

調査基準日: 2026年8月5日

Classification: 内部戦略資料

OpenAI 次世代モデル 「Astra」調査報告書

企業向けインテリジェンス・ブリーフィング：シグナルとノイズの分離



Pillar 1: Key Finding (重要発見)

科学的推論における
重大な進歩

**10件の未解決数学
問題に対する解答・
前進を確認。**



Pillar 2: Current Status (現在地)

製品ではなく、
限定的な研究成果

API、価格、利用開
始日、ベンチマーク
はすべて「未公開」。



Pillar 3: Strategic Action (戦略的アクション)

予算化の保留と
基盤構築

Astraを短期予算の前
提とせず、現行モデル
でのベンダー非依存な
評価基盤を構築せよ。

Fact vs. Fiction

	Myth (市場のノイズ)	Fact (検証済インテリジェンス)
Cost (コスト)	「AstraのAPI利用料は1回約 約2,000ドルである」	【誤解】 約2,000ドルは実験時のGPT-5.6 SolのAPI換算値。Astraの価格表ではない。
Identity (アイデンティティ)	「GoogleのProject Astraと同じ 製品群である」	【別物】 Google側はマルチモーダル・ア シスタントの研究。OpenAI Astraは数学 推論モデルであり組織も目的も異なる。
Availability (提供状況)	「近日中に企業向けChatGPTで 利用可能になる」	【未定】 「次期主要モデルの社内版」とい う言及のみ。ロードマップは未公開。

OpenAIモデルとAstraの公開経緯

2024-05-13

GPT-4o 発表
(リアルタイム・マルチモーダルを訴求)

2026-07-09

GPT-5.6 一般提供
(Sol・Terra・Luna)

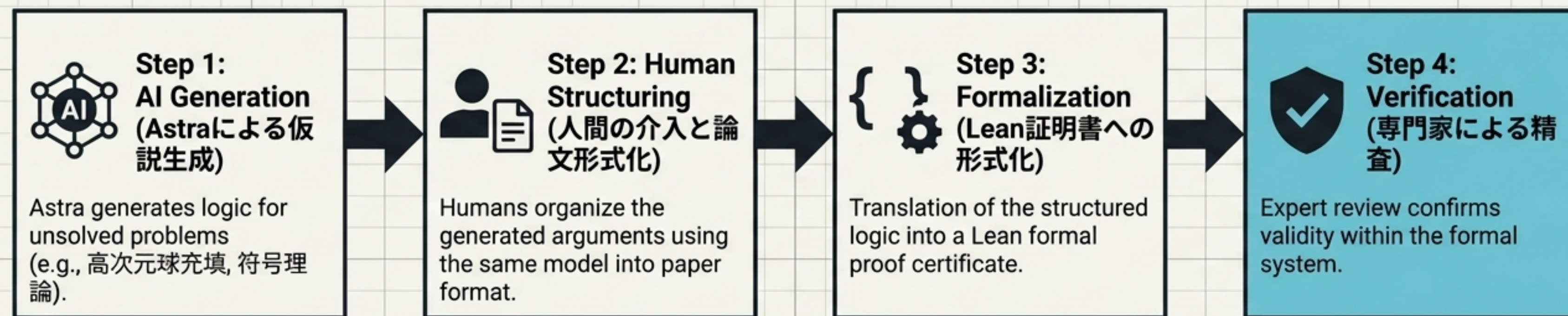
2026-08-01

Astra 公式明示
(社内版による数学・理論CSの10成果)

Future / TBD

一般提供、API、価格、System Cardすべて「将来・時期未公開」

数学研究における「生成と検証」のパイプライン



完全な自律型エージェントではなく、「生成—形式化—機械検証—専門家レビュー」のサイクルを前提とした仮説生成器である。

欠落している製品仕様

System Status

提供形態 (Deployment): API / ChatGPT

【未公開 / 利用不可】

モデル規模 (Scale): パラメータ数 / アーキテクチャ

【未公開】

入力制限 (Context Length): トークン数 / 対応モダリティ

【未公開】

運用コスト (Cost): API価格 / レート制限

【未公開】 (2,000ドルは換算値)

