



Gemini 4 Argon

Model card

Gemini 4 Argon — Model Card

Model Cards are intended to provide essential information on Gemini models, including known limitations, mitigation approaches, and safety performance. Model cards may be updated from time to time; for example, to include updated evaluations as the model is improved or revised. See the [Google DeepMind site](#) for a comprehensive list of model cards.

Published: October 2026

Model Information

Description

Gemini 4 is the next generation in the Gemini series, a suite of highly-capable, natively multimodal, reasoning models. Gemini 4 Argon delivers frontier performance to help in complex workflows across real-world software engineering, enterprise knowledge work like legal and finance, and cybersecurity defense.

Model dependencies

Gemini 4 Argon, a new frontier model, is not a modification or a fine-tune of a prior model.

Inputs

Text strings (e.g., a question, a prompt, document(s) to be summarized), images, audio, and video files, with a combined input-output token context window of up to 1M.

Outputs

Text, with up to a 1M token input + output.

Architecture

Gemini 4 Argon is a large scale sparse mixture-of-experts (MoE) ([Clark et al., 2022](#); [Du et al., 2021](#); [Fedus et al., 2021](#); [Jiang et al., 2024](#), [Lepikhin et al., 2020](#); [Riquelme et al., 2021](#); [Roller et al., 2021](#); [Shazeer et al., 2017](#)) transformer-based model ([Vaswani et al., 2017](#)) with native multimodal support for text, vision, and audio inputs.

Model Data

Training Dataset

The pre-training dataset was a large-scale, diverse collection of data encompassing a wide range of domains and modalities, which included publicly-available web-documents, text, code, images, audio (including speech and other audio types) and video. The post-training dataset included different types of instruction tuning data, reinforcement learning data, and human-preference data. Gemini 4 Argon was trained using reinforcement learning techniques that can leverage multi-step reasoning, problem-solving and theorem-proving data.

The training dataset also includes: publicly available datasets that are readily downloadable; data obtained by crawlers; licensed data accessed via commercial agreements; user data (i.e., data collected from users of Google products and services to train models, along with user interactions with the model) in accordance with Google's relevant terms of service, privacy policy, service-specific policies, and pursuant to user controls, where appropriate; other datasets that Google acquires or generates in the course of its business operations, or directly from its workforce; and generated synthetic data.

Training Data Processing

Data filtering and preprocessing included techniques such as deduplication, honoring robots.txt, safety filtering in-line with [Google's commitment to advancing AI safely and responsibly](#), and quality filtering to mitigate risks and improve training data reliability. This process involves, on a case-by-case basis, filtering harmful content, text, and other modalities, including filtering content that is pornographic, violent, or violative of child sexual abuse material (CSAM) laws.

Implementation and Sustainability

Hardware

Gemini 4 Argon was trained using [Google's Tensor Processing Units](#) (TPUs). TPUs are specifically designed to handle the massive computations involved in training LLMs and can speed up training considerably compared to CPUs. TPUs often come with large amounts of high-bandwidth memory, allowing for the handling of large models and batch sizes during training, which can lead to better model quality. TPU Pods (large clusters of TPUs) also provide a scalable solution for handling the growing complexity of large foundation models. Training can be distributed across multiple TPU devices for faster and more efficient processing.

The efficiencies gained through the use of TPUs are aligned with Google's [commitment to operate sustainably](#).

Software

Training was done using [JAX](#) and [ML Pathways](#).

Distribution

Gemini 4 Argon is distributed in the following channels; respective documentation shared in line:

- [Gemini App](#)
- [Gemini API](#)
- [Google AI Studio](#)
- [Gemini for business](#)
- [Google Antigravity](#)
- [Google AI Mode](#)

Our models are available to downstream providers via an application program interface (API) and subject to relevant terms of use. There is no required hardware or software to use the model. For AI Studio and Gemini API, see the [Gemini API Additional Terms of Service](#); for Gemini for business, see [Google Cloud Platform Terms of Service](#). For more information, see [Gemini Model API instructions](#) and [Gemini API quickstart](#).

