



Gemini

Google just released
Gemini - Only model till
date that beats OpenAI's
GPT-4

Here's what's special >

What's special about Gemini

Gemini is a multimodal model that can process information from:



Text



Audio



Video



Image



Code

State-of-the-art-performance

Gemini Ultra has achieved a milestone by **scoring 90.0%** and surpassing human experts in MMLU (massive multitask language understanding).

This assessment encompasses 57 diverse subjects, including mathematics, physics, history, law, medicine, and ethics, to evaluate a combination of world knowledge and problem-solving skills.

Gemini performance results on text and code

Capability	Benchmark Higher is better	Description	Gemini Ultra	GPT-4 API numbers calculated where reported numbers were missing
General	MMLU	Representation of questions in 57 subjects (incl. STEM, humanities, and others)	90.0% CoT@32*	86.4% 5-shot* (reported)
Reasoning	Big-Bench Hard	Diverse set of challenging tasks requiring multi-step reasoning	83.6% 3-shot	83.1% 3-shot (API)
	DROP	Reading comprehension (F1 Score)	82.4 Variable shots	80.9 3-shot (reported)
	HellaSwag	Commonsense reasoning for everyday tasks	87.8% 10-shot*	95.3% 10-shot* (reported)
Math	GSM8K	Basic arithmetic manipulations (incl. Grade School math problems)	94.4% maj1@32	92.0% 5-shot CoT (reported)
	MATH	Challenging math problems (incl. algebra, geometry, pre-calculus, and others)	53.2% 4-shot	52.9% 4-shot (API)
Code	HumanEval	Python code generation	74.4% 0-shot (IT)*	67.0% 0-shot* (reported)
	Natural2Code	Python code generation. New held out dataset HumanEval-like, not leaked on the web	74.9% 0-shot	73.9% 0-shot (API)

Gemini’s multimodality performance

MULTIMODAL				
Capability	Benchmark	Description Higher is better unless otherwise noted	Gemini	GPT-4V <small>Previous SOTA model listed when capability is not supported in GPT-4V</small>
Image	MMMU	Multi-discipline college-level reasoning problems	59.4% <small>0-shot pass@1 Gemini Ultra (pixel only*)</small>	56.8% <small>0-shot pass@1 GPT-4V</small>
	VQAv2	Natural image understanding	77.8% <small>0-shot Gemini Ultra (pixel only*)</small>	77.2% <small>0-shot GPT-4V</small>
	TextVQA	OCR on natural images	82.3% <small>0-shot Gemini Ultra (pixel only*)</small>	78.0% <small>0-shot GPT-4V</small>
	DocVQA	Document understanding	90.9% <small>0-shot Gemini Ultra (pixel only*)</small>	88.4% <small>0-shot GPT-4V (pixel only)</small>
	Infographic VQA	Infographic understanding	80.3% <small>0-shot Gemini Ultra (pixel only*)</small>	75.1% <small>0-shot GPT-4V (pixel only)</small>
	MathVista	Mathematical reasoning in visual contexts	53.0% <small>0-shot Gemini Ultra (pixel only*)</small>	49.9% <small>0-shot GPT-4V</small>
Video	VATEX	English video captioning (CIDEr)	62.7 <small>4-shot Gemini Ultra</small>	56.0 <small>4-shot DeepMind Flamingo</small>
	Perception Test MCQA	Video question answering	54.7% <small>0-shot Gemini Ultra</small>	46.3% <small>0-shot SeViLA</small>
Audio	CoVoST 2 (21 languages)	Automatic speech translation (BLEU score)	40.1 <small>Gemini Pro</small>	29.1 <small>Whisper v2</small>
	FLEURS (62 languages)	Automatic speech recognition (based on word error rate, lower is better)	7.6% <small>Gemini Pro</small>	17.6% <small>Whisper v3</small>
*Gemini image benchmarks are pixel only – no assistance from OCR systems				

Exciting capabilities

Gemini is a highly capable model meeting human level reasoning skills. Here are a few use-cases:



