

GENERATIVE AI RISKS

Even though it is incredibly powerful and promising, generative AI can bring about some serious negative consequences alongside positive ones. These undesirable implications should not be immediately interpreted in an apocalyptic manner. Instead, they should rather be timely discussed in an informed way, and adequately regulated and mitigated. So, let us identify and explain some of the most prominent imperfections of generative AI tools, as well as the different types of risk AI carries.

AI tools flaws

The first type of risk relates to all the imperfections stemming from technical characteristics and the quality of the AI tools or source materials they are trained on. Even though they can produce serious risks, these flaws are something that can and is very likely to be improved in future iterations of generative AI tools.

Inaccuracy

Generative AI algorithms and training datasets are not perfect; they make mistakes and cannot be fully reliable. In fact, these tools can very confidently produce false information. The data they rely on may be outdated or faulty, and the AI itself doesn't fact-check or doubt the information it generates. For instance, since Chat GPT has been trained on data up to 2021, a user could ask a question about the president of a certain country that had elections in the meantime, and the AI would confidently provide inaccurate information.

Hallucinations

AI tools also tend to create content even if they do not have actual information in their training base – they basically start making things up or hallucinating. There are examples of chatbots providing answers completely out of touch with reality, summarizing non-existent articles or image generation tools generating various out-of-this-world content.

Bias

Generative AI tools can perpetuate and amplify societal bias because of the data they are trained on. It has the potential to replicate the unfair, racist, sexist, or other biases present in the dataset it learns from and hence further harm marginalized groups.

AI companies are moderating offensive or discriminative content; however, the tools still reflect the source materials they were trained on and, in the end, all-pervading bias in our society. For instance, when asked to create an image of a specific profession, the image generation AI tools might link it to certain gender, race, or similar attributes. So, if asked to generate an image of a CEO, all variations will include middle-aged white men.

Policy risks

The second group of risks is the one tied with potential loopholes, policy gray zones and risks that require adequate and timely regulation and governance.

Data privacy

Generative AI gathers and processes large amounts of data and can potentially memorize and reproduce sensitive information present in the datasets they are trained on. AI tools should not be trusted with confidential and sensitive information and users should refrain from providing personal or business data directly to the tools, as all inputs are used for training and research purposes.

The AI companies do not guarantee data privacy and it is unclear how this data is stored, how it can be further used and/or shared with third parties. If adequate protection measures are not in place, individual or business privacy can be compromised in different ways – misuse of sensitive personal information, identity theft, unauthorized access and more.

Intellectual property issues

Generative AI can create content that closely resembles existing copyrighted works, which raises legal and ethical questions about intellectual property rights and copyright infringement.

For instance, an AI image generator might create digital art that closely resembles famous paintings or imitates famous artists' styles, potentially infringing on the copyrights of the original artists or their work. It also raises the question of whether AI-generated art deserves equal copyright protection as works created by humans and who should be considered the author or the owner of such content.

Malicious use

Bad actors risk

Universal accessibility of AI tools has many benefits, but it also means that anyone can use it maliciously - to generate disinformation or deepfakes, write phishing emails or create malware. Deepfakes and disinformation have been around for some time, but the fact that they are now easily accessible for a wider public and are becoming increasingly realistic, raises significant concerns.

Additionally, in cyberspace, malicious actors can create more advanced and sophisticated attacks, in a faster and cheaper manner. For instance, they can use chatbots to imitate human behavior, leading people to believe they are interacting with a real person and thereby bypassing security measures.

Societal risks

The last group of risks concerns those with broader societal implications and is yet in the realm of predictions and speculations about upcoming trends. As with all technological advancements in history, the wider impact of AI will unfold over time.

Future of work

It is certain that generative AI will have a significant impact on the labor market, however, it is still discussed and predicted what kind of impact we can expect. Most experts predict job changes and displacement across various industries, due to the increased automation that will come with advancements in AI technology.

As generative AI becomes more advanced, it will likely automate and streamline certain tasks that are currently performed by humans, leading to increased efficiency and productivity. On the other hand, it is anticipated that AI upgrades and different types of applications will create entirely new careers and increase demand for some already existing jobs.

Geopolitical implications

Progress in the generative AI field is set to influence the global geopolitical arena as well. Many experts envisage that AI will be at the heart of the 21st century arms race. In their efforts to leverage AI's capabilities for economic benefits, technological superiority and influence, global powers will pursue AI dominance over other actors.

AI's ability to create realistic text, images, and videos becomes an important factor, since this AI-generated content can shape information warfare, influence public opinion, and potentially impact diplomatic relations and power dynamics between countries.