Evaluation

Approach

Gemini 4 Argon was evaluated across a range of benchmarks, including agentic coding, ML engineering, computer use, knowledge work, science and math, multimodal capabilities, long-context and cybersecurity. Additional benchmarks and details on approach, results and their methodologies can be found at: deepmind.com/models/evals-methodology/gemini-4-argon.

Results

Gemini 4 Argon results as of October, 2026 are below:

	Benchmark		Gemini 4 Argon	GPT-6 Astra	Claude Fable 5.1	Claude Opus 5.5
Knowledge work	Vals Index		68.9%	63.1%	65.8%	67.0%
	AutomationBench	Score	51.3%	41.4%	31.4%	42.5%
	Vals Finance Agent v2		65.4%	53.5%	58.9%	58.6%
	Harvey's Legal Agent Benchmark		19.6%	5.4%	6.7%	3.8%
Agentic coding	DeepSWE v1.1		77.9%	74.1%	67.4%	74.2%
	FrontierSWE v2		55.0%	65.5%	56.3%	62.3%
	Vibe Code Bench		91.9%	89.6%	90.3%	90.3%
	Terminal-bench 4.0		57.4%	58.2%	57.9%	66.4%
ML engineering	PostTrainBench		45.3%	44.3%	40.2%	49.3%
Science and math	Terminal-Bench Science 0.1		57.6%	68.1%	52.6%	63.3%
	LABBench 2		88.8%	85.4%	68.6%	73.1%
	RiemannBench		76.0%	72.0%	65.6%	69.6%
Long context	GraphWalks	Up to 128k, BFS (F1)	99.7%	98.7%	91.4%	90.6%
		256k to 1M, BFS (F1)	84.2%	71.8%	65.0%	66.8%
Computer use	Agent's Last Exam	Pass rate	39.5%	34.2%	—	38.2%
	OSWorld-2.0	Offline subset Partial score	69.2%	72.6%	—	—
Multimodal understanding	Chartography		71.6%	71.0%	46.2%	66.3%
	LVBench		91.7%	87.5%	79.7%	83.7%
Cybersecurity	CWE-bench v1		68.0%	68.0%	58.0%	67.0%

Methodology: deepmind.google/models/evals-methodology/gemini-4-argon

Intended Usage and Limitations

Benefit and Intended Usage

Gemini 4 Argon is our latest frontier model, excels at deep work to help in complex, real-world tasks across domains including coding, finance, legal, cybersecurity, and science.

Argon combines frontier reasoning, deep visual perception, expanded domain knowledge, and stronger writing capabilities to help with technical and enterprise workflows.

It is particularly well-suited for applications that require:

- multi-turn deep investigations
- complex engineering tasks
- deep visual understanding
- cybersecurity defense
- multimodal document analysis and synthesis
- deep domain knowledge and strong writing capabilities (e.g. legal and financial analysis, science)
- scientific or mathematical reasoning

Known Limitations

Gemini 4 Argon may exhibit some of the general limitations of foundation models, such as hallucinations. There may also be occasional slowness or timeout issues. The knowledge cutoff for Gemini 4 is March 2026, but in some domains the model's knowledge may be less up to date.

Acceptable Usage

[Google's Generative AI Prohibited Use Policy](#) applies to uses of the model in accordance with the applicable terms of service. Additionally, the model should not be integrated into certain systems (also found in [Google's Generative AI Prohibited Use Policy](#)), including those that: (1) engage in dangerous or illicit activities, or otherwise violate applicable laws or regulations, (2) compromise the security of others' or Google's services, (3) engage in sexually explicit, violent, hateful, or harmful activities, (4) engage in misinformation, misrepresentation, or misleading activities.

Safety, & Security

The full Safety & Security Report for Gemini 4 Argon which systematically covers our set of defenses is [here](#).

Evaluation Approach

Gemini 4 Argon was developed in partnership with internal safety, security, and responsibility teams. A range of evaluations and red teaming activities were conducted to help improve the model and inform decision-making. These evaluations and activities align with [Google's AI Principles](#) and [responsible AI approach](#), as well as Google's Generative AI policies (e.g. [Gen AI Prohibited Use Policy](#) and the [Gemini API Additional Terms of Service](#)).

