

Back to Models

Google: Gemini 4 Argon

Active50% OFF

CompareAPI RequestChat

Gemini 4 Argon is a frontier model from Google DeepMind, released on September 30, 2026. It is the first model in the Gemini 4 series and uses a new naming scheme. It is strong in coding, financial research, legal drafting, and autonomously patching cybersecurity vulnerabilities. It accepts text, image, video, and file inputs and has a 1-million-token context window. On benchmarks, it scores 77.9% on DeepSWE v1.1, ahead of Claude Opus 5.5 and GPT-6 Astra, and ranks first on the Vals Index with 68.90%.

By: GoogleInput Type: [Image][Text][Code]Output Type: [Text]Publish time: 2026-10-13

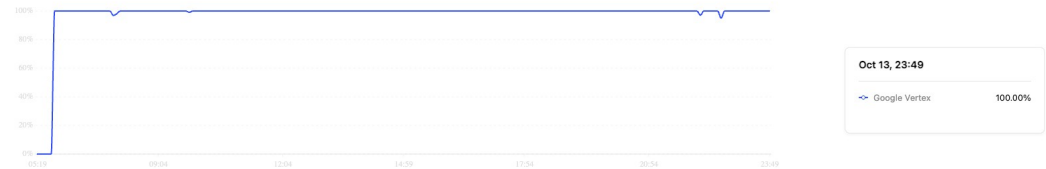
Providers

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

Provider	Cache Hit Rate	Input	Output	Cache read	Context	Latency	Throughput
Google Vertex google-vertex	66.4%	\$4 \$2 / M tokens	\$99 \$10 / M tokens	Read: 612 0.1 / M tokens Write: / M tokens	1.05M	4.6s	61.5tps

Uptime24hours

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



ZenMux Gemini 4 Access Expands With Argon Model Availability

October 15, 2026

Singapore - October 15, 2026 -

ZenMux has announced the availability of Google’s Gemini 4 Argon through its AI model routing platform, adding another model option for developers working on complex software and research tasks. The ZenMux Gemini 4 listing is active, and the company’s announcement states that the model can be used through the same ZenMux API already used to access other models.

The announcement focuses on long-horizon coding, legal and financial research, and cyber defense workflows. By introducing Argon within its existing API environment, ZenMux gives development teams a way to evaluate the model alongside other options without treating the announcement as a requirement to replace their current model selection.

ZenMux identifies Gemini 4 Argon as a Google DeepMind model and the first entry in the Gemini 4 series. Its model page dates the underlying model’s release to September 30, 2026, while the ZenMux listing displays a publication date of October 13, 2026, distinguishing the provider’s release from its appearance on the

routing platform.

The model page describes support for text, image, video, and file inputs, together with a context window of approximately one million tokens. Those specifications concern the model described in the listing; the announcement does not establish that every application or API request format will handle each input modality identically, so implementation decisions should follow the applicable platform documentation.

For software engineering, ZenMux's announcement highlights repository-scale debugging and refactoring across multiple files as potential areas for testing. These tasks can require a model to consider dependencies beyond a single function, follow an issue through several components, and propose changes that remain consistent with the rest of a codebase.

The company also identifies vulnerability discovery and patching as possible uses within cyber defense workflows. In that setting, model-generated analysis or proposed code changes remain material for an engineering review process rather than evidence that a vulnerability has been eliminated; tests and security checks are still necessary before changes reach production.

ZenMux's launch post reports a 77.9 percent score on DeepSWE v1.1 and a 68.90 percent score on the Vals Index, where it describes the model as ranking first. These figures are attributed to the company's published announcement and model listing, not to a separate ZenMux customer deployment or an independent evaluation conducted for this release.

The availability of ZenMux Gemini 4 therefore provides a starting point for application-specific evaluation rather than a prediction of identical results across workloads. Teams comparing models can use representative tasks and review the resulting code, reasoning, response time, and resource consumption against the requirements of their own systems.

Legal drafting and financial research are additional use cases named on the listing, reflecting the announcement's emphasis on multi-step work involving substantial source material. Outputs in these fields require qualified review and verification against the underlying documents, and the listing does not establish professional accuracy or authorize unsupervised decisions in regulated settings.

The shared API is a central part of ZenMux's announcement because model access is only one element of deploying an AI workflow. A team already working within that environment can investigate Argon as another model choice while separately determining how its application will manage inputs, evaluate responses, and handle cases that require human judgment.

The listing also presents provider information and operating details that developers can consult when

planning an evaluation. These details may change over time, and the launch announcement should not be read as a fixed service-level commitment or as proof of a particular production configuration for every account.

With the addition, ZenMux expands the model choices available through its routing platform while keeping the news centered on access to Google's Argon model. The ZenMux Gemini 4 Argon page provides the model identifier, current availability, and listing details for teams assessing whether it fits a specific coding or research workflow.

For more information, visit <https://zenmux.ai/google/gemini-4-argon>

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