Multimodal Foundation Models in the VLAM Program

Tanguy Coenen
Scientific Director at imec

Let's talk about…

- The VLAM program
- Multimodal AI in media
  - Use case: agentic AI in media
  - Use case: virtual production
- Challenges

The VLAM program

The Digital Sovereignty Imperative

Need for Attuned AI

Critical need for culturally and linguistically attuned AI in a global landscape dominated by large international models.

Risk of Bias

Risk of cultural erosion and linguistic bias if we rely solely on dominant, non-local models.

VLAM

Develop robust, culturally attuned foundation models to serve as the backbone for the Flemish ecosystem.

The Foundation: Rich Media Archives

Data is the Core

Unique value from rich digital archives (text, audio, video) of Flemish media houses.

This data enriches open-source international models.

Goal: Create AI that reflects the region’s linguistic and cultural identity.

Strategy: "Adapt, Don't Build" => work on post-training and inference

Building from Scratch

Extremely high cost (billions)

Requires massive compute resources

Slow time-to-market

Difficult to keep pace with global leaders

The VLAM Way: Adapt

Cost-efficient and fast

Uses existing State-of-the-Art (SOTA) models as a base

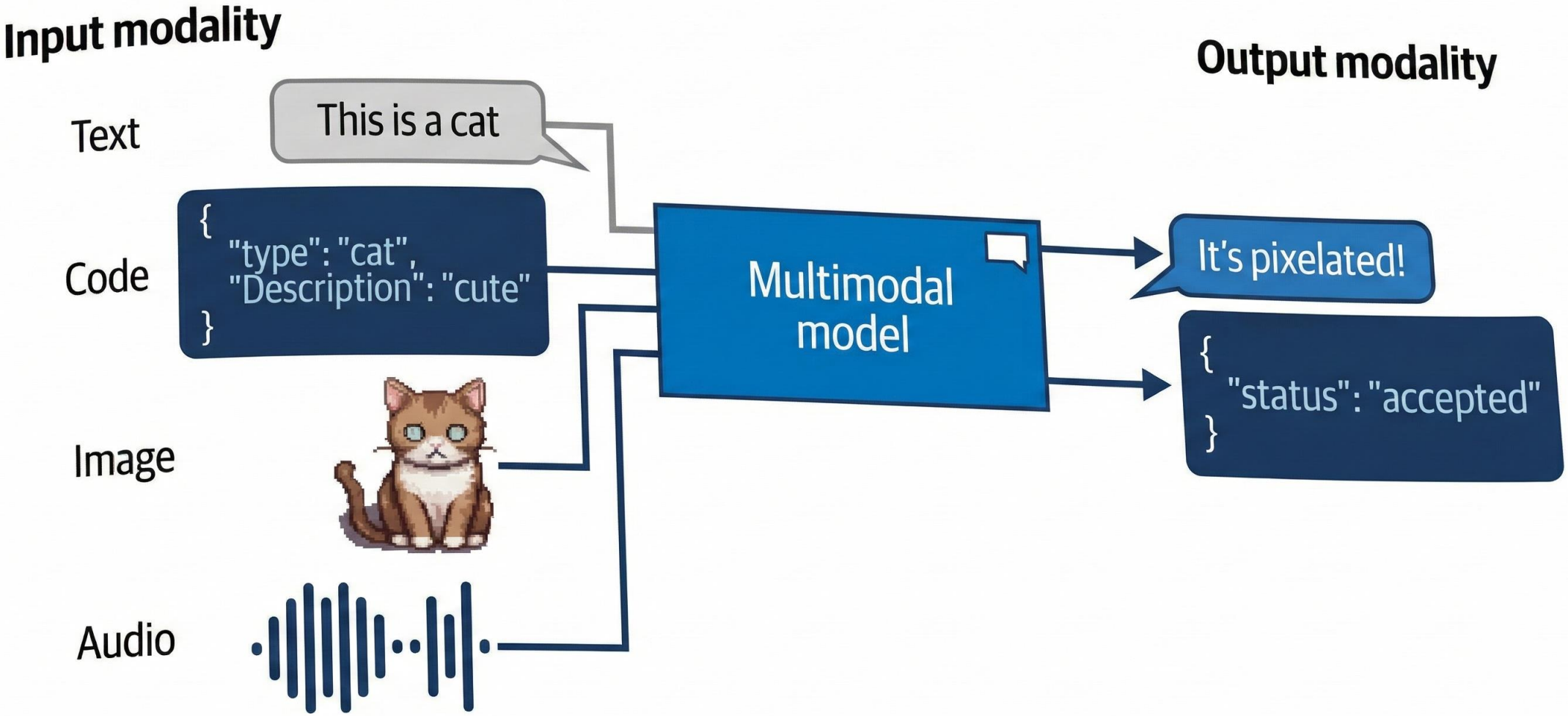
Focuses resources on fine-tuning with unique Flemish data

