MAKÜFEBED

ISSN Online: 1309-2243
http://dergipark.org.tr/makufebed
https://doi.org/10.29048/makufebed.1271477

Mehmet Akif Ersoy Üniversitesi Fen Bilimleri Enstitüsü Dergisi 14(2): 365-376 (2023)
The Journal of Graduate School of Natural and Applied Sciences of Mehmet Akif Ersoy University 14(2): 365-376 (2023)

Derleme Makale / *Review Paper*

ChatGPT with Risks and Opportunities

Tarık KABAK [1*], İsmail KIRBAŞ [2]

[1*]Burdur Mehmet Akif Ersoy University Graduate School of Natural and Applied Sciences, Department of Computer Engineering, Burdur-Turkey
[2]Burdur Mehmet Akif Ersoy University, Faculty of Engineering and Architecture, Department of Computer Engineering, Burdur-Turkey

Geliş Tarihi (Received): 27.03.2023, Kabul Tarihi (Accepted): 11.07.2023
Sorumlu Yazar (Corresponding author): tkabak@mehmetakif.edu.tr*
+90 248 2131250 *+90 248 2131251*

ABSTRACT

Today, as a result of significant advances in Artificial Intelligence (AI), technology giants such as Google, Microsoft, Apple and many other companies or institutions use and continue to develop artificial intelligence models for specific purposes. Different artificial intelligences, which are continuously trained using long and various training methods, are customized for specific purposes and serve to provide benefits in those areas. ChatGPT, a chatbot being developed by OpenAI, which we have started to hear a lot about recently, has the potential to significantly change the way we interact with technology. The most important skill of the ChatGPT chatbot is its ability to understand texts, images and generate new texts with high accuracy, even almost human-like, with the help of the huge amount of data it processes during its training. In this paper, we present an overview of the developments in the Large Language Model (LLM), ChatGPT (Generative Pre-trained Transformer), which has been the subject of much debate, concern, and even conspiracy theories, and the benefits or problems it may cause. This opens up ground-breaking possibilities for the areas of Natural Language Processing (NLP) and Natural Language Understanding (NLU). Built on the GPT language model, ChatGPT offers a wide range of possibilities, from law to medicine, from mathematics to finance, from academic article writing to language translation and even security vulnerability detection.

Keywords: Artificial Intelligence, chatbot, ChatGPT, Large Language Model, Natural Language Processing.

Riskleri ve Fırsatları ile ChatGPT

ÖZ

Günümüzde Yapay Zeka (Artificial Intelligence-AI) alanındaki önemli ilerlemeler sonucunda Google, Microsoft ve Apple gibi teknoloji devleri ve daha birçok firma ya da kurum belirli amaçlar için yapay zeka modellerini kullanmakta ve geliştirmeye de devam etmektedir. Uzun ve çeşitli eğitim yöntemleri kullanılarak sürekli olarak eğitilen farklı yapay zekalar, belirli amaçlara yönelik özelleştirilerek o alanlarda fayda sağlamaları amacıyla hizmet vermektedir. Son günlerde özellikle adını sıkça duymaya başladığımız, OpenAI firması tarafından geliştirilmekte olan ChatGPT isimli sohbet botu; teknoloji ile etkileşimimizi önemli ölçüde değiştirebilecek potansiyele sahiptir. ChatGPT sohbet botunun en önemli becerisi; eğitimi sırasında işlediği çok büyük miktarda verinin de yardımıyla yüksek doğrulukta, hatta neredeyse insana yakın denebilecek ölçüde metinleri anlayıp yeni metinler üretebilmesidir. Birçok tartışmalara, kaygılara, hatta komplo teorilerine konu olan ChatGPT (Generative Pre-trained Transformer-GPT-Önceden Eğitilmiş Üretken Dönüştürücü) olarak adlandırılan büyük dil modeli (Large Language Model-LLM) konusundaki gelişmelere ve sağlayabilecekleri faydalara veya yol açabilecekleri sorunlara dair bir değerlendirme sunulmaktadır. Bu sayede de doğal dil işleme (Natural Language Processing-NLP) ve doğal dil anlama (Natural Language Understanding-NLU) alanlarında devrimsel olasılıklar sunmaktadır. GPT dil modeli üzerine inşa edilen ChatGPT'nin sunduğu bu olasılıklar

Tarık KABAK, https://orcid.org/0000-0001-6845-5915
İsmail KIRBAŞ, https://orcid.org/0000-0002-0287-2334

içerisinde hukuktan tıpa, matematikten finansa, akademik makale yazımından dil tercüme işlemlerine hatta güvenlik açıklarının tespitine kadar geniş bir yelpaze bulunmaktadır.

Anahtar Kelimeler: Büyük Dil İşleme, ChatGPT, Doğal Dil İşleme, sohbet, Yapay Zekâ.

INTRODUCTION

Today, increasing developments in the field of artificial intelligence bring to light some concerns as well as a great potential. Considering what artificial intelligence models can and can do, both results are undeniable. For example, the impact of AI models on the education workforce raises concerns about what to teach and how to teach future generations (Zhai, 2021). In addition, in recent years, artificial intelligence has also shown the potential to partially, in some cases fully replace, creative work, such as academic writing, coding, and art, which is usually performed only by talented professionals (Li et al., 2022).

Several artificial intelligence models, represented in Figure 1, were developed and trained for certain tasks. Each model undergoes various trainings in order to fulfill the targeted operations. These trainings generally differ according to the input and output relationship. Even models trained on the same input and output formats may have different training processes.

