

Issues associated with data generation in artificial intelligence

Tamara Gairabekova ^{1*}, *Timur Aygumov* ², and *Elena Belashova* ³

¹ Kadyrov Chechen State University, Grozny, Russia

² Dagestan State Technical University, I. Shamil Av. 70, Dagestan, Russia

³ Kazan National Research Technical University named after A. N. Tupolev-KAI, Kazan, Russia

Abstract. Artificial intelligence is the late popular technology that can be found in any possible field. The spread of this technology like a virus is due to the improvement of technology and the methods and the wide knowledge of humanity. The technology is conquering areas that one not even imagine and it is turning them into refined products. However, there are always some issues that occur at the beginning or in the process. In this case, the second one is the case. The issue with technology, in some areas, is that it can collapse on itself. The overgeneration of data could lead to unwanted outcomes. Therefore, the work will illustrate how artificial intelligence, in some cases, can be destroying itself. Meaning losing the quality that it had at the beginning of its use.

1 Introduction

That latest so-called revolution in algorithmic improvement led to dramatic changes to the world we know. It introduces new paths to solving existing issues. But most importantly these tools eased some processes for humans. These tools immediately took the mark and reformed some old methods to new shapes. Most of these tools are used daily and some are working nonstop without anyone's Knowledge. The driver of these tools is cleverly written algorithms or another favorable name in the media is neural networks. These algorithms are built in a way that they could mimic some humans' capabilities. In some cases, they are far ahead of what humans could achieve, and in some, they lack in many aspects.

The very nature of these algorithms is hidden in their cleverly written code. The idea that these neural networks could somehow threaten humanity is groundless. These algorithms work within the given boundaries and it is impossible that they somehow will get the idea of self-consciousness and start acting on their own. Even combined with other technology it is capable of handling the tasks that it was programmed to. Of course, there is a chance that the technology could go on a different path, but it is because of the errors that were not considered when writing the code or some outer influence. It is a fact that on its own the technology collapses and runs into issues. Therefore, the article will focus on how the technology (artificial intelligence) can destroy itself or at least affect negatively its quality.

* Corresponding author: sti_ing@mail.ru

Artificial intelligence or neural networks is used, as it was mentioned before, in many aspects of human life. Industries utilize them a great deal. Every technology or electronic device has them embedded. Software and apps are using them to generate and edit media files. The most common examples are ChatGPT, Bard AI, Midjourney, and so on. These tools are so popular due to their ability to generate data depending on the input values. The issue arises when the same data is processed over and over. The initial meaning is lost and the quality too. The article will look into the idea of how can this technology run on issues that it introduces itself [1, 2].

2 The information generated

2.1 Text

The most solid outcome of the technology is text. The generation of text is more refined if compared to other generated pieces. This technology is capable of giving results in the most unimaginable way. Most of the time it outruns the expectations of the users. It is not the first time that we have this kind of tool in our hands. In the last two decades, google was the one, which provided the necessary information once the request was made. Of course, the results were just possible links to the information. As years passed the algorithms were improved and, hence, it better understands the human request. But with the improvement of algorithmic components different ideas and paths emerged. The most notable tool to occur from this development was ChatGPT. This tool elevated the idea behind information search to the next level. Its capabilities are growing daily. Therefore, it has become the favorite tool to be used by the masses to generate the necessary text. It is capable of generating fake information as well as gathering data from all over the fine resources. There is a chance that the results it found not always will be correct as it uses the data that has been fed to it by the developers. It can generate all types of text depending on the request. For example, different styles can be generated such as poetic, scientific, grammatically incorrect, formal, and many other styles. It can easily change or rephrase partially or fully the generated text. It can merge two texts together or expand one that is too small for some reason. Make an accent on some parts of the text and respect to results it will generate further. It can easily generate from a short note create a full story and reverse the process. The capabilities of this tool are immense and will continue further. There are different tools such as ChatGPT, for instance, Bart ai, which is the product of Google. In some cases, the second mentioned will provide better results and, in some cases, the first one will give a better outcome [3, 4].

