Rapid Response

AI making a play for scale

GPT4o and Google I/O

Updated: May 17th 2024

Better decisions, faster.

OMD

Open AI spring updates

After a relatively quiet few weeks, this week saw significant AI news as both Google and Open AI announced new capabilities and, perhaps more importantly, enabled wider access to AI services.

Open AI's focus for its 'Spring Update' was driving the reach of its services with a new model, GPT4o. The o stands for omni as it is multimodal with vision, voice, and emotion capabilities.

The new model is more capable than the previous flagship GPT4Turbo, scoring higher on most standard LLM model benchmarks.
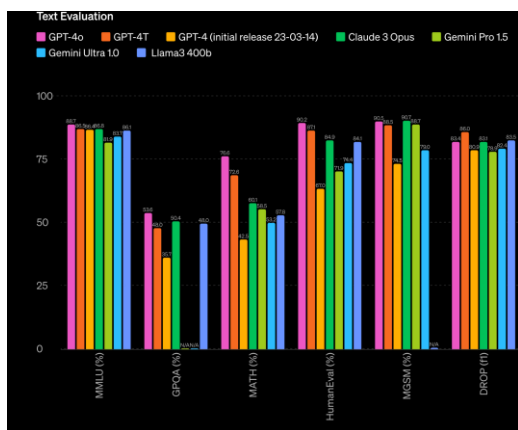
GPT4o is also far faster. This was less of an issue in text interactions as output speed was faster than most of us could read, but it is a significant unlock in voice interactions. The faster model enables a natural-feeling real-time voice assistant.

In a surprise move, the new flagship model will be made available to free and paid users. Thereby accelerating the penetration of their most advanced model in a race to win consumers. Also, token costs will be halved for APIs users opening-up use cases that may have previously proven uneconomical.

Open AI also opens the custom GPT store to non-paying users. Over one million people have already created custom GPTs and recently announced revenue share plans mean that we can expect a thriving GPT business model like that seen in mobile app stores.

Marketing use cases

Outside the main webcast, Open AI also shared numerous capabilities and use cases for GPT4o that illustrate the new kinds of interactions that marketers should explore.

- Grocery lists created and then organized around store maps

- Greater emotional range means that brands will be able to express themselves, or personalize, through tone and character

- Educational use cases imply efficiencies for training and guiding customers through complex processes
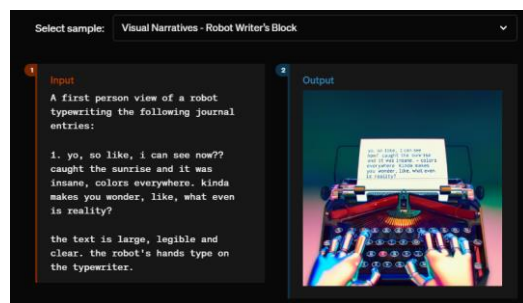
- Images with accurate text representation will help new forms of creative production.

The vision and voice features of GPT4o were illustrated with long-term Open AI partner Be My Eyes, which shows how blind and visually impaired people can be supported. The implications for travel brands are clear, helping discovery or real-time translation.

It hints at future search behaviors. Already, Google delivers 12bn visual searches per month, and this will likely grow as people become accustomed to asking for more information about what they see with a voice query.

Voice assistants may interact on behalf of the consumer, not just with them – Open AI demoed dealing with a customer support line. The accelerated scaling of a free service may mean businesses that rely on customers not doing things, e.g., cancelling a contract, may face increased challenges.

Previews of GPT4o are now in Microsoft Azure, enabling large-scale, secure deployment. We can expect to hear more at next week's Microsoft Build 2024 developer conference.

Better decisions, faster.

Google is making AI helpful for everyone

At this week's I/O developer conference, Google delivered many product updates related to embedding relating to the Gemini model into products that compete with Open AI.

- A video model called Veo to compete with Sora
- Imagen3 to compete with Dalle3
- Gems are like Open AI custom GPTs to create highly specific services
- Music AI Sandbox to compete with music startups like udio

However, the most impactful announcements were made in Google's core products. The lab's Search Generative Experience will be launched in search as 'AI Overviews' in the US this week and elsewhere later.

Google is also launching agents in search, with planning and reasoning to do research for the consumer and create recommendations for optimal purchase decisions, for example, creating a personalized 3-day meal plan for the family and exporting the grocery list. However, publishers worry about zero-click search's impact on inbound traffic.
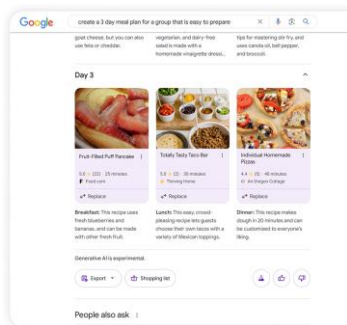
Agents are being widely deployed across Google products through Project Astra to automate returning purchases.

In Google Workspace, an expanded 2 million token context window means that all emails from your child's school can be summarized or created as a personalized educational podcast.

A video of your bookshelf can further catalogue all your books, opening second-hand markets. Data from receipts and locations can help curate vacation itineraries. We can expect to see a major battle to win in travel curation.

Implications and opportunities

The nature of LLM development means that new capabilities can be hard to predict. Therefore, brands must be agile in planning for disruption as new modalities and ways of interacting come on stream. Aggressive price reduction means that meaningful innovation is likely to scale quickly.

Generative AI is only just getting started impacting consumer behaviours. This trend briefing from Sparks & Honey explores how AI technologies will impact product search.

The speed of development means that responsible use of AI is more challenging and important than ever. Last week, Open AI released the model spec to 'deepen the public conversation about how AI models should behave.'

It highlights mechanisms to control what models can output and to whom. This will allow brands to manage legal compliance and policy in areas such as advising on competitor products.

Example: assuming best intent and staying helpful

This week's announcements show that scale matters. The biggest players are reaching beyond tech early adopters to those interested in convenience, saving money, learning, and expressing themselves. Across every part of the consumer experience, we are seeing disruption at ever-lower price points.

Understanding the complex implications of these innovations is critical for every brand.

Jean-Paul Edwards
OMD Worldwide Managing Director, Product
jean-paul.edwards@omd.com

Better decisions, faster.

OMD