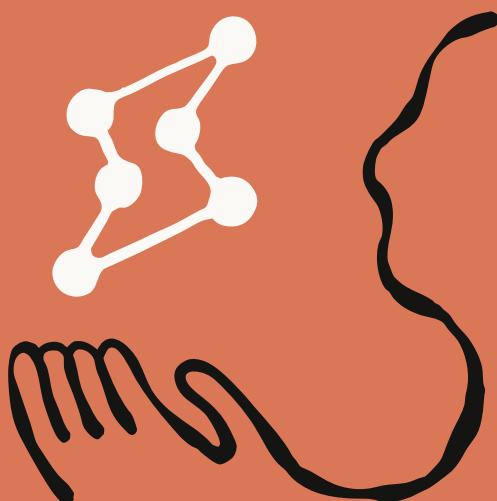


Product

# Introducing Claude Haiku 4.5

Oct 15, 2025 • 2 min read

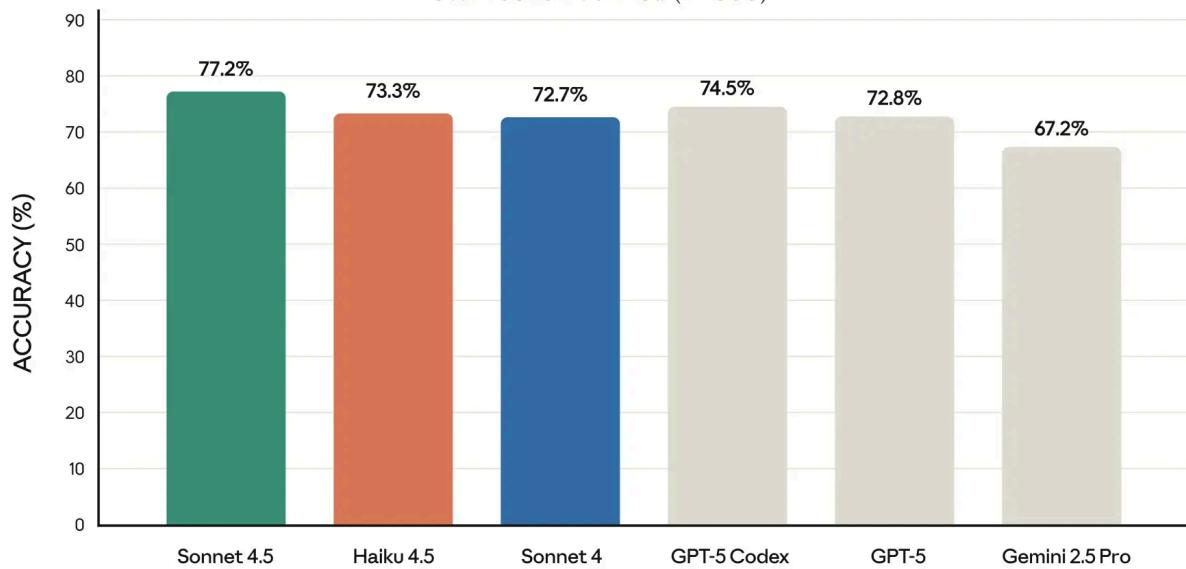


Claude Haiku 4.5, our latest small model, is available today to all users.

What was recently at the frontier is now cheaper and faster. Five months ago, Claude Sonnet 4 was a state-of-the-art model. Today, Claude Haiku 4.5 gives you similar levels of coding performance but at one-third the cost and more than twice the speed.

## Software engineering

SWE-bench Verified (n=500)



Claude Haiku 4.5 even surpasses Claude Sonnet 4 at certain tasks, like using computers. These advances make applications like [Claude for Chrome](#) faster and more useful than ever before.

Users who rely on AI for real-time, low-latency tasks like chat assistants, customer service agents, or pair programming will appreciate Haiku 4.5's combination of high intelligence and remarkable speed. And users of Claude Code will find that Haiku 4.5 makes the coding experience—from multiple-agent projects to rapid prototyping—markedly more responsive.



Claude Sonnet 4.5, released [two weeks ago](#), remains our frontier model and the best coding model in the world. Claude Haiku 4.5 gives users a new option for when they want near-frontier performance with much greater cost-efficiency. It also opens up new ways of using our models together. For example, Sonnet 4.5 can break down a complex problem into multi-step plans, then orchestrate a team of multiple Haiku 4.5s to complete subtasks in parallel.

Claude Haiku 4.5 is available everywhere today. If you're a developer, simply use `claude-haiku-4-5` via the Claude API. Pricing is now \$1/\$5 per million input and output tokens.

## Benchmarks

	Claude Sonnet 4.5	Claude Haiku 4.5	Claude Sonnet 4	GPT-5	Gemini 2.5 Pro
Agentic coding <i>SWE-bench Verified</i>	<b>77.2%</b>	73.3%	72.7%	<b>72.8%</b> GPT-5 (high) <b>74.5%</b> GPT-5-Codex	<b>67.2%</b>
Agentic terminal coding <i>Terminal-Bench</i>	<b>50.0%</b>	41.0%	36.4%	<b>43.8%</b>	25.3%
Agentic tool use <i>t2-bench</i>	Retail <b>86.2%</b> Airline <b>70.0%</b> Telecom <b>98.0%</b>	Retail <b>83.2%</b> Airline <b>63.6%</b> Telecom <b>83.0%</b>	Retail <b>83.8%</b> Airline <b>63.0%</b> Telecom <b>49.6%</b>	Retail <b>81.1%</b> Airline <b>62.6%</b> Telecom <b>96.7%</b>	—
Computer use <i>OSWorld</i>	<b>61.4%</b>	50.7%	42.2%	—	—
High school math competition <i>AIME 2025</i>	100% (python) <b>87.0%</b> (no tools)	96.3% (python) <b>80.7%</b> (no tools)	70.5%	<b>99.6%</b> (python) <b>94.6%</b> (no tools)	88.0%
Graduate-level reasoning <i>GPQA Diamond</i>	<b>83.4%</b>	73.0%	76.1%	85.7%	<b>86.4%</b>
Multilingual Q&A <i>MMMLU</i>	89.1%	83.0%	86.5%	<b>89.4%</b>	—
Visual reasoning <i>MMMU (validation)</i>	77.8%	73.2%	74.4%	<b>84.2%</b>	82.0%

Claude Haiku 4.5 is one of our most powerful models to date. See footnotes for methodology.

