

OpenAI's latest models: o3, o4-mini, and GPT-4.1

April 22, 2025



OpenAI has released a host of new models and tools, including GPT-4.1, new reasoning models o3 and o4-mini, and Codex CLI. These models represent significant advancements in AI capabilities, with improved reasoning, tool use, and performance across various tasks. The new models can agentially use and combine every tool within ChatGPT—including web search, Python analysis, visual reasoning, and image generation.

The details:

Model capabilities: For the first time, OpenAI's reasoning models can agentially use and combine every tool within ChatGPT—including searching the web, analysing uploaded files with Python, reasoning about visual inputs, and generating images.

Pricing details: o3 is priced at \$10/million input tokens and \$40/million for output tokens, with a 75% discount on cached input tokens. o4-mini is more affordable at \$1.10/million for input and \$4.40/million for output.

Performance metrics: The models feature a 200,000 token context window, 100,000 max output tokens, and a May 31st, 2024 training cut-off (same as the GPT-4.1 models).

New feature: A new capability is that the OpenAI API can now optionally return reasoning summary text, providing more transparency into the model's thinking process.

Why it matters for project delivery professionals

For Project Managers, these advancements represent a significant opportunity to enhance project delivery workflows. Project Managers can leverage these models to automate documentation, risk assessment, and

status reporting.

Software Developers will benefit from improved code generation and debugging capabilities, potentially reducing development cycles by 30-40%. System Architects can integrate these models into existing workflows to create more intelligent systems that adapt to changing requirements with minimal human intervention. Teams should evaluate how these capabilities might reshape their technical roadmaps and skill development priorities.

Rabbit Hole:

[Read the OpenAI statement](#)

What are early perceptions? Is it a step closer to AGI?