

## Scenario

RAG reranking (maximum accuracy)

---

RAG reranking (budget or self-hosted)

---

RAG reranking (multimodal documents)

---

Enterprise generation (open weights,  
EU governance)

---

Enterprise generation (long context,  
multilingual, fine-tuning)

## Best-fit model

zerank-2

---

BGE Reranker

---

Jina Reranker (m0)

---

Mistral Large 3

---

AI21 Jamba

## Reasoning

Leaderboard-leading NDCG@10 score,  
faster median latency than Cohere

---

Free, open-source, lightweight — trades  
some accuracy for zero licensing cost

---

Native text and image processing

---

Open weights, self-hostable, EU-oriented  
governance credentials

---

256K context, nine-language support,  
self-serve fine-tuning path