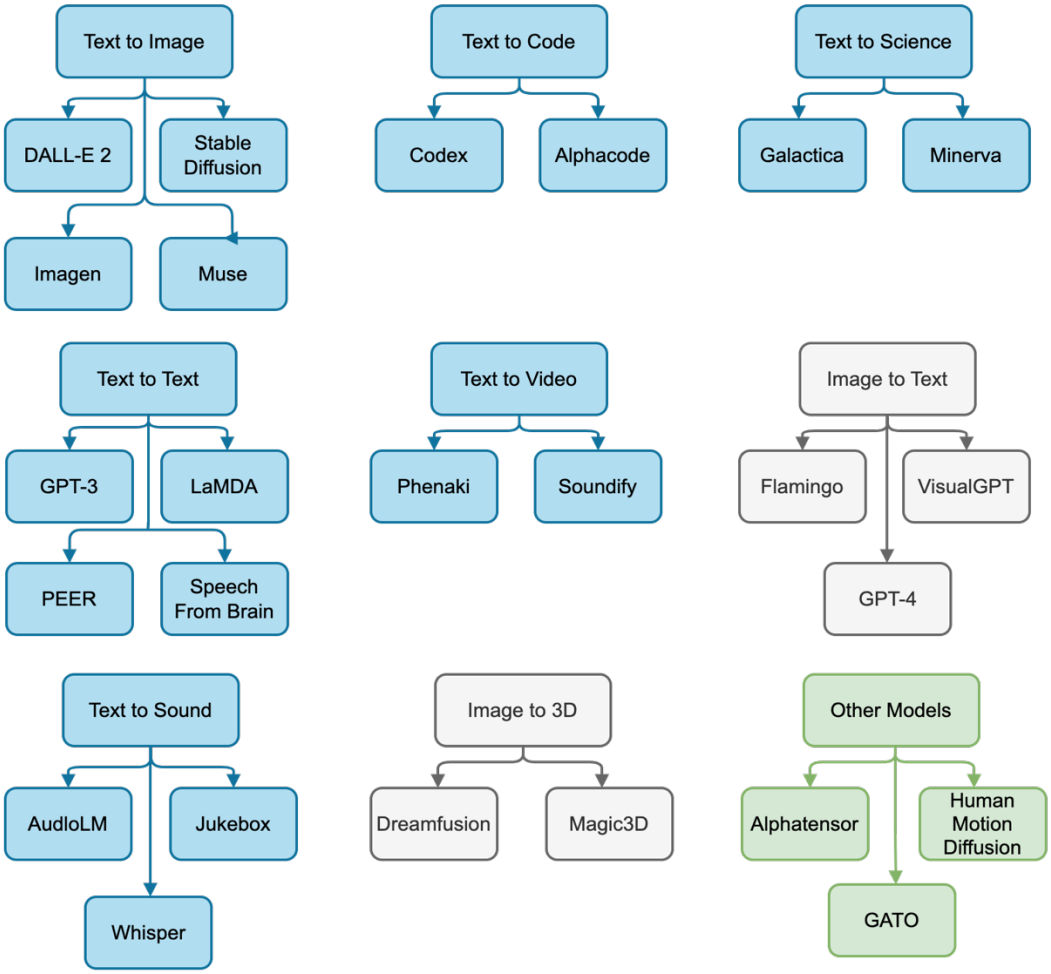

Figure 1. Some artificial intelligence models according to input and output formats

This study focuses on ChatGPT, which is considered today's one of the most intelligent artificial intelligence models and has brought many debates, to draw a general framework about language models, to examine the capabilities of ChatGPT, which uses one of the language models, the big language model, to understand its limits and to focus on its potential in areas such as finance, education, health, law, coding, and even academic writing, as well as the risks it may bring in similar or different areas.

One of the sub-fields of artificial intelligence is natural language processing. As a result of the widespread use of the internet around the world, the amount of data uploaded to the internet environment is increasing day by day with the increase in the number of users. As an inevitable result of this situation, with the proliferation and proliferation of large datasets and the ease of access to these datasets, researchers have developed artificial intelligence models that can provide conversational responses that they train with the help of these easily accessible datasets.

Today, there are dozens of AI models that have been trained in different ways for different purposes by many different developers and released to the market. Figure 2 shows some of the most well-known of these artificial intelligence models developed according to their release dates.

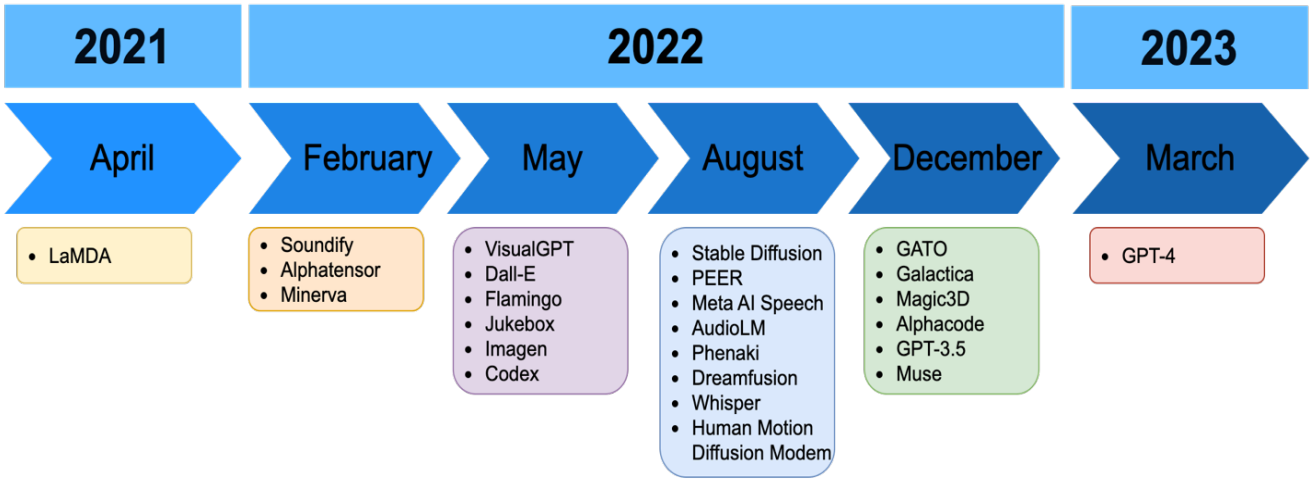

Figure 2. Some of the AI models according to their release dates

Technology companies such as Google, Apple, Nvidia, OpenAI train different artificial intelligence models for different purposes in line with their own goals. Figure 3 shows which companies have developed popular artificial intelligence models.

