Announcing the Latest Addition to Our Leading Multimodal Models – Reka Core

Global / San Francisco, CA. April 15, 2024 – Today, Reka is excited to announce the launch of Reka Core ("Core"), our largest and most capable model yet. Core is a frontier-class multimodal language model capable of text, image, video, and audio input. We trained the model efficiently and from scratch on thousands of GPUs over a period of several months. Various tests of its performance show that it is on par with or better than leading models in the industry today.

Model Highlights

Core is competitive with models from OpenAI, Anthropic, and Google across key industry-accepted evaluation metrics. Given its footprint and performance, on a total cost of ownership basis, Core delivers outsized value. The combination of Core's capabilities and its deployment flexibility unlocks vast new use cases.

Capabilities

1. Multimodal (image and video) understanding. Core is not just a frontier large language model. It has powerful contextualized understanding of images, videos, and audio and is one of only two commercially available comprehensive multimodal solutions.

2. 128K context window. Core is capable of ingesting and precisely and accurately recalling much more information.

3. Reasoning. Core has superb reasoning abilities (including language and math), making it suitable for complex tasks that require sophisticated analysis.

4. Coding and agentic workflow. Core is a top-tier code generator. Its coding ability, when combined with other capabilities, can empower agentic workflows.

5. Multilingual. Core was pretrained on textual data from numerous languages. It is fluent in English as well as several Asian and European languages.

6. Deployment Flexibility. Core, like our other models, is available via API, on-premises, or on-device to satisfy the deployment constraints of our customers and partners.

While we release a first version today, we expect Core—along with our other models—to continue to break performance barriers as it undergoes further training. While all built with our multimodal focus, we developed our suite of models to help us serve a wide range of unique use cases in the most efficient and effective way possible.
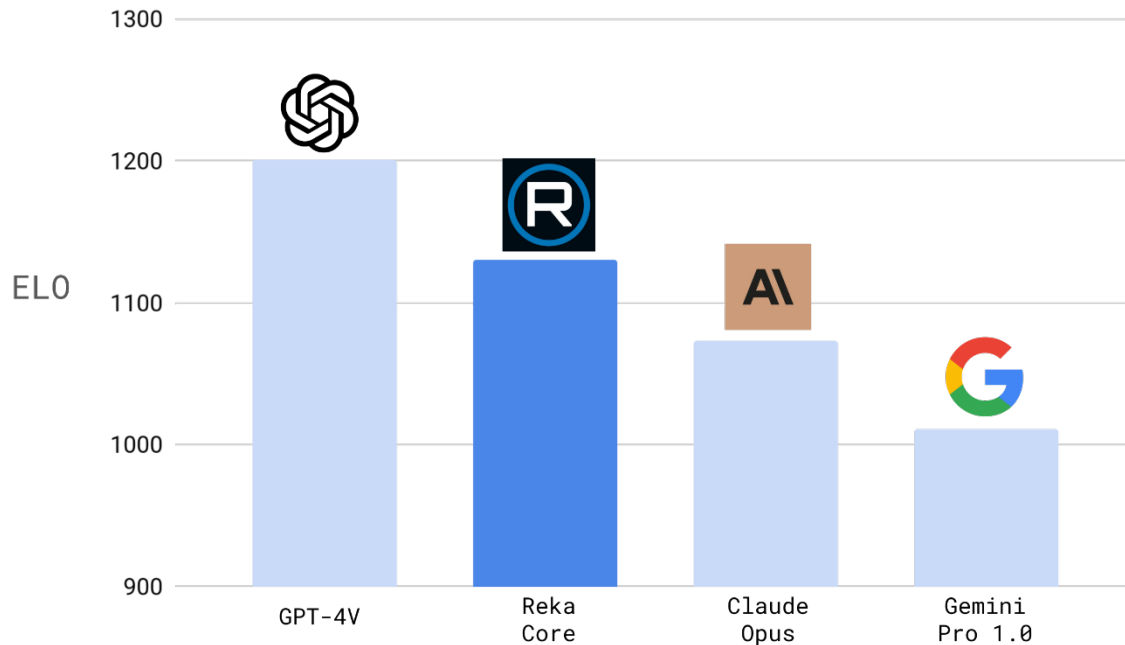
Benchmarking

Core outperforms Claude-3 Opus on our multimodal human evaluation conducted by an independent third party, surpasses Gemini Ultra on video tasks, a key focus given our multimodal approach, and is comparable to GPT-4V on MMMU. On language tasks, Core is competitive with other frontier models on well-established benchmarks.

