

GPT-5 is finally here — What To Expect and What Not To Expect



[Don Lim](#)



Photo by [Mariia Shalabaieva](#) on [Unsplash](#)

Microsoft engineers are set to host Orion (GPT-5) on Azure as early as November 2024. Although it won't be available to the general public, its deployment marks a major AI milestone with far-reaching implications.

When OpenAI announced its work on GPT-5 around March 2023, it raised major concerns, with some prominent figures signing a petition to halt its development. Confirmed signatories include AI pioneer Yoshua Bengio (Turing Award winner), AI researchers Stuart Russell and Gary Marcus, Steve Wozniak (Apple co-founder), Andrew Yang (former U.S. presidential candidate), and Rachel Bronson (President of the Bulletin of the Atomic Scientists — against humanity-ending nuclear war). Elon Musk (Tesla, Twitter, and SpaceX CEO and former OpenAI co-founder) also joined, as did Emad Mostaque (Stability AI CEO).

While GPT-4 marked a significant advancement in AI capabilities, it still has notable limitations. It has not led to widespread job displacement or signaled the downfall of humanity, which suggests that people who signed the petition may be misguided.

The Looming Shadow of Orion: A Cautious Optimism

More than a year has passed since GPT-4 was released, and concerns about GPT-5, codenamed Orion, have lessened. Orion is finally ready to be introduced to the world. Unlike previous releases, it initially won't be widely available through ChatGPT. Instead, OpenAI plans to grant early access to companies they closely collaborate with, allowing

them to develop their own products and features. Although Orion is considered GPT-4's successor internally, it's unclear if it will be publicly named GPT-5.

A report claims GPT-5 will be 100 times more powerful, though the specifics of this "power" are uncertain. However, there are still limitations it will never overcome.

First, let's examine the persistent issues in the latest GPT version.

GPT-4o's Limitations and Concerns:

1. **Language Limitations:** Despite GPT-4o's improvements in multilingual capabilities, it still faces challenges with non-English languages, particularly those with limited high-quality data, resulting in inadequate response and translation.
2. **Data Quality Issues:** A notable concern is the subpar quality of training data, especially in languages like Chinese, which can lead to models prone to providing inaccurate or hallucinatory responses due to the prevalence of low-quality content.
3. **Factual Accuracy:** The model may occasionally provide incorrect or misleading information, even with specific prompts, raising concerns about its credibility and reliability.
4. **Partial Compliance:** LLMs fulfill some, but not all, parts of a request when the user includes a long set of instructions despite all instructions being clear and simple. For AI, this is a complex and novel set of instructions. The model might complete parts of the instruction that it recognizes from training, but it might skip or simplify other parts. Upon feedback, it may correct the missing part of the request but omit parts it previously completed correctly.
5. **Resource Intensive:** The substantial computational resources required to train and run large language models like GPT-4o limit accessibility and raise environmental concerns due to the high energy consumption.
6. **Social Bias:** GPT-4o can inadvertently perpetuate biases present in the training data, resulting in unfair and discriminatory outputs that can sway opinions and reinforce existing social issues.
7. **Anthropomorphization:** The tendency to attribute human-like qualities to AI models using terms like "reasoning" and "understanding" can lead to unrealistic expectations and misunderstandings. While this can let consumers understand LLMs more easily, it's essential to acknowledge that these models operate by predicting words based on patterns, rather than truly perceiving or understanding prompts like humans.

8. **Insufficient External Red Teaming:** The current group of 70+ experts may not be enough to adequately address the diverse needs of potential users and potential risks associated with the technology.
9. **Lack of Transparency:** OpenAI does not provide sufficient information about key aspects such as sustainability metrics, upcoming features, release dates, or product roadmaps.
10. **Limited Access to Advanced Models:** It is likely that only a select group of users will have access to the most powerful versions of the model, or they may be unaffordable for many users due to high costs.

Orion: A Step Forward or More of the Same?

OpenAI's plans for Orion are shrouded in mystery, with conflicting reports about its release date and capabilities. (Unfortunately, inconsistency can be intriguing to some people) While it's expected to address some of GPT-4o's shortcomings, certain fundamental challenges may persist.

