Learn about Claude Opus 4.7: The latest AI model from Anthropic, just released.

Claude Opus 4.7 is Anthropic's most powerful model available to date. It boasts high autonomy and performs particularly well in long-term agentic tasks, cognitive tasks, visual tasks, and memory tasks.

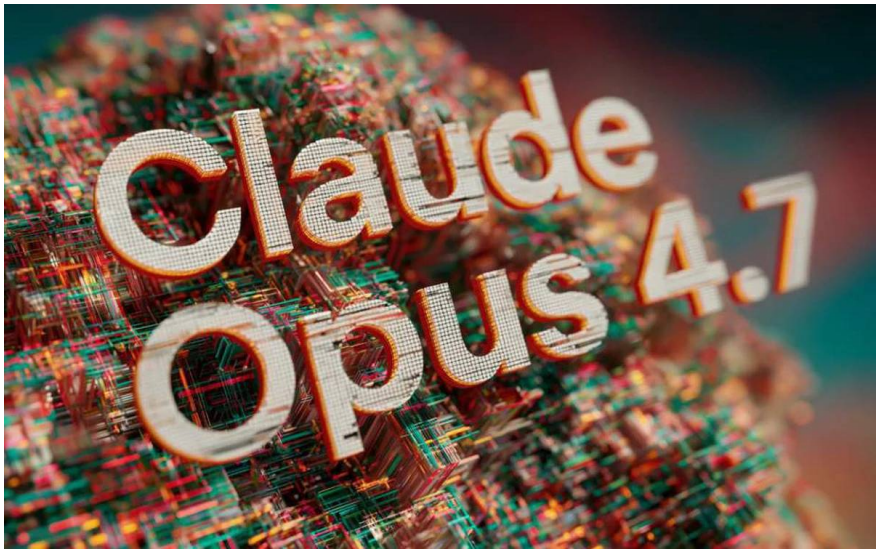
Claude Opus 4.7 is Anthropic's most powerful model to date. It boasts high autonomy and performs particularly well in long-term agentic tasks, cognitive tasks, visual tasks, and memory tasks. The following article summarizes all the new features of Claude Opus 4.7 upon its release.

Claude Opus 4.7: The latest AI model from Anthropic has just been released.

Model	API Model ID	Describe
Claude Opus 4.7	`claude-opus-4-7`	Anthropic's most powerful model currently available for complex reasoning and agentic programming.

Claude Opus 4.7 supports a 1 million token context window, a maximum output of 128,000 tokens, Adaptive Thinking mode, and the same toolset and platform features as Claude Opus 4.6 .

New features of Claude Opus 4.7



Supports high-resolution images.

Claude Opus 4.7 is Anthropic's first Claude model to support high-resolution images. The maximum image resolution has increased to 2576px / 3.75MP (up from the previous limit of 1568px / 1.15MP). This change will improve performance for demanding image processing tasks, and is especially important for computing and workflows that understand screenshots/objects/documents.

In addition, operations such as mapping coordinates to images are now simpler – the model's coordinates are 1:1 scaled with the actual pixels, so no scaling factor needs to be calculated.

High-resolution images use more tokens. If higher image resolution is unnecessary, reduce the image resolution before sending it to Claude to avoid increased token usage.

In addition to resolution, the Claude Opus 4.7 also offers improvements in:

1. Lower-level cognition - pointing, measuring, counting, and similar tasks.
2. Image localization - improves the ability to locate and detect natural image boundary boxes.

New xhigh effort level

The effort parameter allows you to adjust Claude's intelligence based on token cost, trading capabilities for faster speed and lower cost. Start with the new xhigh effort level for programming and agent use cases, using the minimum high effort level for most intelligence-sensitive use cases.

Task budgets (beta)

Claude Opus 4.7 introduces Task Budgets. Task Budgets provide Claude with a rough estimate of the number of tokens needed to target for a complete agent loop, including inference, tool call, tool result, and final output. The model will see a running countdown and use it to prioritize work and efficiently complete tasks when the budget is used up. To use it, set the title to beta task-budgets-2026-03-13 and add the following to your output configuration:

```
response = client.beta.messages.create( model="claude-opus-4-7", max_tokens=1280
```

You may need to experiment with different task budgets for your use case. If the model is given a task budget that is too restrictive for a particular task, it may complete the task less thoroughly or refuse to perform the task at all.

For unlimited automated tasks where quality is more important than speed, a task budget should not be set. Allocate task budgets to workloads where you need to model and limit the scope of the work within a token limit. The minimum value for a task budget is 20,000 tokens.

