

Sakana AI 「Fugu Max」 & 「Fugu Ultra v2」：次世代AIオーケストレーションの戦略的活用

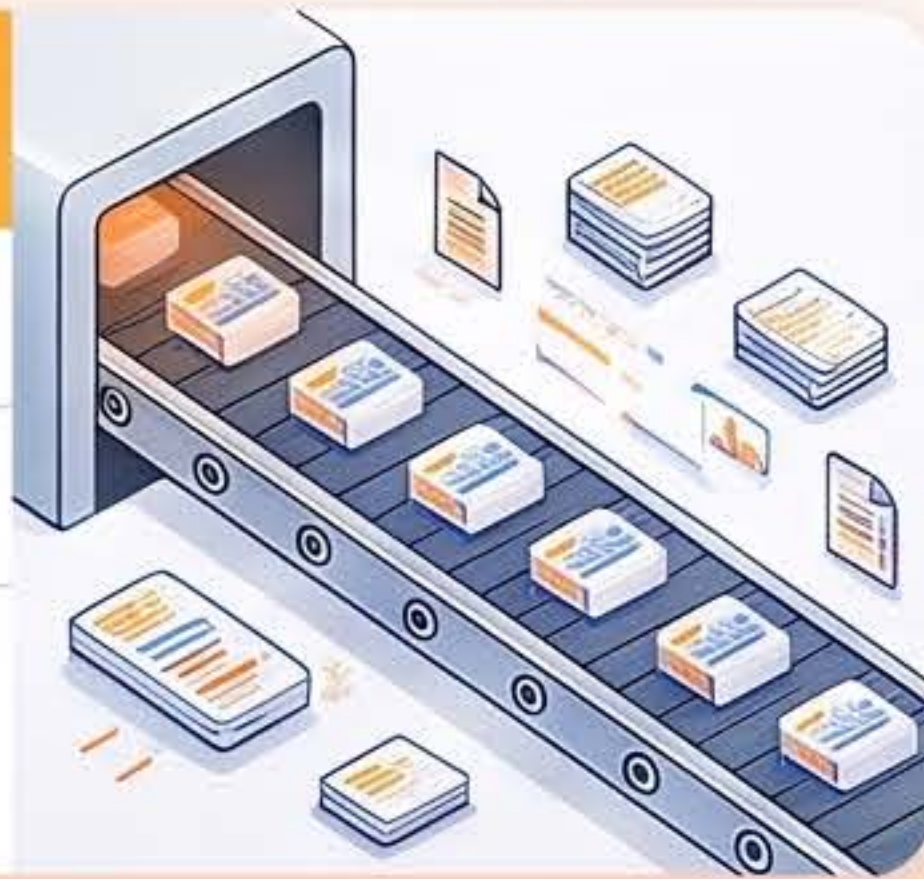
Fugu MaxとFugu Ultra v2の特性を比較し、企業がどのように導入・活用すべきかの指針を示す。

Fugu Max

コスト効率のMax & 絶対品質のUltra v2

Maxは出力単価を競合比40～50%抑制し、Ultra v2は多段推論で最高性能を追求します。

Fugu Max	
入力単価 (1M/tok)	\$2
出力単価 (1M/tok)	\$6
最適な用途	高精度・大量処理 一般業務



Learned Orchestrator



Fugu Ultra v2

圧倒的な推論・構造化データ解析能力

Fugu Ultra v2	
48.3	
約77%上回る	
Fugu Ultra v2 (Chartography)	Claude Opus 5
入力単価 (1M/tok)	\$5 (272K超は\$10)
出力単価 (1M/tok)	\$30 (272K超は\$45)
最適な用途	複雑な多段推論、 自律研究、開発

