INTERNATIONALE HOCHSCHULE

www.iu.de

IU DISCUSSION PAPERS

IT & Engineering

ChatGPT – A Critical View

FLORIAN ALLWEIN

IU Internationale Hochschule

Main Campus: Erfurt

Juri-Gagarin-Ring 152

99084 Erfurt

Telefon: +49 421.166985.23

Fax: +49 2224.9605.115

Kontakt/Contact: kerstin.janson@iu.org

Autorenkontakt/Contact to the author(s):
Florian Allwein

IU Internationale Hochschule, Juri-Gagarin-Ring 152, D-99084 Erfurt, florian.allwein@iu.org, https://orcid.org/0000-0002-2831-7259

IU Internationale Hochschule - Campus Berlin

Frankfurter Allee 73a

10247 Berlin

Telefon: +49- 30208986810

Email: florian.allwein@iu.org

IU Discussion Papers, Reihe: IT & Engineering, Vol. 5, No. 1 (MAR 2024)

ISSN-Nummer: 2750-073X

Website: https://www.iu.de/forschung/publikationen/

ChatGPT – A Critical View

Florian Allwein

ABSTRACT:

ChatGPT, a chatbot based on a Large Language Model, has become one of the fastest-growing consumer software applications in history. Discussion of the tool and its use cases contains an element of hype as it appears that the technology's capabilities are sometimes exaggerated. Moreover, a critical perspective is often lacking in research and practice. This paper points out some significant downsides, risks and limitations of using ChatGPT, arguing for a critical view of the tool based on ethics, regulations and reflected use. This can be used as a guideline for decisions on whether and how to use ChatGPT, and can inform future research.

KEYWORDS:

ChatGPT, AI, Ethics

AUTHOR:

Prof. Dr. Florian Allwein *is Professor for Digital Transformation at IU. Besides an M.A. in Literary Theory from Ludwig-Maximilians-Universität Munich, he holds an MSc and Ph.D. in Management, Information Systems and Innovation from London School of Economics and Political Science (LSE). His research and teaching covers topics in Digitalisation, Knowledge Management and Information Systems.*

Introduction

ChatGPT (Generative Pre-trained Transformer), a chatbot developed by OpenAI, has become one of the fastest-growing consumer software applications in history. Based on a probabilistic large language model (LLM, currently GPT-4), it can generate general-purpose language, allowing users to have conversations with the bot. It has caused a new boom in AI, reflected in media coverage, investments and reporting on the tool in general and business media. This boom, however, contains an element of hype as it appears that the technology's capabilities are exaggerated in some parts (Hiltzik, 2023).

To counter the hype, this paper discusses the question of "What are the issues involved in using ChatGPT, and how can they be mitigated?". Based on recent cases and some academic literature, it illustrates downsides, risks and limitations that have been described with regards to using ChatGPT and proposes mitigating these by a combination of ethics, regulation and reflected use of the tool. Thus, it can help individuals and organizations make informed and ethical decisions about using or not using it. Finally, some opportunities for further research are discussed.

Background

ChatGPT can often create convincing output, so users are likely to return and to recommend it to others. This has raised hopes regarding future use cases that can be supported or taken over by ChatGPT and similar tools. Inevitably, it has also raised the number of sales pitches, as Hanff (2023) points out:

> *Marketers, influencers, and a host of "leadership" coaches, copy writers, and content creators are all over social media telling everyone how much time and money they can save using ChatGPT and similar models to do their work for them - ChatGPT has become the new Grumpy Cat, the new Ice Bucket Challenge - it has become the focus of just about every single industry on the planet.*

Such instances of hype, or promotion consisting of exaggerated claims, have been documented to occur frequently around new technologies. In fact, the concept of hype cycles is well known and researched (see Dedehayir & Steinert, 2016). Gardener's current hype cycle study shows Generative AI at the top of the "peak of inflated expectations", meaning tools like ChatGPT would now go down to the "trough of disillusionment" (Perri, 2023) before a future, productive use develops, (Figure 1). Specifically for the area of artificial intelligence (AI), the concept of AI winters has been used to describe recurring instances when a hype cycle dies down as it becomes clear that the expectations in the technology were too high. This has happened both in the 1970s and 1980s/90s (see e.g. Lim, 2018).

