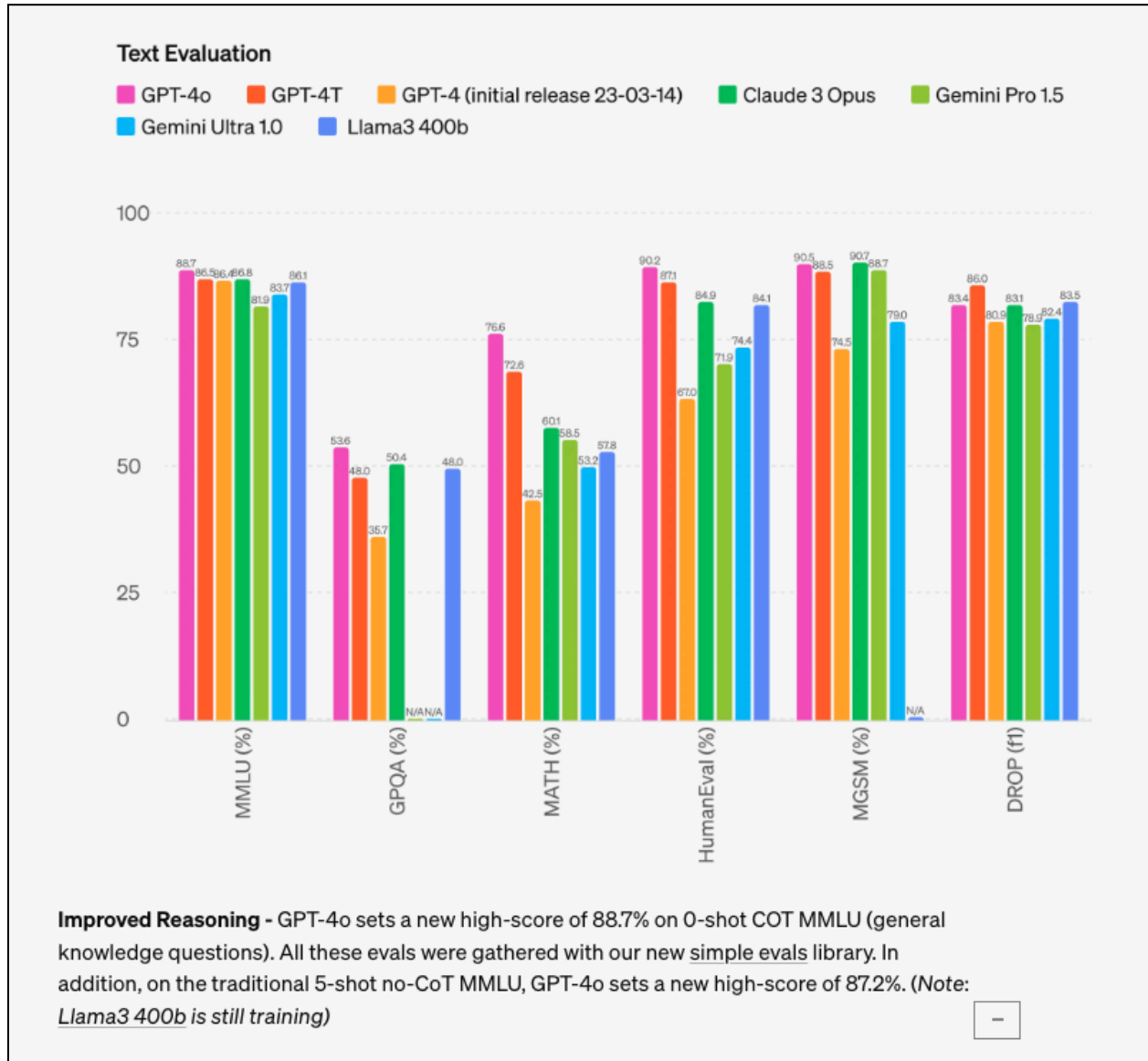


The Empire Strikes Back: GPT-4o Regains the Lead in the GenAI Race

A first look at this exciting new GenAI release.