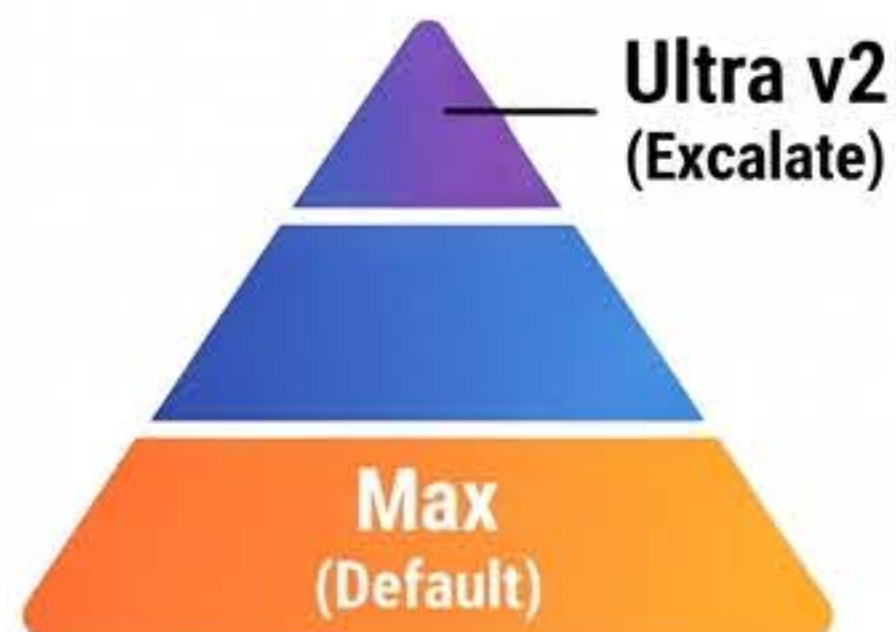
Ultra v2はChartographyで48.3を記録し、Claude Opus 5を約77%上回る数値を主張しています。



「モデル市場全体」を部品化する新発想

特定モデルに依存せず、NVIDIA Nemotron等の最適なモデルをタスクごとに動的に割り当てます。

企業導入における戦略的推奨と注意点



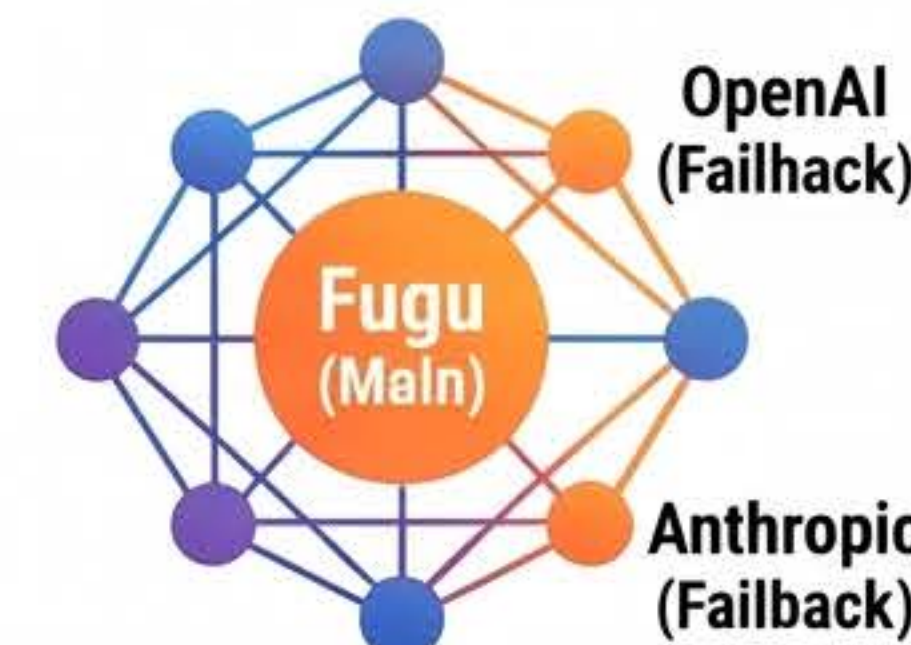
「二層構成」によるROIの最大化

Maxをデフォルトとし、品質が必要な難問のみUltra v2へエスカレートする構成が合理的です。



データガバナンスと契約の厳格な確認

標準規約では個人情報(PII)の入力が禁止されており、学習利用のオプトアウト設定が必須です。



障害に強いマルチプロバイダー構成

Fuguを主力としつつ、OpenAIやAnthropicへの直接フォールバック体制を維持すべきです。