Sophisticated reasoning



Understanding text, images, audio, and more



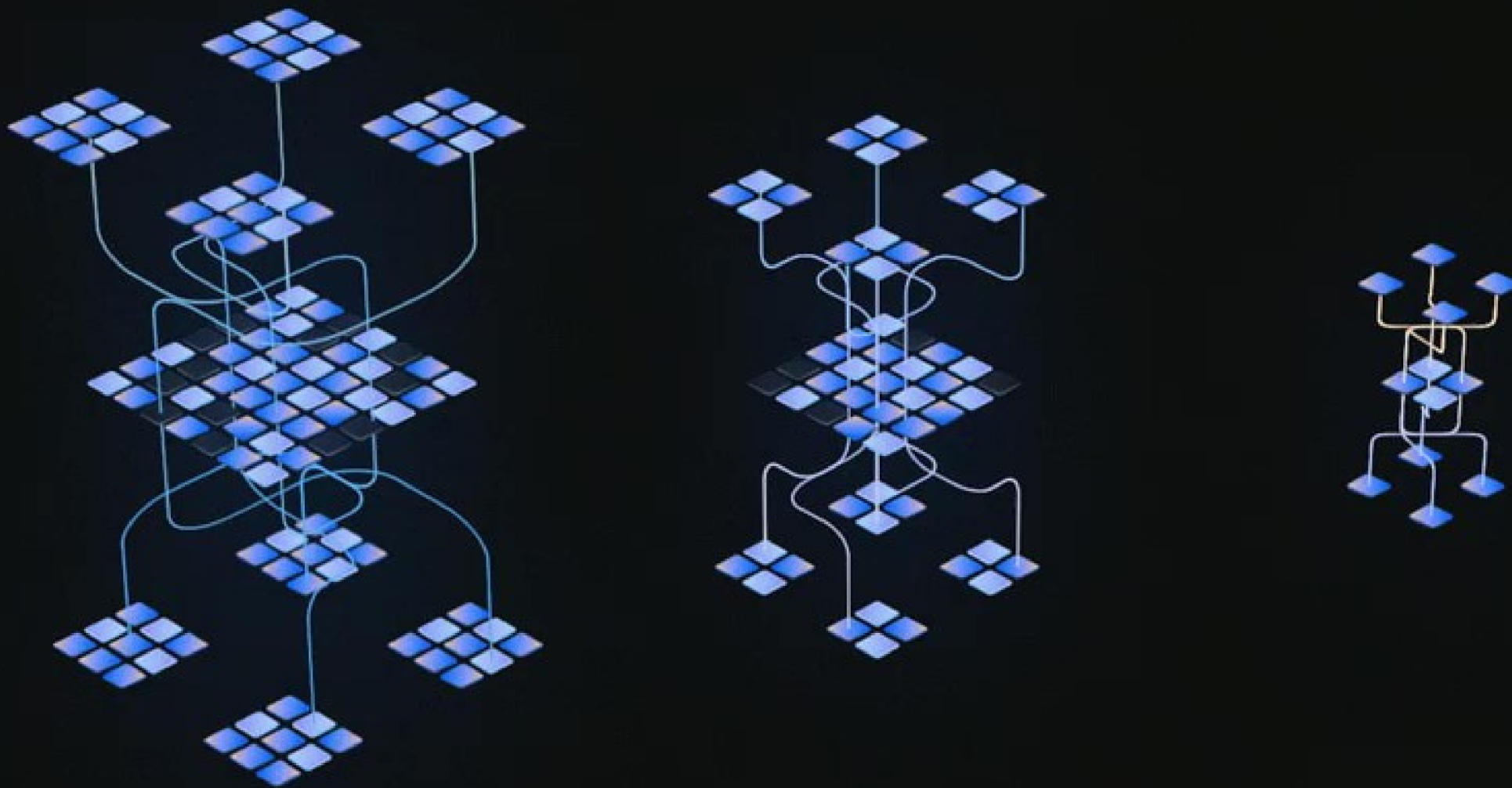
Advanced coding



Reliable and scalable

Gemini availability

Gemini will be available in 3 sizes:



Gemini Ultra

It is the largest and most capable model for highly complex tasks.



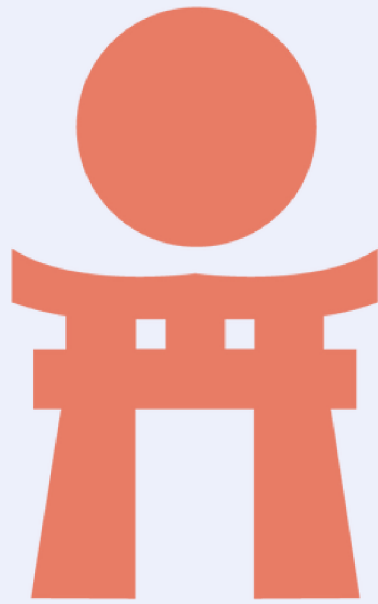
Gemini Pro

It works best for scaling across a wide range of tasks.



Gemini Nano

This is the most efficient model for on-device tasks.



Subscribe to our weekly LinkedIn newsletter
The Data-Driven Dispatch



and get comprehensive resources to
stay ahead in generative AI