

	<i>Gemini 1.5 Flash</i>	<i>Gemini 1.5 Pro</i>	Gemini 2.0 Flash-Lite	Gemini 2.0 Flash	Gemini 2.5 Flash	Gemini 2.5 Pro
Input modalities	Text, Image, Video, Audio	Text, Image, Video, Audio	Text, Image, Video, Audio	Text, Image, Video, Audio	Text, Image, Video, Audio	Text, Image, Video, Audio
Input length	1M	2M	1M	1M	1M	1M
Output modalities	Text	Text	Text	Text, Image*	Text, Audio*	Text, Audio*
Output length	8K	8K	8K	8K	64K	64K
Thinking	No	No	No	Yes*	Dynamic	Dynamic
Supports tool use?	No	No	No	Yes	Yes	Yes
Knowledge cutoff	November 2023	November 2023	June 2024	June 2024	January 2025	January 2025

Table 1 | Comparison of Gemini 2.X model family with Gemini 1.5 Pro and Flash. Tool use refers to the ability of the model to recognize and execute function calls (e.g., to perform web search, complete a math problem, execute code). **currently limited to Experimental or Preview, see Section 2.7. Information accurate as of publication date.*

helpfulness and general tone compared to their 2.0 and 1.5 counterparts. In practice, this means that the 2.5 models are substantially better at providing safe responses without interfering with important use cases or lecturing end users. We also evaluated Gemini 2.5 Pro’s Critical Capabilities, including CBRN, cybersecurity, machine learning R&D, and deceptive alignment. While Gemini 2.5 Pro showed a significant increase in some capabilities compared to previous Gemini models, it did not reach any of the Critical Capability Levels in any area.

Our report is structured as follows: we begin by briefly describing advances we have made in model architecture, training and serving since the release of the Gemini 1.5 model. We then showcase the performance of the Gemini 2.5 models, including qualitative demonstrations of its abilities. We conclude by discussing the safety evaluations and implications of this model series.

2. Model Architecture, Training and Dataset

2.1. Model Architecture

The Gemini 2.5 models are 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) transformers (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 of Gemini 2.5 compared to Gemini 1.5 Pro (see Section 3). Despite their overwhelming success, large transformers and sparse MoE models are known to suffer from training instabilities (Chowdhery et al., 2022; Dehghani et al., 2023; Fedus et al., 2021; Lepikhin et al., 2020; Liu et al., 2020; Molybog et al., 2023; Wortsman et al., 2023; Zhai et al., 2023; Zhang et al., 2022). The Gemini 2.5 model series makes considerable progress in enhancing large-scale training stability, signal propagation and optimization dynamics, resulting in a considerable boost in performance straight out of pre-training compared to previous Gemini models.