{"} augment code

“

warp

“



## Safety evaluations

We ran a detailed series of safety and alignment evaluations on Claude Haiku 4.5. The model showed low rates of concerning behaviors, and was substantially more aligned than its predecessor, Claude Haiku 3.5. In our automated alignment assessment, Claude Haiku 4.5 also showed a statistically significantly lower overall rate of misaligned behaviors than both Claude Sonnet 4.5 and Claude Opus 4.1—making Claude Haiku 4.5, by this metric, our safest model yet.

Our safety testing also showed that Claude Haiku 4.5 poses only limited risks in terms of the production of chemical, biological, radiological, and nuclear (CBRN) weapons. For that reason, we've released it under the AI Safety Level 2 (ASL-2) standard—compared to the more restrictive ASL-3 for Sonnet 4.5 and Opus 4.1. You can read the full reasoning behind the model's ASL-2 classification, as well as details on all our other safety tests, in the [Claude Haiku 4.5 system card](#).

## Further information

Claude Haiku 4.5 is available now on Claude Code and our apps. Its efficiency means you can accomplish more within your usage limits while maintaining premium model performance.

Developers can use Claude Haiku 4.5 on our API, Amazon Bedrock, and Google Cloud's Vertex AI, where it serves as a drop-in replacement for both Haiku 3.5 and Sonnet 4 at our most economical price point.

For complete technical details and evaluation results, see our [system card](#), [model page](#), and [documentation](#).

## Methodology

- **SWE-bench Verified:** All Claude results were reported using a simple scaffold with two tools—bash and file editing via string replacements. We report 73.3%, which was averaged over 50 trials, no test-time compute, 128K thinking budget, and default sampling parameters (temperature, top\_p) on the full 500-problem SWE-bench Verified dataset.
  - The score reported uses a minor prompt addition: "You should use tools as much as possible, ideally more than 100 times. You should also implement your own tests first before attempting the problem."
- **Terminal-Bench:** All scores reported use the default agent framework (Terminus 2), with XML parser, averaging 11 runs (6 without thinking (40.21% score), 5 with 32K thinking budget (41.75% score)) with n-attempts=1.
- **T2-bench:** Scores were achieved averaging over 10 runs using extended thinking (128k thinking budget) and default sampling parameters (temperature, top\_p) with tool use and a prompt addendum to the Airline and Telecom Agent Policy instructing Claude to better target its known failure modes when using the vanilla prompt. A prompt addendum was also added to the Telecom User prompt to avoid failure modes from the user ending the interaction incorrectly.
- **AIME:** Haiku 4.5 score reported as the average over 10 independent runs that each calculate pass@1 over 16 trials with default sampling parameters (temperature, top\_p) and 128K thinking budget.
- **OSWorld:** All scores reported use the official OSWorld-Verified framework with 100 max steps, averaged across 4 runs with 128K total thinking budget and 2K thinking budget per-step configured.
- **MMMLU:** All scores reported are the average of 10 runs over 14 non-English languages with a 128K thinking budget.
- All other scores were averaged over 10 runs with default sampling parameters (temperature, top\_p) and 128K thinking budget.

All OpenAI scores reported from their [GPT-5 post](#), [GPT-5 for developers post](#), [GPT-5 system card](#) (SWE-bench Verified reported using n=500), and [Terminal Bench leaderboard](#) (using Terminus 2). All Gemini scores reported from their [model web page](#), and [Terminal Bench leaderboard](#) (using Terminus 1).



News

**Anthropic officially opens Tokyo office, signs Memorandum of Cooperation with the Japan AI Safety Institute**

Oct 29, 2025

News

## Advancing Claude for Financial Services

Oct 27, 2025

News

## Seoul becomes Anthropic's third office in Asia-Pacific as we continue our international growth

Oct 23, 2025



### Products

[Claude](#)

[Claude Code](#)

[Claude and Slack](#)

[Claude in Excel](#)

[Max plan](#)

[Team plan](#)

[Enterprise plan](#)

[Download app](#)

[Pricing](#)

[Log in to Claude](#)

### Models

[Opus](#)

[Sonnet](#)

[Haiku](#)

## **Solutions**

AI agents  
Code modernization  
Coding  
Customer support  
Education  
Financial services  
Government  
Life sciences

## **Claude Developer Platform**

Overview  
Developer docs  
Pricing  
Amazon Bedrock  
Google Cloud's Vertex AI  
Console login

## **Learn**

Courses  
Connectors  
Customer stories  
Engineering at Anthropic  
Events  
Powered by Claude  
Service partners  
Startups program

## **Company**

Anthropic  
Careers  
Economic Futures

[Research](#)

[News](#)

[Responsible Scaling Policy](#)

[Security and compliance](#)

[Transparency](#)

## **Help and security**

[Availability](#)

[Status](#)

[Support center](#)

## **Terms and policies**

[Privacy choices](#)

[Privacy policy](#)

[Responsible disclosure policy](#)

[Terms of service: Commercial](#)

[Terms of service: Consumer](#)

[Usage policy](#)

© 2025 Anthropic PBC