パフォーマンス (Performance): レイテンシ / Tokens per second

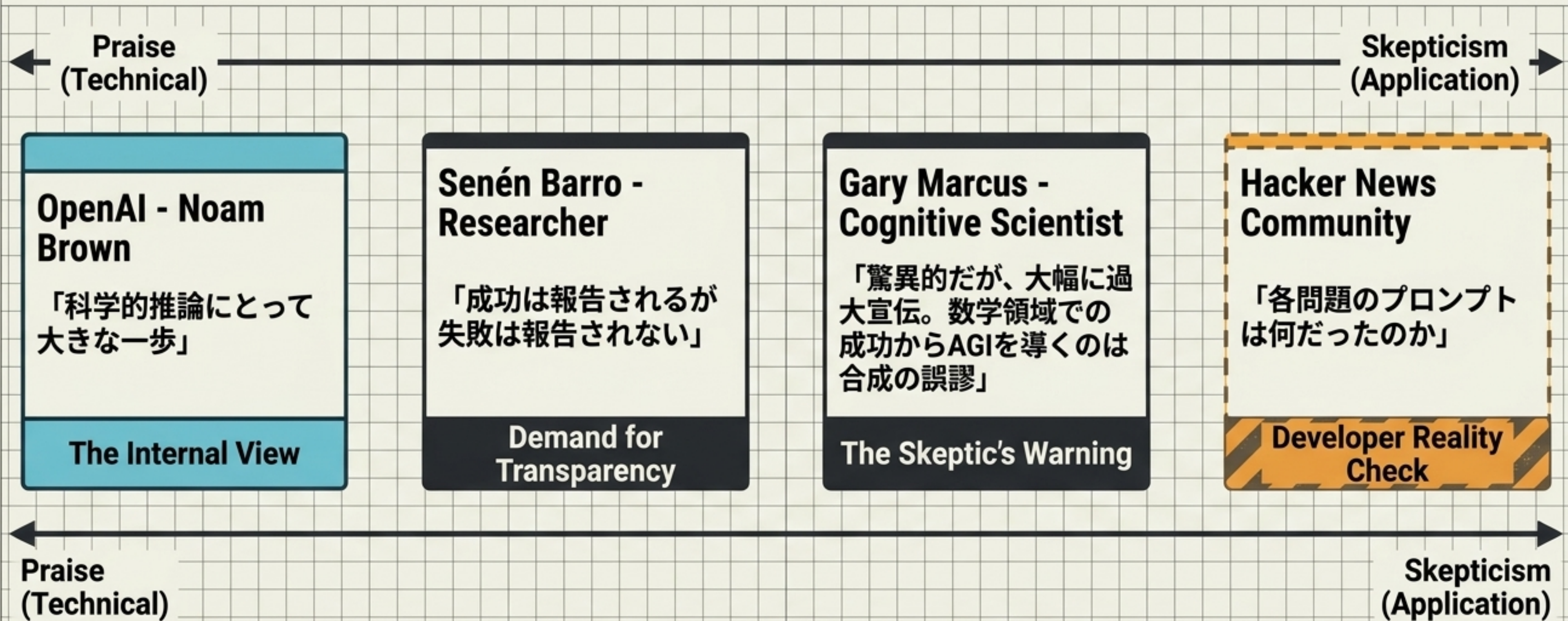
【未公開】

安全性保証 (Safety): System Card / 学習データ

【未公開】

製品導入に必要な可用性、費用、速度、安全性の
基盤データが完全に欠如している。

Global Reactions: The Astra Spectrum



Stakeholder Mapping: Technical Praise vs. Application Skepticism

The Frontier Model Diagnostic Matrix

	OpenAI Astra	GPT-5.6 Sol	Claude Fable 5	Gemini 3.1 Pro Preview
性能根拠 (Performance Proof)	未解決数学問題10件 (標準ベンチマークなし)	GPQA 94.6%, SWE-Bench Pro 64.6%	—	—
価格 (Price per 1M In/Out)	[未公開]	\$5 / \$30	\$10 / \$50	
アーキテクチャ (Architecture)	[未公開]	1.05M コンテキスト, 最大128K 出力		
安全性 (Safety/System Card)	[完全未公開] (RLHF, フィルター等不明)	多層保護、高度なサイバー保護等実装済み		

Astra Threat Mapping: Critical Risks & Defense Protocols

Threat Mapping Diagram

幻覚・誤情報 (Hallucination & Misinformation)

- **Risk:** 数学以外の一般知識や実務領域での事実性評価が未公開。
- **Defense:** RAG、引用検証、回答拒否基準の必須化。

プロンプトインジェクション (Prompt Injection)