Ensures technological relevance

Core Component: The Foundation Base Model

- **Base Models:** We start with powerful, open-source international foundation models as our technological core.
- **Flemish Data:** We enrich this base with our unique, high-quality multimodal media archives.
- **Process:** A continuous process of fine-tuning, retrieval-augmentation, and grounding ensures deep cultural relevance and accuracy.

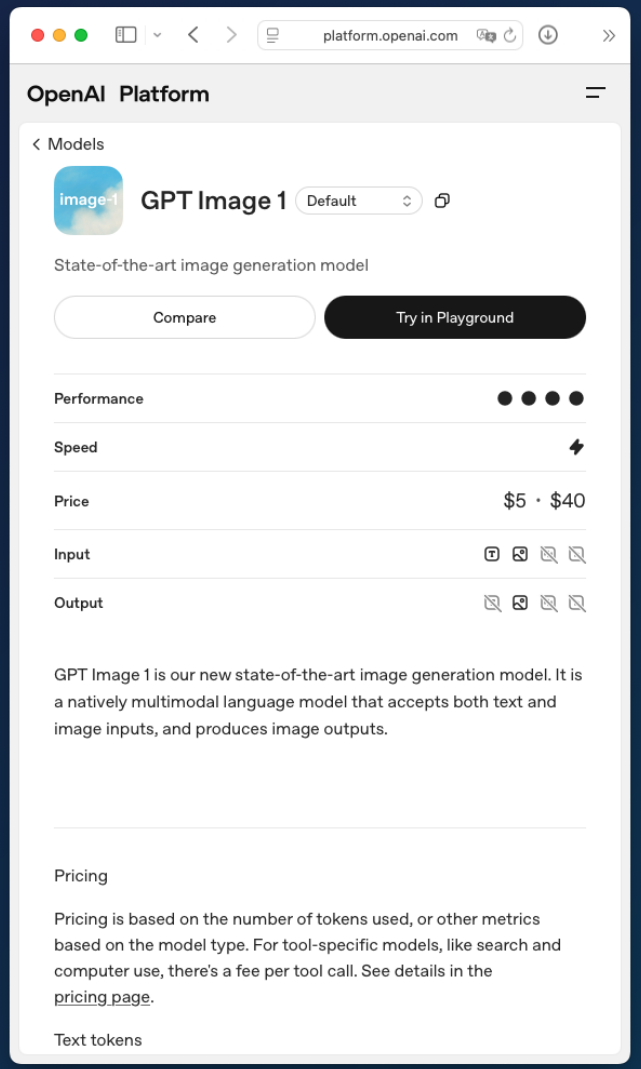
Multimodal AI in media

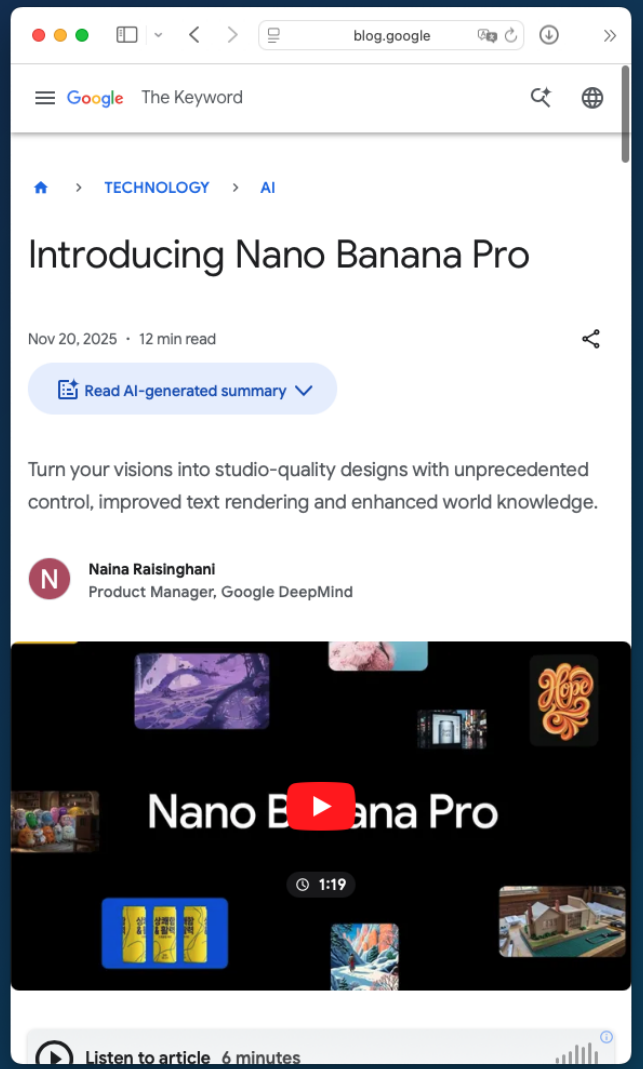


Source: https://www.oreilly.com/library/view/hands-on-large-language/9781098150952/

