

# モデル比較一覧

モデル名	概要・発表日	技術的特徴	価格	用途	参考文献
GPT-4.5	2025年2月27日発表。GPT-4改良版。知識量と対話品質が向上。	大規模事前学習、マルチモーダル対応（画像解析含む）。	Input: \$75/M Output: \$150/M	創造的文章生成、長文要約、大量資料の解析。	<a href="#">公式発表</a> <a href="#">参考記事</a>
OpenAI o1/o3-mini	2024年後半～2025年初頭に発表。推論特化型モデル。	CoT、可変推論パラメータ。	o1: Input: \$15/M, Output: \$60/M o3-mini: Input: \$1.10/M, Output: \$4.40/M	複雑な数学・プログラミング問題、段階的推論が必要なタスク。	<a href="#">o1</a> <a href="#">o3-mini</a>
Anthropic Claude3.7/ClaudeCode	2025年2月24日発表。ハイブリッド推論モデル。拡張思考モードで内部思考の可視化が可能。	拡張思考モード、内部思考ログの可視化、エージェント型コーディング支援機能（ClaudeCode）。	Input: \$3/M Output: \$15/M	対話、コーディング支援、根拠提示によるレポート生成。	<a href="#">公式発表</a> <a href="#">ClaudeCode</a>
xAI Grok3	2025年2月19日発表。高性能推論と最新情報アクセス（DeepSearch機能搭載）。	強化学習による推論、リアルタイムWeb検索(DeepSearch)機能、STEM分野向け最適化。	※参考:無料~\$30/月	STEM分野の難問解決、SNSデータ解析、リアルタイム情報の取得。	<a href="#">公式発表</a> <a href="#">参考記事</a>
Google Gemini2.0	2024年末～2025年初頭に投入。マルチモーダル&エージェント機能を搭載。	テキスト、画像、音声の統合処理、超長文コンテキスト（最大1Mトークン）、外部ツール連携機能。	Flash: Input: \$0.1/M Output: \$0.4/M	Web検索、情報収集・分析。	<a href="#">公式発表</a> <a href="#">参考記事</a>
DeepSeekR1	2025年1月23日発表。オープンソースでMITライセンス。低コストかつ高性能なモデル。	Mixture-of-Experts(MoE)構造、自己改善機構、効率的なGPU使用。	Input: \$0.14/M Output: \$2.19/M	大量データ処理、オンプレAI構築、カスタムモデル作成。	<a href="#">公式発表</a> <a href="#">参考記事</a>

# 2025年 最新生成AIモデル 調査レポート

## 対象モデル：

- GPT-4.5
- o3-mini / o1
- Claude 3.7 / Claude Code
- Grok 3
- Google Gemini 2.0
- DeepSeek R1

調査は各モデルの発表概要、アーキテクチャ・性能、コスト・運用上の特長、用途（特に金融リサーチへの応用展望）を比較・考察しています。

# はじめに

- 2025年2月以降に注目された生成AIモデル群
- 各モデルはそれぞれ独自の強み・技術革新を実現
- 生成AI市場は価格低下と性能向上のスパイラルへ
- 金融リサーチなど実務分野での応用が加速中

# OpenAI GPT-4.5

## 概要:

- 2025年2月27日発表、GPT-4の改良版
- 知識量の拡大と応答の自然さ向上

## 技術的特長:

- 大規模な事前学習と高度な最適化
- マルチモーダル対応（画像解析含む）

## 性能・用途:

- 創造的文章生成、長文要約に最適
- 金融資料の大量解析など、膨大なコンテキスト処理に強み

## 課題:

- 非常に高価、応答遅延もやや長め

# OpenAI o1 / o3-mini

## o1:

- 推論特化型モデル（段階的思考・チェイン・オブ・ソート採用）
- 複雑な数学・プログラミングタスクに優れる

## o3-mini:

- 小型ながら高精度な推論性能
- コスト効率に優れ、低遅延で高速応答

## 用途:

- 多段階推論が必要なシナリオ（金融トレンド予測、サプライチェーン解析等）

# Anthropic Claude 3.7 / Claude Code

## Claude 3.7:

- 2025年2月24日発表、ハイブリッド推論モデル
- 拡張思考モードで内部思考の可視化が可能

## Claude Code:

- 同時発表のコーディング支援ツール
- ターミナル操作で自律的にコード生成・編集

## 用途:

- 高度な対話・コーディング支援、説明責任が求められる業務
- 金融レポート生成など、根拠提示による信頼性向上

# xAI Grok 3

## 概要:

- 2025年2月19日発表、Elon Musk率いるxAIの最新モデル
- “Think”モード搭載で強化された推論能力

## 技術的特長:

- リアルタイムWeb検索（DeepSearch）機能統合
- STEM分野（数学・コーディング）で高評価

## 用途:

- 最新情報を必要とするタスク、SNSデータ解析
- 低価格サブスクリプションで個人利用にも好適

# Google Gemini 2.0

## 概要:

- 2024年末～2025年初頭に投入、Google DeepMindの次世代モデル
- マルチモーダル対応 & エージェント機能が大きな特徴

## 技術的特長:

- テキスト、画像、音声をネイティブに処理
- 超長文コンテキスト（最大1Mトークン）とツール連携

## 用途:

- 検索・情報分析、デジタルアシスタント
- 複数データソース統合で金融レポート作成支援に有用



# DeepSeek R1

## 概要:

- 2025年1月23日発表、オープンソースの大規模モデル
- MITライセンスで公開、低コスト運用が魅力

## 技術的特長:

- Mixture-of-Experts (MoE) 構造と自己改善機構
- 限られたGPUで大規模モデルを実現

## 用途:

- 大量データ処理、社内専用AI構築に最適
- 金融機関向けのオンプレAIシステム構築に貢献

# 各モデルの比較と市場への影響

## 性能面:

- GPT-4.5 ・ Gemini 2.0 : 総合知識とマルチモーダル力
- Claude 3.7 ・ Grok 3 : 高度な推論 ・ コーディング能力
- o1/o3-mini : 段階推論で難問解決
- DeepSeek R1 : 低コストでカスタム可能

## コスト・運用:

- DeepSeek R1が圧倒的な低価格
- 高性能モデルは計算資源 ・ コスト面で課題あり

## 戦略:

- 各社が異なる技術アプローチで競争を激化
- 競争は品質向上と価格低下を促進中

# 金融業界への展望

- **業務効率化:** AIが大量の財務資料や市場ニュースを自動解析
- **レポート生成:** 自動要約と根拠提示でアナリスト支援
- **小規模企業も活用可能:** 低コストモデルで中小金融機関のデータ解析が可能に
- **リスク管理:** AIの判断に対する最終チェック体制の構築が鍵

# 今後の展望

- **技術革新と価格低下:** 各社の競争により性能向上とコスト削減が促進。
- **AIの民主化:** オープンソースモデルの進展により、中小企業や研究機関も高性能AIを活用可能に。
- **金融リサーチ分野:** AIが大量データの解析とレポート自動生成を支援し、アナリストの付加価値創出を促進。
- 人間とAIの**協働による新たな意思決定体制**が構築される未来が近づいています。