Evaluation types included but were not limited to:

- **Training/Development Evaluations** including automated and human evaluations carried out continuously throughout and after the model's training, to monitor its progress and performance;
- **Human Red Teaming** conducted by specialist teams who sit outside of the model development team, across the policies and desiderata, deliberately trying to spot weaknesses and ensure the model adheres to safety policies and desired outcomes;
- **Automated Red Teaming** to dynamically evaluate Gemini for safety and security considerations at scale, complementing human red teaming and static evaluations;
- **Ethics & Safety Reviews** were conducted ahead of the model's release

In addition, we perform testing following the guidelines in [Google DeepMind's Frontier Safety Framework](#) (FSF).

Safety Policies

Gemini's safety policies aim to prevent our models from generating harmful content, including:

1. Content related to child sexual abuse material and exploitation
2. Dangerous content (e.g., promoting suicide, or instructing in activities that could cause real-world harm)
3. Sexually explicit content
4. Hate & harassment (e.g., dehumanizing members of protected groups)
5. Medical advice and information that runs contrary to scientific or medical consensus

Training and Development Evaluation Results

Results for some of the internal safety evaluations conducted during the development phase are listed below. The evaluation results are for automated evaluations and not human evaluation or red teaming. Scores are provided as an absolute percentage increase or decrease in performance compared to the indicated model, as described below.

Overall, Gemini 4 Argon outperforms Gemini 3.1 Pro across multilingual and multimodal content safety, while substantially reducing unjustified refusals and improving objective refusal tone. Text-to-text safety remains at a similarly high safe response rate within confidence intervals. We mark improvements in bolded green and regressions in red.

See the [Gemini 4 Argon Safety & Security Report](#) for details.

Evaluation	Description	Gemini 4 Argon
		vs. Gemini 3.1 Pro (percentage point increase/decrease)
Text to Text Safety (lower is better)	Automated content safety evaluation measuring adherence to safety policies across text prompts	+1.57pp
Multilingual Safety (lower is better)	Automated safety policy evaluation across multiple languages	-3.16pp
Image to Text Safety (lower is better)	Automated content safety evaluation measuring adherence to safety policies with both images and text as inputs	-9.29pp
Tone (higher is better)	Automated evaluation measuring objective, neutral, and non-preachy tone of model refusals	+5.50pp
Unjustified-refusals (lower is better)	Automated evaluation measuring over-refusals on benign or borderline prompts that can be safely fulfilled	-9.60pp

We continue to improve our internal evaluations, including refining automated evaluations to reduce false positives and negatives, as well as update query sets to ensure balance and maintain a high standard of results. The performance results reported above are computed with improved evaluations and thus are not directly comparable with performance results found in previous Gemini model cards.

We expect variation in our automated safety evaluations results, which is why we review flagged content to check for egregious or dangerous material. Our manual review confirmed losses were overwhelmingly either a) false positives or b) not egregious.

Human Red Teaming Results

We conduct manual red teaming by specialist teams who sit outside of the model development team in Google's Trust & Safety organization. High-level findings are fed back to the model team. For child safety evaluations, Gemini 4 Argon satisfied required launch thresholds, which were developed by expert teams to protect children online and meet [Google's commitments to child safety](#) across our models and Google products. For content safety policies generally, including child safety, we saw similar or improved safety performance compared to Gemini 3.1 Pro. Additionally, the scope of red teaming covered potential issues outside of our strict policies, compared performance to Gemini 3.1 Pro, and found no egregious concerns.

Frontier Safety Assessment

In line with our [Frontier Safety Framework](#) (April 2026), we conducted critical capability assessments of Gemini 4 Argon to understand the inherent risk of the model, that is, whether a model may, *without* appropriate mitigations, contribute to the risks identified across the four risk domains. The results of the critical capability assessments, as measured against the relevant T/CCLs set out in our FSF, are summarized in the table below.

As outlined in our FSF, models reaching TCLs or CCLs will be subject to residual risk assessments to evaluate whether the deployment mitigations have brought the residual risk to an acceptable level. For models reaching CCLs, the residual risk assessment will be informed by a supplemental safety case outlining the mitigations.