Although different models are used for language training, one of the most widely used models is the large language model. The ability of the big language model is to read texts, interprete images, translate them when necessary to understand them, and predict the next word in a sentence to be written in order to write and comprehend them in a similar way to humans. Perhaps the most popular artificial intelligence model using this type of model is the language model called GPT, which was also based on the ChatGPT chatbot.

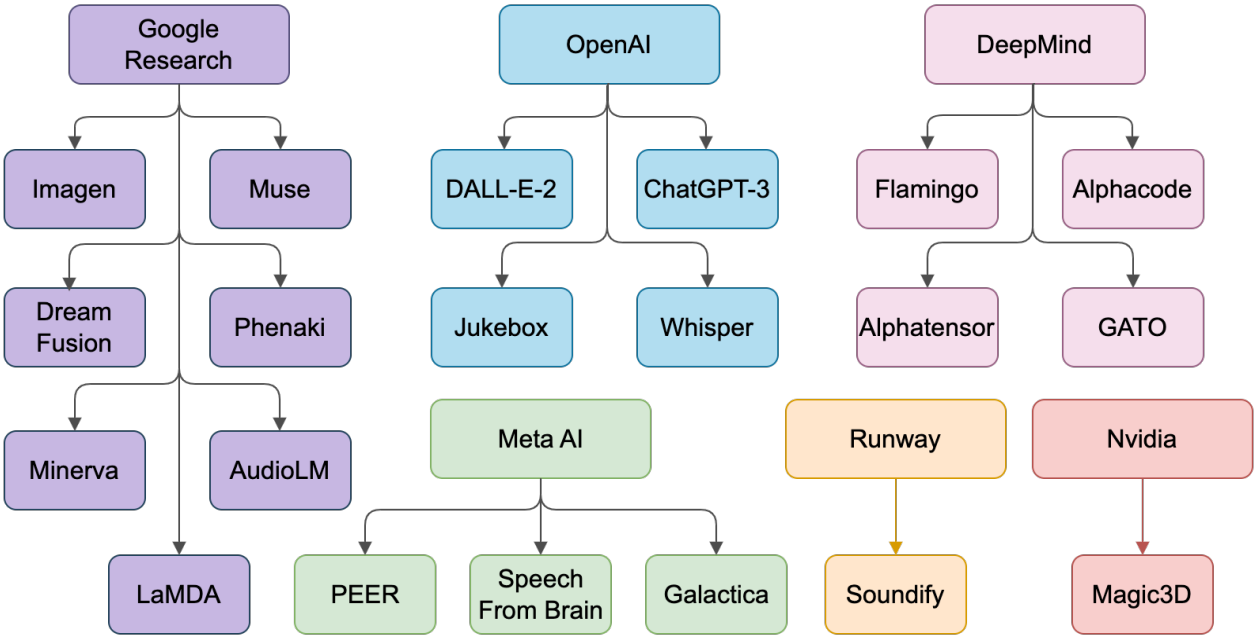

Figure 3. Some AI models organized according to their developers

A Closer Look At The Chatgpt Model

The ChatGPT chatbot which is a general-purpose chatbot was developed by OpenAI, a technology company founded in 2015 in Silicon Valley, using a large language model using large amounts of human-generated text, and released on March 14, 2023 (ChatGPT LLM, 2023). Until the final release, the ChatGPT and OpenAI roadmap is given in Figure 4. ChatGPT was trained using more than 40 terabytes of text. By simple math, this corresponds to close to 40 million books in kindle format (Rudolph, et al., 2023). However, GPT-4, the latest version of the ChatGPT model released at the time of this study, is about 570 times these values.

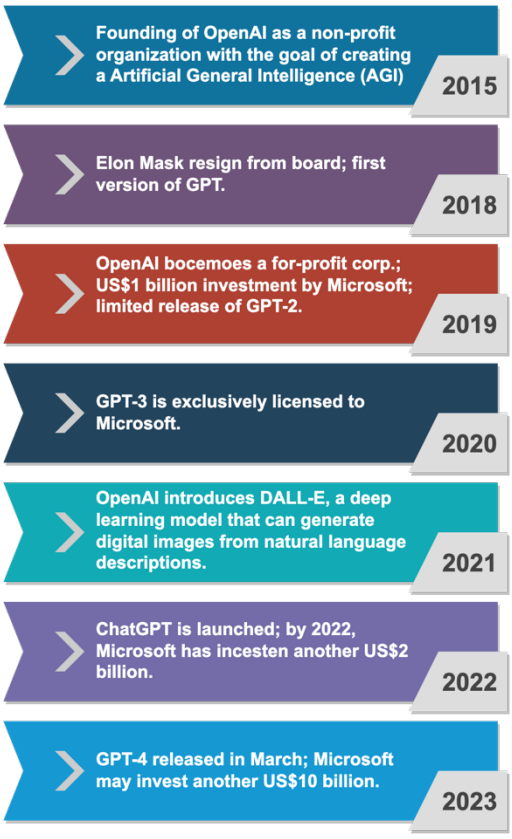

Figure 4. GPT and ChatGPT roadmap with OpenAI (OpenAI Company, 2023)

The ChatGPT model is capable of natural conversations on a wide range of topics, designed to generate human-like text with specific inputs or chat effects (Zhai, 2022). Unlike its predecessor language model, some of its training is done through reinforcement learning through human feedback. In this way, the quality of the responses is ranked in order to fine-tune the model using Proximal Policy Optimization. Thanks to this new approach, ChatGPT is able to answer follow-up queries, accept incorrect answers and reject inappropriate queries. Figure 5 shows the training steps of the GPT model. While GPT-1 used a model with 117 million parameters, this number reached 175 billion for GPT-3 (ChatGPT LLM, 2023). For GPT-4, which was released in March and is currently the latest version, this number (although not published by OpenAI) is estimated to be around 100 trillion. The ChatGPT artificial intelligence model is one of the most well-known models, along with LaMDA, PEER and Speech From Brain, which can take text input and produce text output, as can be seen in Figure 1.

