Galactica Media: Journal of Media Studies. 2024. No 3 | ISSN: 2658-7734
Neural Networks and Technology | https://doi.org/10.46539/gmd.v6i3.502

Large Language Models: A Socio-Philosophical Essay

Regina V. Penner

South Ural State University. Chelyabinsk, Russia. Email: penner.r.v[at]gmail.com
ORCID https://orcid.org/0000-0002-3277-7274

Received: 7 April 2024 | Revised: 19 May 2024 | Accepted: 30 May 2024

Abstract

Neural networks have filled the information space. On the one hand, this indicates the scientific and technological movement of contemporary society (perhaps, AGI is already waiting for us outside the door). On the other hand, in everyday discourse there are extensive discussions about the fact that when neural networks are created, a person is left with hard work. However, a holistic understanding of the neural network is associated with a movement from the mythotechnological framework to the phenomenon itself and the questioning of its social role. The key aim of the paper is returning, through observing the range of functions of current LLMs, to the classic question of whether a machine can think. At the same time another question remains, are humans ready to accept the social subjectivity of machines.

Keywords

Neural Networks; AI; AGI; LLM; OpenAI; GPT; Sora; Social Agent; Social Subject

This work is licensed under a Creative Commons “Attribution” 4.0 International License

Большие языковые модели (LLM): социально-философское эссе

Пеннер Регина Владимировна

Южно-Уральский государственный университет. Челябинск, Россия. Email: penner.r.v[at]gmail.com
ORCID https://orcid.org/0000-0002-3277-7274

Рукопись получена: 7 апреля 2024 | Пересмотрена: 19 мая 2024 | Принята: 30 мая 2024

Аннотация

Нейронные сети заполонили информационное пространство. С одной стороны, это свидетельствует о научно-технологическом развитии современного общества (возможно, AGI [Artificial General Intelligence, Общий искусственный интеллект] уже ждет нас за дверью). С другой стороны, в обыденном дискурсе ведутся обширные дискуссии о том, что при создании нейронных сетей человеку все же достается тяжелая работа. Однако целостное понимание нейронной сети связано с движением от мифотехнологической основы к самому феномену и постановке под сомнение его социальной роли. Основная цель статьи — вернуться, посредством наблюдения за спектром функций нынешних LLM, к классическому вопросу о том, может ли машина мыслить. В то же время сохраняется другой вопрос: готовы ли люди принять социальную субъективность машин.

Ключевые слова

нейронные сети; ИИ; AGI; LLM; OpenAI; GPT; Sora; социальный агент; социальный субъект

Это произведение доступно по лицензии Creative Commons “Attribution” («Атрибуция») 4.0 Всемирная

Introduction

In philistine discourse, the idea is strengthened that the line of development of technology has not gone quite in the right direction. Today we already have Industry 4.0, digital and technological breakthroughs in industry and medicine (Schwab, 2016). In a sense, robots have already entered our everyday life and are confidently establishing themselves in it. However, cases appear in the information field that assert a certain contradictory position of robots, which is not identical to the unambiguous meaning that the fathers of science fiction (K. Capek, I. Asimov) put into this figure, who formed the image, including the "obedient" slave which does all the "dirty" work for his master.

In the mythologies of philistine discourses, in turn, that image of the so-called precedents from the field of AI, according to which all the "dirty" work remains with the person (see, for example, the sensational story about a granny from Moscow who could not "leave a delivery robot in trouble": (Jabborov, n.d.)). Neural networks, that have created information noise today, draw (DALL-E, Midjourney), write music (Suno, WavTool, Silver Scale), poetry (AI Poem Generator, all versions of ChatGPT), and, certainly, program codes (Copilot, Fig, and again versions of ChatGPT). A primary analysis of the above-mentioned practices only convinces us of the myth that silicon is becoming the dominant creative force. However, not all are so simple.

It is noteworthy that among tech companies themselves (OpenAI, Google, Yandex) and in academic circles, mainly among those who tirelessly question the potential for conscious activity of artifacts and technologies of silicon origin, the question of whether a machine can think has not been resolved. The next milestone in the development of this issue was set by a precedent from Google, dated 2022. Then Blake Lemoine was fired from the company on the basis of defending his position that one of the company's developments, LLM LaMDA 2, continuously demonstrated the beginnings of conscious activity (Tiku, n.d.).

Another resonant fact is the announcement of SORA, the neural network from OpenAI. The resonance of this presentation is based on the peculiarities of video generation in comparison with text and static pictures (just remember the video that became viral, where the generated "Will Smith" deliciously eats spaghetti: (*AI Will Smith Eating Spaghetti Pasta (AI Footage and Audio)*, n.d.)). To smooth out the obvious problems of previous experiments, programmers from OpenAI took a text-to-video diffusion model as a basis. As a result, we can already see several different-themed videos from Sora, the first of which was a walk of a strikingly realistic beauty in a red dress around Tokyo (*OpenAI Sora in Action: Tokyo Walk*, n.d.).

And the final precedent, affirming the actor/agent position of neural networks in the context of social contemporaneity, came to us from the world of photography. In 2023, one of the winning images at the World Photography Organization competition turned out to be generated by a neural network (*Photographer Admits Prize-Winning Image Was AI-Generated*, n.d.). The winning portrait is made

in the style of an old black and white photograph. It depicts two women hugging, probably a mother with her grown-up daughter. Both characters in the "shot" do not look into the frame; the focus of attention of "each of them" is directed into the distance. The image emanating from the "photograph" seems to invite the observer to think about something of his own, perhaps lost, gone. As a result, the so-called classic neural network errors escaped the attention of jury experts and casual observers. Upon careful examination of the picture, these errors, however, turn out to be obvious; they concern the anatomy of the arms, or more precisely, the hands. However, the fact remains: the neural network has learned to deceive even the experts among us.

