Cohere Inc. is a Canadian multinational technology company focused on artificial intelligence . Cohere specializes in large language models and AI products for regulated industries, particularly the finance, healthcare, manufacturing, and energy fields, as well as the public sector. Cohere was founded in 2019 by Aidan Gomez , Ivan Zhang, and Nick Frosst and is headquartered in Toronto and San Francisco , with offices in Montreal , London , and New York City .

History

In 2017, a team of researchers at Google Brain introduced the transformer machine learning architecture in " Attention Is All You Need ," which demonstrated state-of-the-art performance on a variety of natural language processing tasks. [5] [6] [7] In 2019, Aidan Gomez, one of its co-authors, along with Nick Frosst, [8] another researcher at Google Brain, founded Cohere with Ivan Zhang, with whom Gomez had done research at FOR.ai. [9] [10] All of the co-founders attended University of Toronto . [11]

Gomez is the company's CEO. [10] In December 2022, Martin Kon, the former CFO of YouTube , became president and COO. [12]

In November 2021, Google Cloud announced that they would help power Cohere's platform using their robust infrastructure, and Cloud's TPUs would be used by Cohere for the development and deployment of their products. [13] [14]

In June 2022, Cohere launched Cohere For AI, a nonprofit research lab and community dedicated to contributing open-source, fundamental machine learning research. It is led by Sara Hooker , a former research scientist at Google Brain. [15]

In December 2022, Cohere released a multilingual model for understanding text that would work with over 100 languages, to help users search for documents by meaning instead of with keywords. This type of process was not previously widely available in languages other than English. [12]

On September 12, 2023, it was announced that Cohere had become one of fifteen tech companies to agree to sign on to a voluntary White House commitment on measures for testing, reporting, and research on the risks of AI. [16] On September 27, 2023, it was announced that Cohere had also signed Canada's voluntary code of conduct for AI, to promote the responsible development and management of advanced generative AI systems. [17] In May 2025, Cohere acquired Ottogrid, a Vancouver-based platform that develops enterprise tools for automating certain kinds of high-level market research. [18]

In 2024, Cohere opened a new office in New York City. [19] [20] [21] In 2025, they later expanded to Montreal, Canada. [22]

In June 2025, it was announced that Cohere had partnered with the governments of Canada and the United Kingdom to expand the use of AI in the public sector. [23]

Products

Considered an alternative to OpenAI , [24] Cohere is focused on generative AI for the enterprise, building technology that businesses can use to deploy chatbots, search engines, copywriting, summarization, and other AI-driven products. [25] [26] Cohere specializes in large language models : AI trained to digest text from an enterprises' internal data or publicly available sources like

the internet to understand how to process and respond to prompts with increasing sophistication. [12]

The Cohere platform is available through API as a managed service, through platforms such as Amazon SageMaker and Google's Vertex AI. It can be used for tasks such as writing copy, moderating content, classifying data, and extracting information. It is cloud agnostic and not tied to a particular cloud service. [10]

Cohere's generative AI technology is embedded into several Oracle products, and its chat capabilities are embedded into Salesforce products. [24]

In March of 2025, Cohere's nonprofit research lab introduced Aya Vision, an AI model that can describe images, translate text, and summarize information. The model is free to use for research but is not permitted to be used for commercial purposes. [27]

Partnerships

On June 13, 2023, Oracle announced a partnership with Cohere to provide generative AI services to help organizations automate end-to-end business processes. As a result, Cohere's technology is integrated into Oracle's business applications, including Oracle Fusion Cloud, Oracle NetSuite, and Oracle industry-specific applications. [28] McKinsey announced a collaboration with Cohere, to help organizations integrate generative AI into their operations on July 18, 2023. [29] Cohere also partnered with software company LivePerson to offer customized large language models for businesses in 2023. [25] here

In 2024, Cohere partnered with Fujitsu in co-developing Takane, a Japanese LLM. [30] [31]

In January 2025, in partnership with RBC , Cohere introduced North for Banking, a secure generative AI platform designed to enhance productivity and data security in financial services. [32] A partnership with LG developed a Korean LLM and customized the North platform for the specific needs of Korean companies. [33] In May 2025, Cohere partnered with SAP to integrate its AI models into SAP's Business Suite and make them available through SAP AI Core's generative AI hub. [34] Simultaneously, Cohere announced a collaboration with Dell Technologies, making Dell the first infrastructure provider to offer Cohere North, the company's secure AI workspace platform, for on-premises deployment. [35] [36] Cohere also entered the healthcare sector through a partnership with Ensemble Health Partners, a revenue cycle management company, to deploy agentic AI solutions for healthcare providers' administrative workflows. [37] [38]

