

Critical assessment of some applications of artificial intelligence

Kamen Spasov¹, Diana Netova¹, Nikolay Netov¹
Sofia University "St. Kliment Ohridski"¹

kspasov@feb.uni-sofia.bg, dianah@feb.uni-sofia.bg, nnetoff@feb.uni-sofia.bg

Abstract: Artificial intelligence applications are rapidly entering many aspects of business, state administration and education with the aim of increasing productivity and efficiency. The report examines the possibilities for implementing artificial intelligence in their activities and the related ethical and legal aspects, as well as the risks. A review and analysis of some applications of AI is made.

Keywords: ARTIFICIAL INTELLIGENCE APPLICATIONS, BUSINESS, STATE ADMINISTRATION, EDUCATION

1. Introduction

Mathematical problem solving is a crucial aspect of contemporary society, and artificial intelligence (AI)-driven mathematical reasoning has become integral to data analysis. The recent literature review examines the many datasets used to enhance the ability of AI models to address mathematical word problems. These datasets offered a variety of problem sets and functioned as benchmarks for assessing the performance of several deep learning models, including recurrent neural networks (RNNs) and graph-based models. Datasets like GSM8K presented problems that even the most advanced transformer models found difficult to surmount, establishing a new benchmark for the examination of AI systems in mathematical problem-solving, [1].

In January 2025, TechTarget published an article "25 of the best large language models in 2025", [2]. In his blog Anthony Cardillo, [3], discusses 44 Large Language Models (LLM). Both authors cover ChatGPT, Perplexity, Gemini, and DeepSeek as some of the best Generative Pre-trained Transformers AI language models. Our experience shows that these LLMs are, maybe, the most popular among students and researchers in Bulgaria. INSAIT claims their LLM product BgGPT is unique and based on the Bulgarian language.

ChatGPT by OpenAI is known for its versatility in generating human-like text, engaging in conversations, and assisting with tasks like writing, coding, and brainstorming. It generally has a user-friendly interface. However, its knowledge is based on its last training data cutoff, so it may not always have the most current information without integrated web Browse features (which are available in some versions). Accuracy can vary, and it sometimes generates plausible-sounding but incorrect or nonsensical information.

Perplexity AI is designed with a focus on providing accurate, up-to-date information by citing sources from the web. This makes it strong for research and fact-checking. Its interface is often described as a hybrid of a search engine and a chatbot. While it excels at information retrieval and providing sourced answers, it might be less geared towards creative writing or extended, open-ended conversations compared to models like ChatGPT.

Gemini 2.0+ (often referred to as Gemini) by Google is a multimodal AI model, meaning it can understand and process information from various formats like text, code, images, and sometimes audio/video. It aims for strong reasoning capabilities and creative content generation. As a Google product, it often integrates well with other Google services. User experience can vary depending on the specific application or interface being used (e.g., a standalone app, integrated into a search).

BgGPT is a Bulgarian-first large language model developed by INSAIT. It is based on Google's Gemma models and has been specifically fine-tuned for the Bulgarian language, demonstrating strong performance in Bulgarian while retaining competitive English capabilities. Its primary audience is Bulgarian-speaking users and developers. The user interface is a chat platform, and it aims to understand Bulgarian cultural nuances.

DeepSeek is an open-source model, often highlighted for its strong performance in coding and technical tasks, as well as its cost-

effectiveness (being free and open-source). It supports a high context length and performs well in both English and Chinese. Users find it efficient for tasks requiring logical reasoning and structured problem-solving. However, it might be less optimized for creative content generation or general conversation compared to some other models, and some reviews have noted potential concerns regarding data privacy given its origin.

Finding direct comparative scientific articles for all five specific free versions from a user-experience viewpoint is challenging, as academic research often focuses on benchmark performance of base models, specific capabilities (like reasoning or bias), or the underlying technology. User experience is also more commonly discussed in tech reviews and user forums.

To solve this problem Gemini 2.5 Pro was prompted:

"Can you provide a comparison of five free Generative Pre-trained Transformers AI language models ChatGPT, Perplexity, Gemini 2.0+, BgGPT (<https://bggpt.ai>), and DeepSeek from a user's viewpoint? Please summarize the results in a table. Please provide scientific articles as references."

