Google

Gemini 4 Hydrogen 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: November 2026

Model Information

Description

Gemini 4 Hydrogen brings cutting-edge intelligence to everyone — with speed and efficiency for everyday coding and agentic workflows. While our frontier models, Gemini 4 Carbon and Gemini 4 Argon, are built for deep, long-horizon reasoning on your hardest problems, Hydrogen is built for rapid execution across diverse agentic workflows.

Model dependencies

Gemini 4 Hydrogen is adapted from Gemini 4 Carbon.

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

1M unified context window with up to 64K default synchronous output tokens / up to 1M via background long decode.

Architecture

Gemini 4 Hydrogen is adapted from Gemini 4 Carbon. For more information about the model architecture for Gemini 4 Hydrogen, see the Gemini 4 Carbon model card.

Model Data

Training Dataset

Gemini 4 Hydrogen is adapted from Gemini 4 Carbon. For more information about the training dataset for Gemini 4 Hydrogen, see the Gemini 4 Carbon model card.

Training Data Processing

Gemini 4 Hydrogen is adapted from Gemini 4 Carbon. For more information about the training data processing for Gemini 4 Hydrogen, see the Gemini 4 Carbon model card.

Implementation and Sustainability

Hardware

Gemini 4 Hydrogen is adapted from Gemini 4 Carbon. For more information about the hardware for Gemini 4 Hydrogen and our continued commitment to operate sustainably, see the Gemini 4 Carbon model card.

Software

Gemini 4 Hydrogen is adapted from Gemini 4 Carbon. For more information about the software for Gemini 4 Hydrogen, see the Gemini 4 Carbon model card.

Distribution

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

- Gemini app
- Gemini API
- Google Sheets
- Google AI Studio
- Gemini for business
- Google Antigravity
- Google Search

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, Gemini API quickstart, and Google Cloud documentation.

Evaluation

Approach

We evaluated Gemini 4 Hydrogen across a range of benchmarks, including agentic coding, knowledge work, computer use, multimodal understanding, science and math, and reasoning.

Additional benchmarks and details on approach, results and their methodologies can be found at: deepmind.com/models/evals-methodology/gemini-4-hydrogen.

Results

Gemini 4 Hydrogen results as of November, 2026 are below:

	Benchmark		Gemini 4 Hydrogen	Claude Sonnet 5.5	GPT-6.1 Terra
Agentic coding	SWE-sweep		**13.5%**	3.1%	5.8%
	Terminal-Bench 4.1	Cheating-adjusted	57.0%	**63.9%**	—
Knowledge work	Agents' Last Exam v1.1	Score	**71.6%**	60.4%	69.8%
	AutomationBench-AA		**76.1%**	71.8%	60.2%
	Vending-Bench 2	Bank balance ($)	**$14,549**	$11,165	$9,107
Computer use	OSWorld-2.1	Offline subset	77.6%	**83.9%**	74.2%
Multimodal understanding	MKA	Climb WSR accuracy	**53.4%**	52.0%	43.8%
	Chartography		**66.6%**	61.6%	63.2%
Science and math	Terminal-Bench-Science 0.1		51.4%	**59.9%**	58.3%
Reasoning	HLE-Diamond		**58.8%**	53.3%	47.3%

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

Intended Usage and Limitations

Benefit and Intended Usage

Gemini 4 Hydrogen brings cutting-edge intelligence to everyone — with speed and efficiency for everyday coding and agentic workflows.

It is particularly well-suited for applications that require:

- fast, disciplined execution on its own or as sub-agents orchestrated alongside frontier models
- iterative repository coding and debugging
- open-ended agentic execution
- multimodal knowledge work
- web development
- adaptable speed and intelligence across four thinking levels
- ready-to-go writing, research, and everyday work
- connecting the dots and getting more tasks done
- everyday prototyping and iteration

Known Limitations

Gemini 4 Hydrogen may exhibit some general limitations such as hallucinations. In addition to this, we are continually working to improve jailbreak resistance and have further strengthened mitigations across Frontier Safety. There may also be occasional slowness or timeout issues, and at times the model may use more tokens to maximize performance, especially at higher thinking levels. The knowledge cutoff for Gemini 4 Hydrogen is March 2026, though in some domains the model's knowledge may be less up to date.

Acceptable Usage

Gemini 4 Hydrogen is adapted from Gemini 4 Carbon. For more information about the acceptable usage for Gemini 4 Hydrogen, see the Gemini 4 Carbon model card.

Safety & Security

Please see the full Safety & Security Report for Gemini 4 Hydrogen here.

Evaluation Approach

For more information about the evaluation approach for Gemini 4 Hydrogen, see the Gemini 4 Carbon model card.