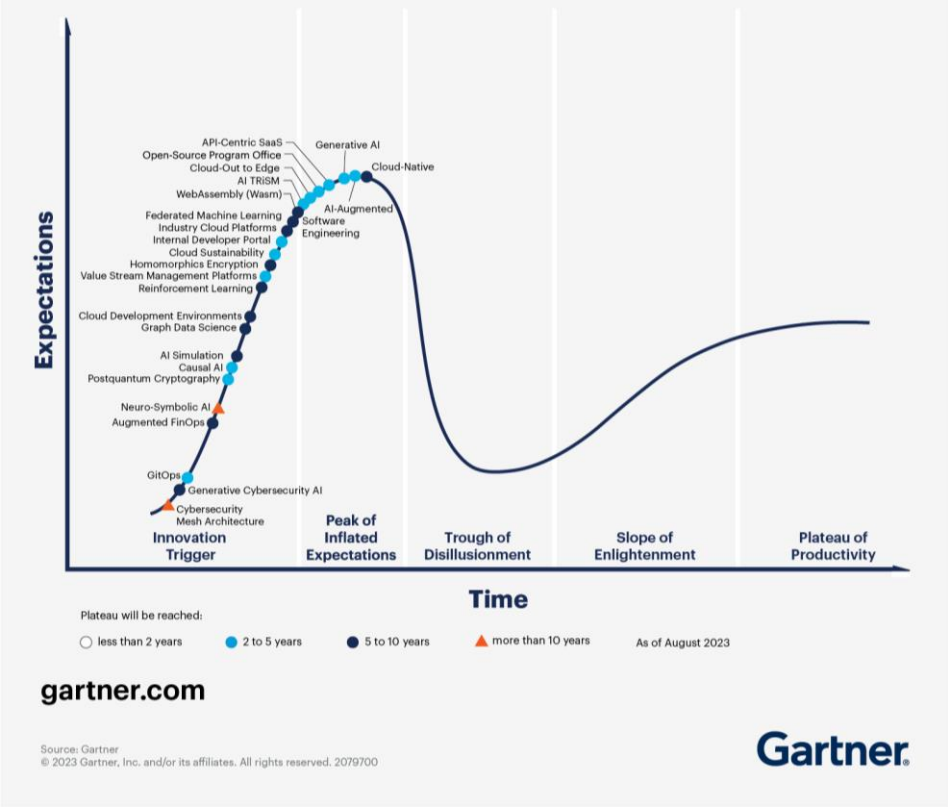

Figure 1 Hype Cycle for Emerging Technologies, 2023 (Perri, 2023)

One example for an exaggerated perception of ChatGPT's capabilities is the paper by Dell'Acqua et al. (2023). It seems to prove that ChatGPT leads to increased productivity and quality of results, as

illustrated in Figure 2. A superficial look at the results seems to show that consultants using ChatGPT in their job complete more tasks at higher quality, which fits into the exaggerated expectations around the tool's capabilities.

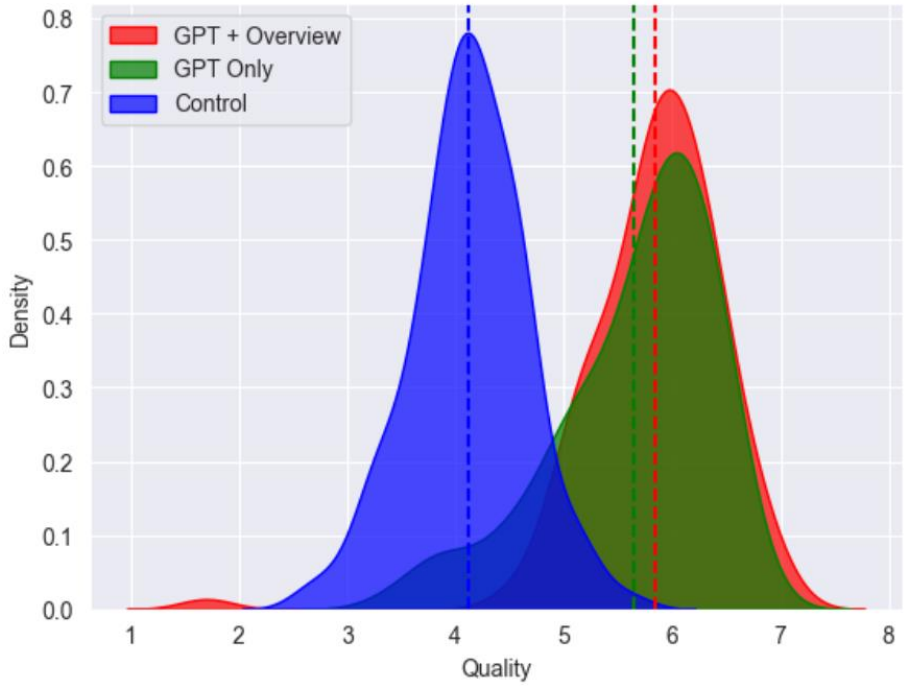

Figure 2 Performance Distribution - Inside the Frontier (Dell'Acqua et al. 2023)

This is, however, a limited view of the argument actually made by Dell'Acqua et al. (2023) – The point of the paper is that such increases are only possible for some tasks and that it is not easy to predict which tasks these might be (see Figure 3):

> *„Our results demonstrate that AI capabilities cover an expanding, but uneven, set of knowledge work we call a "jagged technological frontier." Within this growing frontier, AI can complement or even displace human work; outside of the frontier, AI output is inaccurate, less useful, and degrades human performance. However, because the capabilities of AI are rapidly evolving and poorly under-stood, it can be hard for professionals to grasp exactly what the boundary of this frontier might be" (Dell'Acqua et al. 2023, p. 1)*

This shows that it is important to critically analyse the results of such studies and to consider where the tool can and cannot be useful in a given context. For example, ChatGPT finds it very difficult to adequately deal with questions where the correct answer changes with time or with region, e.g. legal questions.

