MAI-Image-2.5 Model Card

Date: June 2, 2026

Model Summary

Developer	Microsoft Corporation Authorized Representative: Microsoft Ireland Operations Limited (MIOL) 70 Sir John Rogerson's Quay, Dublin 2, D02 R296, Ireland
Description	MAI-Image-2.5 is a diffusion-based generative model designed for both high-quality text-to-image synthesis and precise, controllable image-to-image editing. It progressively transforms random noise into a coherent image aligned with a given text prompt, leveraging a flow-matching loss to learn a continuous transformation between the noise distribution and the data distribution. MAI-Image-2.5 excels at precise, surgical edits with consistency — enabling users and developers to make targeted object edits, adapt layouts, update text, clean up artifacts like motion blur, and preserve visual consistency across iterations.
Model architecture	Diffusion-based generative architecture for text-to-image synthesis and image-to-image editing
Parameters	2×10^{10} (20B) non-embedding parameters
Inputs	Text input; Image input (for editing workflows)
Context length	32K tokens
Outputs	Image output. Maximum total pixel count: 1,048,576 (equivalent to 1024×1024). Either dimension can exceed 1024 as long as the total stays within the limit.
Public data summary	Public data summary available here
Training dates	April 17, 2026 – Ongoing
Release date	June 2, 2026
Release date in the EU	June 2, 2026
License	Various product and service terms where the model is deployed, such as those for Azure AI Foundry and MAI Playground.

Model dependencies	N/A
Additional related assets	N/A
Acceptable use policy	As described in the License section above. Subsequent integrations of MAI-Image-2.5 may be subject to different terms of service and policies.

Model Overview

MAI-Image-2.5 is a diffusion-based generative model designed for both text-to-image synthesis and controllable image-to-image editing. It operates by progressively transforming random noise into a coherent image that aligns with a given text prompt. This approach leverages a flow-matching loss to learn a continuous transformation between the noise distribution and the data distribution, ensuring stable and efficient training.

MAI-Image-2.5 was rebuilt from the ground up for multimodal editing workflows, with stronger visual context understanding and coherence. It reasons across objects, scene structure, lighting, scale, and spatial positioning to produce consistent edits — even from ambiguous prompts. The model gracefully handles multiple constraints at once, including layout preservation, object changes, text updates, and contextual adaptation.

It supports reliable object removal, replacement, attribute changes, inpainting, and image enhancement without destabilizing composition or layout. The model also maintains strong visual consistency across iterative edits.

This combination of flow-matching objectives and diffusion inference enables the model to produce high-quality, diverse images that maintain strong alignment with the input text, making it suitable for creative generation, design tasks, production editing workflows, and multimodal applications.

Alignment Approach

Our alignment objective for MAI-Image-2.5 is to reduce the generation of harmful or inappropriate (e.g., violent, gory, sexual) images, even when requested by a user. We took a defense-in-depth approach, applying mitigations to the data during model development, and deploying the model with additional safety mitigations. The initial release of MAI-Image-2.5 is through integrations with Microsoft products and services, such as Azure AI Foundry and MAI Playground.

Usage

Primary Use Cases

MAI-Image-2.5 is a general-purpose text-to-image and image-to-image generative model, intended for creative generation, design tasks, and production editing workflows. The model is particularly capable in the following areas:

- **Text-to-image generation:** Generates high-quality images from natural language prompts, enabling users to translate textual descriptions into visually coherent outputs suitable for a wide range of creative and design use cases.
- **Image editing:** Supports precise, surgical edits to existing visual assets, including object removal, replacement, attribute changes, inpainting, text updates, and artifact cleanup (e.g., motion blur), while preserving composition and layout.
- **Photorealistic & cinematic quality:** Produces highly realistic imagery with strong lighting, reflections, natural environments, and fine visual detail.
- **High-fidelity portraits:** Generates expressive, natural-looking portraits with accurate facial structure, lighting, and texture.
- **Product, branding & commercial design:** Well suited for product imagery, marketing visuals, brand assets, and commercial creative workflows.
- **Text rendering:** Improved rendering of text within generated images, including labels, posters, packaging, and signage.
- **Creative range:** Generates visually diverse outputs across styles and compositions, reducing repetitive or templated results.
- **Professional-grade outputs:** Trained with carefully curated datasets and evaluated against real creative use cases to support production-quality design workflows.

