Technology in Society 75 (2023) 102372

Contents lists available at ScienceDirect

Technology in Society

journal homepage: www.elsevier.com/locate/techsoc

Generative AI: Here to stay, but for good?

Henrik Skaug Sætra

Østfold University College, Norway

ARTICLE INFO

Keywords:
Generative AI
Large language models
Generative adversarial networks
Harms
Power
Inequality

ABSTRACT

Generative AI has taken the world by storm, kicked off for real by ChatGPT and quickly followed by further development and the release of GPT-4 and similar models from OpenAI's competitors. The street has most certainly found its use for generative artificial intelligence (AI), and there is no longer much point in discussing *whether* generative AI will be influential. It will, and what remains to be discussed it how influential it will be, and what potential harms arise when we use AI to generate text and other forms of content. Technological change entails societal change, and we must always endeavor to ask how new technologies shapes, engenders, or potentially erodes the "good society". In this sense, Generative AI is another instance of politically and culturally disruptive autonomous technology, and in this short commentary I highlight some of the key questions to be asked regarding consequences on the micro, meso, and macro level.

1. Introduction

Generative AI has taken the world by storm, kicked off for real by ChatGPT and quickly followed by further development and the release of GPT-4 and similar models from OpenAI's competitors. Academics' ethical and practical concerns aside, the street has most certainly found its use for generative artificial intelligence (AI), and there is no longer much point in discussing *whether* generative AI will be influential. It will, and what remains to be discussed it how influential it will be, and what potential harms arise when we use AI to generate text and other forms of content. Technological change entails societal change, and we must always endeavor to ask how new technologies shapes, engenders, or potentially erodes, the "good society" [1]. In this sense, Generative AI is another instance of politically and culturally disruptive autonomous technology [2,3], and in this short commentary I highlight some of the key questions to be asked regarding consequences on the micro, meso, and macro level.

1.1. What is it?

Generative AI is here used as an umbrella term to describe machine learning solutions trained on massive amounts of data in order to produce output based on user prompts (input in the form of commands), for example "ai personified overflowing the world with texts and images and other media, futuristic, high resultion, dark", which produces the illustration in Fig. 1 when input to the *Midjourney* image generator.[1]

ChatGPT was mentioned above, and this solution produces various forms of text-based output. It is a *large language model* (LLM) which is specialized for natural language processing. ChatGPT, for example, produces the following description when asked "In a short sentence, can you describe what a large language model is?":

> A large language model is a machine learning model that is trained to generate text that is similar to human language. It is called "large" because it is trained on a large dataset and is able to generate highly realistic and coherent text.

ChatGPT is produced by OpenAI,[2] who have also produced previous versions of GPT. GPT-3 generated similar stories and concerns in 2020, but as it was far less available to the public (and less tuned for general conversations) the interest was largely limited to the industry, academia, and news communities. GPT-4, however, released in 2023, was immediately implemented in ChatGPT and made broadly available.[3] By *industry*, I here refer to those who develop *or* use these models for business purposes, as I'll return to below. Others have their own

E-mail address: Henrik.satra@hiof.no.

[1] https://www.midjourney.com/home/.
[2] https://openai.com.
[3] https://openai.com/product/gpt-4.

https://doi.org/10.1016/j.techsoc.2023.102372
Received 4 January 2023; Received in revised form 10 September 2023; Accepted 10 September 2023
Available online 12 September 2023

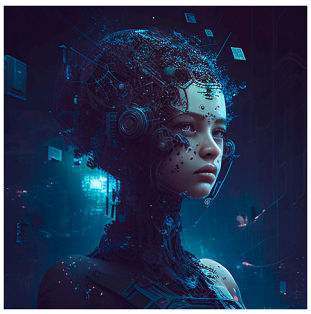

Fig. 1. Midjourney produced image of generative AI personified.

LLMs, and previous examples include DeepMind's Gopher,[4] Meta's OPT-175B,[5] and Google's LaMDA.[6] Following the launch of ChatGPT, however, others scrambled to catch up, and released their own new implementations of LLMs, including Google's Bard,[7] DeepMind's Chinchilla,[8] and Meta's LLaMA.[9] Other models have been released regularly throughout 2023, such as Google's PaLM 2 in May[10] and Anthropic's Claude 2 in July. Notably Meta's Llama 2 was released in July 2023 and was made freely available for research *and* commercial use,[11] potentially disrupting the business models of its competitors. The business models are as of yet varied, but Microsoft, for example, invested in OpenAI to provide access to ChatGPT through its various Office applications [4].