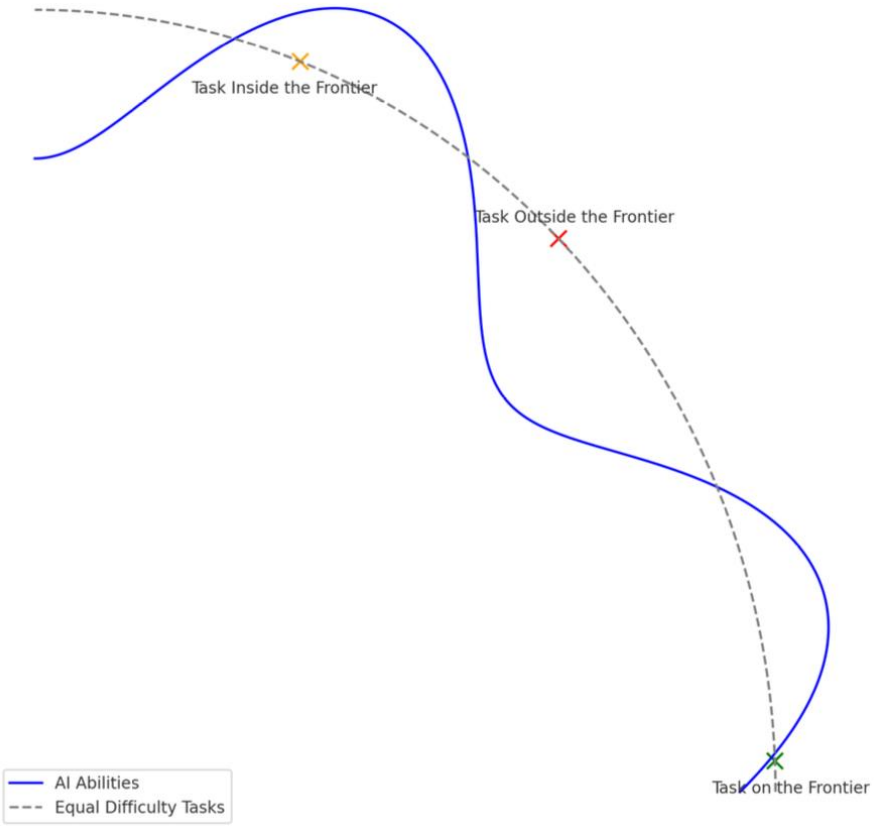

Figure 3 Jagged Frontier of AI Capabilities (Dell'Acqua et al. 2023)

Downsides of ChatGPT

To address the question of "What are the issues involved in using ChatGPT, and how can they be mitigated?", this paper will outline negative aspects often ignored in the use and discussion of ChatGPT. Firstly, using the tool carries significant downsides or negative aspects, from security violations to energy usage and ethical concerns. These are widely acknowledged but seem to have no effect on the discourse.

LEAKING CONFIDENTIAL DATA

Issues around the leaking of confidential data by ChatGPT have been reported soon after the launch of the tool. One study found that one third of respondents were using ChatGPT and similar tools at work without their companies' approval (Bird, 2023). In the process, they often fed confidential documents into the tool, which initially added them to the corpus, so that they could be used for output for other users (DeGeurin, 2023; Vincent, 2023). This can now be mitigated by disabling the chat history in ChatGPT.

A more salient issue is exactly which data has been crawled by ChatGPT, and whether its use of this data is legal. In a case that seems relatively clear-cut, Getty Images is currently suing Stable AI, whose tool Stable Diffusion is using AI to generate images, as some of its output clearly included watermarks by Getty. Similar issues may well be present in ChatGPT (The Economist, 2023a), although spotting them might be more difficult, since texts from the Internet may not be watermarked, and are easier to modify, making detection more difficult. This issue is similar to that of Google crawling and indexing the web

for display in its search results. This has largely been addressed using mechanisms like robots.txt files, which indicate which parts of a website can be crawled. However, since ChatGPT's usage of content is significantly different, it is unclear whether it would be legal even if it did consider limitations set up in robots.txt files.