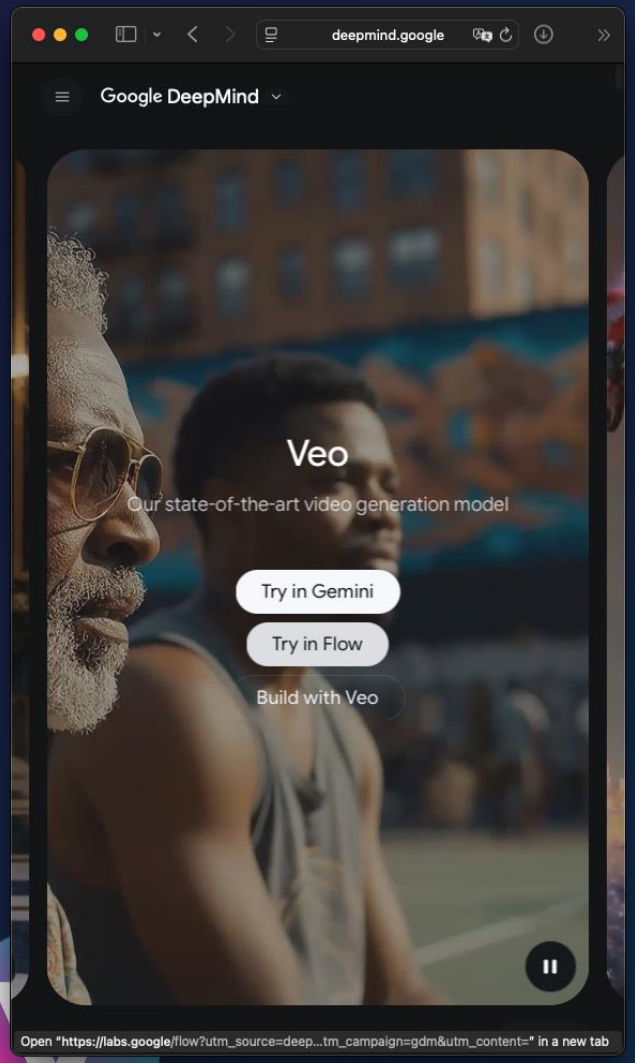
Image

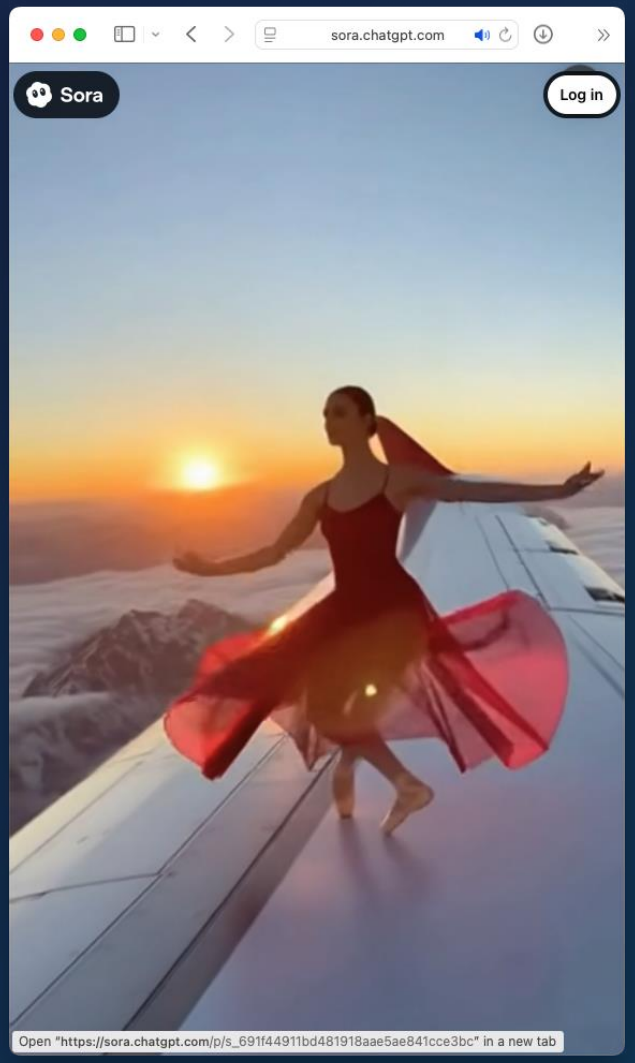

bfl.ai
Black Forest Labs - Frontier AI Lab
FLUX Playground - Black Forest Labs
Black Forest Labs
FLUX.1 Kontext
Learn more
Show more details