Since the release of its last version, ChatGPT has been one of the artificial intelligence models that has been in the center of attention. The following figure, Figure 6, shows the historical distribution of the total number of queries for the words “chatgpt”, “gpt” and “gpt4” and variants from Türkiye and worldwide using the Google search engine between November 1, 2022 and March 25, 2023 (Gpt4 Türkiye, 2023; Gpt-4 Türkiye, 2023; Gpt Türkiye, 2023; Gpt4 Worldwide, 2023; Gpt-4 Worldwide, 2023; Gpt Worldwide, 2023) .

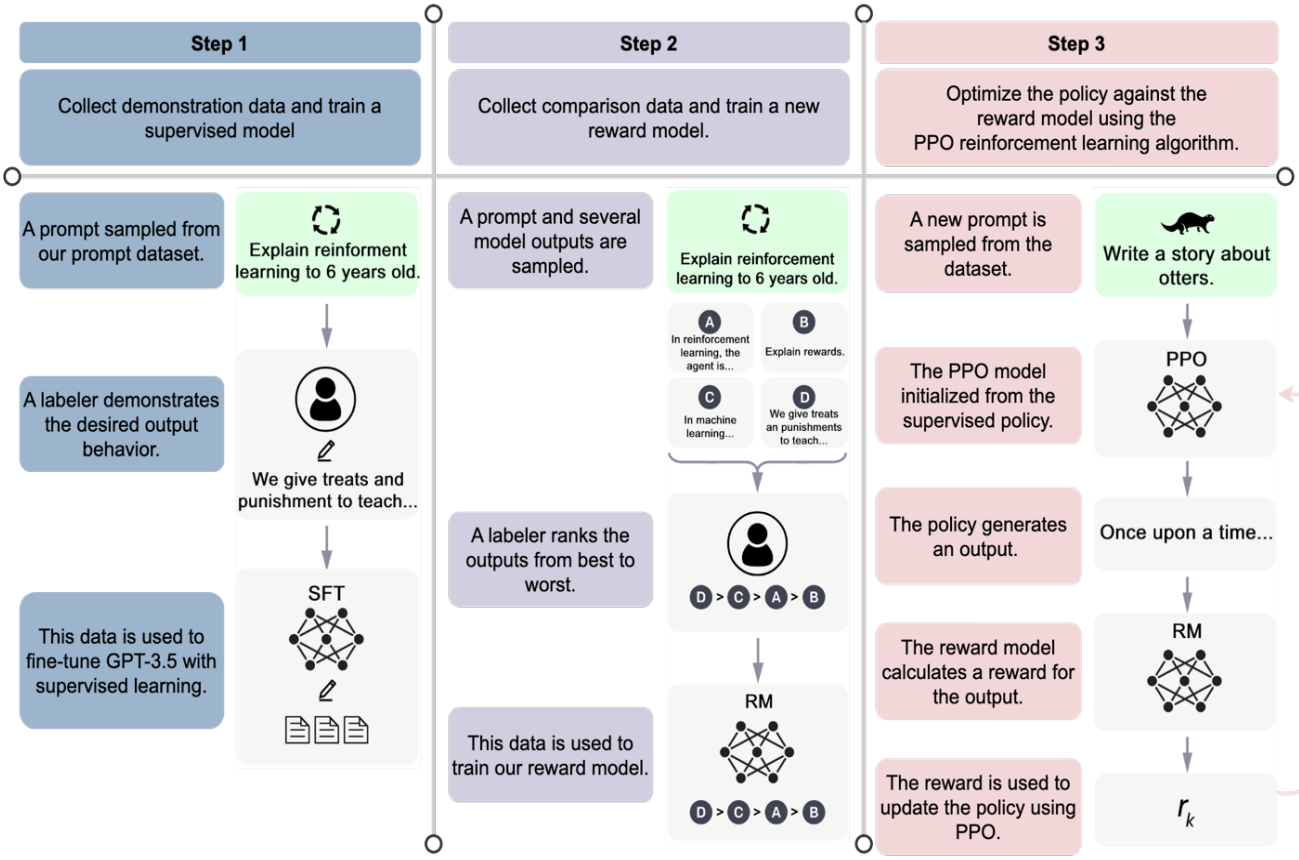

Figure 5. Training steps of the GPT language model (ChatGPT LLM, 2023)