2.2 Image

For the moment the ChatGPT tool is expanding its capabilities introducing more add-ins to its product. For example, in the latest version 4.0. can generate photos by the given prompts. Its capabilities to generate images through add-ins such as DALEE is mesmerizing. It will generate the image through simple keywords. If provided more description the image on the exit will be repeating in itself all the given inputs. The technology can even understand the image that has been fed to it. It can describe it in detail. Even tell the characteristics and the effect that has been added to the image. Some of the images are so realistic that it is hard to distinguish between generated and real ones. The other tool that is in the market is Midjourney. This tool similar to ChatGPT can generate data with some input values. This tool is more focused on images while ChatGPT focuses on text. Midjourney was the first tool powered by artificial intelligence that attracted some attention. Due to its capability to generate any image with simple prompts, it became a sensation in the space of the internet

and quickly moved to news, which gave it a ground to quickly build its path to the masses. Again, many similar tools can perform the same task and differ a little bit, but get not as far as this tool if compared to quality and prompts understanding [5].

2.3 Video and others

The other thing that neural networks can imitate is video. This is the most complex information for the technology to imitate or generate. Videos include different elements that need to synchronize to the overall idea. Therefore, the outcome of such a generation could look like a mess or episodes from a dream. Although, shortly there will be some progress as it is already getting better. There are even some advertisements and movie introductions done using this technology. With the clever edition and using other materials and technology, the result feels solid. There are other things that this technology can do except text, photos, and videos. The first thing that comes after text, photos, and videos is code. Neural networks are themselves a code that is written in a way that could generate or process the information. The technology can write simple programs and run them working. Sometimes it can introduce a mess or incoherent code. With a more precise direction and knowledge, one can take usefulness from it. Another way it is not as perfect as text generation. Other thing that this technology can do is websites. From the provided information, like text or raw material, or feeding it with the website (screen or link) that one wants to imitate, this technology is capable of providing the ready design of the site that can be further refined by other tools [6].

3 Overgeneration

3.1 Blurriness

The following paragraph will look at the issues that are introduced by the overgeneration of data. The most common path that this scenario leads is to difficulties in recognizing the real and quality data. This technology brings unclarity to the available data, as it makes it hard to distinguish generated data from others as the technology advancement, in some cases, can generate the data that is better in many criteria. Therefore, generated data appeals to have smooth angles whereas real data can be lost in the sea of chaotically generated data [7].

3.2 Loop collapse

Collapse of the loop of the generative tool. When iteration is introduced in a closed loop where information is scooped from the previously generated data. This scenario presents in itself unwanted outcomes. The results generated from this type of loop can lose the initial aim or meaning. It also will lose the main idea which will lead to a generation of data that is no longer needed. There is a great chance that the information will lose not just the initial meaning but the sense in general. The generative tool that is stacked on the loop generation will lack access to new or necessary information, that would enrich the data in the end process, resulting in a meaningless outcome [8, 9].

3.3 Space junk

The generated information is stored on different servers, devices, hardware, etc. There is no certainty that every generated information will end up being used for a crucial task. Therefore, most of the space of storage apparatuses will be taken by the generated data. Not all of the data that has been generated is essential. Most of the data from generative tools will

end up floating in servers with no clear purpose. These generated data serve no good for the servers, users, or future data [10].

3.4 Classification issues and misleading

The one thing the technology is good at is mixing unmixable. The technology can easily generate, for instance, text about viruses and how they are dangerous for humanity. If adjusted with extra prompts to generate text where viruses are furniture the technology will easily generate the data. The technology is also capable of mixing different styles together when generating images. This type of generation can easily confuse the viewer or mislead believing that this type of mixture exists. The all above issues of overgeneration are mostly concerned with the uselessness of the junk portion of generated information that leads to unwanted results like loss of quality, misleading, space occupation, toOl collapse, and many more [11].

4 Solution

The first and main solution for the technology is to understand its own data. The technology must recognize the data it generates. With such understanding, it can mark the data that it generated or partially improved and, hence, make decisions on further steps. Another way of improving is improving. It might sound odd, but improvement in this area can solve some issues related to the generation of information in a closed loop. A cleverly written algorithm can solve the problem and push the limits of the technology to a new level. Recognition of the problems related to generated data could improve the outcome. If the technology can analyze generated information, after generation and then improve it depending on the occurred issues, it could lead to better results [12].

