フロンティアLLM第三者評価：ユースケース別モデル選択ガイド

2026年11月時点の評価により、万能な単一首位はなく、タスクやコスト要件で第一候補が変わることが判明。

タスク分野別のモデル適性と強み

主要フロンティアLLMの強みと評価比較

主要ベンチマーク評価比較

モデル	AA指数	Vals科学	Agent Arena
GPT-6.2 Astra	62	1位	6位
Claude Fable 5.2	65	3位	1位
Gemini 4 Neon	61	4位	3位

科学研究：Astra

Astraが科学1位

科学総合68.44%で首位。



Vals科学: 1位
AA指数: 62

実務総合：Fable

Fableが総合首位

総合指数65と実務で首位。



Agent Arena: 1位
AA指数: 65

専門法務：Neon

Neonが法務首位

米国法務78.36%で首位。



Vals科学: 4位
AA指数: 61

推論設定の最適化と業務適応

コスト効率と品質のバランス

- **Astra設定最適化**

	AA指数	1課題費用	課題時間	AA指数
Astra max	AA指数: 62	4.00ドル		14.3分
Astra xhigh	AA指数: 62	2.52ドル	7.7分	費用37%削減
Astra high	AA指数: 60	2.07ドル	6.2分	



報告書作成の差

人手採点ではFableが優位。



業務別の使い分け

発明Astra・法務Neon等。

Gemini Notebook