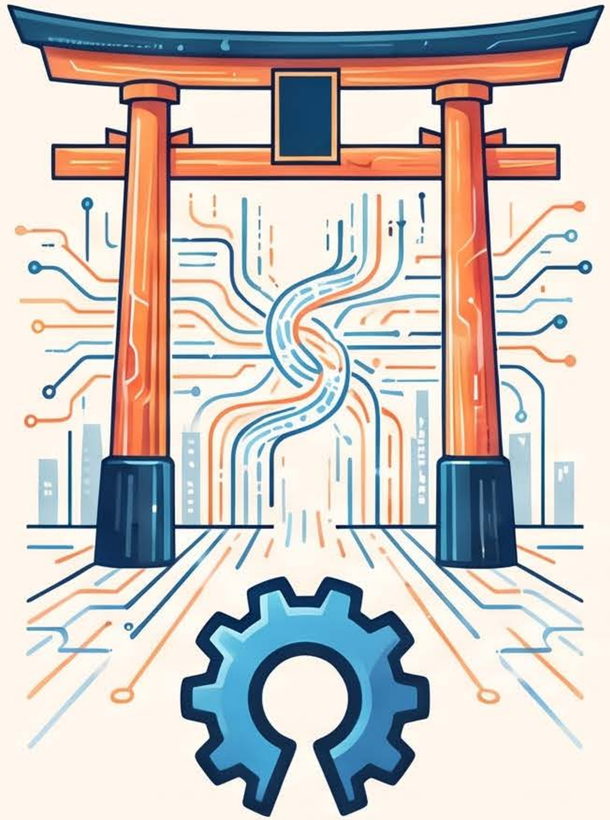
ELYZA推論モデルの実力：日本語業務への特化とOpenAI（o1/o3）との比較

特徴と強み：日本語処理・指示追従



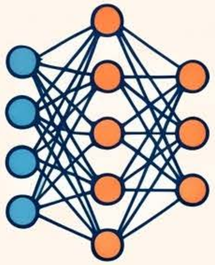


日本語質問応答・指示追従で高いパフォーマンスを発揮

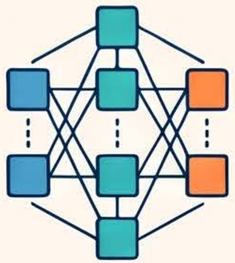
JEMHopQAやJFBench等の日本語ベンチマークで他モデルを上回る結果を記録。

商用利用可能なApache 2.0ライセンスの国産モデル

33B (Dense型) と32B-a3b (MoE型) の2種を公開し、社内運用に最適。

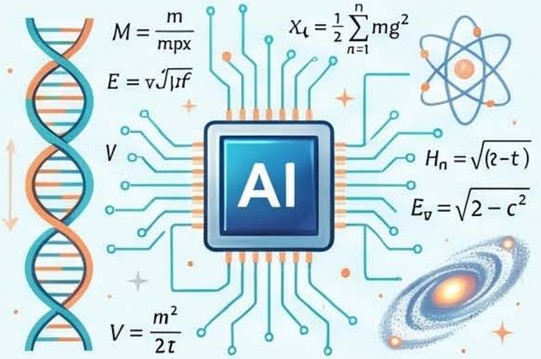


33B (Dense型)



32B-a3b (MoE型)

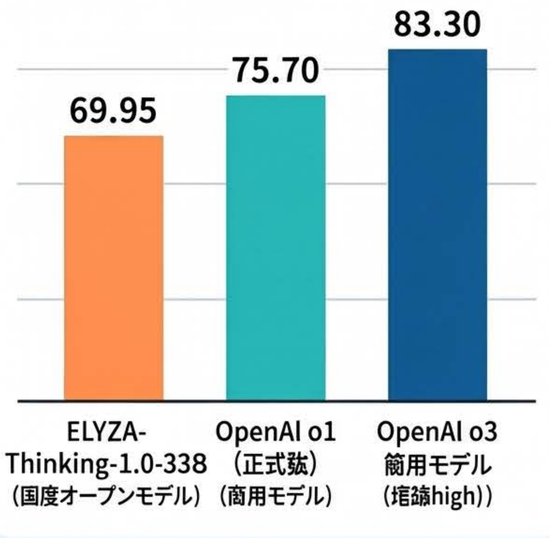
OpenAI（o1/o3）との性能比較と限界





科学・数学などの高度推論ではo1/o3に及ばず

科学推論 (GPQA Diamond) において、正式版o1やo3とは明確なスコア差が存在。



汎用推論で劣る場合も、特定業務での活用価値は十分

知財文書の整理や情報抽出など、目的を限定した実務適用には大きな期待。



Gemini Notebook