This isn't a hard limit; it's a hint that the model is aware of. This is different from `max_tokens`, which is a hard limit per request on the number of tokens generated (`max_tokens` isn't passed to the model and the model doesn't know about it), while `task_budget` is an advisory limit across the entire automated loop. Use `task_budget` when you want the model to self-regulate and `max_tokens` as a hard limit per request to restrict usage.

Changes that affect backward compatibility.

These backward compatibility changes only apply to the Messages API. If you are using Claude Managed Agents, there are no API changes that affect backward compatibility with Claude Opus 4.7.

Extended thinking budgets have been removed.

Extended thinking budgets were removed in Claude Opus 4.7. Setting thinking: {"type": "enabled", "budget_tokens": N} will return a 400 error. Adaptive thinking is the only thinking mode enabled and in Anthropic's internal reviews, it consistently performs better than Extended thinking.

```
# Before (Opus 4.6) thinking = {"type": "enabled", "budget_tokens": 32000} # Aft
```

Adaptive thinking is disabled by default on Claude Opus 4.7. Requests without a thinking field will run without any thinking process. Specify `thinking: {type: "adaptive"}` to enable this mode.

Sampling parameters have been removed.

Starting with Claude Opus 4.7, setting `temperature`, `top_p`, or `top_k` to any value other than the default will return a 400 error. The safest workaround is to omit these parameters entirely from requests and use prompts to guide the model's behavior. If you are using `temperature = 0` to ensure determinism, note that it never guarantees identical output.

Thinking content is ignored by default.

Starting with Claude Opus 4.7, Thinking content is ignored from the response by default. Thinking blocks still appear in the response stream, but their thinking field will be empty unless the caller explicitly selects to enable it. This is a silent change—no errors are thrown—and response latency will be slightly improved. If you need inference output, you can set `display` to `summarized` and select to enable it again with a single line change:

```
thinking = { "type": "adaptive", "display": "summarized", # or "omitted" (defaul
```

If your product conveys inference to the user, the new default setting will display as a long pause before output begins. Set "display": "summarized" to restore the display progress during inference.

Update the token counting method.

Claude Opus 4.7 uses a new tokenizer, contributing to improved performance across many tasks. This new tokenizer can use approximately 1 to 1.35 times more tokens when processing text compared to previous models (up to ~35% more, depending on content), and /v1/messages/count_tokens will return different token counts for Claude Opus 4.7 compared to Claude Opus 4.6. The token usage efficiency of Claude Opus 4.7 can vary depending on the workload shape. Prompt, task_budget, and effort interventions can help control costs and ensure appropriate token usage. Keep in mind that these controls may come at the cost of model intelligence.

Anthropic recommends updating the max_tokens parameters to allow for more space, including compression triggers. Claude Opus 4.7 provides a 1 million-line context window at the standard API price without an additional charge for long contexts.

Improved capabilities of Claude Opus 4.7

Intellectual work

Claude Opus 4.7 demonstrates significant improvements in the tasks of intellectual workers, especially when the model needs to visually verify its own output:

1. Editing .docx and .pptx files - improved in creating and self-checking tracked changes and slide layout.
2. Graph and image analysis - improved by calling programming tools with image processing libraries (e.g., PIL) for graph and image analysis, including data reproduction at the pixel level.

If existing prompts have mitigation measures in these areas (e.g., "check slide layout carefully before returning"), try removing that structure and re-establishing the baseline.

Memory

Claude Opus 4.7 is better at writing and using memory based on the file system. If an agent maintains a draft table, note file, or structured memory store throughout its turns, that agent will improve its note-taking capabilities and leverage those notes in future tasks. To provide Claude with managed temporary memory without building it itself, use the client-side memory engine.

Behavior change

These aren't changes that disrupt the API, but they might require a quick update.

1. Adhering to precise instructions is more effective, especially at lower effort levels. The model will not automatically generalize an instruction from one item to another and will not infer requirements you haven't specified.
2. The response length is adjusted according to the perceived complexity of the task, rather than defaulting to a fixed length.

3. Fewer tool calls by default, more reasoning. Increased effort will increase tool usage.
4. The tone is more direct and subjective, with less focus on authenticity and fewer emotional expressions compared to the warmer style of Claude Opus 4.6.
5. Provide more frequent progress updates to users throughout longer agent traces. If you've added a frame to force temporary status messages to be displayed, try removing it.
6. Fewer subagents are generated by default. This can be controlled via the prompt.
7. Real-time cybersecurity protection measures: Requests related to prohibited or high-risk topics may result in rejection. For legitimate security jobs, apply for the Cyber ??Verification Program.

If you are using Claude Code or Agent SDK, Claude's API skills can automatically apply the migrations to Claude Opus 4.7 to your source code.

You finished reading the article "**Learn about Claude Opus 4.7: The latest AI model from Anthropic, just released.**" 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.

© 2019 TipsMake.com