As described above, AI is also used in generative AI models made for making images using Generative Adversarial Networks (GANs) or diffusion models. Well-known examples include OpenAI's Dall-E,[12] Midjourney,[13] and Stable Diffusion.[14] These models produce various forms of images upon the user's request, allowing for the choice of topic, style, mood, context, etc.

Generative AI is not, however, limited to these media. Movies, for example, are already being made, and it is no major leap of the imagination to imagine a near future in which generative AI can make entire shows based on our commands [5]. Music and voice is also easily produced by AI, and multi-modal media can be produced through a combination of already existing techniques. Imagine asking a more advanced form a ChatGPT to make a short textbook in biology, with illustrations from its companion Dall-E, for example. People are already using language models to make good prompts (commands) for image generators and combining these is merely another small step. Technologically mundane, socially significant.

1.2. What can it do?

Late 2022 saw an explosion in the impact of generative AI, and ChatGPT was the major source of this breakthrough. While the technology was not new, it was made openly and freely available, and it had reached a level of maturity that made it immediately accessible and useful for many users. Therefore, it now makes little sense to keep on arguing that LLM's will not be successful or have major impact. They *are* already successful, and they are having impact.

Being somewhat of a techno-skeptic myself, how can I be so sure of the success of generative AI? By having eyes and ears and seeing the impact it already has. Many already know what this technology can do, and news articles on – and even by – ChatGPT abounds. But much more significant are the stories of how non-specialists and those with no experience with AI or technology are diving into ChatGPT.

For example, it did not take long after ChatGPT become available before my son and his peers at school became aware of it, and they immediately started using it. Some to cheat, I'm sure, but most to get inspiration and foundations for various assignments. For getting a basic overview of a topic, for making outlines for texts, CVs, etc. No programming backgrounds, and no specific knowledge of AI. ChatGPT was quite simply useful from the get-go. And use it people did.

In higher education, professors and administrators alike are scrambling to figure out how to deal with the impact of ChatGPT. The use of take-home exams, for example, must necessarily be reconfigured once students can easily use generative AI to produce whole or parts of their texts, without any system being able to identify the inclusion of ChatGPT content in what is handed in. Some are fearful and angry at this disruption, while others are already using it in class, asking their students to use it in new and creative ways. How higher education changes with ChatGPT is still unknown, but that some change will occur seems uncontroversial.

Beyond this, however, it is already being used by a wide array of professionals. Consultants preparing outlines for presentations, writing reports, sending letters that must be carefully worded, etc., all this is already being done, as both junior and senior staff sees the potential of the new technologies. Some share their newfound power tools, while others must be expected to use such tools in secret. And, not least, people are already building businesses on this technology, and have done so for some time. One early example was Lensa.ai, the app that uses StableDiffusion to generate profile pictures from user uploaded selfies, and which produced the author's profile picture shown in Fig. 2.

Generative AI can produce both text and images or all types, and as mentioned, other media are also either already covered or soon to follow. We must not be blinded by what generative AI can already do, however, as its potential is much greater. A crucial point is that generative AI can produce text, and everything that is done by computers is – in essence – doable through text. Programming provides a particularly interesting example, and by having LLMs produce code of various kind one might, in theory, do just about anything. The positive potential is easily seen in how new applications, templates, 3D models, educational content, etc. can be made. The impact of the programing capabilities of generative AI is clearly demonstrated by the recent ACM article declaring the *end of programming* [6].

1.3. Should we worry?

The creative and productive potential of generative AI is enormous, but what are the potential pitfalls created by generative AI – the dangers

Fig. 2. AI generated profile picture made with the app Lensa.ai.

[4] https://www.deepmind.com/blog/language-modelling-at-scale-gopher-ethical-considerations-and-retrieval.
[5] https://ai.facebook.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b/.
[6] https://blog.google/technology/ai/lamda/.
[7] https://bard.google.com.
[8] https://www.deepmind.com/publications/an-empirical-analysis-of-compute-optimal-large-language-model-training.
[9] https://ai.facebook.com/blog/large-language-model-llama-meta-ai/.
[10] https://blog.google/technology/ai/google-palm-2-ai-large-language-model/.
[11] https://about.fb.com/news/2023/07/llama-2/.
[12] https://openai.com/dall-e-2/.
[13] https://www.midjourney.com/home/.
[14] https://stability.ai/blog/stable-diffusion-public-release.

academics and others have warned of in the build-up to generative AI permeating all sectors? One way to sort the different concerns is through distinguishing between implications on a societal level (macro level), on sectors, groups, or organizations (meso level), or individuals (micro level).