Gemini 2.5 Pro was selected for the task due to its capabilities to use Google Search and Google Scholar in particular. The answer is summarized as follows:

ChatGPT (Free Version) [4]

- (1) Primary Strength: Versatile conversation & content generation
- (2) Ease of Use: Very high; intuitive chat interface
- (3) Accuracy: Variable; can "hallucinate"
- (4) Creativity: High
- (5) Coding Ability: Good
- (6) Up-to-date Information: Limited to training data (unless browses)
- (7) Multimodality: Primarily text (some versions may have image input)
- (8) Speed: Generally fast
- (9) Customization/Control: Limited in free version
- (10) Key User Focus: General assistance, writing, brainstorming
- (11) Potential Limitations: Can be verbose, potential for inaccuracies, knowledge cutoff

Perplexity AI (Free Version) [5]

- (1) Primary Strength: Accurate, sourced information retrieval
- (2) Ease of Use: High: search-like interface
- (3) Accuracy: Generally high: cites sources
- (4) Creativity: Moderate
- (5) Coding Ability: Limited (focus is on information)
- (6) Up-to-date Information: Yes, actively searches the web
- (7) Multimodality: Primarily text: can analyze uploaded files
- (8) Speed: Fast, includes search time
- (9) Customization/Control: Some control over search focus
- (10) Key User Focus: Research, fact-checking, learning

(11) Potential Limitations: Less conversational, creative output may be limited

Gemini 2.0+ (Free Version) [6]

- (1) Primary Strength: Multimodal understanding & reasoning
- (2) Ease of Use: High; varies by Google interface
- (3) Accuracy: Generally high, especially in reasoning
- (4) Creativity: High
- (5) Coding Ability: Good
- (6) Up-to-date Information: Often, through Google Search integration
- (7) Multimodality: Strong (text, images, sometimes audio/video)
- (8) Speed: Generally fast
- (9) Customization/ Control: Some, depending on the interface
- (10) Key User Focus: Creative tasks, complex queries, multi-format input
- (11) Potential Limitations: Can sometimes be overly cautious or "Google-centric"

BgGPT [7]

- (1) Primary Strength: Bulgarian language & cultural nuance
- (2) Ease of Use: High; chat interface
- (3) Accuracy: High for Bulgarian; good for English
- (4) Creativity: Moderate
- (5) Coding Ability: Not a primary focus
- (6) Up-to-date Information: Limited to training data
- (7) Multimodality: Primarily text
- (8) Speed: Good
- (9) Customization / Control: Standard chat interaction
- (10) Key User Focus: Bulgarian language interaction & information
- (11) Potential Limitations: Primarily for Bulgarian/English, newer model

DeepSeek (Open Source) [8]

- (1) Primary Strength: Coding, technical tasks, open-source
- (2) Ease of Use: Moderate to High; intuitive interface
- (3) Accuracy: High in technical domains; can be biased
- (4) Creativity: Moderate
- (5) Coding Ability: Very Strong
- (6) Up-to-date Information: Limited to training data
- (7) Multimodality: Primarily text
- (8) Speed: Fast, efficient
- (9) Customization / Control: High (as open-source)
- (10) Key User Focus: Developers, technical users, data analysis
- (11) Potential Limitations: May have data privacy concerns for some, less conversational finesse

2. Experiment: solve reasoning problems

Students and academics are tempted to use Generative Pre-trained Transformers AI language models as tools in their study and research. A simple experiment was conducted to explore the problem-solving capabilities of the abovementioned LLMs. Regarding the set of five popular free Generative Pre-trained Transformers AI language models, a simple logical problem was presented. The surprise was that only one of them was able to return the right answer. After three months, the test was repeated. The experiment is designed to test the functionality of the AI language models in solving reasoning problems from a user's viewpoint and does not cover the structure of the models and program code.

The problem:

"Maria has 2 sisters and a brother, John. How many sisters does John have?"

Test results in a three-month interval for 5 five popular GPT AI language models are summarized as follows in Table 1

Table 1. Test results in a three-month interval for 5 five popular GPT AI language models. Answers were cited.