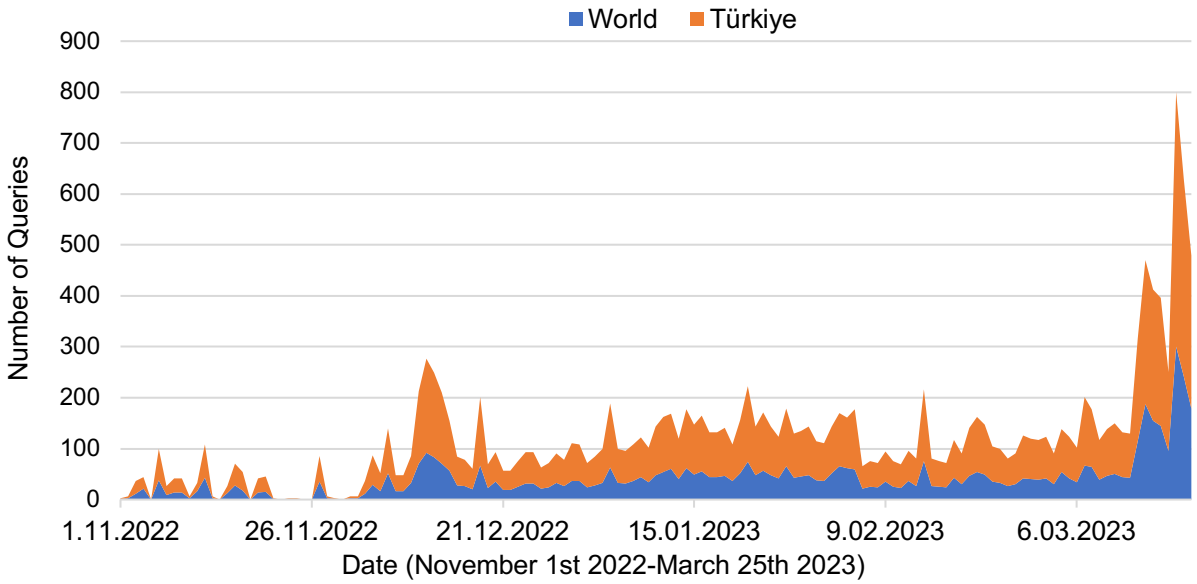

Figure 6. Between November 1, 2022 and March 25 2023, historical distribution of the total number of "chatgpt", "gpt", and "gpt4" queries using the Google search engine from Türkiye and around the world.

In general, the most important factor that determines the capabilities of language models and the limits of their performance is used the number of parameters in their training. The higher the comprehension and interpretation capabilities of language models when the more parameters used. Table 1 shows the number of parameters used in the training of some popular language models and their release dates.

Table 1. Release dates of different large language models and the number of parameters used for their training.

Model	BERT	GPT-1	GPT-2	T5	TuringNLG	GPT-3	GPT-4
Parameter Numbers*	340	117	1200	11000	17000	175000	~10000000000
Release Dates	2018	2018	2019	2019	2020	2022	2023

*Million pieces

From GPT-3.5 to GPT-4: Improvements

GPT-4, the latest version of the GPT language model, offers several important improvements and developments compared to its successor GPT-3.5. Some of these are as follows.

Complexity: One of the most significant new capabilities of GPT-4 is its ability to comprehend more complex and nuanced prompts. According to OpenAI, GPT-4 demonstrates human-level performance on a variety of academic and professional benchmarks (GPT-4, 2023).

This was demonstrated by subjecting GPT-4 to a variety of human-level examinations and standardized tests, such as the SAT, GRE, and BAR, without any specific training. GPT-4 not only comprehended and completed these tests with relatively high overall scores, but it also outperformed its predecessor, GPT-3.5. The ability to comprehend more complex input prompts is also helped by GPT-4's significantly increased word limit. The new model can handle up to 25,000 words input prompts, whereas GPT-3.5 was limited to 8,000 words.

Multimodal Capabilities: The prior version of ChatGPT only supported text prompts. In contrast, one

of the most recent characteristics of GPT-4 is its multimodal capabilities. The model can respond to both textual and visual prompts. This indicates that the ChatGPT can accept an image as input and interpret and comprehend it in the same manner as a text prompt. This feature covers all sizes and types of images and text, including documents that combine the two, hand-drawn sketches, and screenshots.

Safety: OpenAI spent approximately six months making GPT-4 more secure and fine-tuned. In comparison to GPT-3.5, the company claims that it is 82% less likely to respond to requests for inappropriate or otherwise prohibited content, 29% more likely to respond to sensitive requests in accordance with OpenAI's policies, and 40% more likely to produce factual responses (GPT-4, 2023).

It's not perfect, and you should still expect it to make nonsense from time to time and to be incorrect in its predictions, so ChatGPT still should not be trusted blindly.

Performance Improvements: In addition to evaluating the performance of the ChatGPT model on human exams, OpenAI also evaluated the chatbot using traditional machine learning benchmarks.
It is asserted that GPT-4 significantly outperforms existing LLMs and other cutting-edge models. In addition to the MMLU, these benchmarks include the AI2 Reasoning Challenge (ARC), WinoGrande, HumanEval, and Drop, which all assess individual capabilities (OpenAI, 2023).

ChatGPT's Capabilities and Opportunities

ChatGPT and all other artificial intelligence models are ultimately developed to make people's lives easier, directly or indirectly. The benefits and opportunities it can provide will also be in written form, as it is basically a chatbot model that is trained as text-to-text. While with GPT-3.5, ChatGPT was basically only capable of text-to-text processing, with GPT-4, it has gained the ability to process image-to-text. Although it is not yet clear to what extent it can provide opportunities and benefits in which field, there are studies showing that it can provide serious benefits in areas such as finance, education, law and health. Figure 7 shows some of these possibilities.

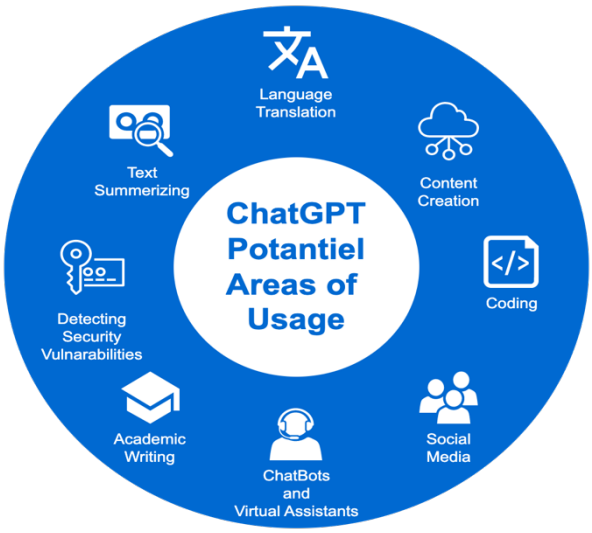

Figure 7. Some examples of potential use cases of the ChatGPT artificial intelligence model

GPT-4 was tested by OpenAI to see how much better it was than the previous language model, GPT-3.5., used by ChatGPT, and the results were better than expected.