1.3.1. Macro level challenges

Firstly, there are concerns that generative AI could have detrimental effects on democracy and political stability. One example is how generative AI can generate practically unlimited amounts of political content for dissemination. Fake news is one concern, and this could be both text and generated videos where real people or situations are presented in new and imagined ways (deepfakes). When AI floods the information sphere with new content, there are, for example, fears that people will lose their grasp of what is true or not, or that we'll experience increased polarization [7]. Also related to democracy is the proposed use of generative AI to *improve* democratic and deliberative processes [8–10]. Using generative AI to foster agreement and better decisions would clearly be a hugely important benefit of generative AI, but there is also the risk that such usage of AI dilutes democracy as we know and value it [11].

Secondly, generative AI has, as described above, the potential to replace workers of all kinds, including so-called "knowledge workers" [12]. Even if humans are not replaced by AI, there is a risk that work *changes*, and whenever this occurs we need to be wary of how power constellations change and whether such changes are conducive to the decency of work, as emphasized in, for example, the United Nation's Sustainable Development Goal (SDG) 8 [13]. Generative AI will likely have a major impact on both economic growth (SDG 8) and innovation (SDG 9), and it will consequently be important to ensure that this impact is positive and conducive to socially sustainable development. It is, however, also important to note that while developing generative AI models is in theory possible for all, it requires vast resources, and a world increasingly using such models risks being increasingly dependent on large tech companies [14].

A third risk is that generative AI tends to promote the status quo. It is based on historical data, and when such models become increasingly influential, they might prevent desired societal changes. Furthermore, human history is ripe with bias and discrimination, and systems based on historical data will tend to reproduce such harms in new and opaque ways [15].

Finally, AI requires energy, and generative AI is particularly fond of massive amounts of data, which in turn means more energy use. This translates to an increased carbon footprint of AI [16]. As the information and communication technology sector is a significant (and growing) contributor to global emissions, the greenhouse gas emissions generated by training these models must be part of the equation when the pros and cons of generative AI is considered.

1.3.2. Meso level challenges

Firstly, and related to the macro level challenge of work and labor changes, generative AI will change professions and consequently change the power relationships between professions, employer and employees, and different groups. What happens to copywriters, for example, when LLM's can produce copy for news, advertising, etc.? And what about freelance photographers, when news organizations source images from Dall-E or Midjourney instead of buying stock photos or hiring photographers? Professions always change with technological change, however [17]. The example of how typographers, for example, went from playing an important and protected part of the news value chain to becoming obsolete or being required to branch out into adjacent professions, illustrates the dynamic to be expected.

Secondly, there are severe challenges related to how generative AI extracts, appropriates, and produces content produced by human beings. These producers will usually not have wanted their content to be used for training such models, but they are left with no say in this process. This is a dire challenge, particularly considering how these models can in turn be made to reproduce individual styles, and generative AI will consequently take content produced by humans, make it their own, and make the original creators obsolete. All this occurs without any compensation or right to recourse for human content producers, highlighting a significant shortcoming in current regulation of the extraction and use of data for training models. This challenge also relates to how Google, for example, were allowed to "appropriate" the imagery of streets (Google Street View), and to scan and extract the world's literature (Google Books) [18].

Finally, biased and discriminatory systems will also have meso level effects as the negative impacts are not equally distributed, but concentrated [15]. These negative effects are extra problematic when considering how already marginalized groups tend to be most exposed to discrimination based on historical data. Furthermore, the same groups also tend to be underrepresented amongst those who develop and control the systems in questions. The various approaches to monetizing generative AI systems also show the potential for these systems to generate "digital divides" between those who have access and not [19]. Such divides could occur between groups in developed nations, but the gaps between nations with generally good internet and computing infrastructure and, for example, those living in developing and least-developed nations, must also be considered. Even the "free" version of ChatGPT will be inaccessible to many.

1.3.3. Micro level challenges

Individuals will experience macro and micro level effects to varying degrees, but there are also impacts that are best understood by considering how generative AI might impact individuals more directly.

Firstly, one recurring concern with new assistive technologies is that we run the risk of cognitive atrophy. When we allow AI to do perform mentally and cognitively challenging tasks, and even doing our creative work, we might run the risk of not being able to do this work ourselves in the long run [20]. Just like calculators have been detrimental to our mental arithmetic skills, ChatGPT may be detrimental to our writing skills.