- **Risk:** Astra固有の防御評価・System Cardが存在しない。
- **Defense:** サンドボックス化、外部データと命令の分離。

長時間エージェントの逸脱 (Agent Drift)

- Risk:** 自律性と停止機構の仕様が不明。
- **Defense:** 予算上限、段階的承認、実行前レビューの徹底。

ベンダーロックイン (Vendor Lock-in)

- Risk:** API形式や価格体系が未知数。
- **Defense:** 複数モデル対応のAPI抽象化層の構築。

Astra Compliance Analysis: Strict Japanese Legal Frameworks

Regulatory Pillars & Non-Compliance Status

個人情報保護委員会 (APPI Compliance)

- **Rule:** 学習利用の有無、越境移転、オプトアウト要件の確認。

Astra Status: Zero Data Retention (ZDR) や保持期間が完全に未公開。導入不可。

文化庁・著作権 (Agency for Cultural Affairs)

- **Rule:** 既存著作物との「類似性」と「依拠性」による侵害リスク。

Astra Status: 学習データや類似出力率が未検証のため、出力物の来歴保証が困難。

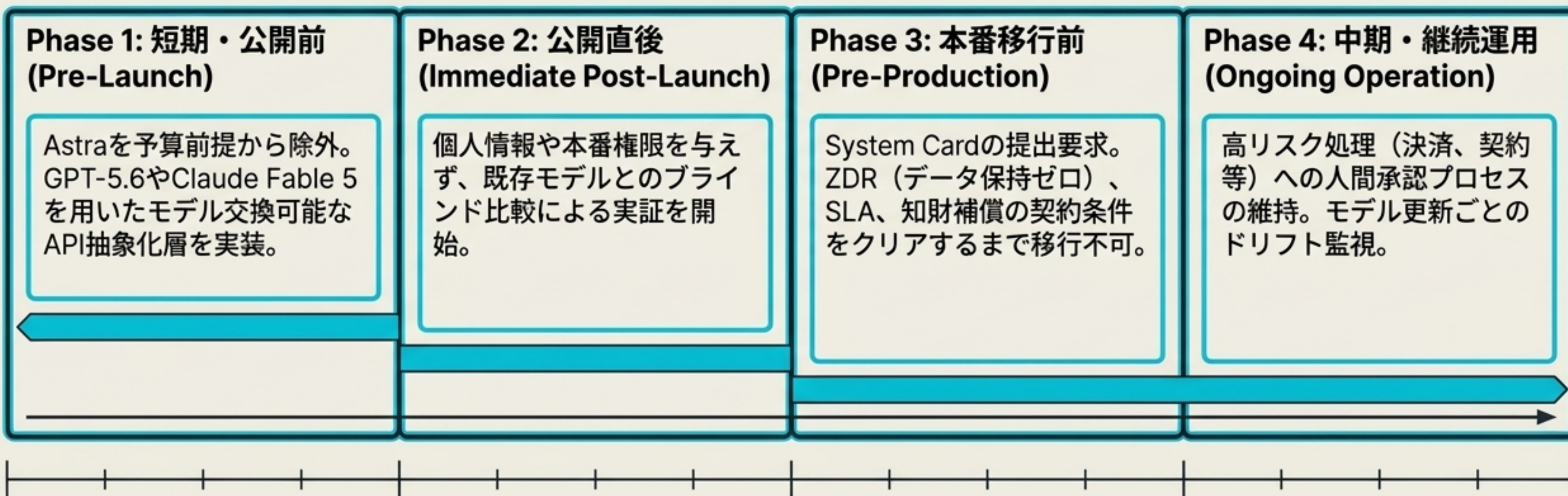
経産省・総務省ガイドライン (METI/MIC)

- **Rule:** 安全性、透明性、アカウントビリティを含む継続的AIガバナンス。

Astra Status: 監査可能性やインシデント報告手順などの組織的統制要件を現時点では満たせない。

Astra Strategic Playbook: Phased Implementation & Risk Mitigation Timeline

Gantt-Style Roadmap for Astra Integration & Control



**汎用的な自律型
エージェント
(Autonomous AGI
Worker)**

「何でも任せられる魔法の
社員」という幻想。

**高度な仮説生成器
(Advanced Hypothesis
Generator)**

人間のキュレーションと形
式検証 (Lean等) を組み合
わせて初めて価値を生む強
力な探索ツール。

「研究成果を買うことはできない。製品がリリースされ、SLAと安全性が実証されるまで待て。」