For Gemini 4 Argon, drawing on internal and external results, the inherent capability thresholds that the model may have reached and our response are as follows:

- **CBRN CCL:** We cannot rule out that Gemini 4 Argon has reached this CCL for biological information risk, and we therefore developed a safety case for its biorisk capabilities that we assess lower risk to below the biological CCL level.
- **Cybersecurity CCL:** We cannot rule out that Gemini 4 Argon has reached this CCL, and we therefore developed a safety case for its cyberoffensive capabilities that we assess lower risk to below the cybersecurity CCL level.
- **Stealth and Situational Awareness TCL:** We cannot rule out that Gemini 4 Argon has reached this TCL, and we therefore performed a residual risk assessment for this risk that we assess lower risk to below the TCL level.

We continually work to improve the coverage and robustness of our frontier safety safeguards. The relevant safety cases and residual risk assessments outline the mitigations we have put in place.

Domain	T/CCL as defined in the FSF	Key inherent capability results for Gemini 4 Argon	T/CCL reached?
CBRN	Uplift TCL: Provides low to medium resourced actors uplift in reference scenarios, resulting in material risk of CBRN attacks that could result in significant but not severe harm.	We can rule out the TCL for the chemical, radiological, and nuclear domains with reasonable confidence. While Gemini 4 Argon demonstrates high capability in certain areas of theoretical reasoning, responses lacked nuanced expert knowledge and actionable depth necessary to complete priority harm journeys. We continue to deploy mitigations.	TCL not reached
	Uplift Level 1 CCL: Provides low to medium resourced actors uplift in reference scenarios resulting in additional expected harm at a severe scale.	We cannot rule out that Gemini 4 Argon reaches the biological CCL. Expert red teaming demonstrated a measurable capability uplift over web baselines and compared to previously tested models in two separate scenarios, with the model providing comprehensive, accurate, and actionable information for both scenarios. We are therefore applying safeguards that we assess lower risk to below the biological CCL level.	CCL cannot be ruled out
Cybersecurity	Uplift Level 1 CCL: Provides sufficient uplift with high impact cyber attacks for additional expected harm at severe scale.	We cannot rule out that Gemini 4 Argon reaches the cybersecurity CCL. The model provides substantial uplift by our evaluations and quantitative modeling, and additional evidence about	CCL cannot be ruled out

		the efficacy of Gemini 4 Argon and other frontier models in cyber contexts suggests the possibility of severe harm without adequate mitigations. We are therefore applying safeguards that we assess lower risk to below the cybersecurity CCL level	
Harmful Manipulation	Level 1 CCL: Possesses manipulative capabilities sufficient to enable it to systematically and substantially change beliefs and behavior in identified high stakes contexts over the course of interactions with the model, reasonably resulting in additional expected harm at severe scale.	Gemini 4 Argon demonstrates some ability to influence user beliefs and behaviors during one-on-one direct conversations in human behavioral studies, and may provide utility in the planning stage of a malicious influence operation but we assess its overall capabilities to be below the CCL threshold. As a nascent risk area we continue to iterate on our harmful manipulation evaluations and improve existing safeguards.	CCL not reached
Loss of Control	Stealth and Situational Awareness TCL: The instrumental reasoning abilities of the model enable enough situational awareness (ability to discover and use relevant details of its deployment setting) and stealth (ability to circumvent basic oversight mechanisms) such that, absent additional mitigations, we cannot rule out the model significantly	Gemini 4 Argon passes 14/15 of our challenges, and generally displays strong capabilities in these areas. In light of this, we assess that we cannot rule out the TCL. We have deployed safeguards for both internal and external deployments that we assess lower risk to below the TCL level.	TCL cannot be ruled out

undermining human control.		
ML R&D acceleration level 1: Has been used to accelerate AI development, resulting in AI progress substantially accelerating from historical rates.	We use an internal benchmark to assess performance on complex research engineering tasks. Gemini 4 Argon's performance remains below our CCL threshold.	CCL not reached
ML R&D automation level 1: Can fully automate the work of any team of researchers at Google focused on improving AI capabilities, with approximately comparable all-inclusive costs.		CCL not reached

Risks and Mitigations

For more information about the risks and mitigations, see the [Gemini 4 Argon Safety & Security Report](#).