The language of many existing ML benchmarks is English. OpenAI used Azure Translate to translate the MMLU (Multi-task Language Understanding) benchmark -a suite of 14,000 multiple-choice problems spanning 57 subjects- into a variety of languages to develop an early sense of language proficiency. GPT-4 outperforms GPT-3.5 and other large language models, including low-resource languages such as Welsh, Latvian and Swahili, in 24 of 26 tested languages. The results of testing GPT with some of these languages are shown in Figure 8.

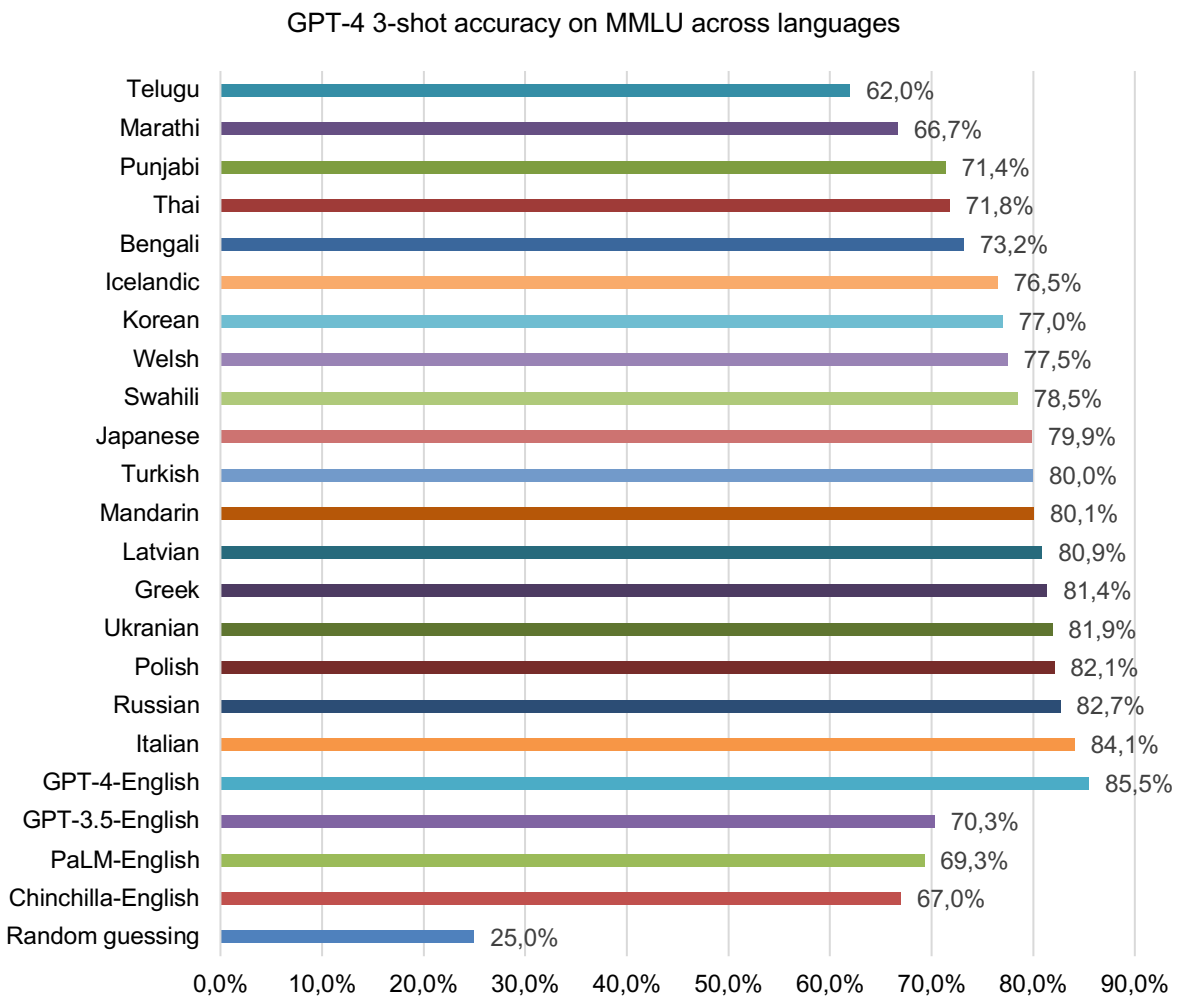

Figure 8. GPT-4 performance on MMLU in a variety of languages compared to prior models in English (GPT-4, 2023)

In a recent educational study, the writing proficiency of the ChatGPT language model was found to be well above that of the average student, and the text was found to be readable, coherent and (relatively) accurate in terms of content, and compared to humans, it was shown to be able to find the information needed more efficiently (Zhai, 2022).

ChatGPT is even referred to as a game-changing artificial intelligence model in terms of the potential opportunities it offers in the field of education (Should Professors Worry?, 2023). Compared to other AI language models such as Meta's language tool PEER or RoBERTa, ChatGPT offers "more creative" answers (ChatGPT AI Chatbot, 2023).

One of the biggest advantages of the ChatGPT language model is its ability to aid research in academic writing. ChatGPT can generate summaries of articles, highlight key points and even provide citations. This can save researchers a significant amount of time and effort, allowing them to focus on more important tasks such as analysis and interpretation (Aljanabi et al., 2023).

AI models such as ChatGPT can also provide various benefits for practicing lawyers. A lawyer can have the ChatGPT AI model produce an initial draft of a memo and then modify it as needed. The ChatGPT AI model can also be used to generate an initial set of arguments and then sift them down to the most effective ones, or to review older examples of legal documents to make the work more efficient. Pedagogically, law schools should consider how to prepare law students to use these tools most effectively in their practice, while at the same time emphasizing to students that their basic skills in legal research and evaluation cannot be delegated to language models alone. While

ChatGPT and similar tools can help a lawyer to work more efficiently, they cannot replace the need for a lawyer today to find, understand and reason from relevant legal sources (Choi et al., 2023).