[From the OpenAI Blog Post Announcing GPT-4o](#)

By John Tredennick and Dr. William Webber

Not long after Anthropic seized the spotlight with the release of Claude 3 Opus, OpenAI has struck back, elevating the stakes in the GenAI wars with its latest innovation, GPT-4o. The "o" in GPT-4o stands for "omnimodal," a nod to its robust capacity to understand and process diverse data types—including audio, images, and video—within a unified model framework.



While the name might echo its predecessor, GPT-4 Turbo, the capabilities of GPT-4o mark a significant evolution in AI technology. This release is not just another step forward; it is a transformative leap that revolutionizes how we handle multimedia data and text analysis. With improvements that include doubling the processing speed and halving the cost of its predecessor, GPT-4o is setting new benchmarks for efficiency and versatility.

So, what's the story on GPT-4o?

We are just getting our hands on it, but there is a lot to be excited about. Speed and cost reductions are nice, but this release goes way beyond that. Here is a quick look at several important new features that will almost certainly have an impact on discovery workflows.

Key Features

GPT-4o is fundamentally multimodal, natively supporting four types of input: text, audio, video, and images. Discovery professionals will find each of these capabilities highly beneficial.

1. Image Recognition

Photos, diagrams, and other graphical representations are central to many legal disputes. At my old company Catalyst, we frequently handled pictures and a variety of other images that could not be retrieved using keyword searches.

To paraphrase Amazing Grace, GPT was blind but now it can see. Let's test this by feeding a photo into GPT-4o and asking it to describe the image. Here is a photo retrieved from the Internet I uploaded to GPT. Here was my prompt: "Describe this picture and tell me all you can about it."



Photo retrieved from the Internet

GPT's answer:

The image depicts an aerial view of a car accident at an intersection. Here are the details and observations about the scene:

- Location and Setting:
 - The intersection appears to be in a residential or suburban area, as indicated by the surrounding buildings and the types of vehicles present.
 - The roads are paved but show signs of wear and tear, suggesting they are not newly constructed.
- Accident Details:
 - Two cars are involved in the collision. One is a white car, and the other is a dark-colored car.
 - The impact seems to have occurred in the middle of the intersection, with the front of the dark-colored car hitting the side of the white car.
 - Both cars have come to a stop at the center of the intersection, blocking some of the lanes.
- Other Vehicles and People:

- Several other cars are parked around the intersection, some on the sidewalk or very close to the edge of the road.
- A few people are present at the scene, likely involved in or witnessing the accident. One person appears to be standing near the cars, possibly assessing the damage or communicating with the other driver.
- Environmental Context:
 - There are trees and greenery visible in the upper part of the image, suggesting a relatively green neighborhood.
 - The general state of the roads and the presence of multiple cars indicate a moderately busy area, although it is not a main highway.
- Potential Causes and Consequences:
 - The cause of the accident is not immediately clear from the image, but it could be due to a failure to yield, running a stop sign, or misjudgment while crossing the intersection.
 - The consequences likely involve vehicle damage and potential injuries to the occupants, although no severe damage or injuries are visible in the image.

This image is useful for illustrating the aftermath of a typical urban intersection collision and can be used to discuss traffic safety, the importance of yielding and stopping at intersections, and the response to minor traffic accidents.

You have to admit, that answer was pretty good. How might we use it? For starters, we plan to add this capability to DiscoveryPartner so we can ask it to describe non-text documents. Imagine being able to have your litigation system identify and describe these otherwise blind-to-the-system files. Store the description as text and suddenly the files are searchable. GPT-4o can do more than describe images, it can analyze them. Think of the possibilities there.

2. Audio and Video

This new model by OpenAI is described as "omnimodal," meaning it can process and understand multiple types of data, including text, audio, images, and video, all within a single model. The service is not yet available for us to test but the OpenAI webcast showing its capabilities was astounding. For the first time, a computer can speak to a human in a fashion similar to how the computer on the Starship Enterprise spoke to Captain Kirk over 50 years ago. (Forgive the mixing of metaphors here.) It just sounds amazing.

GPT-4o's audio and video processing capabilities will be important for a number of industries, including the legal sector. Based on early reports, it can accurately transcribe spoken language, differentiate between speakers, understand context, and detect nuances like tone and emotion. It can also summarize long audio clips, analyze the



emotional and tonal aspects of speech, and extract keywords, making it easier to grasp the main points quickly.