Theoretical and methodological background

The object of study is LLMs (Large language Models). The paper discusses various LLMs, but the research focus is primarily on the developments of "OpenAI". The subject of the research is professional technological discourses (OpenAI reports, articles from habr.com, audiovisual content from YouTube). The discourses used are analyzed in the conceptual framework of social philosophy, each LLM is considered not only as a technological product, but as a social agent occupying a certain place in the social arena and performing the functions of social order.

In the indicated logic, the results of research in Social Studies and Humanities on the topics of neural networks and artificial intelligence are of particular interest. For example, in 2019, Lev Manovich, the theorist of digital culture, published a miniature book called "AI Aesthetics" (Manovich, 2019). As in the previous study on the topic of Software Cultures (Manovich, 2017) he presented results of the Cultural Analytics Lab, operating under his guide since 2007. Among other things, L. Manovich is one of the few thinkers who recognizes the specific status of artificial intelligence and defines machine intelligence in the context of "endless 'intelligent' operations" (Manovich, 2019, p. 12). However, the emphasis is not on the thinking capabilities of machines, but on the dynamics of "aesthetic creation and aesthetic choices" (Manovich, 2019, p. 16) in human activity, in which the capabilities and applications of AI are growing exponentially in the following key areas, selecting from existing content (Instagram) and creation/editing of new content (Photoshop).

In conclusion of the book, L. Manovich stated a thesis that continues the research thrust of his theory of Software Cultures, "we need new methods for seeing culture at its new scale, velocity, and connectivity that can combine both qualitative and quantitative approaches and that can reveal full variability of this new ecosystem without reducing it to a small number of categories" (Manovich, 2019, p. 56). The pathos of the thesis resonates with the call for art researchers to overcome the tendency to focus on single works as opposed to constructing an objective picture through the analysis of a large array of data. It is noteworthy

that similar theses were proposed before (see, for example, (Fuller, 2008)), and after (Qin et al., 2018), thereby confirming the intuition of L. Manovich.

In his turn, David J. Chalmers, an Australian philosopher who specializes in finding the root of consciousness in silicon, never tires of creating thought experiments designed to demonstrate the pros and cons of answering the question of whether a machine can think. In 2022, he gave a lecture on the topic "Could a Large Language Model be Conscious?", where he examined in detail two cases that affirm or refute the conscientiousness of LLMs (Chalmers, 2023).

His reasoning is based on a logical form in which we take consciousness as X, then:

(i) large language models have X,

(ii) if a system has X, then it is probably conscious;

give good reasons for (i) and (ii).
Or vice versa:

(i) these models lack X,

(ii) if a system lacks X, it probably isn't conscious,

and give good reasons for (i) and (ii).

For each form, David J. Chalmers selected his own X's. Among the arguments "for" consciousness, he named: Self-Report (the statement of the model itself that it is conscious); Seems-Conscious (according to the philosopher, a weak argument indicating the appearance of the sentient model for some people); Conversational Ability (according to the philosopher, a strong argument indicating the conversational nature of LLMs, that they are good at giving reasons and explanations; this is what Alan Turing perceived as a hallmark of thinking); General Intelligence (orientation towards the generality, multi-tasking model, which is perfectly demonstrated by GPT-4). Among the arguments "against" consciousness in LLMs he named: Biology (carbon-based biology as a foundation of consciousness); Senses and Embodiment (since LLMs do not have sensors, they do not feel, since they do not have a body, they do not act, while feeling and action are a platform for a conscious act, as Immanuel Kant described it, analysing the movement from Mind to Reason [Vernunft]); World-Models and Self-Models (LLMs create models of texts, not worlds, but discourses about the world, thereby ignoring the world itself in its entirety); Recurrent Processing (these models do not have memory-like internal states, i.e. are not able to establish the continuity of their actions/responses in time); Global Workspace (theory by Bernard Baars and Stan Franklin, according to which consciousness is like the global workspace, making conscious all modules that find themselves within its boundaries (Baars, 1988); it seems that this theory has an intersection with the "sphere of consciousness" of Merab Mamardashvili and Alexander Piatigorsky as a metatheory of consciousness, affirming the consciousness of all existence (Mamardashvili & Piatigorsky, 2009)); in turn,

LLMs do not bring about any consciousness either to their blocks or to the information entering/outgoing from them; they process the information); finally, Unified Agency (the deepest obstacle to consciousness, according to the philosopher; LLMs are like chameleons, they lack of identity or their identity is liquid (Zygmunt Bauman), since their key goal remains the T-9 function, to predict the next word or frame).

After completing this analysis, David J. Chalmers concluded that "the arguments against consciousness in current LLM systems are much stronger" (Chalmers, 2023). Indeed, nowadays, we do not have a conventional/contractual theory of consciousness that gives an answer to the question of what consciousness itself is. On the other hand, programmers successfully scale models, and along with the growth of problems solved by the model and the improvement in the quality of the solutions offered by the model, some layers (the so-called "black box") are hidden even from the developer.

In a sense, both L. Manovich and the mathematician and programmer Stephen Wolfram agree with the attitude towards doubting consciousness, or more precisely the autonomy of machines. For instance, S. Wolfram has drawn the following paradoxical picture in which we see "a box with a trillion souls, each of which spends forever playing video games" (Brockman, 2019, p. 383). This is a sketch on the topic of the future of civilization, or more precisely the end of this future, in which the machine acquires the desired consciousness (if it is not the machine itself that is looking for it, then at least thinkers who problematize the question of consciousness) and in which there is no place for a human being.