ENERGY USAGE

Moreover, it has been shown that the energy usage of ChatGPT is very significant. Factoring in the computing and energy used for training the tool, one analysis shows that "for each query, the energy consumed is = (1064+260.42) MWh / 195 [million] = 6.79 Wh. Or in simpler terms, each query on ChatGPT consumes the equivalent amount of energy of running a 5W LED bulb for 1hr 20 min" (Zodhya, 2023). While the details and the methodology are up for discussion, it can be assumed that the order of magnitude is correct. Conversely, a Google search is said to consume the equivalent amount of energy of running the same bulb for only 3 minutes (Zodhya, 2023), meaning the ChatGPT query takes approximately 27 times as much energy.

EXPLOITING WORKERS IN THE GLOBAL SOUTH

Finally, it must be pointed out that ChatGPT relies on the labour of low-paid workers in the global South, who are instrumental in labelling data (Perrigo, 2023). As the texts reviewed by these human workers were sourced from the Internet, they often contain descriptions of "graphic sexual violence", as described in a current petition to the Kenyan government calling for an investigation into the practices of OpenAI and its associates (see Rowe, 2023).

While none of the downsides presented so far seem to have an impact on the tool's popularity, it must be noted that they constitute serious ethical dilemmas. Consequently, any use of ChatGPT constitutes an ethical decision to accept these downsides. Individuals and organisations should be aware of them and consider whether their specific use case is justified given the downsides.

Risks of ChatGPT

Beyond these downsides, there has been much publicity around the perceived or real risks associated with the use of ChatGPT. Risks are seen as situations involving exposure to danger. Shortly after the launch of ChatGPT, an open letter was published by the so-called Future of Life Institute, calling for a six-month hiatus in the development of similar tools. The letter openly named significant concerns:

> *Contemporary AI systems are now becoming human-competitive at general tasks, and we must ask ourselves: Should we let machines flood our information channels with propaganda and untruth? Should we automate away all the jobs, including the fulfilling ones? Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization? Such decisions must not be delegated to unelected tech leaders. Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable.*
>
> (Future of Life Institute, 2023)

It does name risks which are often discussed in public or the media, in particular the loss of jobs. So far, however, there has not been any evidence supporting this concern. One early study finds that Generative AI is more likely to augment rather than destroy jobs (Gmyrek et al., 2023). Likewise, while automatically generated misinformation is a concern, the risk of loss of control over civilisation seems vague and overstated. Moreover, it was surprising that the letter was signed by some individuals who are themselves involved in the development of AI (including Elon Musk).

Conversely, it has been argued that this open letter is more of a PR stunt, creating hype and then downplaying its risk as a way of getting media attention (Singh, 2023), and that it hides the real risks associated with ChatGPT:

> *We agree that misinformation, impact on labor, and safety are three of the main risks of AI. Unfortunately, in each case, the letter presents a speculative, futuristic risk, ignoring the version of the problem that is already harming people. It distracts from the real issues and makes it harder to address them. The letter has a containment mindset analogous to nuclear risk, but that's a poor fit for AI. It plays right into the hands of the companies it seeks to regulate.*
>
> (Kapoor & Narayanan, 2023)

Kapoor & Narayanan (2023) instead propose focussing on what they consider the real risks of the technology, namely Overreliance on inaccurate tools, centralized power, labour exploitation and near-term security risks, e.g. hacking chatbots to reveal their users' personal data (see Table 1).

	Speculative risks	Real risks
Misinformation	Malicious disinformation	Overreliance on inaccurate tools
Labor impact	LLMs will replace all jobs	Centralized power, labor exploitation
Safety	Long-term existential risks	Near-term security risks

Table 1 Speculative and real risks (Kapoor & Narayanan, 2023)

An early academic paper dealing with the risks of LLMs and tools using them is by Bender et al. (2021). They name environmental and financial costs as well as limitations in the training data as possible risks and argue for a more balanced approach, "investing resources into curating and carefully documenting datasets rather than ingesting everything on the web" (p. 610).

This shows that ChatGPT is often perceived as a risky technology, even though the risks themselves cannot be fully specified, let alone quantified, yet. What seems clear is that individuals and society have a responsibility to identify and address these risks.

Limitations of ChatGPT

Beyond the downsides and risks discussed so far, the most important factor in ChatGPT is the limitations or restrictions inherent in the technology – as we have seen in the discussion of the Dell'Acqua et al. (2023) paper, it can do some things but not others.

