

About Cohere

Cohere is the leading AI platform for enterprise. We build world-class large language models (LLMs) that allow computers to search, understand meaning, and converse in text. Our models are uniquely suited to the needs of business, providing ease of use and strong security and privacy controls across multiple deployment options.

Embeddings Models

Cohere Embed is a representative model which translates text into numerical vectors that models can understand. We provide industry-leading English and multilingual models (100+ languages) for a range of use cases, including:

- **Semantic search:** Enables powerful search applications when coupled with a vector database, with excellent relevance based on search phrase meaning
- **Text classification:** Supports intent recognition, sentiment analysis, and advanced analysis, like automatic text clustering and naming, and it is fine-tunable by domain
- **Search engine for RAG** (retrieval augmented generation). Finds and retrieves most relevant information from connected enterprise data sources for RAG use cases
- **Legacy search improvement:** Greatly improves search relevance without replacing legacy tools using Cohere's Rerank model and a few lines of code

Generative Models

Cohere Command is a text generation model, available in two different sizes, that is highly customizable for business use cases, including:

- **Text generation:** Builds articles, product descriptions, and more based on user prompts
- **Text summarization:** Summarizes key ideas and themes from diverse long-form text, such as news, investor reports, legal, or medical information
- **RAG:** Generates text responses to prompts and user requests using accurate, up-to-date information from connected enterprise data sources
- **Chat:** Powers knowledge agents and chatbots that maintain conversational context, provide accurate responses using RAG, citing sources if necessary

The Cohere Difference

- **Customization (fine-tuning):** Cohere offers sophisticated fine-tuning tools and capabilities that enhance model performance for specific business tasks or domain knowledge, giving customers superior performance at industry-leading inference cost
- **Scalability:** Cohere's models are packaged with inference engines that deliver better runtime performance at a lower cost than open-source equivalents
- **Flexible deployment:** Models can be accessed through a SaaS API, cloud services (e.g., OCI, AWS SageMaker, Bedrock), and private deployments (VPC and on-prem)
- **Privacy:** Customer data is never used in training base models, and customers have complete control over customization and model inputs/outputs
- **Model integrity:** Models are trained from scratch from known, purchased, or public domain data sources and are subject to extensive adversarial testing and bias mitigation

To learn more about Cohere or discuss your business needs, **contact us at: cohere.com/contact-sales**