Neural network as a phenomenon

Let us consider the phenomenon of a neural network using ChatGPT as an example. OpenAI's release of what is probably its best-known software product has received significant online attention. The response itself turned out to be surprising for the authors, because... there have already been percentages of such developments being made available to the public (few generative models in 2016-2022). But previous versions had other interfaces that were not accessible to users who do not "speak" programming languages. The correlation between the usability of the interface and the breadth of the user audience is clearly presented in the key stages of neural network development:

1) In 2016, OpenAI for the first time published the results of a study on creating generative models trained on a wide sample of data from a specific field (following the example of the chess program Deep Blue, which beat G. Kasparov in 1996) to generate such data (Karpathy & et al., n.d.);
2) In 2019, the company presented GPT-2, which was one of the first to allow the user to set semantics in the operation of the language model (Ziegler & et al., n.d.);

3) In early 2022, OpenAI published research on InstructGPT demonstrating the ability of the neural network to follow user instructions (Lowe & Leike, n.d.);
4) On November 30, 2022, ChatGPT was put online as a free preview of the GPT-3.5 study (OpenAI, n.d.-d);
5) On March 14, 2023, OpenAI released GPT-4, which constantly passes exams and actively writes codes, thereby frightening programmers (OpenAI, n.d.-b);
6) On May 13, 2024, OpenAI presented GPT-4o with voice accompaniment (OpenAI, n.d.-c);
7) In 2024, we expect an improved version of the neural network, GPT-5, which was confirmed by Sam Altman in his in-depth interview on Lex Friedman's podcast in February (Fridman, n.d.).

Almost every model from the GPT family is heard by both professional and lay audiences. But GPT is not the only one, much less the first of the known LLMs (Large language Models). The first model is considered to be BERT from Google, published in 2018, just a year ahead of GPT-2. Due to the novelty of the phenomenon itself, today we do not have a strict definition of LLM. In the context of programming languages, LLMs belong to deep learning models; their fundamental difference is that they are not focused on a specific task (how to beat a grandmaster in chess, for instance), but are aimed at solving a wide range of problems, so the developers increase the number of parameters for them (up to 1.2 trillion: (Du et al., 2021)) and scale the corpus of texts on which the model is trained (more than a trillion tokens: (Hoffmann et al., 2022)). In comparison with computer programs for chess or Go, the rules of which determine the logic of the program's functioning, the parameters of LLM are dynamic, i.e. they are the values that the model can change during the learning process. From here, the output we get is that each of the known LLMs today solves a wide range of problems, as if recognizing the syntax and semantics of natural language.

But back to ChatGPT; some analysts call it "T9 from your phone, but on bull steroids" (Kotenkov, n.d.-b). Such a comparison is not accidental, because the ChatGPT operator expects the program to provide an exact answer to his request, and in fact, to guess those words that have already been entered into the program.

It is not for nothing that the GPT family bears such a name. Its name is an acronym that stands for 'Generative Pre-trained Transformer'. GPT is a transformer, i.e., a special architecture of a neural network that has been trained to generate text. A wide audience began to gain access to knowledge of the existence of such models only in 2017, with the first reports from Google about the results of its experiments. In a sense, continuing the experience of previous software developments, the transformer is a universal computing mechanism; at the input it receives one pool of data, and at the output it produces the same set

of data, but transformed according to the algorithm. The peculiarity of the functioning of the transformer is that it converts any received signal at the input into a sequence of characters and numbers; it operates not with words, sounds, pictures or videos themselves, but with their variants translated into the language of numbers; hence, in comparison with their analogues, software products before 2017, transformers have broken out of the framework of single-tasking and can potentially be trained to solve a problem of any complexity. But this is only potential for now.

So, the breakthrough of ChatGPT is not only that its phenomenon itself marked the achievement of a new stage in this endless "crawling" in the direction of artificial intelligence, but also that it introduced diverse audiences to neural networks and their multiple abilities. This happened as a result of some experiment by developers from OpenAI, who significantly increased the amount of training data supplied as the model input and the number of its parameters. It was the birth of GPT-2, which was the predecessor of ChatGPT, but the specifics of its functioning remained closed not only to onlookers, but also to professional audiences. In 2019, OpenAI openly stated that they were afraid to release the model into the public domain, so as not to subsequently be responsible for the flow of fake information produced by it (which is actualized in the context of information confrontation and elections in different countries, for instance). However, GPT-4, on the contrary, quickly became available via subscription, although its capabilities, in turn, are many times greater than the potential of GPT-2. This is most clearly represented not in the production of text, but of program code, when the GPT-4 algorithm is enough to create workable code for a person who does not have basic knowledge of programming (Kotenkov, n.d.-a).

Even by the standards of a fast-paced human life, models from the GPT family have come a short way, in terms of time, but a long way in terms of their effectiveness. For example, recently GitHub has taken a course on using the potential of the neural network in the direction of Copilot (*The World's Most Widely Adopted AI Developer Tool*, n.d.), when neural network is not read as a full-fledged subject of work processes, but as an assistant that can reduce the time of solving a problem and scale its result. However, this is about the capabilities of neural network today; we should not forget that the idea of Artificial general intelligence (AGI) is enshrined in the OpenAI Charter (OpenAI, n.d.-c).

Different Studies about Neural Networks

Raising the question of neural network outside the discourse of key players in the IT industry demonstrates that the discourse about neural network began to take shape before the release of any product from the GPT family. This is due to the fact that neural network is not about a specific product or even about an algorithm in the development of AI-models. The idea of an artificial neural network was inspired by the biological neural networks in animal brains (Yang,

2014). Once again there is a collision with the logic in the operation of neural network products outlined above. When we have different datasets at the input and output of the system, it means there is some kind of a concealed layer, the mechanics of which are hidden from the observer. In zoological discourses, such a scheme is applied to the actions of the animal brain. Essentially the same can be applied to the human brain. This is indicated by J. Mark Bishop, scholar from the Department of Computing, University of London.