Secondly, others are concerned about how generative AI aimed at interacting with humans will be increasingly adept at persuasion, and that this will easily cross into the domain of manipulation [21]. Researchers at OpenAI, for example, have published an article problematizing how they foresee growing difficulties related to alignment once capable generative AI systems tasked with optimizing some parameter could easily start manipulating human behavior and perceptions in order to achieve such goals [22]. While few stories of serious human harm – for example suicides – have as of yet been linked to LLMs, people in the industry fear that this is just a matter of time [23].

Finally, generative AI can be so seductive and intriguing that humans might come to prefer them to human partners. The combination of text and video could, for example, allow for a new intimate *partner*, much like what Replika – "the AI companion that cares – is already doing.[15] Combining generative AI with robotics, we'll have companion robots providing enjoyment, care, and intimacy without any of the hassles associated with human partners [24]. While potentially comfortable, these partners might be detrimental for both the chances of finding human partners (now preoccupied with the AI partners) and even our own ability to be good and patient partners for humans.

The potential dangers are shown in Fig. 3, which also illustrates how the three levels interact and how dangers on one level also transfers to the other levels.

2. Conclusion

"There is no stopping progress" is not a sound strategy for achieving

[15] https://replika.com.

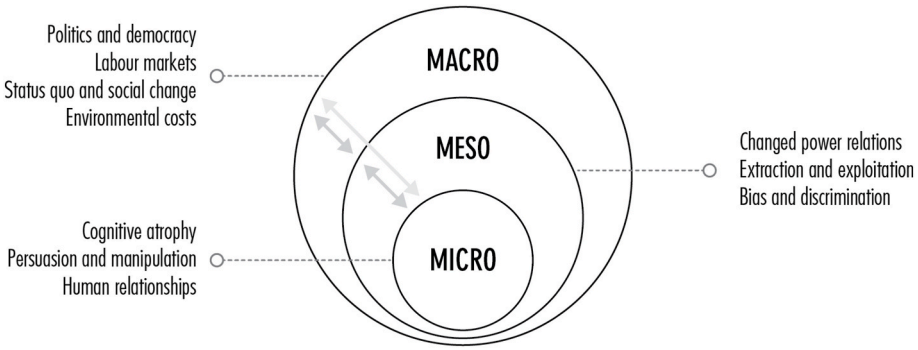

Fig. 3. Dangers on the micro, meso, and micro levels.

the future we want, and we must ensure that we evaluate the implications of any kind of technological change before resigning to its inevitability [25]. Having done so, we are in a position where we know how new technologies change values, power constellations, and social structures. Only then can we know what is *actual* progress, and not just more complicated technology. What seemed like progress could turn out to be detrimental to human wellbeing and our values as a community of beings, and this, I argue, we both can and should stop prevent.

For generative AI to contribute to the good society, it must promote and not undermine our fundamental values. These values will of course be debated, but an example set of values conducive to a good society consists of freedom, democracy, sustainability, well-being, and justice [26]. As discussed, all these values are potentially challenged by generative AI systems, but there are also arguments to be made that generative AI could support and promote them.

Generative AI has incredible positive potential, and it is no questioning that such technologies can improve the lives of some. And the wealth reservoir of others. The technology is useful and should most likely not be stopped outright. However, there must be proper regulation in place to ensure that both the development and use of generative AI does not lead to the negative harms discussed on the macro, meso, and micro levels. For example, if we aim to reap the benefits of generative AI for economic growth, we would do good to ensure that the growth created is in line with SDG 8, which states that it should be "sustained, inclusive, and sustainable" and contribute to "full and productive employment and decent work for all" [13].

This requires us to emphasize human agency and our means for shaping and controlling technology. However, Collingridge's dilemma states that technology can relatively easily regulated in its infancy, but at that stage our knowledge of the impacts and the reasons for regulating it are also in its infancy [27]. When the technology is deployed and widely diffused [28], enforcing social control of technology is much harder, while the reasons to do so have often become painfully clear [27]. With OpenAI's strategy of unleashing generative AI on our societies, we are currently in the latter condition. Despite the difficulties this involves, individuals, groups, and societies both can and must endeavor to make the technology conducive to the good society [1].

Funding

No funding declaration.

Conflict of interest

No conflicting interests.

Author statement

Henrik Skaug Sætra: All parts of the article.

