ELYZA-Thinking-1.0: 国産LLMの強みとグローバル性能比較

日本語・実務運用における「強み」

日本語タスクで海外競合モデルを凌駕





日本語複数文書QA
(JEMHopQA)

商用利用可能な完全オープンモデル



Apache 2.0 License

無償公開、データ持ち出し制限
のある自社環境で運用可能



低コストで効率的な開発・学習基盤



事後学習から評価までの
人間介入工数を1〜2人日程度に
抑えて構築を完了



高度推論における「課題と性能比較」

数学・論理推論における海外勢との格差

ELYZA-Thinking-33b



AIME（競技数学）

OpenAI o1



OpenAI o1/o3に
大きく水を開けられている

コーディング・ツール呼び出しの弱点



LiveCodeBenchで60.23%となり、
自律的なにエージェント操作や
複雑な実装で課題を残す



汎用推論競争ではなく「実務運用」へ最適化



Compact Size

33Bというコンパクトなサイズで、
自社管理下での低コストかつ
セキュアな運用を実現



主要ベンチマーク正解率比較

	ELYZA-Thinking-33b	OpenAI o1 (High/Full)
AIME 2024/2025	68.85%	83.3%
GPQA Diamond	69.95%	78.0%
MATH-500	96.00%	94.8%〜96.4%

Gemini Notebook