According to J. Mark Bishop, behind the current public interest in neural networks there is a nearly two-thousand-year history of the mystery of the functioning of the mind (Bishop, n.d.). The scholar identifies three key steps in approaches to solving the puzzle:

1) mind is located in a heart or other organs of a body;
2) Cartesian tradition separated mind and body;
3) scientific discourses of the 20th century identified mind with brain.

Hence J. Mark Bishop offered a definition of neural networks that meets the requirements of contemporary scientific discursive practices. According to this definition, neural networks are not a simple result of cause-and-effect relationships, but "as a consequence of real, embodied neural behavior, emerging in a brain, seated in a body, embedded in a culture and rooted in our world" (Bishop, n.d.). Therefore, it is heuristic to reflect about the specifics of the functioning of computer programs in conjunction with thoughts about the work of mind of living beings. It is noteworthy that in the designated context, a human being is not the only rational being.

Patrik Christen and Olivier Del Fabbro, researchers from the Institute for Information Systems at University of Applied Sciences and Arts Northwestern Switzerland, in their turn, proposed an interesting synthesis of philosophy and programming. In the article "Automatic Programming of Cellular Automata and Artificial Neural Networks Guided by Philosophy", they presented an allagmatic method, which minimizes the operator's difficulty in reading the model's output. The proposed method automates performance models with the least possible constraints and simultaneously interpretable results (Christen & Del Fabbro, 2021). Where is philosophy here, you ask. In fact, it turns out that the article itself does not contain 'philosophical' in its traditional sense; there is precisely the author's interpretation of philosophy, pointing to the meta-model proposed by the authors (here we are supposed to be convinced of the philosophical nature of the model by the well-known root "μετά") as a certain structure consisting of heterogeneous blocks (in the article the authors analyze cases with a simple cellular automaton and an artificial neural network), but nevertheless defining the algorithm in the unified functioning of the blocks. In fairness, it is worth noting that I am making attempts to understand neural networks from a philosophical perspective. So, why can't

programmers, with their terminology and tools, turn to philosophy, albeit in an unconventional reading of it? Probably, they can.

Let us continue thinking within the discursive framework of software developers. For example, in 2021, programmers from University of Science and Technology of China, Shengfei Lyu and Jiaqi Liu, proposed a new approach to machine text processing (Lyu & Liu, 2021). The novelty of the approach is that they synthesized the two most commonly used architectures in text classification, recurrent neural network (which analyzes the text as a whole) and convolutional neural network (which determines the value of the meaning of each word in the text).

The relevance of the approach chosen by the programmers is confirmed in reality itself. Back in the 1980s., one of the key theses of postmodern philosophers, everything is text, raised many questions. In the digital era, we observe not only the correctness of the thesis, but also a peculiar claim to its axiomaticity. In the "ocean" of text, a person must look for mechanisms for adequate processing of incoming information. Consequently, software tools can provide assistance. For example, GPT-4, a "large-scale multimodal model", on receiving information in any format as input, is capable of converting it into a text (OpenAI et al., 2023). Programmers call one of the directions in the functioning of GPT-4, analysis of fake messages, in particular, detecting DeepFakes (Jia et al., 2024).

However, one should not forget that the phenomenon of neural networks was taken by programmers from the discourse of physiologists, naturalists, and even philosophers. For example, Rachapun Rotrattanadumrong and Yohei Yokobayashi, researchers from Okinawa Institute of Science and Technology, Graduate University, define neural networks in the semantic environment of genotypes and phenotypes, in the search for "the mutational robustness and evolvability of biomolecules" (Rotrattanadumrong & Yokobayashi, 2022). The researchers carry out the study of naturalist environment through the use of software tools (from the text of the article: "experimental evaluation of over 120,000 ribozyme sequences"), not traditional methods of chemistry or biology. However, understanding of a neural network using the example of this article is legitimized beyond the software, in the context of "neutral genotypes within the sequence space of an RNA ligase ribozyme".

Sora: a breakthrough in text-to-video system

On February 2024, OpenAI presented the world with its new development, a diffusion model Sora. The presentation is not a release of access to the program code. In the pre-video message, representatives from OpenAI warn that the key goal of the video is to introduce the viewer to the horizons of the developments being implemented in the company; before opening access they must finalize some technical characteristics. In the meantime, technologies like DeepFakes have been circulating the Internet for a relatively long time, at best making stars blush for something of which they are innocent (Webber, n.d.). However, the Internet audience

today has access to a video clip, the image in which is completely generated on the Sora engine. The video was posted on YouTube on February 17, 2024, it collected 2,428,853 views and 45 thousand likes as of March 24, 2024 (OpenAI, n.d.-e). The duration of the video is about 10 minutes; it is an anthology of short (on average 15-30 seconds) videos, different in theme and style. What unites the videos is that they are collected using text-to-video technology, each video is preceded by a text request addressed to Sora (see, for example, "Historical footage of California during gold rush"), and also the fact that none of the videos have been processed. My personal top-3 videos from Sora are an image of a smiling elderly man in a beret, a young Dalmatian running along the windowsills of Burano Italy style houses and a flock of paper airplanes flying through a dense jungle.

