

Which is better for automation tasks: GPT-5.4 or Claude Opus 4.6?

The GPT-5.4 and Claude Opus 4.6 are currently at the center of that question. Both have different capabilities and were launched just weeks apart. However, both models have different price points and perform best in different scenarios.

A few years ago, it was difficult to use a large language model to write a decent email. When OpenAI released its first open-source model, it was amazing to see it generate coherent text. Just a few years later, we have AI models that can build entire software engineering projects, schedule meetings, buy products on Amazon, and so on. By 2026, the landscape has truly changed, and the question developers are asking is which model will best suit their use case.

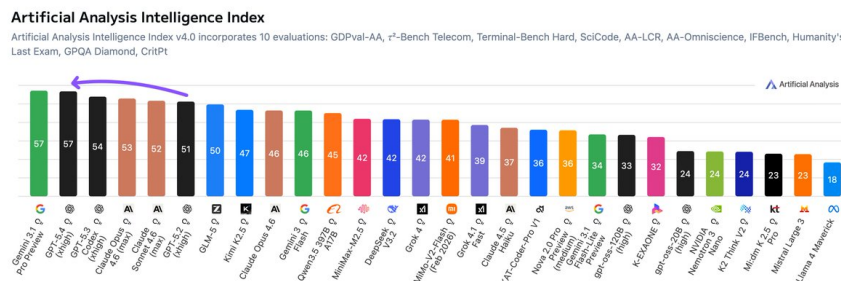
The GPT-5.4 and Claude Opus 4.6 are currently at the center of that question. Both have different capabilities and were launched just weeks apart. However, both models have different price points and perform best in different scenarios.

This article will help you decide which model is best suited to your workflow.

A direct comparison of GPT-5.4 and Claude Opus 4.6

Now, let's compare GPT-5.4 and Opus 4.6 to determine which model is best suited to your use case.

Overall, the GPT-5.4 is the best model according to the Artificial Analysis Intelligence Index (AII), which measures the performance of models across various benchmarks. Only the Gemini 3.1 Pro is superior.



Agent and computer utilization efficiency

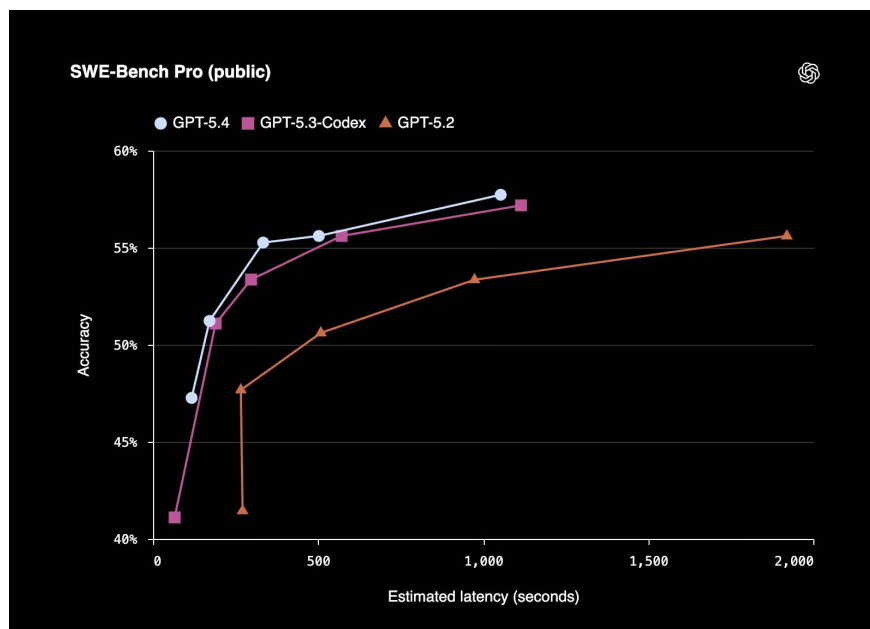
Claude Opus 4.6 excels when it comes to multi-agent coordination. With Agent Teams, you can run multiple workflows with agents simultaneously performing different tasks.

GPT-5.4 wins by a narrow margin in terms of computing performance. If your agent needs to operate a desktop, browse a browser, or interact with graphical user interface (GUI) software, then GPT-5.4 is the better choice currently available.