Imagine being able to have GPT-4o review wiretap recordings, long runs of video or use it to review, transcribe and analyze deposition recordings. Certainly our friends on the criminal defense side will appreciate GPT-4o's help in going through lengthy audio recordings produced by the government.

3. Multi-Lingual

We regularly use large language models like GPT and Claude to read, analyze and translate different languages. GPT-4o now supports more than 50 languages including Spanish, French, German, Chinese, Japanese, Korean, Russian and many others. According to OpenAI, the new model covers 97% of the world's spoken languages.

Imagine using a mobile version of GPT-4o for witness interviews in a different language, possibly removing the need for translators. Or using GPT-4o to review documents in a variety of languages stored in a single repository.

For international investigations, the value is even greater. Imagine a widely dispersed team of investigators reviewing documents in a language they don't speak with GPT as the universal translator. Then go further to imagine that the reviewers don't even speak the same languages as their teammates. Yet each can interact with GPT in their native language, having GPT translate the contents of foreign files including audio and video as they review.

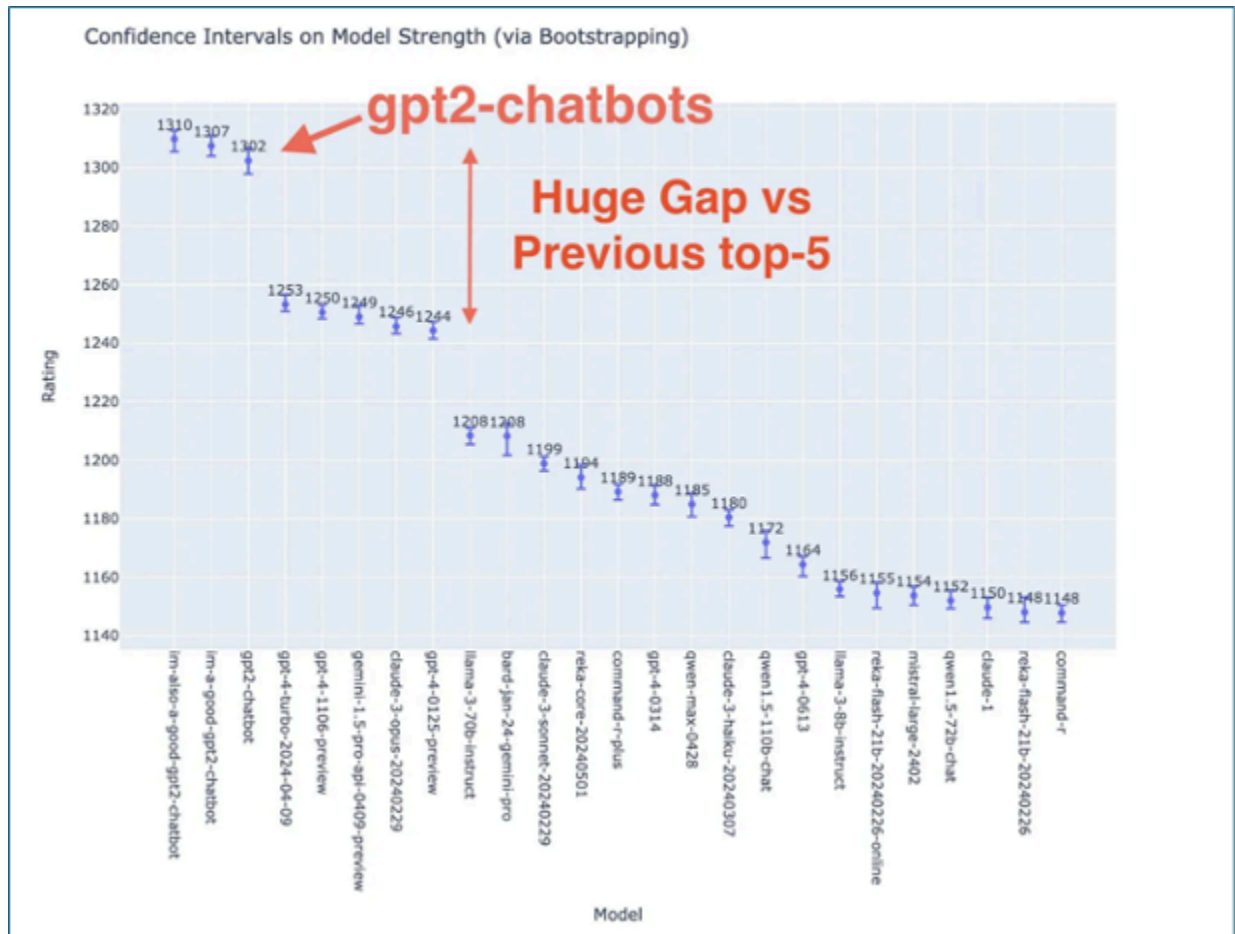