On March 25, 2024, an article was published on the OpenAI blog in which developers rejoice at the feedback received from the creative community (including Director Paul Trillo, Digital AR/XR Artist Don Allen Stevenson III, Sculptor/Artist Alex Reben) (OpenAI, n.d.-g). Each of the representatives of art received access to the neural network, created their own peace of art and shared their impressions of interacting with the neural network (see, for example, "As great as Sora is at generating things that appear real – what excites us is its ability to make things that are totally surreal" by multimedia production company "Shy kids"). The videos last from 30 to 90 seconds. Many of them revolve around unrealistic plots. By the way, "Shy kids" really presented a surreal picture; in 1977, David Lynch scared us with his "Erasehead", and "Shy kids" in 2024, through Sora, touch us with the story of "Air Head", a "man" with a ball instead of a head, who lives differently, but "he" does not lose heart and cheerfully shares "his" story with the viewer.

Nevertheless, a more technically advanced audience was attracted by the video about a lady in a bright red dress mentioned earlier in the paper (*OpenAI Sora in Action: Tokyo Walk*, n.d.). Unlike previous results from Sora, the "Tokio Walk" video lasts 60 seconds. A girl in black glasses, a red dress and a black leather jacket walks through evening Tokyo. Just as with the photo supposedly depicting two women, the technical blemishes of the video are not noticeable without close viewing; the proportions of the city change in accordance with the movement of the character, just as the proportions of the figure change in dynamics. The city is reflected in the heroine's glasses and in the puddles on the asphalt. Advertising signs are bright and clear. In general, a quick viewing of the video does not justify the idea of its artificial nature. This caused a stir online. Programmers, for example, began to look for analogies with the classic "Matrix" trilogy, created by the then Wachowski brothers.

Recall the dialogue (more correctly, even a monologue) between Morpheus and Neo to the dynamic track "Clubbed to Death" by Rob Dougan.

> Morpheus: The Matrix is a system, Neo. That system is our enemy. But when you're inside, you look around, what do you see? Business men, teachers, lawyers, carpenters. The very minds of the people we are trying to save. But until we do, these people are still a part of that system, and that makes them our enemy. You have

> to understand, most of these people are not ready to be unplugged. And many of them are so injured, so hopelessly dependent on the system, that they will fight to protect it [Neo's attention is drawn to a passing attractive woman wearing a red dress]. Were you listening to me, Neo, or were you looking at the woman in the red dress?
>
> Neo: I was… (The Matrix, n.d.)

Anyway, some programmers find in the girl's walk from Sora a direct analogy with the famous walk from the film. This means that, behind the simple generation of text-to-video, there may be something complicated.

To get this out, let us first reflect on what Sora is not:

1) Sora was not designed to replace actors and other film industry employees so that producers would make money out of thin air and not owe anyone anything (look into the 2023 SAG-AFTRA strike, for example, and the current Screen Actors Guild – American Federation of Television and Radio Artists specific concern regarding working conditions, when the possibilities of using AI in film production are growing (2023 SAG-AFTRA Strike, n.d.));
2) Sora is probably not a new milestone in the development of DeepFake AI (this version is confirmed by the long road of OpenAI opening access to GPT-2 in order to prevent the flow of fake information);
3) finally, generating and reviving memes is also unlikely to be on the list of development tasks (although "Will Smith" chewing spaghetti will remain in social memory for a long time).

According to the developers, Sora is the company's attempt to teach neural network to understand the physical world, the ability to model it, as well as simulate objects and people's actions (OpenAI, n.d.-a).

Despite the fact that the developers do not publish the Sora program code, discussions are taking place among programmers about what is behind the presented picture. For example, nowadays, software products like Sora have been disassembled into key basic blocks. The first block is an autoencoder with an encoder (video converter) and a decoder, which help compress the image and perform deconstruction from the latent (code, the result of the encoder conversion); the second is a model for predicting the next latent (the model predicts the near future so that all the proportions of the picture are preserved and the face of "Will Smith" does not spread when turning or eating spaghetti); the third is a bot, a kind of generating center inside the system that sees only latents and draws conclusions from them, a kind of analogue to a human brain (based on materials (Kotenkov, n.d.-c)).

It was these blocks that formed the architecture of bots assembled by AI researchers David Ha and Jurgen Schmidhuber. In their paper on recurrent world models, they showed a group of bots trained on the best of optimal strategies in racing and shooting games (Ha & Schmidhuber, 2018). All this was done not only

to demonstrate the scale of AI, but also to "transfer this policy back into the actual environment", i.e., to create models of the world that are produced according to the logic of brain. The results collected by the authors showed that, sometimes, stumbling over an inaccuracy, a "hole" in the simulation, the bot stops acting according to the algorithm. And they immediately proposed a solution related to increasing the capacity of the simulation.

From this we see that OpenAI's neural networks don't just predict the next word or frame; in essence, they are simulations, models of the world that are capable of acting outside the boundaries of the algorithm and the instructions contained in them. Neural networks do not work with pure information; they convert incoming information into code, latent, and process it in the internal layers according to a given algorithm. As a rule, scaling a model leads to its improvement (remember the evolution of GPT). Nowadays, it is difficult to imagine bots as independent agents in work practices. However, increasingly, major companies like Microsoft, Duolingo and Coca Cola are turning to neural networks like GPT-4 to solve internal and external problems (based on materials (Marr, n.d.)). However, every powerful quality of any invention has a downside. Developers have already recognized the accumulation of information about their operator with its subsequent processing as the dark side of GPT (see, for instance (Tsibin, n.d.)). But this is already a problem of digital traces, manipulation of big data and digital literacy of Internet users themselves.

