

GPTProto

Search models...

Dashboard
Models
Features
Resources
Pricing
47% bonus
English
Sign Up

Home / Model / Google / gemini-4-argon

Gemini 4 Argon

Playground

Access Google's Gemini 4 Argon through GPTProto for text reasoning, software engineering, long-horizon agents, and structured analysis. Call gemini-4-argon with one GPTProto key and one shared balance across 200+ models, with API rates 40% below Google's direct pricing.

MODALITIES

T
I
O
E
M
→
T

INPUT / OUTPUT

Promo 0.5x
40% off

\$2.4 / 1M
\$4 official price

/

\$12 / 1M
\$20 official price

CACHE READ

40% off

\$0.12 / 1M
\$0.2 official price

Top up \$100 now → *free Wan 3.0 credits*

Oct 11 - Oct 31 - Limited quota - First come, first served

Top up now →

\$ Pricing

Related Models

Examples

About

FAQ

Blog

Gemini 4 Argon pricing

Chat, coding agents & document work. Priced per 1M tokens — input, cached input and output are billed separately. GPTProto is 40% below official rates.

SCENARIO	GOOGLE LIST	OPENROUTER	GPTPROTO	YOU SAVE / MO
Personal 100M tokens / mo (4.8M cached)	\$53.76	\$56.72	\$30.72	-\$23.04 = \$276.48 / yr
Team 100M tokens / mo (48M cached)	\$537.60	\$567.17	\$307.20	-\$230.40 = \$2764.80 / yr
Business 500M tokens / mo (240M cached)	\$2688.00	\$2835.84	\$1536.00	-\$1152.00 = \$13824.00 / yr

Monthly cost by source

OpenRouter
\$567.17

Google
\$537.60

GPTProto
\$307.20

Save \$230.40 / month
Annual savings = \$2764.80 - pay as you go

Top up \$500 →

GPTProto Gemini 4 Argon Joins Platform's API Model Catalog

October 16, 2026

Hong Kong - October 16, 2026 -

GPTProto has expanded its public model catalog with Google's Gemini 4 Argon, adding an API option for text reasoning, software engineering, and multi-step agent workflows. The GPTProto Gemini 4 landing page brings together model specifications, integration considerations, and task-specific routes for developers assessing the addition within existing applications.

The catalog lists access through a GPTProto API key and a shared platform balance, placing Argon alongside other supported models. The addition gives teams already using GPTProto another model to evaluate within their account, while the accompanying page explains technical differences that matter when moving from earlier model integrations.

The main landing page focuses on text input and text output rather than presenting every supported modality as the same request. It directs image, video, file, and audio analysis to dedicated task pages, helping distinguish the model's broader input capabilities from the particular integration covered by the parent page.

A central specification is the shared context window, which GPTProto describes as 1,048,576 tokens across input, thinking, and generated output. This means a large prompt does not leave an additional, equally sized allowance for the response: source material, reasoning, and generation draw on one combined budget.

That distinction affects how development teams plan repository analysis, document synthesis, and other tasks with extensive supporting material. An application may need to reserve space for both reasoning and the final response rather than filling the available context with files before generation begins.

The GPTProto Gemini 4 Argon page also describes low, medium, and high thinking levels, with medium identified as the default. According to the listing, increasing the reasoning level can involve additional latency and token consumption, making the setting a workload-specific choice rather than a universal recommendation for every request.

The landing page separates ordinary response limits from long decode, describing continuation and background execution for unusually large outputs. It also cautions that Google-native parameters should not be assumed to map directly into GPTProto request bodies, so the applicable platform documentation remains the reference for implementation.

For software teams, the listed applications include examining architecture, dependencies, tests, and behavior across files before proposing a migration plan or patch. A practical evaluation can compare whether generated changes satisfy repository tests and review criteria, not simply whether a response contains a plausible explanation or a large amount of code.

The page additionally describes agent-oriented capabilities such as function calling and structured outputs, alongside other tools associated with the model. These capabilities provide components for workflows that plan, act, and check results, but application developers still determine tool permissions, state management, validation, and recovery from failed steps.

For document-heavy analysis, GPTProto identifies financial, legal, compliance, and operational research as possible applications. The relevant integration question is whether the system can produce reviewable evidence or structured findings from the supplied material; professional conclusions still require source verification and appropriate human review.

GPTProto's landing page presents comparative benchmark results with an explicit distinction between provider-reported evaluations and GPTProto testing. Rather than declaring a single model superior for every task, it notes that different models lead different evaluations, supporting a comparison based on representative prompts, tools, timeouts, and acceptance criteria.

The release of the catalog entry adds a model-specific reference for teams planning that comparison. Context allocation, thinking settings, input routing, and expected output length can be reviewed together before an application sends production traffic, giving the listing a practical role beyond identifying the model by name.

The GPTProto Gemini 4 Argon page and public catalog now provide the starting points for reviewing current access and implementation details. GPTProto's documentation describes unified access to multiple AI providers, while the new entry sets out the particular constraints and workflows associated with Argon's text API use.

For more information, visit <https://gptproto.com/model/google/gemini-4-argon>

About GPTProto:

GPTProto is a unified AI API platform that brings together access to text, image and video models from multiple providers. Its website provides model information, integration guidance and account access for developers evaluating AI applications.

###

For more information about GPTProto, contact the company here: GPTProto Tabe Tsang +44 7724693132 gptproto.official@gmail.com Unit 1022a, Beverley Commercial Centre, 87-105 Chatham Road South, Tsim Sha Tsui, Hong Kong

GPTProto

GPTProto is an AI model API platform giving developers access to 200+ text, image, video, and audio models including Claude, GPT, Gemini, Kimi, GLM, Seedance, Kling, and Seedream through OpenAI-compatible and other documented API formats.

Website: <https://gptproto.com/>

Email: gptproto.official@gmail.com

Phone: +44 7724693132