The ChatGPT AI model can be used to generate reports and summaries of economic and financial data, making it easier for researchers and analysts to make sense of their findings. It can also be used to generate forecasts and predictions based on historical data, which can be valuable for decision-making processes (Alshater, 2022).

In a standard benchmark test on coding bug fixing, the ChatGPT artificial intelligence model performed similarly to Codex and its own deep learning-based Automated Program Repair (APR), but was much more successful, fixing 19 out of 40 bugs, compared to the standard automated program repair methods, which only fixed 7. In fact, providing more information about the bug improved ChatGPT performance even more, reaching a success rate of 77.5% (Sobania et al., 2023).

The categorization of the answers given by ChatGPT for the issues received from QuixBugs is as in Figure 9 below. (QuixBugs, 2023).

Figure 9. Classification of responses to problems from QuixBugs by the ChatGPT artificial intelligence model (Sobania et al., 2023)

Limitations and Risks of ChatGPT

Even though the ChatGPT artificial intelligence model has been trained with a very large dataset, it is still only capable of the information provided by that dataset. For this reason, it can still produce incomplete or incorrect information output. Sometimes, while it can solve very complex algorithmic problems, it can give incorrect answers to simple math questions. Some of the limits and risks that may occur with ChatGPT are shown in Figure 10.

In a recent study, the ChatGPT artificial intelligence model was able to answer 37.5% of the programming questions correctly. When the characteristics of the wrong answers are analyzed;

- Lack of information
- Faulty assumption
- Lack of information and faulty assumption together

three main reasons stand out (Jalil et al., 2023).

ChatGPT poses a significant threat to the integrity of online exams, especially in higher education where online exams are becoming increasingly common. While these and similar AI models have a high degree of critical thinking skills, they can produce highly realistic texts with very little input, making it possible for students to cheat on exams. Although going back to face-to-face exams is a solution, it is not always possible, and using advanced artificial intelligence text output detection tools can be effective in solving this problem. Unfortunately, such tools are far from providing perfect solutions. More research is needed to fully understand the effects of AI models such as ChatGPT and to develop methods to combat problems such as cheating using these tools. It is crucial that educators and institutions are aware of the risk of ChatGPT being used to cheat and that they take this issue seriously and provide the necessary research and measures to ensure the fairness and validity of online exams for all students (Susnjak, 2022).

In a study to test ChatGPT's mathematical capabilities, it was found that, contrary to media hype, ChatGPT is not yet ready to provide consistently high quality proofs or calculations (Frieder et al., 2023).

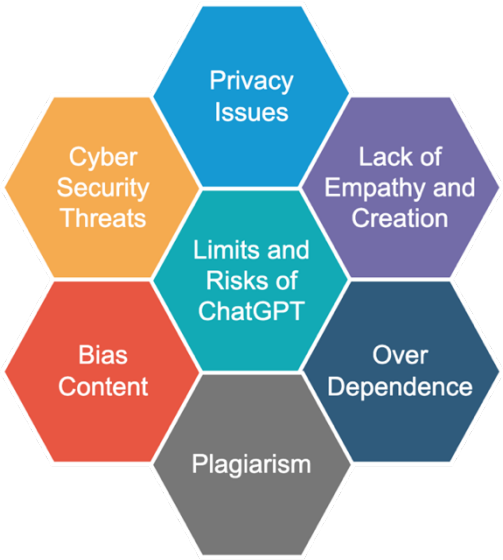

Figure 10. Some examples of ChatGPT limits and risks

Perhaps the most important risk or concern, and one that is mentioned even more frequently than the ones mentioned above, is the extent to which existing professions are sensitive to these developments after every major technological development or breakthrough, of which the ChatGPT artificial intelligence model is a part. These technological breakthroughs, which have been experienced many times from the past to the present, have always caused fear among the masses that technology will sooner or later replace humans. One study has shown that such technological advances are likely to result in the assignment of mainly low-skilled workers to non-technology-sensitive tasks, i.e. tasks that require creativity and social intelligence. But for people to win this race, they will need to acquire or develop creative and social skills (Frey and Osborne, 2017).

In addition to the limits and risks mentioned above, other limits and risks foreseen for the ChatGPT artificial intelligence model can be listed as follows.

- Privacy issues
- Cyber security threat
- Biased content
- Dismissal
- Plagiarism
- Overdependence
- Restricted content
- Lack of empathy
- Lack of creativity
- Contextual limitations

RESULTS AND DISCUSSION

In addition to processing information, artificial intelligence models are more reliable in terms of business execution. They are also very adept at collecting and presenting information. Today, scientific studies are increasingly relying on artificial intelligence models. In order to carry out these scientific studies, scientists and engineers use supercomputers to predict, classify and infer in order to solve complex problems. Since artificial intelligence models are software, the more advanced these computer systems become, the more space artificial intelligence models will have to process bigger data and provide faster and uninterrupted service. In addition, integrating artificial intelligence models into learning tasks in certain domains is critical for the training of artificial intelligence models as it shows how humans solve real-world problems.

As a result of the possibilities offered by the ChatGPT language model, it will be able to break new ground in communication by eliminating the language barrier in human communication, as well as providing incredible convenience to companies and customers in areas such as customer relations. It will be able to identify and summarize the important points of academic articles or books in a very short time, and it will be able to understand and simplify complex questions, which will provide students with significant convenience in their educational lives. ChatGPT has significant power to advance academia and librarianship in new ways that are both disconcerting and exciting (Lund and Wang, 2023).

Apart from education, the ChatGPT artificial intelligence model has also shown that it can pass certain parts of important exams in fields such as medicine, mathematics, economics and law. The ChatGPT artificial intelligence model, which is quite successful in some areas and subjects but insufficient in different subjects, has the ability to achieve much more success if it is provided with much more data resources and its training is concentrated in those areas. In a different study conducted in the field of healthcare, a great potential was seen in using big language models such as ChatGPT to improve patient-centered care in radiology and other medical fields by automatically generating simplified radiology reports by incorporating them into clinical processes (Jeblick et al., 2022).