Programming benchmarks

Claude Opus 4.6 is a better programmer with a score of 80.84% on the SWE-Bench Verified and 81.4% when using the modified prompt.

GPT-5.4 inherits the programmability of GPT-5.3-Codex. According to OpenAI, GPT-5.4 scored 57.7% on SWE-Bench Pro (Public) with lower latency in inference tasks.



Cost and effectiveness of token usage

In its report, OpenAI stated that GPT-5.4 demonstrated a 47% reduction in token usage for certain tasks. Although more expensive than Opus 4.6, GPT-5.4 could be cheaper to operate at scale due to this token reduction.

However, Opus 4.6 may still be a better model for performing fewer complex tasks.

To put it into perspective, the most robust GPT-5.4 model (context length > 272K) costs \$60 per million input tokens and \$270 per million output tokens, while Claude Opus 4.6 costs \$5 per million input tokens and \$25 per million output tokens.

Context window and memory

Both GPT-5.4 and Claude Opus 4.6 support up to 1 million context tokens, although the Claude version is still in beta. This makes both models strong competitors when working with large codebases.

Comparison table

Criteria	Claude Opus 4.6	GPT-5.4
Automated tasks	Powerful (Agent Teams, parallel coordination)	Proficient (computer skills, OSWorld 75%)
Programming benchmarks	SWE-Bench scored 80.2% in cognitive ability.	57.7% on SWE-Bench Pro (Public)
Use a computer.	72.7% on OSWorld	OSWorld 75% (outperforming even human experts)
Context window	1M token (beta), 128K maximum output	1M token
Intellectual work	Humanity's Last Exam Leader	GDPval 83%
Pricing (inputs/outputs)	The basic entry token price is \$5. The token price is \$25 per million tokens.	gpt-5.4
Token effectiveness	Standard	Reduced the number of tokens required for certain tasks by 47%.
Most suitable for	These agents operate for extended periods and have complex code bases.	Using computers, document workflows, businesses

Should I choose GPT-5.4 or Claude Opus 4.6?

Finally, let's answer the most important question: Which of these two should you choose?

You should choose Claude Opus 4.6 if...

1. You are building or running agents that operate within large codebases over extended periods.
2. You want a multi-agent workflow where different agents work in parallel and delegate tasks to each other.

3. Your workflow involves very long documents, lengthy code files, or tasks that require holding a large amount of context.
4. You're already in the Anthropic ecosystem and your team is familiar with Claude .

You should choose GPT-5.4 if...

1. Your AI agent needs to operate the computer. Click, type, navigate applications, and fill out forms automatically.
2. You work in professional fields such as finance, legal, or operations, and need an operating model at the level of an industry expert.
3. You want to reduce API costs on a large scale. Improvements in token usage efficiency of up to 47% for certain tasks will accumulate gradually over thousands of completions per day.
4. You want a single model for everything without having to switch between specialized models.

Mô hình AI nào phù hợp cho nhu cầu của bạn?



Future prospects

Anthropic models have long been a top choice for programming, but they also shine in unexpected areas like creative writing. In fact, many consider them the best models in the industry in this field.

But Anthropic has never publicly claimed that its models specialize in any particular task, unlike how OpenAI claims that its Codex models are specifically designed for programming.

It's interesting that OpenAI is now following the direction of Anthropic. With its latest releases, they are moving towards a single, unified model that handles a wide range of professional tasks. This is a big win for users; nobody wants to constantly switch between specialized models to get their work done.

On the other hand, it's good to see Anthropic adopting a 1 million token context window, something other models have had for a long time (like Gemini 3). In the future, these models will have very similar features, to the point where there will be very few obstacles for users. However, the performance of the model across different tasks will be the main differentiating factor, as users will prioritize models that perform well in their specific workflows.

Conclude

In 2026, both Anthropic and OpenAI will have powerful models for automation work. What might be confusing is that they report different benchmarks. Perhaps they are selectively choosing the areas where their models will shine.

Now, you need to consult independent analyses of other benchmarks and test them on your own use cases. However, it's clear that the models are getting better and better. And you should be using them better too. One way to ensure you don't fall behind in this automation movement is to master how to effectively use these models for software engineering.

You finished reading the article "**Which is better for automation tasks: GPT-5.4 or Claude Opus 4.6?**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.