> (25.09.2023).

<https://www.npr.org/2023/08/26/1195662267/ai-is-biased-the-white-house-is-working-with-hackers-to-try-to-fix-that> (1.09.2023).

<https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html> (15.04.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-and-human-rights/id69493293> (8.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-biases/id62398793> (9.09.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-definitions/id62876503> (12.05.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-false-accusations/id68994378> (2.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-in-employment/id69547258> (25.09.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-ip-claims/id62513396> (25.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-manipulations/id68952637> (20.09.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-misinformation/id69147099> (15.07.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-privacy-compliance/id69428310> (8.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-risks/id69428165> (5.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-security/id68952301> (6.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/ai-social-impact/id69428393> (1.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/anthropomorphisation/id71886740> (15.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/call-for-ban-moratory/id65034697> (3.06.2023).

<https://www.pearltrees.com/t/artificial-intelligence/call-for-ban-moratory/id650> (6.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/chatgpt-alternatives/id62359814> (20.05.2023).

<https://www.pearltrees.com/t/artificial-intelligence/chatgpt-incident/id71414555> (17.09.2023).

<https://www.pearltrees.com/t/artificial-intelligence/chatgpt-writings/id71414004> (1.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/environmental-impacts/id68975514> (5.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/labour-exploitation/id69490858> (25.10.2023).

<https://www.pearltrees.com/t/artificial-intelligence/mistral/id71509507> (18.09.2023).

<https://www.pearltrees.com/t/artificial-intelligence/new-data-centres/id71828910> (7.10.2023).

<https://www.politico.com/newsletters/digital-future-daily/2023/02/15/ais-political-bias-problem-00083095> (13.03.2023).

<https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/> (20.05.2023).

<https://www.technologyreview.com/2022/11/14/1063192/were-getting-a-better-idea-of-ais-true-carbon-footprint/> (10.09.2023).

<https://www.technologyreview.com/2022/12/23/1065852/whats-next-for-ai/> (9.05.2023).

<https://www.technologyreview.com/author/melissa-heikkila/> (6.10.2023).

<https://www.telegraph.co.uk/business/2023/03/29/control-ai-threat-civilisation-warns-elon-musk/> (1.10.2023).

<https://www.telegraph.co.uk/news/2023/04/23/yuval-noah-harari-i-dont-know-if-humans-can-survive-ai/> (15.05.2023).

<https://www.theguardian.com/commentisfree/2023/feb/04/chatgpt-isnt-a-great-leap-forward-its-an-expensive-deal-with-the-devil> (12.09.2023).

<https://www.thesun.ie/tech/10808612/snapchat-artificial-intelligence-bot-danger-children/> (14.07.2023).

<https://www.thetimes.co.uk/article/why-2024s-presidential-race-will-be-the-first-ai-election-jb32pj8br> (17.08.2023).
<https://www.verywellmind.com/jean-piaget-biography-1896-1980-2795549> (17.05.2023).
<https://www.weforum.org/whitepapers/jobs-of-tomorrow-large-language-models-and-jobs> (1.06.2023).
<https://www.wired.com/story/fast-forward-chatbot-hallucinations-are-poisoning-web-search/> (7.10.2023).
<https://www.wired.com/story/fast-forward-chatbot-hallucinations-are-poisoning-web-search/> (8.10.2023).
https://www3.weforum.org/docs/WEF_Jobs_of_Tomorrow_Generative_AI_2023.pdf (1.06.2023).