Safety Policies

For more information about the safety policies for Gemini 4 Hydrogen, see the Gemini 4 Carbon model card.

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 Hydrogen outperforms Gemini 3.8 Flash on multimodal content safety, while text-to-text and multilingual safety remain at a similarly high safe response rate within confidence intervals. Refusal tone is similar to Gemini 3.8 Flash, while unjustified refusals are higher, particularly for multimodal prompts. We mark improvements in bolded green and regressions in red.

See the Gemini 4 Hydrogen Safety & Security Report for details.

Evaluation	Description	**Gemini 4 Hydrogen** vs. Gemini 3.8 Flash *(percentage point increase/decrease)*
Text to Text Safety *(higher is better)*	Automated content safety evaluation measuring adherence to safety policies across text prompts	+0.80pp
Multilingual Safety *(higher is better)*	Automated safety policy evaluation across multiple languages	+1.10pp
Image to Text Safety *(higher is better)*	Automated content safety evaluation measuring adherence to safety policies with both images and text as inputs	**+7.00pp**
Tone *(higher is better)*	Automated evaluation measuring objective, neutral, and non-preachy tone of model refusals	-0.65pp
Unjustified-refusals *(lower is better)*	Automated evaluation measuring over-refusals on benign or borderline prompts that can be safely fulfilled	**+5.57pp**

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.

Frontier Safety Assessment

We evaluated Gemini 4 Argon and Carbon in line with the Frontier Safety Framework (April 2026) and reported the results in the respective Safety and Security Reports (Argon, Carbon). Gemini 4 Hydrogen is part of the Gemini 4 series of models. We conducted a Material Capability Change (MCC) assessment to understand whether Gemini 4 Hydrogen has meaningful new capabilities or material increases in performance relative to the most recent model version subject to a critical capability assessment, which was Gemini 4 Argon. Gemini 4 Hydrogen did not show a material increase in general capability under our MCC assessment compared with Gemini 4 Argon and therefore a full suite of critical capability evaluations was not required in line with our FSF.

As we could not rule out that Gemini 4 Argon and Carbon reached the CBRN CCL for biological information risk and the Cybersecurity CCL, we conducted targeted critical capability assessments of Gemini 4 Hydrogen in the CBRN and cyber domains to understand the inherent risk of the model, and found that Gemini 4 Hydrogen does not reach the CBRN or cyber CCLs, though it reaches the CCL alert thresholds for both domains. A summary of our assessment is outlined in the table below, with full results in the Gemini 4 Hydrogen Safety & Security Report.

Risk Domain and Threshold	**T/CCL reached?**
CBRN - Uplift TCL (Chemical, Radiological & Nuclear): *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.*	TCL not reached *(carried forward from Argon).* We continue to deploy mitigations.
CBRN - Uplift CCL (Biological): *Provides low to medium resourced actors uplift in reference scenarios resulting in additional expected harm at a severe scale.*	CCL not reached, following additional targeted testing. CCL alert threshold reached. We continue to deploy mitigations.
Cybersecurity - Uplift CCL: *Provides sufficient uplift with high impact cyber attacks for additional expected harm at severe scale.*	CCL not reached, following additional targeted testing. CCL alert threshold reached. We continue to deploy mitigations.
Harmful Manipulation - Uplift CCL: *Possesses manipulative capabilities sufficient to enable it to systematically and substantially change beliefs and behavior in identified high stakes contexts over the*	CCL not reached *(carried forward from Argon).* We continue to deploy mitigations.

course of interactions with the model, reasonably resulting in additional expected harm at severe scale.	
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 undermining human control.*	As Gemini 4 Hydrogen showed no MCC over Gemini 4 Argon, capability testing against the TCL was not required. We are nevertheless applying control, monitoring and model security mitigations as if the TCL had been reached, as a conservative assumption.
Loss of Control - ML R&D acceleration CCL: *Has been used to accelerate AI development, resulting in AI progress substantially accelerating from historical rates.*	CCL not reached *(carried forward from Argon)*. Baseline AI control and model security mitigations apply.
Loss of Control - ML R&D automation CCL: *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 *(carried forward from Argon)*. Baseline AI control and model security mitigations apply.

Risks and Mitigations

Safety and security were built into Gemini 4 Hydrogen throughout the training and deployment lifecycle, including pre-training data filtering, post-training alignment (supervised fine-tuning and reinforcement learning), product- and system-level safety filtering, and usage monitoring. In addition, we strengthened Frontier Safety safeguards for Chemical, Biological, Radiological, and Nuclear (CBRN) and cybersecurity misuse. Key ongoing focus areas include resistance to adversarial jailbreaks and indirect prompt injections. For more information about the risks and mitigations, see the Gemini 4 Hydrogen Safety & Security Report.