5 Future perspectives

Similar to any other technology it is clear that the technology will be improving. Even with its present tendency, one can easily say it is improving daily in some aspects while some are far behind. More refined artificial intelligence will be prone to provide more accurate results. It is suggested that the technology will level up in the near decade. The stage of the technology that we are using is called weak Artificial intelligence and the one that will come after is mediocre, which is believed to be in the same plane as our capabilities. The third stage is strong Artificial intelligence which in theory must exceed the capabilities of humans. Therefore, the technology will be evolving and, hence, some of its issues will be solved. The technology will be improving accordingly to other technological progress and derive its capabilities from technological combinations. Although, there is a great chance that some other technological revolution might overshadow the technology. It is clear that the frequency of technological revolutions is increasing and the chance of a new one is inevitable [13, 14].

6 Conclusion

To conclude, the following work was done to present the issue of over generation of information. Artificial intelligence is one of the most important technologies that is driving most industries and will increase its importance further by improving itself. Technology is also embedded in our lives without our knowledge and its issues must be understood and brought to light. The work briefly explained the terminology of neural networks and also outlined its importance in our daily lives. The main focus of the work was put on two

paragraphs, which are related to the data the technology can generate and the idea of overgeneration and its negative effect. Also, this work has brief coverage of the solutions to the issues and future perspectives of the technology.

References

1. C. Rios-Campos, J. D. C. L. Viteri, E. A. P. Batalla, J. F. C. Castro, J. B. Núñez, E. V. Calderón., F. J. G. Nicacio M. Y. P. Tello, Generative Artificial Intelligence. South Florida Journal of Development, **4(6)**, 2305–2320 (2023)
2. J. L. Ordoñez-Avila, and M. G. Martínez-Rangel, Propuesta de Modelo de Sistemas Abiertos e Inteligentes Agricultura en Fincas Cafetaleras dentro de la Cosecha Proceso. II LACCEI International Multiconference on Entrepreneurship, Innovation and Regional Development - LEIRD 2022
3. C. Kumar, "Artificial Intelligence: Definition, Types, Examples, Technologies," 31 August 2018. [Online].
4. I. Magomedov, E. Belashova and M.-D. Bersanov, *Review article: enhancing the power of artificial intelligence in mechanical design*, International Scientific Siberian Transport Forum – TransSiberia. Vol. **402** (2023)
5. I. A. Magomedov, Z. A. Dzhabrailov and A. A. Elmurzaev, Journal of Physics: Conference Series, **1889**, (2021)
6. S. N. Cherny and R. F. Gibadullin, "*The Recognition of Handwritten Digits Using Neural Network Technology*", International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), Sochi, Russian (2022)
7. I. A. Magomedov, T. G. Aygumov and N. I. Pikuleva, Possibilities of deterioration of artificial intelligence. 2nd International Conference on Environmental Sustainability Management and Green Technologies, vol **451**. (2024)
8. A. K. Rakhmatullin & R. F. Gibadullin, *Synthesis and analysis of elementary algorithms for a differential neural computer*. Lobachevskii Journal of Mathematics, vol. **43(2)**, pp. 473-483. (2022)
9. B. Cheatham, K. Javanmardian, and H. Samandari, *Confronting the risks of artificial intelligence*. McKinsey (2019)
10. V. A. Gerasimov, M. G. Nuriev and D. A. Gashigullin, International Russian Automation Conference (2022)
11. I.A. Magomedov, K.V. Mashukov, E.S. Kremleva, Future trends in artificial intelligence that could pose a threat to humanity. E3S Web of Conferences. Vol. **451**, 05012 (2023)
12. M. Nuriev, R. Zaripova, S. Sitnikov, & R. Nurgaliev, *The principles of building a parallel program for steganographic file protection*. In E3S Web of Conferences. Vol. **458**, p. 09027 (2023)
13. J. Shabbir & T. Anwer, *Artificial Intelligence and its Role in Near Future*. ArXiv. **14(8)**, pp. 1–11, (2015)
14. I. Magomedov, A. Bagov and A. Tkachenko, *Future perspectives of artificial intelligence*, International Scientific and Practical Conference “Development and Modern Problems of Aquaculture. Vol. **84** (2024)