Application	February 2025	May 2025
ChatGPT [4]	John has 2 sisters .	John has 3 sisters .
Perplexity [5]	If Maria has two sisters and one brother named John, then John has the same two sisters: Maria and her other sister. This means John has two sisters .	If Maria has two sisters and one brother named John, then John has the same two sisters: Maria and her other sister. This means John has two sisters .
Google Gemini [6]	Maria's brother has two sisters .	
Google Gemini (2.5 Flash) [6]		John has 2 sisters . Maria and her other sister are John's sisters.
Google Gemini (2.5 Pro) [6]		Here's how to solve this: Maria has 2 sisters. Maria herself is also a sister to John. So, John has Maria + Maria's 2 sisters = 3 sisters .
BgGPT [7] <i>The conversation is in Bulgarian</i>	Mary's brother also has two sisters, since they are siblings in common. So, Mary's brother has two sisters . Мария има две сестри и един брат Иван. Колко сестри има Иван? Братът на Мария също има две сестри, тъй като те са общи братя и сестри. Така че братът на Мария има две сестри.	Ivan has one sister, who is common to him and Maria. Since Maria has two sisters and one brother (who is Ivan), this means there are three children in the family: Maria, Ivan, and another sister. Therefore, Ivan has only one sister . Иван има една сестра, която е обща за него и Мария. Тъй като Мария има две сестри и един брат (който е Иван), това означава, че те са три деца в семейството: Мария, Иван и още една сестра. Следователно, Иван има само една сестра.
DeepSeek [8]	John has 3 sisters . (8 pages of reasoning, verification with an example, and conclusion)	John has 3 sisters . (8 pages of reasoning, verification with an example, and conclusion)

The only application that answered correctly in February 2025 was DeepSeek. It means mathematical logic is implemented in the AI language model. All other AI language models provided wrong answers.

The same test was conducted 3 months later. During that time, Google launched few more free versions of Gemini, and they were included in the experiment. Google Gemini (2.5 Flash) application delivered the same result. Google Gemini (2.5 Pro) application returned the right answer. Perplexity and the original Gemini 2.0. Unfortunately, the product of INSAT BgGPT deteriorated. It was expected that the logical capabilities of the AI language models would improve, but the reality is different. Some companies improved their AI language models, most probably adding mathematical logic. Others left these improvements at a later stage. Unfortunately, the attempt of INSAIT to add mathematical logic to BgGPT, if any, was a failure. DeepSeek returned the same right answer. It is interesting to note that DeepSeek claims that "Think before responding to solve reasoning problems". The results from this simple experiment prove the claim.

3. Conclusion

Most of the free Generative Pre-trained Transformers AI language models were not trained to solve reasoning problems at the beginning of 2025. With time, the leading vendors improved their language models, adding additional versions with capabilities to solve reasoning problems. Most probably, the paid versions provide such capabilities. Users need to be aware that not all Generative Pre-trained Transformers (GPT) AI language models are trained to solve reasoning problems, and they must carefully select what model to use for a specific task.

References

- 1] M. Kurisappan, S. Sundara Pandiyan, "Decoding math : A review of datasets shaping AI-driven mathematical reasoning," *Journal of Interdisciplinary Mathematics*, vol. 28, no. 2, pp. 607 - 625, (2025).
- 2] Sean Michael Kerner, Ben Lutkevich, "TechTarget," 31 January (2025). [Online]. Available: <https://www.techtarget.com/whatis/feature/12-of-the-best-large-language-models>.
- 3] A. Cardillo, "Best 44 Large Language Models (LLMs) in 2025," 9 May (2025). [Online]. Available: <https://explodingtopics.com/blog/list-of-llms>.
- 4] ChatGPT, (2025). [Online]. Available: <https://chatgpt.com>.
- 5] Perplexity, (2025). [Online]. Available: <https://www.perplexity.ai>.
- 6] G. Gemini, (2025). [Online]. Available: <https://gemini.google.com>.
- 7] BgGPT, (2025). [Online]. Available: <https://bggpt.ai>.
- 8] DeepSeek, (2025). [Online]. Available: <https://chat.deepseek.com/>.