In July 2025, Cohere announced a partnership with Bell Canada to provide AI services to government and enterprise customers. Under the agreement, Bell Canada will deploy Cohere's technology on its data center infrastructure and offer AI solutions to clients. The partnership positions the services as a Canadian alternative to international cloud providers, with Bell also implementing Cohere's North platform for internal operations. [39]

Funding

On September 7, 2021, Cohere announced that they had raised $40 million in Series A funding led by Index Ventures ; Index Ventures partner Mike Volpi also joined Cohere's board. [13] [14] The round also included Radical Ventures, Section 32 , and AI-experts Geoffrey Hinton , Fei-Fei Li , Pieter Abbeel , and Raquel Urtasun . [40]

In February 2022, Cohere announced they had raised $125 million in series B funding led by Tiger Global . [41] In June 2023, Cohere raised an additional $270 million in series C funding from investors including Inovia Capital, Oracle , Salesforce , and Nvidia , at a valuation of $2.2 billion. [25] [42]

In March 2024, it was reported that Cohere had an annualized revenue run rate of $22 million and in July 2024 it was reported they closed $500 million in funding from investors including Cisco , AMD and Fujitsu at a valuation of $5.5 billion. [43] [44]

In December 2024, Cohere received $240 million in public funding from the Canadian government as part of the Canadian Sovereign AI Compute Strategy to support the construction of a domestic AI data center . [45]

As of May 2025, Cohere's annual revenue was reported to be $100 million. [46] [47]

In 2025, Cohere raised US$500 million in a funding round led by Radical Ventures and Inovia Capital, achieving a valuation of US$6.8 billion. [48] Following the investment, Inovia's Patrick Pichette was appointed to Cohere's Board of Directors. [49]

Awards

Cohere was listed on the Forbes AI 50 list in four consecutive years, 2022, [50] 2023, [51] 2024, [52] and 2025 [53]

Fortune magazine listed the company on their 50 AI innovators list in 2023. [54]

Cohere was listed on CNBC 's Disruptor list in 2023 [55] and 2024. [56]

Litigation

In February 2025, it was reported that a consortium of 14 publishing companies including The Atlantic , Condé Nast , and Forbes had sued Cohere for copyright infringement, accusing Cohere of using their content for training without permission, as well as displaying considerable portions or entire articles of their content without permission. [57]

See also

Anthropic

OpenAI

Glean Technologies

References

v

t

e

Autoencoder

Deep learning

Fine-tuning

Foundation model

Generative adversarial network

Generative pre-trained transformer

Large language model

Model Context Protocol

Neural network

Prompt engineering

Reinforcement learning from human feedback

Retrieval-augmented generation

Self-supervised learning

Stochastic parrot

Synthetic data

Top-p sampling

Transformer

Variational autoencoder

Vibe coding

Vision transformer

Waluigi effect

Word embedding

Character.ai

ChatGPT

DeepSeek

Ernie

Gemini

Grok

Copilot

Claude

Gemini

Gemma

GPT 1 2 3 J 4 4o 4.5 4.1 OSS 5

1

2

3

J

4

4o

4.5

4.1

OSS

5

Llama

o1

o3

o4-mini

Qwen

Base44

Claude Code

Cursor

Devstral

GitHub Copilot

Kimi-Dev

Qwen3-Coder

Replit

Xcode

Aurora

Firefly

Flux

GPT Image 1

Ideogram

Imagen

Midjourney

Qwen-Image

Recraft

Seedream

Stable Diffusion

Dream Machine

Hailuo AI

Kling

Midjourney Video

Runway Gen

Seedance

Sora

Veo

Wan

15.ai

Eleven

MiniMax Speech 2.5

WaveNet

Eleven Music

Endel

Lyria

Riffusion

Suno AI

Udio

Agentforce

AutoGLM

AutoGPT

ChatGPT Agent

Devin AI

Manus

OpenAI Codex

Operator

Replit Agent

01.AI

Aleph Alpha

Anthropic

Baichuan

Canva

Cognition AI

Cohere

Contextual AI

DeepSeek

ElevenLabs

Google DeepMind

HeyGen

Hugging Face

Inflection AI

Krikey AI

Kuaishou

Luma Labs

Meta AI

MiniMax

Mistral AI

Moonshot AI

OpenAI

Perplexity AI

Runway

Safe Superintelligence

Salesforce

Scale AI

SoundHound

Stability AI

Synthesia

Thinking Machines Lab

Upstage

xAI

Z.ai

Category