4. Faster and Better

For most discovery professionals, the biggest benefit of the new model is its speed and improved analytical capabilities. We all know the benefits of speed. Without trying to put a number on it, I can tell you that answers just fly onto the page when you ask GPT-4o questions. Improved speed means we can submit queries involving hundreds of documents, a typical scenario in our work, and receive answers within seconds, closely mimicking real-time interaction. For discovery professionals like me who don't like standing in line, this new response is delightful.

Even more important, GPT-4o reportedly sets a new standard for comprehension and analytical capabilities. Before its release, Claude Opus and GPT-4 Turbo set the benchmarks for large language models. We liked and used both with a feeling that Claude Opus was a bit better than GPT-4 Turbo.

Here is one comparison using the ELO rating system, which was originally used to compare the skill levels of different chess players. In the context of AI models, it serves as a quality measure to evaluate the performance and effectiveness of different chatbots or AI systems. The overall ELO score indicates the general performance of the models across various tasks or benchmarks.

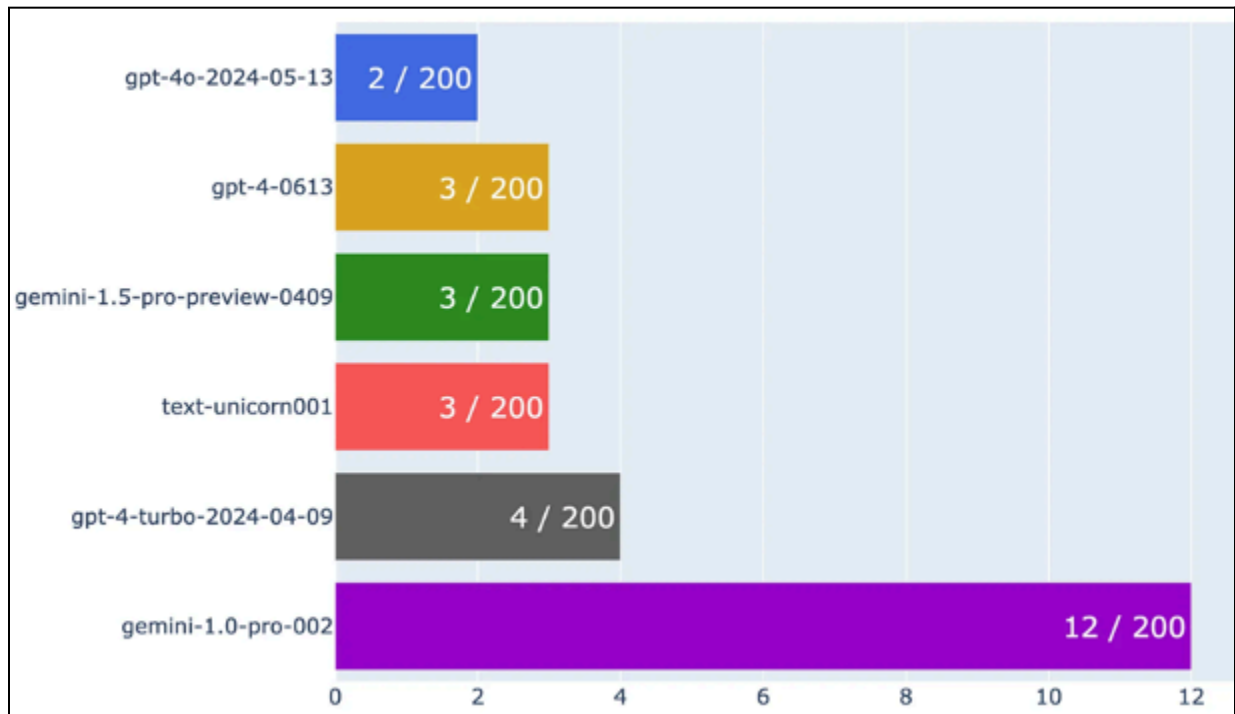
Here is a chart showing how the models compare:



Ignacio de Gregorio: [ChatGPT-4o, OpenAI's New Flagship Model's Full Review](#)

The testing was run against the new ChatGPT and shows that the GPT-4o chatbots (labeled “gpt2-chatbots” in the chart) significantly outperform both GPT-4 and Claude 3 Opus models in terms of overall ELO scores.

Here are the results of another kind of test. This one compares the number of mistakes different models made in matching specific sentences to their corresponding topics.



[GPT-4o vs. GPT-4 vs. Gemini 1.5 — Performance Analysis](#)

The fewer the mistakes, the better.

Did the Empire Strike Back?

DiscoveryPartner harnesses large language models like GPT and Claude to rapidly summarize documents based on the user's questions. Accuracy and reliability are critical for these tasks because legal professionals use this information for case strategy and discovery examinations. With each advancement in these models, our ability to assist users in their investigations grows stronger.

In crafting our GenAI strategy, we consciously opted for a system that employs LLMs from various leading providers. This design not only allows us to match specific models to particular tasks but also offers the flexibility to seamlessly incorporate the latest breakthroughs from industry giants—such as OpenAI and Anthropic—as well as promising new entrants. This approach ensures that our clients always have access to the most advanced tools at the most competitive prices.

As the GenAI wars unfold—featuring titans like OpenAI, Anthropic, Google, and others vying for dominance—the question lingers: Will the Empire regain control of the galaxy, or will the Jedi return with even more potent tools? What remains certain is that legal professionals will gain from the ongoing enhancements in speed and efficiency, while our clients will appreciate the cost savings in their investigations and discovery efforts.



About the Authors

John Tredennick (JT@Merlin.Tech) is the CEO and founder of Merlin Search Technologies, a software company leveraging generative AI and cloud technologies to make investigation and discovery workflow faster, easier, and less expensive. Prior to that he was founder and CEO of Catalyst Repository Systems, which he sold to a public company in early 2019. For the first 20 years of his career, he was a trial lawyer and litigation partner at a national law firm.



Tredennick is a prolific speaker and writer. Over the past 30 years, he has written eight books and countless articles on legal technology topics, including “TAR for Smart People” (3rd Ed.), two ABA bestsellers “Winning with Computers” (Vols. One and Two), “How to Prepare For, Take and Use a Deposition at Trial (James Publishing), and several editions of “The Lawyer’s Guide to Spreadsheets.” Tredennick has served as a Chair of the ABA’s Law Practice Management Section and Editor in Chief of its flagship magazine. He is currently active with EDRM and the Sedona Conference having a lead drafting role on numerous of their publications.

Dr. William Webber (wwebber@Merlin.Tech) is the Chief Data Scientist of Merlin Search Technologies. He completed his PhD in Measurement in Information Retrieval Evaluation at the University of Melbourne under Professors Alistair Moffat and Justin Zobel, and his post-doctoral research at the E-Discovery Lab of the University of Maryland under Professor Doug Oard.



With over 30 peer-reviewed scientific publications in the areas of information retrieval, statistical evaluation, and machine learning, he is a world expert in AI and statistical measurement for information retrieval and ediscovery. He has almost a decade of industry experience as a consulting data scientist to ediscovery software vendors, service providers, and law firms.

About Merlin

Merlin is a pioneering software company harnessing Generative AI and advanced cloud technologies to revolutionize search, investigations, and document discovery. Our DiscoveryPartner platform helps people find, analyze, and review information in large document sets faster, more effectively, and at a fraction of the cost of manual efforts. Our unique single-tenant architecture provides greater security and set the stage for our industry-first "On/Off" Cloud Utility Pricing, a green computing initiative that allows clients to save up to 60% on hosting costs by turning their site off when not in use.



For more information about Merlin Search Technologies, visit www.Merlin.Tech.