The table below provides an in-depth comparison of Core with other leading models.

Evaluation	Reka Core	GPT 4	Claude 3 Opus	Claude 3 Sonnet	Gemini Ultra	Gemini Pro 1.5	Mistral Large	Grok 1.5
Multimodal Input Capabilities	Image, Video, Audio	Image	Image	Image	Image, Video, Audio	Image, Video, Audio	None	Image
MMLU (knowledge)	83.2	86.4	86.8	79.0	83.7	81.9	81.2	81.3
GSM8K (reasoning)	92.2	92.0	95.0	92.3	94.4	91.7	91.2	90.0
HumanEval (coding)	76.8	76.5	84.9	73.9	74.4	71.9	45.1	74.1
GPQA-main (knowledge)	38.2	38.1	50.2	39.1	35.7	41.5	-	-
MMMU (image)	56.3	56.8	59.4	53.1	59.4	58.5	-	53.6
VQAv2 (image)	78.1	77.2	-	-	77.8	73.2	-	-
Perception-test (video)	59.3	-	-	-	54.7		-	-

Source: Reka internal information, company websites, publicly disclosed information, and technical reports. "-" denotes data that is not disclosed or not relevant / applicable due to model capabilities.

Rankings on Human Evaluation (Multimodal)

A higher ELO score represents better performance. Rankings measured in ELO computed from third-party blind human preferences on a diverse test set of multimodal prompts.

Pricing

Core will be available to the public at prices competitive with peers given its performance. Our launch API pricing is $10 per million input tokens and $25 per million output tokens.

Our Partners

In less than a year, Reka has become one of only two developers providing models that allow for comprehensive multimodal input. Its three models allow image, video, and audio input in addition to text. This enables broader and differentiated customer use cases for industries including e-commerce, social media, digital content and video games, healthcare, and robotics, to name a few.

A crucial part of delivering on our mission to make frontier multimodal models that benefit humanity is our various partners. We are proud to count amongst our partners leading global technology platforms and government organizations such as Snowflake, Oracle, and AI Singapore. They enable our customers, organizations, and individuals around the world to benefit from and build with Reka models by democratizing access to multimodal technology.

"We are excited to partner with Reka to bring Reka Core's impressive industry leading performance to customers through Snowflake Cortex," said Baris Gultekin, Head of AI, Snowflake. "Snowflake and Reka are at the forefront of AI innovation, and we're excited to offer the state-of-the-art AI capabilities, all from within Snowflake's built-in security and governance."

"Reka's dynamic multimodal and multilingual models allow enterprises to unlock more value from their data, streamline complex tasks, and realize cost efficiencies," said Greg Pavlik, senior vice president, AI and Data Management Services, Oracle Cloud Infrastructure. "We look forward to collaborating with Reka to bring video, audio and image capabilities to organizations globally."

"One of the goals of AI Singapore is to develop highly capable open-source models for Southeast Asia. We are excited about Reka's new models and look forward to working closely with their series of highly performant general-purpose models to build the next generation of SEA-LION models for the region," said Leslie Teo, Senior Director of AI Products, AI Singapore, a national programme launched by the National Research Foundation, Singapore, to anchor deep national capabilities in AI.

Concluding Remarks

We live in a multimodal world and Reka is dedicated to the development of frontier multimodal models. The launch of Core today marks an important milestone in delivering against our mission. With our full suite of models comprising Edge, Flash, and now Core, we are ready to take on a greater range of challenges. We are excited by what the future holds and look forward to welcoming new partners who want to join us on that journey.

Additional Resources

- Experience firsthand Reka Core, Flash and Edge
- See demonstrations of What Reka Can Do
- Learn more about Reka Model Pricing
- Review the Technical Paper for Reka Core

Contact Info

Reka PR
Eugenie Lamprecht
eugenie@reka.ai

About Reka

Reka is a developer of industry-leading, multimodal, AI models that enable individuals and organizations to deploy generative AI applications. Reka was founded by a team of highly experienced scientists and engineers from DeepMind, Google Brain, and FAIR. For more information or to try out our multimodal models, please visit us at www.reka.ai