

Model versions and lifecycle

This document defines key terms related to the lifecycle stages and important dates for Gemini and embedding models that are available on Google Cloud Vertex AI. It also gives you the recommended upgrades for the models and points you to available migration paths.

Key Terms

Stable model: A publicly released version of the model that is available and supported for production use starting on the release date. A stable model version is typically released with a *retirement date*, which indicates the last day that the model is available. After this date, the model is no longer accessible or supported by Google.

- **Latest stable models:** The latest version within the model family recommended for new and active projects and should be the target for migrations from earlier versions. See [Latest stable models \(#latest-stable\)](#).
- **Retired models:** The model version is past its retirement date and has been permanently deactivated. Retired models are no longer accessible or supported by Google. API requests referencing a retired model ID typically returns a 404 error. See [Retired models \(#retired-models\)](#).
- **Recommended upgrades:** The latest stable model that we recommend switching to. Latest stable models tend to offer better performance and more capabilities as compared to legacy stable models. See the recommended upgrades [Retired models \(#retired-models\)](#) sections.

Latest stable models

The following table lists the latest stable models:

Model ID	Release date	Retirement date
gemini-live-2.5-flash-native-audio	December 12, 2025	December 13, 2026
gemini-2.5-pro	June 17, 2025	June 17, 2026
gemini-2.5-flash	June 17, 2025	June 17, 2026

gemini-2.5-flash-lite	July 22, 2025	July 22, 2026
gemini-2.5-flash-image	October 2, 2025	No retirement date announced
gemini-2.0-flash-001	February 5, 2025	March 3, 2026
gemini-2.0-flash-lite-001	February 25, 2025	March 3, 2026
gemini-embedding-001	May 20, 2025	No retirement date announced
text-embedding-005	November 18, 2024	No retirement date announced
text-embedding-004	May 14, 2024	No retirement date announced
text-multilingual-embedding-002	May 14, 2024	No retirement date announced
multimodalembedding@001	February 12, 2024	No retirement date announced

Migrate to a latest stable model

To learn how to migrate to a latest stable model, see [Migrate to the latest Gemini models](#) (/vertex-ai/generative-ai/docs/migrate). This guide gives you a set of migration steps that aims to minimize some potential risks involved in model migration and helps you use new models in an optimal way.

However, if you don't have time to follow the guide and just need to quickly resolve the errors caused by models reaching their retirement dates, do the following:

1. Update your application to point to the recommended upgrades.
2. Test all mission critical features to make sure everything works as expected.
3. Deploy the updates like you normally would.

Gemini auto-updated aliases

The auto-updated alias of a Gemini model always points to the latest stable model. When a new latest stable model is available, the auto-updated alias automatically points to the new version.

The following table shows the auto-updated aliases for Gemini models and the latest stable models that they point to.

Auto-updated alias	Stable version reference
gemini-2.5-pro	gemini-2.5-pro
gemini-2.5-flash	gemini-2.5-flash
gemini-2.5-flash-lite	gemini-2.5-flash-lite
gemini-2.0-flash-lite	gemini-2.0-flash-lite-001
gemini-2.0-flash	gemini-2.0-flash-001

Retired models

+ The following table lists the retired models (click to expand)

Model ID	Release date	Retirement date	Recommended upgrade
gemini-1.5-pro-001	May 24, 2024	May 24, 2025	gemini-2.0-flash
gemini-1.5-pro-002	September 24, 2024	September 24, 2025	gemini-2.0-flash
gemini-1.5-flash-001	May 24, 2024	May 24, 2025	gemini-2.0-flash-lite
gemini-1.5-flash-002	September 24, 2024	September 24, 2025	gemini-2.0-flash-lite
textembedding-gecko@003*	December 12, 2023	May 24, 2025	gemini-embedding-001
textembedding-gecko-multilingual@001	November 2, 2023	May 24, 2025	gemini-embedding-001
gemini-1.0-pro-001	February 15, 2024	April 21, 2025	gemini-2.0-flash
gemini-1.0-pro-002	April 9, 2024	April 21, 2025	gemini-2.0-flash
gemini-1.0-pro-vision-001	February 15, 2024	April 21, 2025	gemini-2.0-flash
text-bison	May 2023	April 21, 2025	gemini-2.0-flash-lite
chat-bison	May 2023	April 21, 2025	gemini-2.0-flash-lite

code-gecko	May 2023	April 21, 2025	gemini-2.0-flash-lite
textembedding-gecko@002	November 2, 2023	April 21, 2025	gemini-embedding-001
textembedding-gecko@001	June 7, 2023	April 21, 2025	gemini-embedding-001
imagetext	June 7, 2023	September 24, 2025	gemini-2.5-flash-image
imagegeneration@002	August 17, 2023	September 24, 2025	imagen-3.0-capability-001 or imagen-3.0-generate-002
imagegeneration@005	November 22, 2023	September 24, 2025	imagen-3.0-capability-001 or imagen-3.0-generate-002
imagegeneration@006	March 27, 2024	September 24, 2025	imagen-3.0-capability-001 or imagen-3.0-generate-002

What's next

- To learn which regions models are available, see [Deployments and endpoints](#) (/vertex-ai/generative-ai/docs/learn/locations).
- For details about individual models, see [Google models](#) (/vertex-ai/generative-ai/docs/models) and click the model that you want to learn more about.

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](#) (<https://creativecommons.org/licenses/by/4.0/>), and code samples are licensed under the [Apache 2.0 License](#) (<https://www.apache.org/licenses/LICENSE-2.0>). For details, see the [Google Developers Site Policies](#) (<https://developers.google.com/site-policies>). Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2026-01-07 UTC.