

Trial and playground keys land developers on Command, Embed, and Rerank APIs; Production pay-as-you-go converts PLG usage; Model Vault and private deploy harden inference for regulated buyers; North packages agents and RAG as an enterprise workspace sold with AWS, Azure, SageMaker, and Oracle marketplaces.

KEY PARTNERS

- Amazon Bedrock & SageMaker (docs models).
- Microsoft Azure AI Foundry (docs models).
- Oracle Generative AI Service (docs models).
- Cloud/GPU infra for training and Model Vault.
- Enterprise co-marketing (e.g. North for Banking).
- Cohere Labs open-science community (About).

KEY ACTIVITIES

- Train and evaluate Command / Aya / North model families.
- Operate Chat, Embed, Rerank, Parse, Transcribe APIs.
- Run Model Vault dedicated inference + rate limits.
- Build North agents, connectors, and private deploy.
- **Developer education** — docs, playground, Trial keys.
- Enterprise security reviews and vertical solutions.

KEY RESOURCES

- Command family weights (A+, A, R7B, specialty variants).
- Embed v4 + Rerank v4 retrieval stack; Parse/Transcribe.
- North agentic platform + connectors.
- Model Vault dedicated inference operations.
- Enterprise/security positioning and Labs research brand.
- **Investor capital** — About claim ~\$1B; press Series D (SECONDARY).

VALUE PROPOSITIONS

- Enterprise LLMs purpose-built for RAG, tools, and agents.
- **Retrieval stack** — Embed v4 + Rerank v4 (+ Parse).
- Trial free land; Production PAYG for scale.
- Model Vault / private deploy for isolation & performance.
- **North** — governed agents and workplace automation.
- **Multi-cloud** — Cohere API + AWS / Azure / Oracle.

CUSTOMER RELATIONSHIPS

- Self-serve Trial signup and API keys dashboard.
- Production key workflow on Billing & Usage page.
- Docs, cookbooks, and rate-limit transparency.
- Sales/demo motion for North and private deploy.
- Support path for rate-limit increases (docs).
- **Expansion** — API → Model Vault → North inside accounts.

CHANNELS

- **Primary** — cohere.com + dashboard onboarding.
- Developer docs, SDKs, playground / Trial keys.
- AWS Bedrock / SageMaker listings.
- Azure AI Foundry and Oracle GenAI catalogs.
- Direct enterprise sales for North / private deploy.
- Labs research and content marketing.

CUSTOMER SEGMENTS

- Indie developers / prototypes (Trial free).
- Startup & ISV product teams (Production PAYG).
- ML/platform teams needing Embed + Rerank quality.
- Regulated buyers needing Vault / VPC / on-prem.
- Enterprises buying North workplace agents.
- **Verticals** — finance, public sector, tech, energy (site).

COST STRUCTURE

- **GPU / cloud COGS** — train and serve inference.
- Free Trial allowances as demand-gen cost.
- **R&D** — Command family, retrieval, North, Labs.
- **Sales & marketing** — PLG content + enterprise field.
- Support / success and rate-limit operations.
- Compliance & private-deploy / Vault isolation overhead.

REVENUE STREAMS

- **Trial API free** — demand gen (FACT — pricing).
- **Production PAYG** — tokens & Rerank searches (FACT).
- **FAQ rate card** — e.g. Command R+ 08-2024 \$2.50/\$10; Aya \$0.50/\$1.50 (FACT).
- **Model Vault instance fees** — Embed 4 from \$4/hr · \$2,500/mo (FACT).
- **North + private deploy** — enterprise contracts (sales).
- CSP marketplace consumption (AWS / Azure / Oracle).
- **Newer Command A+/Reasoning/Vision** — often contact sales for prod.