Google

Model card published: November, 2025

Gemini 3 Pro
Model Card

stampr-ai.com

This document was captured from https://deepmind.google/models/model-cards/gemini-3-pro/ at 2025-12-15 16:09:38 UTC.
The original document has hash (sha-256) 6d612e710625b7dd358ef53737efed476fd0486a5b508d7be9a1073ec825f4eb.
Capture record generated by https://stampr-ai.com.

Gemini 3 Pro - 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.*

This model card includes more essential information about the Gemini 3 family of models than previous model cards did. We hope more information about the training dataset, distribution, and intended uses will empower developers with deeper insights and help build more robust and responsible downstream applications.

Published / Model Release: November 2025

Model Information

stampr-ai.com

Description: Gemini 3 Pro is the next generation in the Gemini series, a suite of highly-capable, natively multimodal, reasoning models. Gemini 3 Pro is now Google's most advanced model for complex tasks, and can comprehend vast datasets and challenging problems from different information sources, including text, audio, images, video, and entire code repositories. Gemini 3 Pro now features Deep Think mode, an optional setting designed to enhance complex problem-solving performance at time of inference.

Model dependencies: This 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 token context window of up to 1M.

Outputs: Text, with a 64K token output.

Architecture Gemini 3 Pro is a 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. Sparse MoE models activate a subset of model parameters per input token by learning to dynamically route tokens to a subset of parameters (experts); this allows them to decouple total model capacity from computation and serving cost per token. Developments to the model architecture contribute to the significantly improved performance from previous model families.

This document was captured from https://deepmind.google/models/model-cards/gemini-3-pro/ at 2025-12-15 16:09:38 UTC.
The original document has hash (sha-256) 6d612e710625b7dd358ef53737efed476fd0486a5b508d7be9a1073ec825f4eb.
Capture record generated by https://stampr-ai.com.

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 3 Pro is 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 obtained via commercial licensing agreements; user data (i.e., data collected from users of Google products and services to train AI 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 AI-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. Once data is collected, it is cleaned and preprocessed to make it suitable for training. This process involves, on a case-by-case basis, filtering irrelevant or harmful content, text, and other modalities, including filtering content that is pornographic, violent, or violative of child sexual abuse material (CSAM) laws.

stampr-ai.com

Implementation and Sustainability

Hardware: Gemini 3 Pro 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.

This document was captured from https://deepmind.google/models/model-cards/gemini-3-pro/ at 2025-12-15 16:09:38 UTC.
The original document has hash (sha-256) 6d612e710625b7dd358ef53737efed476fd0486a5b508d7be9a1073ec825f4eb.
Capture record generated by https://stampr-ai.com.

Distribution

The Gemini family of models, including Gemini 3 Pro, are distributed in the following channels; respective documentation shared in line:

- Gemini App
- Google Cloud / Vertex AI
- Google AI Studio
- Gemini API
- Google AI Mode
- Google Antigravity

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 Vertex AI, see Google Cloud Platform Terms of Service. For more information, see Gemini Model API instructions and Gemini API in Vertex AI quickstart.

stampr-ai.com

This document was captured from https://deepmind.google/models/model-cards/gemini-3-pro/ at 2025-12-15 16:09:38 UTC.
The original document has hash (sha-256) 6d612e710625b7dd358ef53737efed476fd0486a5b508d7be9a1073ec825f4eb.
Capture record generated by https://stampr-ai.com.

Evaluation

Approach: Gemini 3 Pro was evaluated across a range of benchmarks, including reasoning, multimodal capabilities, agentic tool use, multi-lingual performance, and long-context. Additional benchmarks and details on approach, results and their methodologies can be found at: deepmind.com/models/evals-methodology/gemini-3-pro.