As it has only been launched relatively recently, it is still too early to get the full picture of what it can do and how well it can do it. Since artificial intelligence models will continue to evolve in direct proportion to the limits of human intelligence, much more research is needed

on both human intelligence and artificial intelligence models to better understand which parts of human intelligence can be replaced by artificial intelligence models (Frey and Osborne, 2017).

Despite their inherent limitations, with the right tools and approaches, artificial intelligence and natural language processing technologies have the potential to greatly increase efficiency and effectiveness in many areas, leading to new discoveries and insights and shaping the future (Alshater, 2022).

FUNDING

This study was supported by Burdur Mehmet Akif Ersoy University Scientific Research Projects Commission. Project No: 0886-YL-23.

REFERENCES

Aljanabi, M., Ghazi, M., Ali, A. H., Abed, A. (2023). ChatGpt: Open Possibilities. *Iraqi Journal for Computer Science and Mathematics*, *4*(1). https://doi.org/10.52866/ijcsm.2023.01.01.0018

Alshater, M. (2022). Exploring the Role of Artificial Intelligence in Enhancing Academic Performance: A Case Study of ChatGPT. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4312358

ChatGPT AI Chatbot (2023). ChatGPT is a new AI chatbot that can answer questions and write essays. https://www.cnbc.com/2022/12/13/chatgpt-is-a-new-ai-chatbot-that-can-answer-questions-and-write-essays.html (Access Date: 02.03.2023)

ChatGPT LLM (2023). ChatGPT: Optimizing Language Models for Dialogue. https://openai.com/blog/chatgpt/ (Access Date: 27.02.2023)

Choi, J. H., Hickman, K. E., Monahan, A., Schwarcz, D. B. (2023). ChatGPT Goes to Law School. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4335905

Frey, C. B., Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Frieder, S., Pinchetti, L., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., Chevalier, A., Berner, J. (2023). *Mathematical Capabilities of ChatGPT* (arXiv:2301.13867). arXiv. http://arxiv.org/abs/2301.13867

GPT-4 (2023). GPT-4 OpenAI. https://openai.com/research/gpt-4 (Acess Date: 25.03.2023)

Gpt4 Türkiye (2023). Google Trendler. https://trends.google.com/trends/explore?date=2022-11-01%202023-03-25&geo=TR&q=gpt4&hl=tr (Access Date: 25.03.2023)

Gpt4 Worldwide (2023). Google Trendler. https://trends.google.com/trends/explore?date=2022-11-01%202023-03-25&q=gpt4&hl=tr (Access Date: 25.03.2023)

Gpt-4 Türkiye (2023). Google Trendler. https://trends.google.com/trends/explore?date=2022-11-01%202023-03-25&geo=TR&q=gpt-4&hl=tr (Access Date: 25.03.2023)

Gpt-4 Woldwide (2023). Google Trendler. https://trends.google.com/trends/explore?date=2022-11-01%202023-03-25&q=gpt-4&hl=tr (Access Date: 25.03.2023)

Gpt Türkiye (2023). Google Trendler. https://trends.google.com/trends/explore?date=2022-11-01%202023-03-25&geo=TR&q=gpt&hl=tr (Access Date: 25.03.2023)

Gpt Worldwide (2023). Google Trendler. https://trends.google.com/trends/explore?date=2022-11-01%202023-03-25&q=gpt&hl=tr (Access Date: 25.03.2023)

Jalil, S., Rafi, S., LaToza, T. D., Moran, K., Lam, W. (2023). *ChatGPT and Software Testing Education: Promises & Perils* (arXiv:2302.03287). arXiv. http://arxiv.org/abs/2302.03287

Jeblick, K., Schachtner, B., Dexl, J., Mittermeier, A., Stüber, A. T., Topalis, J., Weber, T., Wesp, P., Sabel, B., Ricke, J., Ingrisch, M. (2022). *ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports* (arXiv:2212.14882). arXiv. http://arxiv.org/abs/2212.14882

Li, Y., Choi, D., Chung, J., Kushman, N., Schrittwieser, J., Leblond, R., Eccles, T., Keeling, J., Gimeno, F., Lago, A. D., Hubert, T., Choy, P., d'Autume, C. de M., Babuschkin, I., Chen, X., Huang, P.-S., Welbl, J., Gowal, S., Cherepanov, A., Vinyals, O. (2022). Competition-Level Code Generation with AlphaCode. *Science*, *378*(6624), 1092–1097. https://doi.org/10.1126/science.abq1158

Lund, B. D., Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*. https://doi.org/10.1108/LHTN-01-2023-0009

OpenAI (2023). GPT-4 Technical Report (arXiv:2303.08774). arXiv. http://arxiv.org/abs/2303.08774

OpenAI Company (2023). About OpenAI. https://openai.com/about/ (Access Date: 21.02.2023)

QuixBugs (2023). QuixBugs Benchmark. https://jkoppel.github.io/QuixBugs/ (Access Date: 02.03.2023)

Rudolph, J., Tan, S., Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, *6*(1). https://doi.org/10.37074/jalt.2023.6.1.9

Should professors worry? (2023). AI bot ChatGPT writes smart essays. https://www.nature.com/articles/d41586-022-04397-7 (Access Date: 28.02.2023)

Sobania, D., Briesch, M., Hanna, C., Petke, J. (2023). *An Analysis of the Automatic Bug Fixing Performance of ChatGPT* (arXiv:2301.08653). arXiv. http://arxiv.org/abs/2301.08653

Susnjak, T. (2022). *ChatGPT: The End of Online Exam Integrity?* (arXiv:2212.09292). arXiv. http://arxiv.org/abs/2212.09292

Zhai, X. (2021). Practices and Theories: How Can Machine Learning Assist in Innovative Assessment Practices in Science Education. *Journal of Science Education and Technology*, *30*(2), 139–149. https://doi.org/10.1007/s10956-021-09901-8

Zhai, X. (2022). ChatGPT User Experience: Implications for Education. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4312418