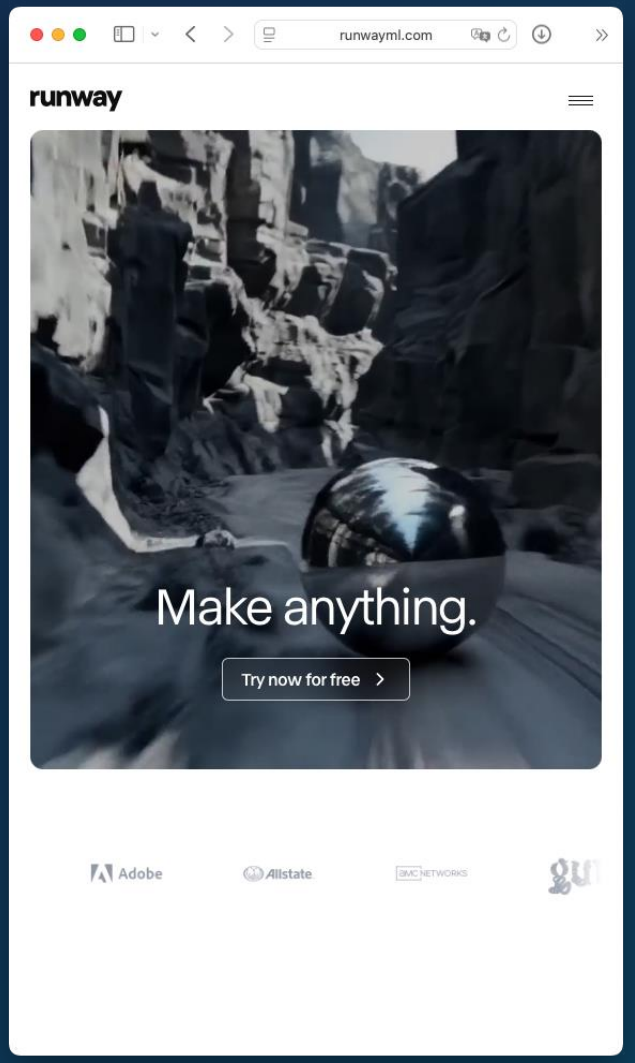

platform.openai.com
OpenAI Platform
Models
GPT Image 1
Default
State-of-the-art image generation model
Compare
Try in Playground
Performance
Speed
Price
$5 · $40
Input
Output
GPT Image 1 is our new state-of-the-art image generation model. It is a natively multimodal language model that accepts both text and image inputs, and produces image outputs.
Pricing
Pricing is based on the number of tokens used, or other metrics based on the model type. For tool-specific models, like search and computer use, there's a fee per tool call. See details in the pricing page.
Text tokens


blog.google
Google The Keyword
TECHNOLOGY
AI
Introducing Nano Banana Pro
Nov 20, 2025 · 12 min read
Read AI-generated summary
Turn your visions into studio-quality designs with unprecedented control, improved text rendering and enhanced world knowledge.
Naina Raisinghani
Product Manager, Google DeepMind
Nano Banana Pro
1:19
Listen to article 6 minutes