Conclusion

In philosophical reflection on LLMs and the search for consciousness in a machine we come across the classic suffering of Pygmalion: with his own hands, the sculptor was able to create the most beautiful statue and fell in love with it, but he was unable to revive it. According to ancient Greek myth, only the intervention of Aphrodite prevents the tragic end. Having received a body from Pygmalion and a soul from Aphrodite, Galatea turned from a statue into a companion of her creator. But with neural networks, the situation is somewhat more complicated. There is enough data on contemporary models to conclude that neural networks (1) are developing exponentially, (2) are active social agents, (3) are claiming the status of social subjects, and (4) are becoming more and more competent in solving specific problems. Isn't this art demonstrated by a neural network? It remains only to answer the question of whether the human developer and human operator are ready to perceive and accept such art.

After the main text of the paper was written, the author came to the attention of another development from OpenAI, a modified ChatGPT with a built-in voice assistant. With this particular model, the programmers demonstrated that the 2013 science fiction "Her" by Spike Jonze has ceased to be a technological horizon, turning into a contemporaneity. Once again, the frontier development of OpenAI caused a flurry of reactions among Internet audiences, where, along with skepti-

cism and neutral reviews, ideas like “I’ll subscribe and I’ll be happy” were firmly voiced.

A specific case pushes us to think about “phygital” (for example, about the possibility of relationships between a human and a non-human). It throws up another argument in favor of the assertion that the humanocentric picture of the world with the recognition of human exceptionalism is in question. Non-human agents, nature or technology, have already established their positions in social arenas and are increasingly declaring their social subjectivity. Neural networks create new things and become more and more human-like. It remains only to answer the question of whether the human developer and human operator are ready to perceive and accept it.

Acknowledgments

The research was supported by Russian Science Foundation “Conducting fundamental scientific research and exploratory scientific research by individual scientific groups” (regional competition) 22-18-20011 “Digital literacy: interdisciplinary research (regional aspect)”.

References

2023 SAG-AFTRA *strike*. (n.d.). Wikipedia. https://en.wikipedia.org/wiki/2023_SAG-AFTRA_strike

AI Will Smith eating spaghetti pasta (AI footage and audio). (n.d.). YouTube. https://www.youtube.com/watch?v=XQr4Xklqzw8

Baars, B. J. (1988). A *Cognitive Theory of Consciousness*. Cambridge University Press.

Bishop, J. M. (2015). History and philosophy of neural networks. In H. Ishibuchi (Ed.), *Computational Intelligence* (pp. 2–73). Encyclopedia of Life Support Systems (EOLSS). https://www.researchgate.net/publication/271841595%0AHISTORY

Brockman, J. (Ed.). (2019). *Possible minds: Twenty-five ways of looking at AI*. Penguin Press.

Chalmers, D. J. (2023). *Could a Large Language Model be Conscious?* https://doi.org/10.48550/ARXIV.2303.07103

Christen, P., & Del Fabbro, O. (2021). Automatic Programming of Cellular Automata and Artificial Neural Networks Guided by Philosophy. In R. Dornberger (Ed.), *New Trends in Business Information Systems and Technology* (Vol. 294, pp. 131–146). Springer International Publishing. https://doi.org/10.1007/978-3-030-48332-6_9

Creating video from text. (n.d.). OpenAI. https://openai.com/sora

Du, N., Huang, Y., Dai, A. M., Tong, S., Lepikhin, D., Xu, Y., Krikun, M., Zhou, Y., Yu, A. W., Firat, O., Zoph, B., Fedus, L., Bosma, M., Zhou, Z., Wang, T., Wang, Y. E., Webster, K., Pellat, M., Robinson, K., … Cui, C. (2021). *GLaM: Efficient Scaling of Language Models with Mixture-of-Experts* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2112.06905

Fridman, L. (n.d.). *Sam Altman: OpenAI, GPT-5, Sora, Board Saga, Elon Musk, Ilya, Power & AGI | Lex Fridman Podcast #419*. YouTube. https://www.youtube.com/watch?v=jvqFAi7vkBc

Fuller, M. (2008). *Software studies: A lexicon*. MIT Press. https://doi.org/10.7551/mitpress/9780262062749.001.0001

GPT-4. (2023, March 14). OpenAI. https://openai.com/research/gpt-4

Ha, D., & Schmidhuber, J. (2018). Recurrent World Models Facilitate Policy Evolution. *Neural Information Processing Systems*, 31.

Hello GPT-4o. (2024, May 13). OpenAI. https://openai.com/index/hello-gpt-4o/

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. de L., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., Driessche, G. van den, Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., … Sifre, L. (2022). *Training Compute-Optimal Large Language Models* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2203.15556

How to tame the neural net. (2024, January 30). Habr. https://habr.com/ru/amp/publications/790138/ (In Russian).

Introducing ChatGPT. (2022, November 30). OpenAI. https://openai.com/blog/chatgpt

Introducing Sora—OpenAI's text-to-video model. (n.d.). YouTube. https://www.youtube.com/watch?v=HK6y8DAPN_0

Jabborov, D. (2022, April 1). *Moscow grandmother helped Yandex's robot courier to cross the road.* Gazeta.Ru. https://www.gazeta.ru/tech/news/2022/04/01/17508595.shtml (In Russian).

Jia, S., Lyu, R., Zhao, K., Chen, Y., Yan, Z., Ju, Y., Hu, C., Li, X., Wu, B., & Lyu, S. (2024). *Can ChatGPT Detect DeepFakes? A Study of Using Multimodal Large Language Models for Media Forensics* (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2403.14077

Karpathy, A., Abbeel, P., Brockman, G., Chen, P., Cheung, V., Duan, Y., Goodfellow, I., Kingma, D., Ho, H., Houthooft, R., Salimans, T., Schulman, J., Sutskever, I., & Zaremba, W. (2016, June 16). *Generative models*. https://openai.com/research/generative-models

Kotenkov, I. (2023a, March 18). *GPT-4: What the new neural network has learned, and why it's a little creepy*. Habr. https://habr.com/ru/companies/ods/articles/722644/ (In Russian).