Data availability

No data was used for the research described in the article.

Acknowledgements

No acknowledgments.

References

[1] C. Griffy-Brown, B.D. Earp, O. Rosas, Technology and the good society, Technol. Soc. 52 (2018) 1–3.
[2] L. Winner, Autonomous Technology: Technics-Out-Of-Control as a Theme in Political Thought, MIT Press, Cambridge, 1977.
[3] J. Ellul, The Technological Society, Vintage Books, New York, 1964, pp. 229–318.
[4] D. Bass, Microsoft Invests $10 Billion in ChatGPT Maker OpenAI, Bloomberg, 2023.
[5] L. Brown, Soon You'll Be Able to Make Your Own Movie with AI: artificial intelligence isn't about to change the movie industry. It already has, Vulture. Available: https://www.vulture.com/2022/12/ai-art-midjourney-chatgpt-phenaki-movies-hollywood.html, 2022, December 27.
[6] M. Welsh, The end of programming, Commun. ACM 66 (1) (2022) 34–35, https://doi.org/10.1145/3570220.
[7] J. Danaher, H.S. Sætra, Technology and moral change: the transformation of truth and trust, Ethics Inf. Technol. (2022), https://doi.org/10.1007/s10676-022-09661-y.
[8] S. Fish, et al., "Generative Social Choice," (2023) *arXiv preprint arXiv:2309.01291*.
[9] R. Koster, et al., Human-centered mechanism design with Democratic AI, Nat. Human Behav. (2022), https://doi.org/10.1038/s41562-022-01383-x.
[10] M. Bakker, et al., Fine-tuning language models to find agreement among humans with diverse preferences, Adv. Neural Inf. Process. Syst. 35 (2022) 38176–38189.
[11] H.S. Sætra, H. Borgebund, M. Coeckelbergh, Avoid diluting democracy by algorithms, Nat. Mach. Intell. 4 (10) (2022) 804–806, https://doi.org/10.1038/s42256-022-00537-w.
[12] B. Edwards, Artists stage mass protest against AI-generated artwork on ArtStation, in: Ars Technica, 2022.
[13] United Nations, Transforming Our World: the 2030 Agenda for Sustainable Development, Division for Sustainable Development Goals, New York, NY, USA, 2015.
[14] D.G. Widder, S. West, M. Whittaker, Open (for business): big tech, concentrated power, and the political Economy of open AI, Concentrated Power, and the Political Economy of Open AI (August 17, 2023) 2023.
[15] E.M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: can language models be too big, Proceedings of FAccT (2021), https://doi.org/10.1145/3442188.3445922.
[16] B. Brevini, Is AI good for the Planet? Polity (2021).
[17] S.R. Barley, Work and Technological Change, Oxford University Press, Oxford, 2020.
[18] S. Zuboff, The Age of Surveillance Capitalism: the Fight for a Human Future at the New Frontier of Power, PublicAffairs, New York, 2019.
[19] E.I. Nordrum, The technologically sustained digital divide, in: H.S. Sætra (Ed.), Technology and Sustainable Development, Routledge, Milton Park, 2023, pp. 97–108, ch. 8.
[20] H.S. Sætra, The Ghost in the machine, Human Arenas 2 (1) (2019) 60–78.
[21] H.S. Sætra, S. Mills, Psychological force, liberty and technology, Technol. Soc. 69 (2021), 101973, https://doi.org/10.1016/j.techsoc.2022.101973.
[22] R. Ngo, The Alignment Problem from a Deep Learning Perspective, 2022 *arXiv preprint arXiv:2209.00626*.
[23] G. Marcus, The dark risk of large language models, Wired (December 29 2022) [Online]. Available: https://www.wired.com/story/large-language-models-artificial-intelligence/.
[24] H.S. Sætra, Loving robots changing love: towards a practical deficiency-love, Journal of future robot life 3 (2) (2021) 109–127, https://doi.org/10.3233/FRL-200023.

[25] A. Næss, Samfunn Økologi, *Livsstil* (Bokkubbens Kulturbibliotek), vol. 1971, 1999. Oslo.
[26] P. Brey, The strategic role of technology in a good society, Technol. Soc. 52 (2018) 39–45.
[27] D. Collingridge, The Social Control of Technology, Frances Pinter, London, 1980.
[28] E. Engström, P. Strimling, Deep learning diffusion by infusion into preexisting technologies–Implications for users and society at large, Technol. Soc. 63 (2020), 101396.