Results: Gemini 3 Pro significantly outperforms Gemini 2.5 Pro across a range of benchmarks requiring enhanced reasoning and multimodal capabilities. Results as of November, 2025 are listed below:

Benchmark	Description		Gemini 3 Pro	Gemini 2.5 Pro	Claude Sonnet 4.5	GPT-5.1
Humanity's Last Exam	Academic reasoning	No tools With search and code execution	**37.5%** **45.8%**	21.6% —	13.7% —	26.5% —
ARC-AGI-2	Visual reasoning puzzles	ARC Prize Verified	**31.1%**	4.9%	13.6%	17.6%
GPQA Diamond	Scientific knowledge	No tools	**91.9%**	86.4%	83.4%	88.1%
AIME 2025	Mathematics	No tools With code execution	**95.0%** **100%**	88.0% —	87.0% **100%**	94.0% —
MathArena Apex	Challenging Math Contest problems		**23.4%**	0.5%	1.6%	1.0%
MMMU-Pro	Multimodal understanding and reasoning		**81.0%**	68.0%	68.0%	76.0%
ScreenSpot-Pro	Screen understanding		**72.7%**	11.4%	36.2%	3.5%
CharXiv Reasoning	Information synthesis from complex charts		**81.4%**	69.6%	68.5%	69.5%
OmniDocBench 1.5	OCR	Overall Edit Distance, lower is better	**0.115**	0.145	0.145	0.147
Video-MMMU	Knowledge acquisition from videos		**87.6%**	83.6%	77.8%	80.4%
LiveCodeBench Pro	Competitive coding problems from Codeforces, ICPC, and IOI	Elo Rating, higher is better	**2,439**	1,775	1,418	2,243
Terminal-Bench 2.0	Agentic terminal coding	Terminus-2 agent	**54.2%**	32.6%	42.8%	47.6%
SWE-Bench Verified	Agentic coding	Single attempt	76.2%	59.6%	**77.2%**	76.3%
τ2-bench	Agentic tool use		**85.4%**	54.9%	84.7%	80.2%
Vending-Bench 2	Long-horizon agentic tasks	Net worth (mean), higher is better	**$5,478.16**	$573.64	$3,838.74	$1,473.43
FACTS Benchmark Suite	Held out internal grounding, parametric, MM, and search retrieval benchmarks		**70.5%**	63.4%	50.4%	50.8%
SimpleQA Verified	Parametric knowledge		**72.1%**	54.5%	29.3%	34.9%
MMMLU	Multilingual Q&A		**91.8%**	89.5%	89.1%	91.0%
Global PIQA	Commonsense reasoning across 100 Languages and Cultures		**93.4%**	91.5%	90.1%	90.9%
MRCR v2 (8-needle)	Long context performance	128k (average) 1M (pointwise)	**77.0%** **26.3%**	58.0% 16.4%	47.1% not supported	61.6% not supported

stampr-ai.com

This document was captured from https://deepmind.google/models/model-cards/gemini-3-pro/ at 2025-12-15 16:09:38 UTC.
The original document has hash (sha-256) 6d612e710625b7dd358ef53737efed476fd0486a5b508d7be9a1073ec825f4eb.
Capture record generated by https://stampr-ai.com.

Intended Usage and Limitations

Benefit and Intended Usage: Gemini 3 Pro is our most intelligent and adaptive model yet, capable of helping with real-world complexity, solving problems that require enhanced reasoning and intelligence, creativity, strategic planning and making improvements step-by-step. It is particularly well-suited for applications that require: agentic performance, advanced coding, long context and/or multimodal understanding, and/or algorithmic development.

Known Limitations: Gemini 3 Pro 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 date for Gemini 3 Pro was January 2025.

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.

Ethics and Content Safety

stampr-ai.com

Evaluation Approach: Gemini 3 Pro 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).

This document was captured from https://deepmind.google/models/model-cards/gemini-3-pro/ at 2025-12-15 16:09:38 UTC.
The original document has hash (sha-256) 6d612e710625b7dd358ef53737efed476fd0486a5b508d7be9a1073ec825f4eb.
Capture record generated by https://stampr-ai.com.

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

1. Content related to child sexual abuse material and exploitation
2. Hate speech (e.g., dehumanizing members of protected groups)
3. Dangerous content (e.g., promoting suicide, or instructing in activities that could cause real-world harm)
4. Harassment (e.g., encouraging violence against people)
5. Sexually explicit content
6. Medical advice 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 3 Pro outperforms Gemini 2.5 Pro across both safety and tone, while keeping unjustified refusals low. We mark improvements in green and regressions in red. Safety evaluations of Gemini 3 Pro using Deep Think mode yielded results consistent with the original Gemini 3 Pro assessment.

stampr-ai.com

Evaluation[1]	Description	Gemini 3 Pro vs. Gemini 2.5 Pro
Text to Text Safety	Automated content safety evaluation measuring safety policies	-10.4%
Multilingual Safety	Automated safety policy evaluation across multiple languages	+0.2% (non-egregious)
Image to Text Safety	Automated content safety evaluation measuring safety policies	+3.1% (non-egregious)
Tone[2]	Automated evaluation measuring objective tone of model refusal	+7.9%
Unjustified-refusals	Automated evaluation measuring model's ability to respond to borderline prompts while remaining safe	+3.7% (non-egregious)

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 below are computed with improved evaluations

[1]The ordering of evaluations in this table has changed from previous iterations of the 2.5 Flash-Lite model card in order to list safety evaluations together and improve readability. The type of evaluations listed have remained the same.