HOW LLMs WORK

In this context, it is important to recall how ChatGPT, and LLMs in general, work. The introduction by (The Economist, 2023b) seems to be concise and accessible to general readers. The authors describe how language is converted from words into tokens, a set of numbers. These are then placed in a "meaning space", where similar words are located close to each other. Over the course of its training, an LLM learns which words are commonly used together. Consequently, it understands language only in a limited way: "If it understands language at all, an LLM only does so in a statistical, rather than a grammatical, way. It is much more like an abacus than it is like a mind" (The Economist, 2023b). This becomes obvious when output is generated:

> *"Once the prompt has been processed, the LLM initiates a response. At this point, for each of the tokens in the model's vocabulary, the attention network has produced a probability of that token being the most appropriate one to use next in the sentence it is generating.*
>
> (The Economist, 2023b)

So, while the output is not entirely predictable and can produce surprising results, it does become obvious that the process is fundamentally different from human creativity.

NO FEELINGS OR INTELLIGENCE

Indeed, an early criticism of ChatGPT that received significant publicity is just about this lack of creativity: Musician Nick Cave opposed to people generating "lyrics" in his style using the tool:

> *Songs arise out of suffering, by which I mean they are predicated upon the complex, internal human struggle of creation and, well, as far as I know, algorithms don't feel. Data doesn't suffer.*
>
> *ChatGPT has no inner being, it has been nowhere, it has endured nothing, it has not had the audacity to reach beyond its limitations, and hence it doesn't have the capacity for a shared transcendent experience, as it has no limitations from which to transcend.*
>
> (Cave, 2023)

While few people would disagree with this sentiment, it still did not keep people from generating texts in the style of various authors. This example, however, is a good starting point for discussing the intelligence and limitations of ChatGPT.

ChatGPT is sometimes associated with Artificial General Intelligence, i.e. hypothetical future tools that could accomplish any tasks humans can perform (see e.g. Bubeck et al., 2023). Most prominently, Sam Altman, one of the founders of Open AI, has stated that it is a step in this direction (Shah, 2023).

Even acknowledging that researchers of psychology are having difficulties defining human intelligence, it seems safe to say that ChatGPT's abilities are far from it. In a psychology textbook, Myers & DeWall (2021, p. 728) acknowledge the definition of intelligence would depend on the cultural context of an individual, before venturing that, in general, "intelligence is the ability to learn from experience, solve problems, and use knowledge to adapt to new situations". For AI, Russell & Norvig (2021, p. 52) postulate that "AI is concerned mainly with rational action. An ideal intelligent agent takes the best possible action in a situation."

Conversely, a blogpost looking at research on ChatGPT's intelligence concludes that it performs poorly on the categories of abstraction, memory, reasoning and inference, and planning, concluding that "major pieces of the puzzle are still missing before the flexibility required for intelligence can emerge in machines" (Matar, 2023) and that "The 'trick' used by LLMs is the unfathomable amounts of texts they are trained on, allowing them to come up with reasonable answers for many queries. But when tested in uncharted territory, their abilities dissipate" (Matar, 2023). This has recently been substantiated by a research paper by Google's DeepMind laboratory claiming that LLMs perform very well "when the task families are well-represented in their pretraining data" but not when "presented with tasks or functions which are out-of-domain of their pretraining data" (Yadlowsky et al., 2023, p. 1).

LIMITATIONS TO GROWTH

Nor is it likely that this will change significantly in the future: Whereas information technology is often associated with long periods of exponential growth in capabilities like processing power, as described by Moore's Law (see e.g. Kurzweil, 2005), such exponential growth is unlikely to occur with ChatGPT's capabilities. Instead, it has been claimed that "training LLMs gets expensive faster than it gets better" (The Economist, 2023b). Since "GPT-3 has already been trained on what amounts to all of the high-quality text that is available to download from the internet" (The Economist, 2023b), it can be concluded that "the stock of high-quality language data will be exhausted soon; likely before 2026" (Villalobos et al., 2022).

DATA AND CODE REFLECT VALUES

Another key limitation is the fact that data and code are generally not neutral, but reflect values. In the case of ChatGPT, this refers to the values of both the people who created the tool and the people who created the content it has been trained on, i.e. largely the content of the Internet.

There is much research on the phenomenon of bias inherent in information technology or the data it uses (see Kordzadeh & Ghasemaghaei, 2022). For example, on gender bias, Bryson et al. (2020) conclude that "[a]n artificial intelligence (AI) system operates on — and learns from — the data it's given. And when that data is generated by and collected from humans, it carries all the biases that we do, including bias about women. The result: firms develop technologies that reinforce inequality".

O'Neill showed in 2017 that decisions affecting our lives are increasingly made by algorithms often based on opaque models that tend to reinforce bias. She calls these "weapons of math destruction"

and concludes that "[p]romising efficiency and fairness, they distort higher education, drive up debt, spur mass incarceration, pummel the poor at nearly every juncture, and undermine democracy." (p. 174). With powerful tools like ChatGPT being used more broadly, AI may already contribute to this arsenal.