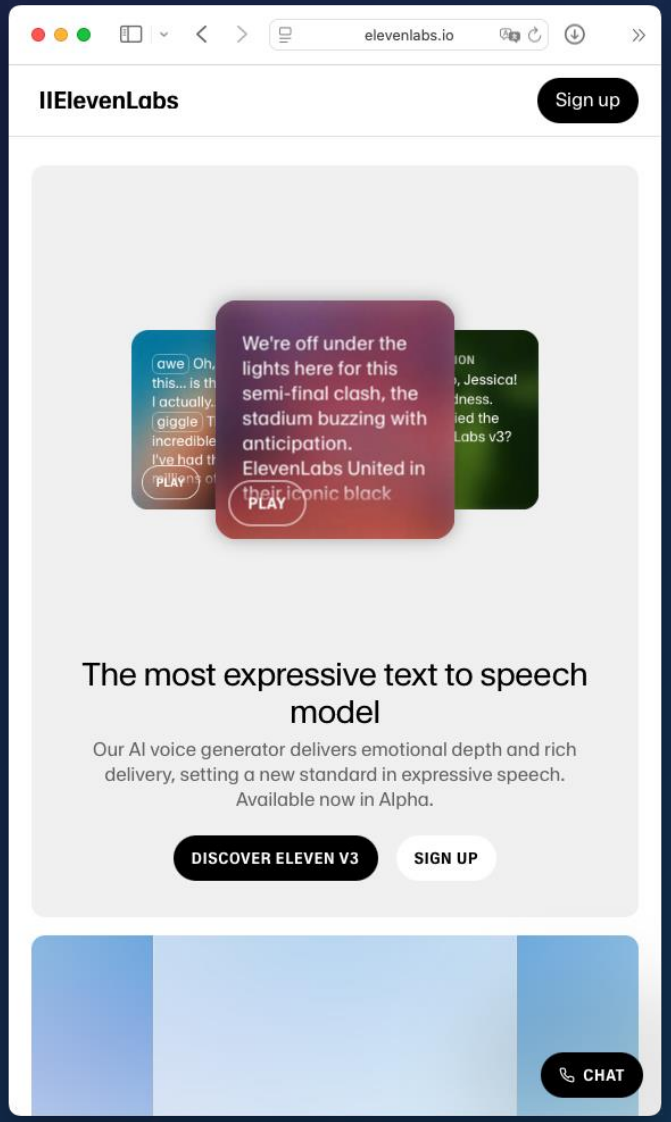
Video

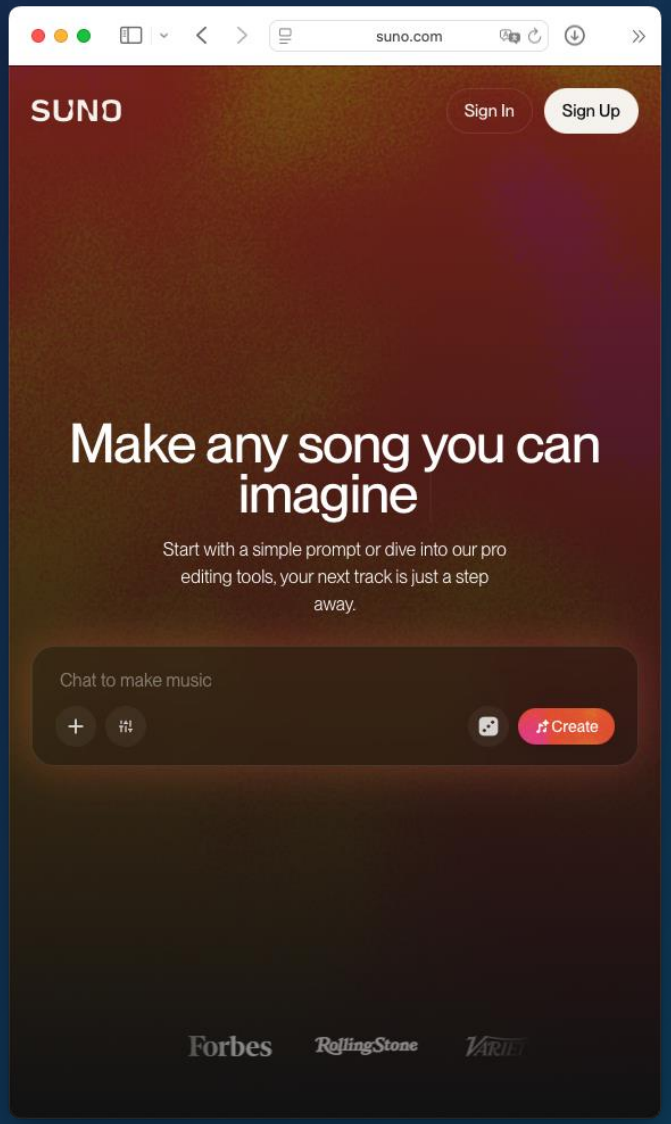

deepmind.google
Google DeepMind
Veo
Our state-of-the-art video generation model
Try in Gemini
Try in Flow
Build with Veo
Open "https://labs.google/flow?utm_source=deep...tm_campaign=gdm&utm_content=" in a new tab


sora.chatgpt.com
Sora
Log in
Open "https://sora.chatgpt.com/p/s_691f44911bd481918aae5ae841cce3bc" in a new tab

