

УДК 681.5

ARTIFICIAL INTELLIGENCE IN MODERN SOCIETY

Д. Р. МАЛЮКОВ

Научный руководитель Т. П. ПИНЧУКОВА

Белорусско-Российский университет

Могилев, Беларусь

Artificial Intelligence (AI) is revolutionizing our world and has the potential to transform our future in unimaginable ways. With each passing day, AI is becoming more advanced and its impact is being felt across multiple domains. AI is already being used in a wide range of fields, including healthcare, finance, transportation, and, most surprisingly, in creative tasks. I will dive into the current state of AI and its impact on our society, as well as its potential implications for the future.

Before we start, I want to explain what AI is. AI is a broad field that involves developing computer systems capable of performing tasks that traditionally require human intelligence, like learning, reasoning, problem-solving, and pattern recognition [1]. Modern AIs are created through training on large amounts of data using machine learning techniques and are based on neural networks. The most important thing is that they are trained, not hard-coded, so humans just create their architecture then feed data into them while neural networks find the patterns themselves and use backpropagation to adjust their behaviour [2].

AI algorithms are trained using large datasets to recognize patterns and make predictions or decisions based on that data. They can also learn from their mistakes and improve their performance over time. Some AI systems are designed for specific tasks, such as playing chess or recognizing speech, while others can learn from experience and adapt to new situations.

I want to discuss the most recent advancements in the AI world, and it's the ability to solve creative tasks. For many years, people thought that AI would first replace low-skilled jobs and blue-collar labour, but as we see now, it's not the case. AI has come for high-skilled creative jobs, breaking the assumption that high-skilled jobs, such as those in the visual arts, were safe from automation. Recently developed AI tools for image generation based on neural networks, such as Midjourney and DALL·E 2, which were launched at the start of 2022, and later in August, Stable Diffusion, showed us that machines can replicate human creativity. At their first launch, they were not that impressive, but just in a matter of half a year, they've made their way from childish paintings to virtually indistinguishable from real photos or highly creative and professional artworks [3].

Of course, such advancements weren't left unnoticed. Some people said that these tools would greatly help artists to improve and explore their own creativity, while others thought that wide adoption of these tools would mean the death of creativity and art in general. The most concerned were the artists themselves, with some of them feeling endangered by the technology which had cut off their main source of income.

A group of artists has filed a class-action lawsuit against the companies behind three AI art generators, accusing them of violating copyright and unfair competition

laws. They said that AI tools were trained not only on free-to-use images but also on copyrighted images and their own paintings, whose style AI can replicate if it is asked to [4]. As of now, this question remains unanswered. Some experts argue that it falls under fair use, while others believe it may be copyright infringement [5].

Just a few months later, after image generators had made their way into people's minds, another previously unseen technology came out – AI conversational chatbots.

ChatGPT is the first conversational AI chatbot that uses natural language processing (NLP) which was publicly released by OpenAI in November 2022. It can carry on a conversation in a natural way, being able to remember context, answer follow-up questions, and follow instructions. These underlying features enable it to respond to questions and compose various written content, including articles, social media posts, essays, code, and emails. Initially, it was built on top of GPT-3.5 foundational large language model (LLM), but now the GPT-4 model is also available [6].

The name of these models suggests that it's not the first generation of GPT LLMs. And indeed, GPT-1 was created in 2018 and GPT-3 in 2021, so what made them so promising only right now? The main reason is reinforcement learning from human feedback (RLHF). In simple terms, RLHF involves training a language model to generate better responses based on feedback from humans. This feedback helps the model understand which responses are more aligned with human preferences, allowing it to improve its performance over time [7].

When AI becomes so advanced and natural in its conversations with humans, it's not surprising that people will use it as their friend or personal psychologist. Some of them claim that ChatGPT is better than any therapist they've visited so far [8]. Although some experts say that it's not healthy to emotionally engage with a machine, and that suggestions made by ChatGPT may be dangerous, others argue that a machine is more patient, attentive, and always willing to help.

Almost every AI discussion today is about morality and ethics. Companies such as OpenAI claim that they are trying to align AI with human values, but they also acknowledge that the alignment problem is not completely solved and becomes harder to tackle as AI gets smarter. Besides, whose values does OpenAI represent [9]? People's attitudes towards AI are based only on their own preferences. There is no right answer to these questions, and when discussion touches on topics such as "Whether AI will kill all of us or not" even researchers don't know, because they create architecture and train these models, but they don't know what the output will be. Basically, at this point, AI development is largely uncontrolled, and results often are unexpected. Because of that, many people are starting to feel existential dread and fear malicious use of AI [10].

In conclusion, the advancement of AI has had a massive impact on various fields. The recent advancements in AI have given rise to some ethical and legal concerns, particularly in the creative industry. While there are concerns about the potential dangers of AI, there is also the potential for AI to be a positive force for good, transforming the way we live and work. And this transformation will only be greater in the future. Considering the pace at which AI is developing, it is no longer a science fiction scenario to see the world being fully automated in five or ten years.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Artificial intelligence (AI) | Definition, Examples, Types, Applications, Companies, & Facts | Britannica [Electronic resource]. – Mode of access: <https://www.britannica.com/technology/artificial-intelligence>. – Date of access: 07.05.2023.
2. Back propagation Algorithm – Back Propagation in Neural Networks [Electronic resource]. – Mode of access: <https://intellipaat.com/blog/tutorial/artificial-intelligence-tutorial/back-propagation-algorithm/?US>. – Date of access: 07.05.2023.
3. Fake Trump arrest photos: How to spot an AI-generated image – BBC News [Electronic resource]. – Mode of access: <https://www.bbc.com/news/world-us-canada-65069316>. – Date of access: 06.05.2023.
4. Artists and Illustrators Are Suing Three A.I. Art Generators for Scraping and ‘Collaging’ Their Work Without Consent [Electronic resource]. – Mode of access: <https://news.artnet.com/art-world/class-action-lawsuit-ai-generators-deviantart-midjourney-stable-diffusion-2246770>. – Date of access: 07.05.2023.
5. AI Trained on Copyrighted Works: When Is It Fair Use? [Electronic resource]. – Mode of access: <https://www.thefashionlaw.com/ai-trained-on-copyrighted-works-when-is-it-fair-use/>. – Date of access: 06.05.2023.
6. ChatGPT – Wikipedia [Electronic resource]. – Mode of access: <https://en.wikipedia.org/wiki/ChatGPT>. – Date of access: 07.05.2023.
7. Reinforcement Learning from Human Feedback (RLHF) – a simplified explanation [Electronic resource]. – Mode of access: <https://www.nebuly.com/blog/reinforcement-learning-from-human-feedback-rlhf-a-simplified-explanation>. – Date of access: 07.05.2023.
8. From a psychological-therapy standpoint, ChatGPT has been an absolute godsend for me [Electronic resource]. – Mode of access: https://www.reddit.com/r/ChatGPT/comments/12cr4cw/from_a_psychologicaltherapy_standpoint_chatgpt/. – Date of access: 07.05.2023.
9. [D] AI ethics research is unethical: MachineLearning [Electronic resource]. – Mode of access: https://www.reddit.com/r/MachineLearning/comments/og48q2/d_ai_ethics_research_is_unethical/. – Date of access: 08.05.2023.
10. AI progression making me weird: singularity [Electronic resource]. – 2023. – Mode of access: https://www.reddit.com/r/singularity/comments/12dy1g1/ai_progression_making_me_weird/. – Date of access: 08.05.2023.