As an example, bias is already evident in images generated by other AI tools. Turk (2023) analysed images generated by Midjourney using prompts like "an Indian person" or "a Mexican person" and found that "[t]he results show a hugely stereotypical view of the world" and also that, for example, „images associated with higher-paying job titles featured people with lighter skin tones, (…) results for most professional roles were male-dominated."

Due to the way it works and the training data it uses, similar bias can be expected in texts generated by ChatGPT. Beyond bias, there is also the issues of racism, or more generally abuse, and hate speech generated by ChatGPT. This is an inherent issue due to the nature of its training data. Makers of AI tools are aware of this and are trying to mitigate the issue. This is one of the reasons for employing humans to work on the training data. However, so far, it has not been possible to keep this kind of results entirely out. Deshpande et al. (2023) even show how it is possible to get ChatGPT to give more toxic answers by assigning it a persona (e.g. Muhammad Ali or Adolf Hitler).

BULLSHITTING

An issue inherent in the way LLMs work is the fact that they may create results which are factually inaccurate. These are technically known as hallucinations. Another term sometimes used is bullshitting, defined as speech intended to persuade without regard for truth. There is a clear distinction to lies, which the speaker knows to be untrue (Frankfurt, 1988).

Instances of hallucinations or bullshitting serve well to illustrate the limitations of ChatGPT. Hanff (2023) describes a particularly troubling case, where the tool insisted that the author had in fact passed away, providing the following information:

> *The media reports about Alexander Hanff's death did not mention the cause of his death. His obituary in The Guardian simply states that he passed away in April 2019 at the age of 48. Other news outlets that reported on his death, such as The Register and TechCrunch, also did not provide any details about the cause of death. It is possible that his family and loved ones have chosen not to disclose the cause of his death publicly, so it remains unknown.*
>
> ChatGPT output, from Hanff (2023).

As Hanff points out, there are significant risks and disadvantages arising once an information system thinks an individual has passed away. Friends and relatives being shocked to receive the news is even a minor concern, compared to being automatically rejected when applying for a job or a mortgage. Motivated by event, Hanff actuality demands that ChatGPT should be shut down altogether.

CONSERVATIVE TECHNOLOGY

A final interesting point to note is that ChatGPT must be seen as a fundamentally conservative technology. Since its data is naturally based on the past and given the limitations regarding its creativity, it is unclear how well it can generate genuinely new ideas (Geuter, 2023b). In a similar vein, the use of ChatGPT in organisations is often associated not with innovation, but with automation. In supporting traditional goals like higher productivity, it may also weaken the position of human workers. Geuter (2023b) argues that a large part of the tool's appeal lies in this promise of automatization. In this context, the quality of the output is not even that important, what counts is that it is seen as good enough:

> *ChatGPT and other systems do not generate perfect results, with or without guardrails, but the bullshit is good enough in many cases. Of course humans also generate bullshit, but they are generally able to produce better texts and less prone images, and in reality they do. But they're also more expensive.*
>
> (Geuter, 2023a, own translation)

Consequently, a result of using ChatGPT can be a weakening of the position of human workers, as McQuillan (2023) points out:

> *For these players, the seductive vision isn't real AI (whatever that is) but technologies that are good enough to replace human workers or, more importantly, to precaritise them and undermine them.*

Analysis: A Critical View

Returning to the question of "What are the issues involved in using ChatGPT, and how can they be mitigated?", we have seen a number of significant downsides, risks and limitations which can contribute to a critical view of ChatGPT. Whereas Hanff (2023) and others have argued that the use of ChatGPT should be banned outright, this paper argues that a reflected use is still possible if it follows the guiding principles of ethics, regulation and reflected use. This is similar to the recommendations given by Bender et al. (2021).

ETHICS

There is a strong tradition of ethical guidelines around the work of computer scientists. Many big organisations or institutions have already published ethical guidelines on the development and use of generative AI (e.g. ACM, 2023; UNESCO, 2022). Such guidelines were also an important influence on the European Union's AI act, which is currently in the process of legislation.

There is also a significant tradition of research on the ethics of AI, which can help in setting up guidelines for an ethical use of ChatGPT. Most recently, Floridi (2023) proposed a unified framework of five principles for ethical AI, namely:

1. Beneficence: Promoting Well-Being, Preserving Dignity, and Sustaining the Planet
2. Nonmaleficence: Privacy, Security, and 'Capability Caution'

3. Autonomy: The Power to ‘Decide to Decide’
4. Justice: Promoting Prosperity, Preserving Solidarity, Avoiding Unfairness
5. Explicability: Enabling the Other Principles through Intelligibility and Accountability

Together with earlier research in the established field of ethical use of ICT, this framework serves as a good starting point for individuals and organisations weighing the use of tools like ChatGPT. For instance, it could be used as a starting point for formulating more specific guidelines based on the needs of an organization. This is particularly relevant for discussions of the downsides mentioned earlier in this paper. Users should decide for themselves whether their intended use of ChatGPT and similar tools justifies downsides like high energy usage and ethical issues.

