TURING

Turing AGI Advancement

Advance Multimodal Intelligence With Real-World Audio, Image, and GUI Data

Train models that see, hear, reason, and act—accurately and in context.

Multimodal models must understand real-world inputs—from structured visuals to accented, noisy speech. High-quality, human-authored data helps models improve grounding, reasoning, and performance across vision-language and audio benchmarks.

What we do

Turing AGI Advancement helps multimodal models make sense of structured, noisy, and multilingual inputs—across vision, speech, and UI.

Multimodal Data Packs (SFT, RL, Evals)

- Curated datasets across real-world formats: images, audio, documents, and UIs, tailored for **enterprise-grade multimodal models**
- Covers single or multi-source **financial charts, STEM diagrams, medical scans, PPT slides, PDF documents, GUI screenshots**, and more
- Supports **captioning, CoT, reasoning-based explanation, editing, rewriting, semantic extraction, and vision-language grounding**
- Created by vetted experts and multilingual trainers in **60+ languages with 1,000s of active contributors**

Audio Intelligence & Speech Modeling

- **Scalable audio workflows** for both speech understanding and generation
- Pre-Training + SFT + RLHF data for **ASR, TTS, voice cloning, speech translation, diarization, and full duplex audio to audio**, including metadata and additional emotion/style/pace/phonetic annotations based on your requirements
- Diarized **multi-speaker data, clean/noisy channels, and edge-case instruction following**
- Delivered in WAV/MP3/FLAC across **60+ locales with phoneme, emotion, and pace tagging**

Evaluation & Feedback for VLMs

- Real-world benchmarks to **test vision-language reasoning in high-stakes use cases**
- **700+ open-ended tasks grounded in business and STEM workflows** with <10% HARD set accuracy across top models
- Pinpoint failures in reasoning across **image charts, financial tables, scan captioning, and GUI text extraction**
- **Scored reward labels and CoT evaluations** for model alignment and fine-tuning

Training Data for Computer Use Agents (CUA)

- Multimodal data pipelines for **browser agents, desktop UIs, and SaaS workflows**
- Includes **screen perception, tool interaction, and long-horizon planning**
- Covers **click/scroll/keyboard event prediction, adversarial injections, and preference detection**
- Includes cloned environments for Salesforce, SAP, Jira, and more—supporting **secure training without ToS violations**

Is your multimodal model ready for real-world complexity?

Run an evaluation or request data packs at https://go.turing.com/llm-multimodality-hub