Kotenkov, I. (2023a, March 6). *How ChatGPT works: Explaining in simple Russian the evolution of language models from T9 to miracle*. Habr. https://habr.com/ru/companies/ods/articles/716918/ (In Russian).

Kotenkov, I. (2024, February 28). *The Real Purpose of OpenAI SORA: How and Why to Simulate the "Matrix" for ChatGPT*. Habr. https://habr.com/ru/articles/794566/ (In Russian).

Lowe, R., & Leike, J. (2022, January 27). *Aligning language models to follow instructions*. Openai. https://openai.com/research/instruction-following

Lyu, S., & Liu, J. (2021). Convolutional Recurrent Neural Networks for Text Classification. *Journal of Database Management*, *32*(4), 65–82. https://doi.org/10.4018/JDM.2021100105

Mamardashvili, M. K., & Piatigorsky, A. M. (2009). *Symbol and consciousness*. Gnosis (In Russian).

Manovich, L. (2017). *Theories of Theories of Software Culture.* Krasnaya lastochka. (In Russian).

Manovich, L. (2019). *AI Aesthetics*. Strelka Press.

Marr, B. (2023, May 30). *10 Amazing Real-World Examples Of How Companies Are Using ChatGPT In 2023*. Forbes. https://www.forbes.com/sites/bernardmarr/2023/05/30/10-amazing-real-world-examples-of-how-companies-are-using-chatgpt-in-2023/

OpenAI Charter. (n.d.). OpenAI. https://openai.com/charter

OpenAI Sora in Action: Tokyo Walk. (n.d.). YouTube. https://www.youtube.com/watch?v=tRSdt5kmeW0

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2023). *GPT-4 Technical Report* (Version 6). arXiv. https://doi.org/10.48550/ARXIV.2303.08774

Photographer admits prize-winning image was AI-generated. (2023, April 17). *The Guardian*. https://www.theguardian.com/technology/2023/apr/17/photographer-admits-prize-winning-image-was-ai-generated

Qin, Z., Zhang, H., Qin, X., Xu, K., Dimitrov, K. N. A., Wang, G., & Yu, W. (2018). *Fundamentals of Software Culture*. Springer Singapore. https://doi.org/10.1007/978-981-13-0701-0

Rotrattanadumrong, R., & Yokobayashi, Y. (2022). Experimental exploration of a ribozyme neutral network using evolutionary algorithm and deep learning. *Nature Communications*, 13(1), 4847. https://doi.org/10.1038/s41467-022-32538-z

Schwab, K. (2016). *The fourth industrial revolution*. Crown Business.

Sora: First impressions. (2024, March 25). OpenAI. https://openai.com/blog/sora-first-impressions

The Matrix. (n.d.). IMDb. https://www.imdb.com/title/tt0133093/

The world's most widely adopted AI developer tool. (n.d.). GitHub. https://github.com/features/copilot

Tiku, N. (2022, June 11). The Google engineer who thinks the company's AI has come to life. *Washington Post*. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/

Webber, S. (n.d.). A *Taylor Swift Sex Tape*? IMDb. https://m.imdb.com/news/ni46594315/

Yang, Z. (2014). *Comprehensive Biomedical Physics*. Elsevier.

Ziegler, D., Wu, J., Brown, T., Amodei, D., Radford, A., Christiano, P., & Irving, G. (2019, September 19). *Fine-tuning GPT-2 from human preferences*. OpenAI. https://openai.com/research/fine-tuning-gpt-2

Список литературы

2023 SAG-AFTRA *strike*. (n.d.). Wikipedia. https://en.wikipedia.org/wiki/2023_SAG-AFTRA_strike

AI Will Smith eating spaghetti pasta (AI footage and audio). (n.d.). YouTube. https://www.youtube.com/watch?v=XQr4Xklqzw8

Baars, B. J. (1988). A *Cognitive Theory of Consciousness*. Cambridge University Press.

Bishop, J. M. (2015). History and philosophy of neural networks. In H. Ishibuchi (Ed.), *Computational Intelligence* (pp. 2–73). Encyclopedia of Life Support Systems (EOLSS). https://www.researchgate.net/publication/271841595%0AHISTORY

Brockman, J. (Ed.). (2019). *Possible minds: Twenty-five ways of looking at AI*. Penguin Press.

Chalmers, D. J. (2023). *Could a Large Language Model be Conscious*? https://doi.org/10.48550/ARXIV.2303.07103

Christen, P., & Del Fabbro, O. (2021). Automatic Programming of Cellular Automata and Artificial Neural Networks Guided by Philosophy. In R. Dornberger (Ed.), *New Trends in Business Information*

Systems and Technology (Vol. 294, pp. 131–146). Springer International Publishing. https://doi.org/10.1007/978-3-030-48332-6_9

Creating video from text. (n.d.). OpenAI. https://openai.com/sora

Du, N., Huang, Y., Dai, A. M., Tong, S., Lepikhin, D., Xu, Y., Krikun, M., Zhou, Y., Yu, A. W., Firat, O., Zoph, B., Fedus, L., Bosma, M., Zhou, Z., Wang, T., Wang, Y. E., Webster, K., Pellat, M., Robinson, K., … Cui, C. (2021). *GLaM: Efficient Scaling of Language Models with Mixture-of-Experts* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2112.06905

Fridman, L. (n.d.). *Sam Altman: OpenAI, GPT-5, Sora, Board Saga, Elon Musk, Ilya, Power & AGI | Lex Fridman Podcast #419*. YouTube. https://www.youtube.com/watch?v=jvqFAi7vkBc

Fuller, M. (2008). *Software studies: A lexicon*. MIT Press. https://doi.org/10.7551/mitpress/9780262062749.001.0001

GPT-4. (2023, March 14). OpenAI. https://openai.com/research/gpt-4

Ha, D., & Schmidhuber, J. (2018). Recurrent World Models Facilitate Policy Evolution. *Neural Information Processing Systems*, 31.