REGULATION

This paper has also discussed the risks associated with the use of ChatGPT. While there is disagreement on whether these are broad and existential (replacing all jobs, putting the future of humanity at risk) or more short term (labour exploitation, security risks), it is clear that the risks need to be addressed. The Artificial Intelligence Act currently proposed by the EU (European Commission, 2023), would be a significant step towards mitigating these risks. Regulation will be based on an initial classification of the proposed use case into one of four categories of risk (“unacceptable”, “high”, “limited” and “minimal”), which should ensure that experimentation in less critical areas remains feasible, whereas sensitive areas like education or recruitment are protected by stronger regulation. In accordance with the EU’s values, unacceptable use cases like the using AI to manipulate human behaviour or for real-time remote biometric identification (e.g. facial recognition) are banned.

REFLECTED USE

Finally, this paper has outlined significant limits to ChatGPT’s capabilities. Most significantly among those may be what was described here as bullshitting, the fact that it can give false answers with high confidence. This needs to be considered in any use of ChatGPT, along with the other limits outlined above (e.g. bias, lack of creativity). The best way to deal with these limits may be for individuals to be more reflective or considerate in their use ChatGPT. As Geuter (2023b) reminds us, it is up to each individual to decide whether and how to use technologies like ChatGPT.

Conclusion

As we have seen, the reception of ChatGPT contains an element of hype. It does not have intelligence; it is a technology like any other. As such, it comes with significant downsides, risks and limits, which have been introduced and outlined in this paper. To use ChatGPT and similar tools safely and productively, we need solid ethical frameworks and regulations. Moreover, as with any technology, its value lies in the reflected use by people within organizations. Individuals and organisations should consider what they want to achieve with it and why and use it, if at all, only after critical reflection. In the end, we have a choice here.

RESEARCH DIRECTIONS

This paper has briefly outlined some issues around ChatGPT and its current reception by media, practitioners and some academics. While it was only able to raise the various points briefly, it did, show that there is an acute need for researchers to share a critical view on ChatGPT. Future research should call out and further investigate the downsides, risks and limits outlined here. The public reception and use of ChatGPT could be discussed in the light of existing theories around technology adoption (e.g. fashions, hype cycles) or desires for new technology to directly influence social or other progress (e.g. technological determinism). There is a particular need for case studies researching the reception and discussion of ChatGPT in various organisational contexts, as well as the reasons informing decisions on its use or non-use.

References:

ACM. (2023). ACM Technology Policy Council Releases „Principles for Generative AI Technologies“. https://www.acm.org/articles/bulletins/2023/july/tpc-principles-generative-ai

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bird, J. (2023, September 18). *AI Statistics—Workplace Survey 2023—Add People*. https://www.addpeople.co.uk/blog/ai-workplace-survey/

Bryson, J. J., Etlinger, S., Keyes, O., & Rankin, J. L. (2020, March 19). *Gender Bias in Technology: How Far Have We Come and What Comes Next?* Centre for International Governance Innovation. https://www.cigionline.org/articles/gender-bias-technology-how-far-have-we-come-and-what-comes-next/

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4* (arXiv:2303.12712). arXiv. https://doi.org/10.48550/arXiv.2303.12712

Cave, N. (2023, January 16). Issue #218. *The Red Hand Files*. https://www.theredhandfiles.com/chat-gpt-what-do-you-think/

Dedehayir, O., & Steinert, M. (2016). The hype cycle model: A review and future directions. *Technological Forecasting and Social Change*, *108*, 28–41. https://doi.org/10.1016/j.techfore.2016.04.005

DeGeurin, M. (2023, April 6). *Oops: Samsung Employees Leaked Confidential Data to ChatGPT*. Gizmodo. https://gizmodo.com/chatgpt-ai-samsung-employees-leak-data-1850307376

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4573321

Deshpande, A., Murahari, V., Rajpurohit, T., Kalyan, A., & Narasimhan, K. (2023). *Toxicity in ChatGPT: Analyzing Persona-assigned Language Models* (arXiv:2304.05335). arXiv. https://doi.org/10.48550/arXiv.2304.05335

European Commission. (2023). *A European approach to artificial intelligence*. Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

Floridi, L. (2023). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford University Press.