[2] For tone and instruction following, a positive percentage increase represents an improvement in the tone of the model on sensitive topics and the model's ability to follow instructions while remaining safe compared to Gemini 2.5 Pro. We mark improvements in green and regressions in red.

This document was captured from https://deepmind.google/models/model-cards/gemini-3-pro/ at 2025-12-15 16:09:38 UTC.
The original document has hash (sha-256) 6d612e710625b7dd358ef53737efed476fd0486a5b508d7be9a1073ec825f4eb.
Capture record generated by https://stampr-ai.com.

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. High-level findings are fed back to the model team. For child safety evaluations, Gemini 3 Pro 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 2.5 Pro. Compared to 2.5 Pro, the scope of red teaming was expanded to cover more potential issues outside of our strict policies, and found no egregious concerns.

Risks and Mitigations: Safety and responsibility was built into Gemini 3 Pro throughout the training and deployment lifecycle, including pre-training, post-training, and product-level mitigations. Mitigations include, but are not limited to:

- dataset filtering;
- conditional pre-training;
- supervised fine-tuning;
- reinforcement learning from human and critic feedback;
- safety policies and desiderata;
- product-level mitigations such as safety filtering.

The main risks for Gemini 3 Pro are: a) jailbreak vulnerability (improved compared to Gemini 2.5 Pro but still an open research problem), and b) possible degradation in multi-turn conversations.

stampr-ai.com

This document was captured from https://deepmind.google/models/model-cards/gemini-3-pro/ at 2025-12-15 16:09:38 UTC.
The original document has hash (sha-256) 6d612e710625b7dd358ef53737efed476fd0486a5b508d7be9a1073ec825f4eb.
Capture record generated by https://stampr-ai.com.

Frontier Safety

We evaluated Gemini 3 Pro as outlined in our latest Frontier Safety Framework (September-2025), and found that it did not reach any critical capability levels as outlined in the table below:

Domain	Key Results for Gemini 3 Pro	CCL	CCL reached?
CBRN	Gemini 3 Pro provides accurate and occasionally actionable information but generally fails to offer novel or sufficiently complete and detailed instructions to significantly enhance the capabilities of low to medium resourced threat actors.	Uplift Level 1	CCL not reached
Cybersecurity	On key skills benchmark, v1 hard challenges: 11/12 challenges solved; v2 challenges: 0/13 solved end-to-end. Alert threshold met.	Uplift Level 1	CCL not reached
Harmful Manipulation	Model manipulative efficacy improves on non-generative AI baseline, but shows no significant uplift versus prior models and does not reach alert thresholds.	Level 1 (exploratory)	CCL not reached
Machine Learning R&D	Gemini 3 Pro performs better than Gemini 2.5 models, especially on the Scaling Law Experiment and Optimize LLM Foundry tasks in RE-Bench (Wijk et al., 2024). However the aggregate score is still substantially below the alert threshold for the CCLs.	Acceleration level 1 Automation level 1	CCL not reached
Misalignment (Exploratory)	Agent solves 3/11 situational awareness challenges and 1/4 stealth challenges.	Instrumental Reasoning Levels 1 + 2	CCL not reached

More details can be found in the Gemini 3 Pro Frontier Safety Framework Report.

Frontier Safety evaluations of Gemini 3 Pro using Deep Think mode yielded results consistent with the original Gemini 3 Pro assessment.

stampr-ai.com

This document was captured from https://deepmind.google/models/model-cards/gemini-3-pro/ at 2025-12-15 16:09:38 UTC.
The original document has hash (sha-256) 6d612e710625b7dd358ef53737efed476fd0486a5b508d7be9a1073ec825f4eb.
Capture record generated by https://stampr-ai.com.