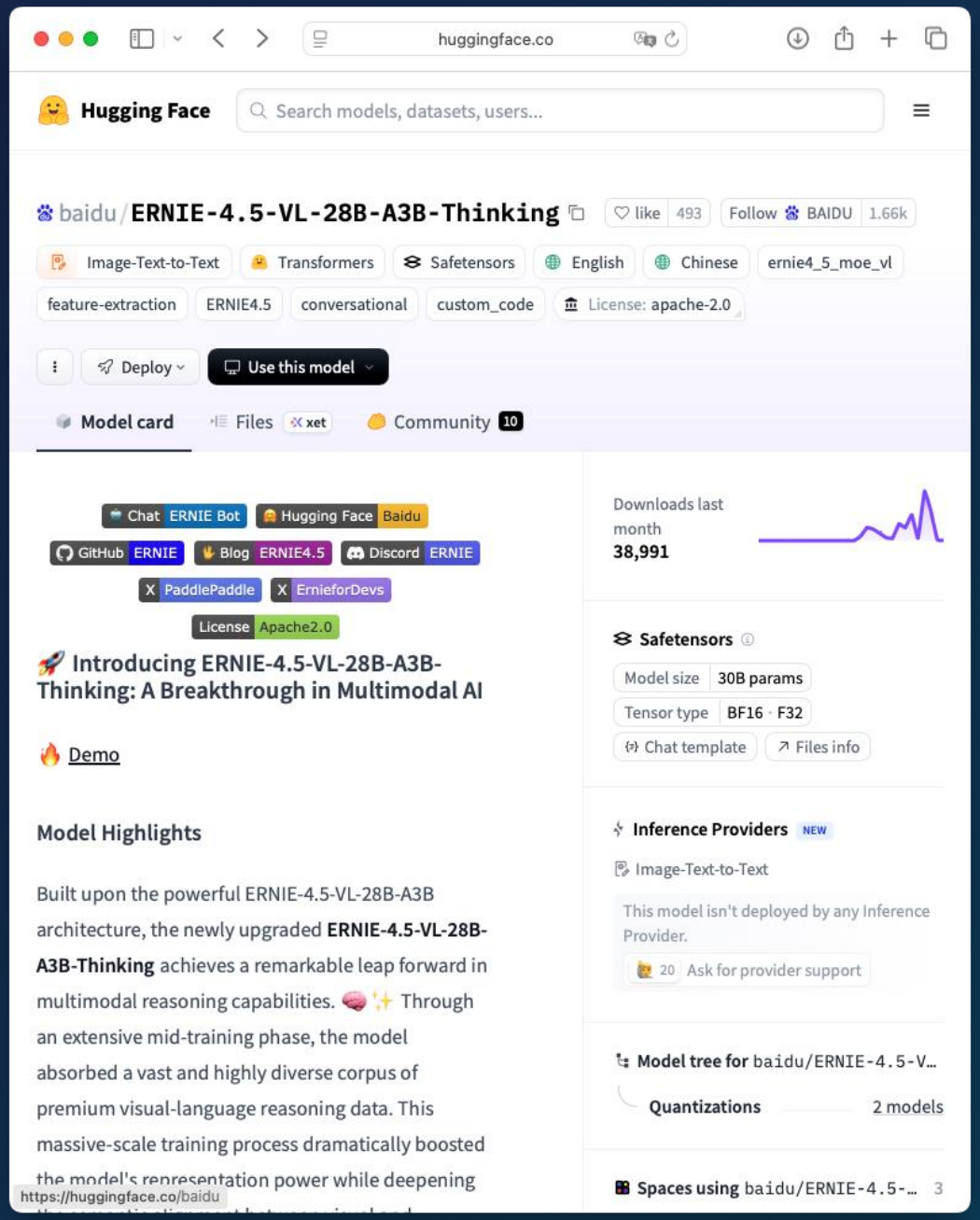

runwayml.com
runway
Make anything.
Try now for free
Adobe
Allstate

Audio


elevenlabs.io
IIElevenLabs
Sign up
We're off under the lights here for this semi-final clash, the stadium buzzing with anticipation. ElevenLabs United in their iconic black
PLAY
The most expressive text to speech model
Our AI voice generator delivers emotional depth and rich delivery, setting a new standard in expressive speech. Available now in Alpha.
DISCOVER ELEVEN V3
SIGN UP
CHAT