Improvements Expected in Orion

1. **Fact-Checking Enhancement:** One of the expected improvements in Orion is better fact-checking capabilities. While current models like GPT-4 occasionally produce inaccuracies or confabulations, Orion aims to minimize these errors by referencing more reliable data sources and refining its response validation processes.
2. **Enhanced Video Processing:** GPT-4 has limited capabilities when it comes to processing and understanding video content, constraining its application in multimedia contexts. Orion, however, is expected to significantly advance video comprehension and processing abilities, potentially allowing for applications in real-time video analysis, summarization, and even generation.
3. **Contextual Memory:** Orion is expected to have an improved memory capacity, allowing it to retain more context across longer conversations. This would help create more coherent, contextually aware interactions, enabling the model to remember details from earlier in a conversation and apply them in relevant ways over extended exchanges.
4. **Advanced Multimodal Capabilities:** While GPT-4 introduced basic multimodal features, allowing it to process both text and images, Orion is likely to expand on this by seamlessly integrating multiple types of input, such as text, image, video, and potentially even audio.
5. **Ethical and Bias Controls:** Future models like Orion are expected to implement more sophisticated mechanisms for reducing bias and handling ethically complex situations

6. **Customization and Fine-Tuning for Businesses:** Orion is likely to allow businesses more granular control in fine-tuning the model to align with their specific needs. OpenAI may provide more advanced APIs and tools for companies to customize language generation, adjust tone, and tailor the model's responses, enabling more versatile applications in customer service, marketing, and other sectors.
7. **Improved Real-Time Language Translation:** Orion is expected to improve on GPT-4's language translation capabilities, aiming for real-time, contextually accurate translations across more languages and dialects.

Ongoing Challenges and Limitations

Several issues are expected to persist in future versions of the GPT model:

1. **Data Quality Concerns:** The model's reliance on internet-scale data, which is often prone to biases and inaccuracies, is a persistent problem.
2. **Non-English Language Challenges:** Despite efforts to improve non-English response quality, limitations are expected to remain due to the scarcity of high-quality training data and skilled trainers.
3. **Partial Compliance:** Even clear, simple instructions can seem complex to LLMs if they contain many constraints, as they match patterns rather than fully understand prompts. Given many constraints, LLMs may follow the closest pattern they recognize, often overlooking some instructions. This limitation is not likely to be resolved.
4. **Resource Intensiveness:** The transformer architecture that powers GPT models requires substantial resources, which is likely to continue. This could lead to environmental disasters, such as accelerated global warming or heightened nuclear waste risks if new data centers rely on nuclear power as reported.
5. **Perpetuation of Social Bias:** The model's training data, which is largely sourced from the internet, may reinforce and perpetuate social biases if high-quality, unbiased sources are not prioritized. This highlights a limitation in OpenAI's data collection practices and the need for more careful curation of training data.

The Path Forward

To address these limitations and ensure the responsible development of AI, several steps may be needed:

1. **A New Language Model:** A new language model is needed. It should be resource-efficient and capable of learning effectively without needing extensive data for every possible use case of the language and all possible scenarios and nuances.

2. **High-Quality Data:** To mitigate biases and inaccuracies, it's crucial to invest in the creation and curation of diverse, high-quality datasets. This involves carefully selecting data sources, removing biases, and ensuring data integrity.
3. **Transparency and Accountability:** OpenAI and other AI developers should prioritize transparency in their research and development processes. By sharing information about their models' limitations, biases, and potential impacts, they can foster public trust and enable informed discussions about AI's role in society.
4. **User Education:** Educating the public about the capabilities and limitations of AI can help manage expectations and avoid misuse.
5. **Regulatory Oversight:** Governments and international organizations play a vital role in regulating AI development and deployment. By establishing appropriate regulations, they can strike a balance between innovation and safety. These regulations should focus on issues like data privacy, algorithmic fairness, and accountability.

To achieve balanced growth, we must ensure that AI development is collaborative, rather than dominated by a select few. Fair competition and transparency are essential in driving progress, rather than resorting to marketing hype. By discussing the challenges and communicating with authenticity, we can harness the power of AI for the benefit of humanity.

Like, subscribe, or share if you want to support content like this!

References:

<https://www.npr.org/2023/03/29/1166896809/tech-leaders-urge-a-pause-in-the-out-of-control-artificial-intelligence-race>

<https://www.theverge.com/2024/10/24/24278999/openai-plans-orion-ai-model-release-december>

<https://techcrunch.com/2024/10/26/openai-denies-its-releasing-a-model-called-orion-this-year/>

<https://openai.com/index/red-teaming-network/>

<https://patriciagestoso.com/2024/05/21/openai-chatgpt-4o-the-good-the-bad-and-the-irresponsible/>

<https://www.technologyreview.com/2024/05/22/1092763/openais-gpt4o-chinese-ai-data/>

<https://deeperinsights.com/ai-blog/the-unspoken-challenges-of-large-language-models>