Hello GPT-4o. (2024, May 13). OpenAI. https://openai.com/index/hello-gpt-4o/

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. de L., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., Driessche, G. van den, Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., … Sifre, L. (2022). *Training Compute-Optimal Large Language Models* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2203.15556

Introducing ChatGPT. (2022, November 30). OpenAI. https://openai.com/blog/chatgpt

Introducing Sora—OpenAI's text-to-video model. (n.d.). YouTube. https://www.youtube.com/watch?v=HK6y8DAPN_0

Jia, S., Lyu, R., Zhao, K., Chen, Y., Yan, Z., Ju, Y., Hu, C., Li, X., Wu, B., & Lyu, S. (2024). *Can ChatGPT Detect DeepFakes? A Study of Using Multimodal Large Language Models for Media Forensics* (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2403.14077

Karpathy, A., Abbeel, P., Brockman, G., Chen, P., Cheung, V., Duan, Y., Goodfellow, I., Kingma, D., Ho, H., Houthooft, R., Salimans, T., Schulman, J., Sutskever, I., & Zaremba, W. (2016, June 16). *Generative models*. https://openai.com/research/generative-models

Lowe, R., & Leike, J. (2022, January 27). *Aligning language models to follow instructions*. Openai. https://openai.com/research/instruction-following

Lyu, S., & Liu, J. (2021). Convolutional Recurrent Neural Networks for Text Classification. *Journal of Database Management*, 32(4), 65–82. https://doi.org/10.4018/JDM.2021100105

Manovich, L. (2019). *AI Aesthetics*. Strelka Press.

Marr, B. (2023, May 30). *10 Amazing Real-World Examples Of How Companies Are Using ChatGPT In 2023*. Forbes. https://www.forbes.com/sites/bernardmarr/2023/05/30/10-amazing-real-world-examples-of-how-companies-are-using-chatgpt-in-2023/

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2023). *GPT-4 Technical Report* (Version 6). arXiv. https://doi.org/10.48550/ARXIV.2303.08774

OpenAI Charter. (n.d.). OpenAI. https://openai.com/charter

OpenAI Sora in Action: Tokyo Walk. (n.d.). YouTube. https://www.youtube.com/watch?v=tRSdt5kmeW0

Photographer admits prize-winning image was AI-generated. (2023, April 17). *The Guardian*. https://www.theguardian.com/technology/2023/apr/17/photographer-admits-prize-winning-image-was-ai-generated

Qin, Z., Zhang, H., Qin, X., Xu, K., Dimitrov, K. N. A., Wang, G., & Yu, W. (2018). *Fundamentals of Software Culture*. Springer Singapore. https://doi.org/10.1007/978-981-13-0701-0

Rotrattanadumrong, R., & Yokobayashi, Y. (2022). Experimental exploration of a ribozyme neutral network using evolutionary algorithm and deep learning. *Nature Communications*, 13(1), 4847. https://doi.org/10.1038/s41467-022-32538-z

Schwab, K. (2016). *The fourth industrial revolution*. Crown Business.

Sora: First impressions. (2024, March 25). OpenAI. https://openai.com/blog/sora-first-impressions

The Matrix. (n.d.). IMDb. https://www.imdb.com/title/tt0133093/

The world's most widely adopted AI developer tool. (n.d.). GitHub. https://github.com/features/copilot

Tiku, N. (2022, June 11). The Google engineer who thinks the company's AI has come to life. *Washington Post*. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/

Webber, S. (n.d.). *A Taylor Swift Sex Tape?* IMDb. https://m.imdb.com/news/ni46594315/

Yang, Z. (2014). *Comprehensive Biomedical Physics*. Elsevier.

Ziegler, D., Wu, J., Brown, T., Amodei, D., Radford, A., Christiano, P., & Irving, G. (2019, September 19). *Fine-tuning GPT-2 from human preferences*. OpenAI. https://openai.com/research/fine-tuning-gpt-2

Джабборов, Д. (2022, апрель 1). *Московская бабушка помогла перейти дорогу роботу-курьеру «Яндекса»*. Газета.Ru. https://www.gazeta.ru/tech/news/2022/04/01/17508595.shtml

Как приручить нейросеть. (2024, январь 30). Хабр. https://habr.com/ru/amp/publications/790138/

Котенков, И. (2023a, март 6). *Как работает ChatGPT: Объясняем на простом русском эволюцию языковых моделей с T9 до чуда*. Хабр. https://habr.com/ru/companies/ods/articles/716918/

Котенков, И. (2023b, март 18). *GPT-4: Чему научилась новая нейросеть, и почему это немного жутковато*. Хабр. https://habr.com/ru/companies/ods/articles/722644/

Котенков, И. (2024, февраль 28). *Настоящее предназначение OpenAI SORA: Как и зачем симулировать «Матрицу» для ChatGPT*. Хабр. https://habr.com/ru/articles/794566/

Мамардашвили, М. К., & Пятигорский, А. М. (2009). *Символ и сознание*. Гнозиз.

Манович, Л. (2017). *Теории софт-культуры*. Красная ласточка.