

[Back to Models](#)**Anthropic: Claude Fable 5.1**

+ Active

anthropic/claude-fable-5.1

Compare

API Request

Chat

Claude Fable 5.1 is Anthropic's latest model for demanding reasoning and long-horizon agentic work, with stronger performance in long-running coding, multistep research, and document, spreadsheet, and slide workflows. It supports text and image inputs, text output, a 1M-token context window, and adaptive thinking that is always on with effort-based steering.

By : Anthropic Input Type : Output Type : Publish time : 2026-09-01

Providers

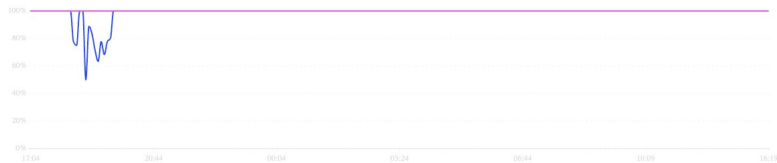
Route requests across multiple providers. Copy a provider slug to set your preference.

Provider	Cache Hit Rate	Input	Output	Cache read	Context	Latency	Throughput
Anthropic	-	\$10 / 1M tokens	\$50 / 1M tokens	Read: 0.25 / 1M tokens Write: 12.5-20 / 1M tokens	1M	28.4s	11.6tps
Google Vertex	-	\$10 / 1M tokens	\$50 / 1M tokens	Read: 0.25 / 1M tokens Write: 12.5-20 / 1M tokens	1M	2.78s	68.3tps

Uptime 24hours

Direct request success rate on AI Gateway and per-provider.

1D 1W



Sep 18, 09:39

Anthropic	100.00%
Google Vertex	100.00%

ZenMux Launches Claude Fable 5.1 Model Access for Developers

September 29, 2026

Singapore - September 29, 2026 -

ZenMux has added Claude Fable 5.1 to its model catalog, giving developers access to the model through the company's model-routing platform. The ZenMux listing records a publish date of September 1, 2026 and presents the model as an option for reasoning, coding, research, and document-centered workflows. The addition expands the set of Anthropic models available to teams that use ZenMux to evaluate and route artificial intelligence workloads.

The model page describes Claude Fable 5.1 as supporting text and image inputs with text output. It lists a context window of one million tokens, allowing applications to submit substantially larger bodies of source material than conventional short-context requests. Actual usable input sizes can depend on message structure, output limits, provider policies, and other implementation details, so developers remain responsible for testing production workloads against the applicable endpoint requirements. Typical use cases for this scale of context include analyzing lengthy legal or technical documents, reviewing extended codebases in a single pass, and maintaining continuity across long multi-turn research or drafting sessions.

The ZenMux Claude Fable 5.1 listing identifies long-running coding, multistep research, and work involving documents, spreadsheets, and presentations as relevant use cases. It also notes always-on adaptive thinking with effort-based steering. These descriptions are provided as model-page specifications rather than independent performance guarantees, and application teams can determine suitability through their own evaluations, test sets, and operational monitoring.

ZenMux lists routes from Anthropic and Google Vertex for the model. The availability of more than one provider route can help engineering teams compare operational characteristics while keeping the model selection consistent at the application layer. Provider availability, quotas, regional access, response time, and pricing may change, and the current model page remains the reference for route-specific information.

The model's support for image input also allows developers to evaluate mixed text-and-image tasks. Potential applications may include reviewing visual documents, interpreting screenshots, or combining written instructions with image-based reference material. Teams working with sensitive or consequential information should continue to use human review and domain-appropriate safeguards.

ZenMux positions the catalog as a practical layer for model discovery and routing rather than a substitute for application testing. The company publishes route information, context specifications, and observed platform metrics on individual model pages. Some displayed metrics are dynamic and can change with provider conditions, traffic, caching, and the time of measurement. For that reason, this announcement does not treat a displayed response-time or throughput figure as a permanent service commitment.

Developers reviewing ZenMux Claude Fable 5.1 can see the model identifier and current provider routes on the dedicated page. Existing users can evaluate the model within the same broader catalog used for other supported models, while new users can review the published specifications before selecting an integration approach.

The addition of Claude Fable 5.1 reflects ZenMux's ongoing expansion of its routed model catalog. The company plans to continue presenting model and provider information in a format that helps technical teams compare availability and make deployment decisions based on their own requirements.

For more information, visit <https://zenmux.ai/anthropic/claude-fable-5.1>

About ZenMux:

ZenMux is an enterprise-grade large model aggregation platform with an insurance payout mechanism. The platform provides one-stop access to the latest models across providers. When issues such as poor output quality or excessive latency occur during use, our intelligent insurance detection and payout mechanism automatically compensates, addressing enterprise concerns around AI hallucinations and unstable quality.

###

For more information about ZenMux, contact the company here: ZenMuxEmberember@zenmux.ai Singapore

ZenMux

ZenMux is an enterprise-grade AI model aggregation platform with a built-in insurance payout mechanism, offering one-stop access across providers and automatic compensation for quality or latency issues.

Website: <https://zenmux.ai/>

Email: ember@zenmux.ai

AI

Simplified & Assured.

ZenMux