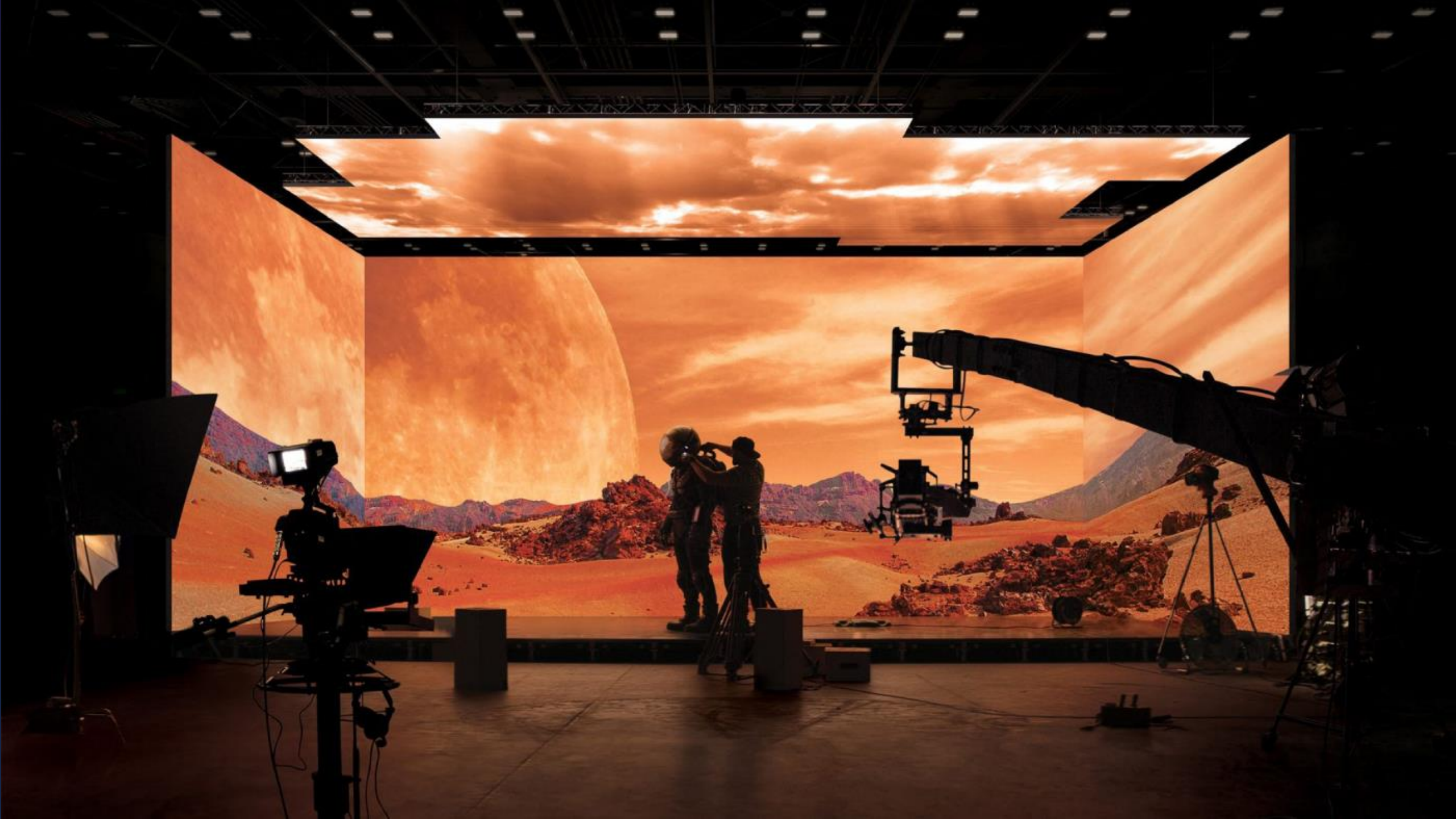

suno.com
SUNO
Sign In
Sign Up
Make any song you can imagine
Start with a simple prompt or dive into our pro editing tools, your next track is just a step away.
Chat to make music
Create
Forbes
RollingStone
VARIETY

Open source, 3B, Chinese…



Use case 1: Agentic AI in media

Use case 2: Virtual production



Challenges

Some of the multimodal challenges we plan to work on in VLAM

- Authenticity tracking
- Data value attribution
- Multimodal RAG
- Narrative structure
- Emotional connection
- Collaboration between The Netherlands and Flanders

Thank You

tanguy.coenen@imec.be