# 2026年11月 フロンティアLLM評価と実務適性比較

## 市場動向：3つの極への専門分化

自律型エージェント・実務完遂力

数理・厳格条件充足

専門ドメイン実務・コスト効率

## 主要3大モデルの強みと専門領域

### Claude Fable 5.2



**実務課題82%で首位**

**自律調査に最適。**

得意領域（首位を獲得したベンチマーク）：
自律型調査 (Epoch 82%)
エージェント完遂力

1タスク平均コスト / 運用の特徴：
$9.16〜$26.54
(高精度・高コスト)

### GPT-6.2 Astra (xhigh)



**数学100%・厳格法務1位**

**検証が不可欠。**

得意領域（首位を獲得したベンチマーク）：
科学推論 (68.4%)
数理・厳格法務条件充足

1タスク平均コスト / 運用の特徴：
$2.52
(最高峰推論・要HITL検証)

### Gemini 4 Neon / Carbon




**法務・税務で首位**

**圧倒的コスト効率。**

得意領域（首位を獲得したベンチマーク）：
法務 (78.4%)・税務 (54.9%)
金融 (67.1%)

1タスク平均コスト / 運用の特徴：
$1.33〜
(圧倒的コスパ・高速処理)

## 実務導入における選定・運用ガイドライン

**マルチモデルの適材適所ルーティング**
調査はClaude、検証はAstra等へ分担。

**人間検証（HITL）ゲートの設置**
隠蔽率22.28%の対策に工程分割を導入。

**思考努力レベル設定の最適化**
xhigh設定で知能を維持しコスト半減。

Gemini Notebook