

Models overview

Claude is a family of state-of-the-art large language models developed by Anthropic. This guide introduces our models and compares their performance.

📄

Copy page

▼

Choosing a model

If you're unsure which model to use, we recommend starting with **Claude Sonnet 4.5**. It offers the best balance of intelligence, speed, and cost for most use cases, with exceptional performance in coding and agentic tasks.

All current Claude models support text and image input, text output, multilingual capabilities, and vision. Models are available via the Anthropic API, AWS Bedrock, and Google Vertex AI.

Once you've picked a model, [learn how to make your first API call](#).

Latest models comparison

Feature	Claude Sonnet 4.5	Claude Haiku 4.5	Claude Opus 4.5
Description	Our smart model for complex agents and coding	Our fastest model with near-frontier intelligence	Premium model combining maximum intelligence with practical performance
Claude API ID	claude-sonnet-4-5-20250929	claude-haiku-4-5-20251001	claude-opus-4-5-20251101
Claude API alias ¹	claude-sonnet-4-5	claude-haiku-4-5	claude-opus-4-5

AWS Bedrock ID	-sonnet-4-5-20250929-v1:0	haiku-4-5-20251001-v1:0	anthropic.claude-opus-4-5-20251101-v1:0
GCP Vertex AI ID	claude-sonnet-4-5@20250929	claude-haiku-4-5@20251001	claude-opus-4-5@20251101
Pricing ²	\$3 / input MTok \$15 / output MTok	\$1 / input MTok \$5 / output MTok	\$5 / input MTok \$25 / output MTok
Extended thinking	Yes	Yes	Yes
Priority Tier	Yes	Yes	Yes
Comparative latency	Fast	Fastest	Moderate
Context window	200K tokens / 1M tokens (beta) ³	200K tokens	200K tokens
Max output	64K tokens	64K tokens	64K tokens
Reliable knowledge cutoff	Jan 2025 ⁴	Feb 2025	May 2025 ⁴
Training data cutoff	Jul 2025	Jul 2025	Aug 2025

1 - Aliases automatically point to the most recent model snapshot. When we release new model snapshots, we migrate aliases to point to the newest version of a model, typically within a week of the new release. While aliases are useful for experimentation, we recommend using specific model versions (e.g., `claude-sonnet-4-5-20250929`) in production applications to ensure consistent behavior.

2 - See our [pricing page](#) for complete pricing information including batch API discounts, prompt caching rates, extended thinking costs, and vision processing fees.

3 - Claude Sonnet 4.5 supports a 1M token context window when using the `context-1m-2025-08-07` beta header. Long context pricing applies to requests exceeding 200K tokens.

4 - **Reliable knowledge cutoff** indicates the date through which a model's knowledge is most extensive and reliable. **Training data cutoff** is the broader date range of training data used. For example, Claude Sonnet 4.5 was trained on publicly available

❶ Models with the same snapshot date (e.g., 20240620) are identical across all platforms and do not change. The snapshot date in the model name ensures consistency and allows developers to rely on stable performance across different environments.

❷ Starting with **Claude Sonnet 4.5 and all future models**, AWS Bedrock and Google Vertex AI offer two endpoint types: **global endpoints** (dynamic routing for maximum availability) and **regional endpoints** (guaranteed data routing through specific geographic regions). For more information, see the [third-party platform pricing section](#).

> **Legacy models**

Prompt and output performance

Claude 4 models excel in:

- **Performance:** Top-tier results in reasoning, coding, multilingual tasks, long-context handling, honesty, and image processing. See the [Claude 4 blog post](#) for more information.
- **Engaging responses:** Claude models are ideal for applications that require rich, human-like interactions.
 - If you prefer more concise responses, you can adjust your prompts to guide the model toward the desired output length. Refer to our [prompt engineering guides](#) for details.
 - For specific Claude 4 prompting best practices, see our [Claude 4 best practices guide](#).
- **Output quality:** When migrating from previous model generations to Claude 4, you may notice larger improvements in overall performance.

Migrating to Claude 4.5

If you're currently using Claude 3 models, we recommend migrating to Claude 4.5 to take advantage of improved intelligence and enhanced capabilities. For detailed migration

Ask Docs

Get started with Claude

If you're ready to start exploring what Claude can do for you, let's dive in! Whether you're a developer looking to integrate Claude into your applications or a user wanting to experience the power of AI firsthand, we've got you covered.

 Looking to chat with Claude? Visit claude.ai!



Intro to Claude

Explore Claude's capabilities and development flow.



Quickstart

Learn how to make your first API call in minutes.



Claude Console

Craft and test powerful prompts directly in your browser.

If you have any questions or need assistance, don't hesitate to reach out to our [support team](#) or consult the [Discord community](#).

Was this page helpful?



Claude API Docs



Solutions

AI agents

Code modernization

Coding

Customer support

Education

Financial services

Government

Life sciences

Partners

Amazon Bedrock

Google Cloud's Vertex AI

Learn

Blog

Catalog

Courses

Use cases

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

Discord

Ask Docs

[Responsible disclosure policy](#)

[Terms of service: Commercial](#)

[Terms of service: Consumer](#)

[Usage policy](#)