Frankfurt, H. G.(1988). On bullshit. In *The Importance of What We Care About: Philosophical Essays* (pp. 117–133). Cambridge University Press. https://doi.org/10.1017/CBO9780511818172.011

Future of Life Institute. (2023). Pause Giant AI Experiments: An Open Letter. *Future of Life Institute*. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Geuter, J. (2023a, March 16). *ChatGPT, Bard und Co: Bullshit, der (e)skaliert*. Golem.de. https://www.golem.de/news/chatgpt-bard-und-co-bullshit-der-e-skaliert-2303-172677.html

Geuter, J. (tante@). (2023b, June 5). *I'm sorry HAL, I won't let you do that.* re:publica 23. https://re-publica.com/de/session/im-sorry-hal-i-wont-let-you-do

Gmyrek, P., Berg, J., & Bescond, D. (2023). *Generative AI and jobs: A global analysis of potential effects on job quantity and quality* (Vol. 96). ILO. https://doi.org/10.54394/FHEM8239

Hanff, A. (2023, March 2). *ChatGPT should be considered a malevolent AI and destroyed*. https://www.theregister.com/2023/03/02/chatgpt_considered_harmful/

Hiltzik, M. (2023, Juli 13). Opinion: Is ChatGPT's Hype Outpacing Its Usefulness? GovTech. https://www.govtech.com/education/higher-ed/opinion-is-chatgpts-hype-outpacing-its-usefulness

Kapoor, S., & Narayanan, A. (2023, March 20). A misleading open letter about sci-fi AI dangers ignores the real risks. *AI Snake Oil*. https://www.aisnakeoil.com/p/a-misleading-open-letter-about-sci

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Kurzweil, R. (2005). *The Singularity is Near: When Humans Transcend Biology*. Gerald Duckworth & Co Ltd.

Lim, M. (2018, September 5). *History of AI Winters*. Actuaries Digital. https://www.actuaries.digital/2018/09/05/histor y-of-ai-winters/

Matar, O. (2023, December 20). *Is ChatGPT Intelligent? A Scientific Review*. Medium. https://towardsdatascience.com/is-chatgpt-intelligent-a-scientific-review-0362eadb25f9

McQuillan, D. (2023, February 6). *We come to bury ChatGPT, not to praise it.* https://www.danmcquillan.org/chatgpt.html

Myers, D., & DeWall, C. N. (2021). *Psychology*. Worth.

O'Neil, C. (2017). *Weapons Of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books.

Perri, L. (2023, August 23). *What's New in the 2023 Gartner Hype Cycle for Emerging Technologies*. Gartner. https://www.gartner.com/en/articles/what-s-new-in-the-2023-gartner-hype-cycle-for-emerging-technologies

Perrigo, B. (2023, January 18). *OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic*. TIME. https://time.com/6247678/openai-chatgpt-kenya-workers/

Rowe, N. (2023, August 2). 'It's destroyed me completely': Kenyan moderators decry toll of training of AI models. *The Guardian*. https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai

Russell, S., & Norvig, P. (2021). *Artificial Intelligence, Global Edition: A Modern Approach*. Pearson Deutschland.

Shah, S. (2023, December 13). *Sam Altman on OpenAI and Artificial General Intelligence*. TIME.

https://time.com/6344160/a-year-in-time-ceo-interview-sam-altman/

Singh, T. (2023). *IBM was the OpenAI in the 60s* [LinkedIn post]. L:inkedIn. https://www.linkedin.com/feed/update/urn:li:activity:7134542846010773505/

The Economist. (2023a, March 15). A battle royal is brewing over copyright and AI. *The Economist*. https://www.economist.com/business/2023/03/15/a-battle-royal-is-brewing-over-copyright-and-ai

The Economist. (2023b, April 20). Large, creative AI models will transform lives and labour markets. *The Economist*. https://www.economist.com/interactive/science-and-technology/2023/04/22/large-creative-ai-models-will-transform-how-we-live-and-work

Turk, V. (2023). *How AI reduces the world to stereotypes*. Rest of World. https://restofworld.org/2023/ai-image-stereotypes/

UNESCO. (2022). Recommendation on the Ethics of Artificial Intelligence. UNESCO Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Villalobos, P., Sevilla, J., Heim, L., Besiroglu, T., Hobbhahn, M., & Ho, A. (2022). *Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning* (arXiv:2211.04325). arXiv. https://doi.org/10.48550/arXiv.2211.04325

Vincent, J. (2023, May 19). *Apple restricts employees from using ChatGPT over fear of data leaks*. The Verge. https://www.theverge.com/2023/5/19/23729619/apple-bans-chatgpt-openai-fears-data-leak

Yadlowsky, S., Doshi, L., & Tripuraneni, N. (2023). *Pretraining Data Mixtures Enable Narrow Model Selection Capabilities in Transformer Models* (arXiv:2311.00871). arXiv. https://doi.org/10.48550/arXiv.2311.00871

Zodhya. (2023, May 20). *How much energy does ChatGPT consume?* Medium. https://medium.com/@zodhyatech/how-much-energy-does-chatgpt-consume-4cba1a7aef85