Distribution Channels

MAI-Image-2.5 is initially available through integrations with Microsoft products and services:

- **Azure AI Foundry** — API access for developers (private preview, expanding to public preview)
- **MAI Playground** — Publicly available site for users to interact with MAI models.
- **Microsoft PowerPoint** — Users can generate presentation-ready visuals and slides from prompts
- **Microsoft OneDrive Photos** — Creative and editing workflows for personal photos

Any future release formats will be accompanied by an update to relevant documentation.

Out-of-Scope Use Cases

MAI-Image-2.5 should not be used to:

- Generate content intended to deceive, mislead, or impersonate real individuals
- Produce content that violates applicable laws or Microsoft's terms of service
- Create harmful, abusive, or policy-violating content

Responsible AI Considerations

Despite technical mitigations such as data filtering and content classifiers applied at the system level, image generation models are assumed to be able to produce harmful or unexpected content based on user requests. Common risk areas associated with image generation models include violent or gory content, sexual content or nudity, depictions of public figures, and replication of trademarked or other protected material.

MAI-Image-2.5 includes layered safety guardrails, with prompt and output filtering to help detect and block harmful, abusive, or policy-violating content. These mitigations are designed to support safer use across consumer and developer workflows.

Data Overview

Training, Testing, and Validation Datasets

The training dataset consists of paired images and text descriptions, where each caption provides a detailed account of the visual content. The data spans a broad, general-purpose domain including everyday objects, natural scenes, people, and abstract concepts, making it suitable for open-domain text-to-image generation and image editing.

The dataset combines visual data (images) with natural language text, enabling multimodal learning. These characteristics — broad domain coverage and multimodal pairing — are directly aligned with the model's purpose: to generate high-quality, semantically aligned images from text prompts and to perform controllable edits across a wide range of scenarios.

To read more about the data used to train MAI-Image-2.5, please see the public data summary.

Quality and Performance Evaluation

MAI-Image-2.5 achieves a top-3 ranking on both the Arena text-to-image and image-editing Arena leaderboards, on par with GPT-Image-1.5 and Nano Banana Pro 2K.

The model was evaluated by human raters alongside comparable models and across a range of capability areas. Specifically, raters were tasked with selecting a preferred model output, in different topic areas based on real user intents (for example, "product/branding," "cartoon," "photorealistic") and by reference to the output's alignment with the prompt intent as well as the output's visual appeal. This resulted in an Elo score calculation.

Text-to-Image Evaluation

Category	MAI-Image-2.5	MAI-Image-2	MAI-Image-1
Photorealistic & Cinematic Imagery	1247	1201 ± 12	1104 ± 5
Product, Branding, Commercial Design	1263	1191 ± 11	1085 ± 5
3D Imaging & Modeling	1254	1184 ± 22	1096 ± 8
Cartoon, Anime & Fantasy	1268	1186 ± 14	1100 ± 5
Art	1256	1191 ± 18	1104 ± 7
Portraits	1261	1201 ± 17	1095 ± 6

Text Rendering	1278	1186 ± 12	1069 ± 5
Overall	1254	1190 ± 8	1093 ± 4

Safety Evaluation and Red-Teaming

Image generation models are known to produce potentially harmful or unexpected content based on user requests, with common risk areas including violent or gory content, sexual content or nudity. In addition to technical work on the model such as data filtering, we evaluated the model focusing on safeguards in place in the product deployments.

Our Red Teams conducted multiple rounds of adversarial testing of MAI-Image-2.5 to emulate real-world adversaries across a spectrum of skill levels, ranging from straightforward prompting to advanced attack methodologies. AIRT developed attack strategies spanning multiple harm categories and engaged subject matter experts to evaluate the model against these risk areas.

Evaluation was carried out in two phases: pre-mitigation and post-mitigation. The assessment followed a break-fix cycle in which the Red Team identified vulnerabilities and shared them with the model development team for remediation. The updated model was then re-evaluated in subsequent rounds to assess the effectiveness of mitigations and to identify any remaining weaknesses.

This model card will be updated as training completes and evaluation data becomes available.