Gemini 4 Hydrogen 調査レポート

― 実在・内容・評価・評判 ―

調査時点:2026 年 11 月 13 日(日本時間)

Claude Fable 5.2

要旨

「Gemini 4 Hydrogen」は実在する。リークや誤報ではない。 Google が 2026 年 11 月 12 日(米国時間、日本時間 11 月 13 日未明)に公式発表し、同日から提供している Gemini 4 ファミリーの高速・実行用モデルである(モデル ID は gemini-4-hydrogen)[1][2][3]。実質的には Gemini 3.8 Flash の後継にあたる。

- **実在・公式:** Google 公式ブログ(Koray Kavukcuoglu 名義)、Google DeepMind のモデルカード、Google Cloud ドキュメントのいずれでも確認できる[1][2][3]。「Hydrogen」は発表前には社内コードネームとしてリークされていたが、そのまま正式な製品名になった[4]。ファミリーは Carbon(最上位)、Argon、Hydrogen(高速・軽量)の 3 モデルである。
- **評価:** Artificial Analysis Intelligence Index の独立計測では、既定の high で 51、xhigh で 54。上位モデルの Argon(50)を上回り、Claude Sonnet 5.5 と GPT-6.1 Sol(各 55)に 1 点差まで迫った[5]。一方、GPT-6.1 Terra は同じ 51 点をより低い 1 タスクあたりコスト(0.96 ドル)で出している[5]。
- **評判:** 速度と実務での使い勝手は概ね好評である。批判は、紛らわしい命名、3.8 Flash の現行価格の 2 倍という単価、エージェント作業での「やり過ぎ」の 3 点に集まっている[6][7]。
- **留意点:** Legal Research Bench で「同率 1 位」という Google の主張は、Vals.ai の公開データでは 6 位(53.85%)であり食い違う[8]。法務・知財の用途では自社データでの検証が欠かせない。

本書では、「一次ソースで確認できた事実」「二次情報・報道」「分析・推測」を区別して記述する。第 5 章は分析・推測である。

1. 前提の真偽検証(一次ソースで確認)

1.1 公式発表の確認

- **公式ブログ:** 「Introducing Gemini 4 Hydrogen」。著者は Koray Kavukcuoglu(Google DeepMind SVP 兼 Google Chief AI Architect)、公開日時は 2026 年 11 月 12 日 18:00 UTC。米国太平洋時間では 11 月 12 日 10:00、日本時間では 11 月 13 日 03:00 にあたる[1]。開発者向けブログと Google 検索向けブログも同時に公開された[9][10]。
- **モデルカード:** Google DeepMind のモデルカード一覧に「Gemini 4 Hydrogen — Updated 12 November 2026」として掲載され、PDF 版も公開されている[11][2]。
- **クラウドドキュメント:** Google Cloud(Gemini Enterprise Agent Platform、旧 Vertex AI 系)に gemini-4-hydrogen のモデルページと開発者ガイド(移行ガイド)があり、一般提供(GA)として扱われている[3][12]。
- **公式 X:** News from Google、Google DeepMind、Google AI Studio の各アカウントが発表を投稿している[13]。

1.2 名称の経緯(二次情報)

- 10 月 27 日、フォロワー約 1,500 人で一次資料を出した実績のない「White」という X アカウントが、「Gemini 4 Flash が Hydrogen というコードネームで社内に現れた」と投稿したのが始まりとされる[4]。
- 11 月 4 日頃から、テスターが「AI Studio と LMArena(現 Arena)で A/B テスト中」「5 つのチェックポイントがある」と報告した。Arena 上では「gemini-3.8-flash」という表示名で出ていたとされる[4]。
- 11 月 11 日、Business Insider が、社内開発プラットフォーム「Jetski」で社員向けテストが行われていると報じた[14]。
- その翌日、コードネームがそのまま製品名として発表された[1]。

1.3 表記揺れと注意点

- 「Gemini 4 Flash」という正式名称は存在しない。Google は今回、Flash の名称を使っていない。「Gemini 4.0 Hydrogen」も公式表記ではない。

- 無関係の同名モデルに注意が必要である。Deep Principle 社のオープンウェイト科学モデル「Hydrogen-1.0」は別物である。
- 発表前の解説記事は「未発表・非公式」と書いており、現在は古い情報である[15]。

2. 内容:位置づけと仕様

2.1 主要仕様

項目	内容
位置づけ	日常のコーディングとエージェント作業向けの「daily driver」。Carbon と Argon は深い長期推論を担うフロンティアモデル[1]
出自	「Gemini 4 Carbon から適応(adapted from)」。アーキテクチャや学習データの詳細は Carbon のモデルカード参照とされ、パラメータ数などは非公開[2]
入力	テキスト、画像、音声、動画(Google Cloud では PDF や Office 文書も直接扱える)[2][3]
コンテキスト	入出力合算で最大 1M(1,048,576)トークン。同期出力は既定で最大 64K(65,536)、バックグラウンドの長文生成では最大 1M[2]
思考レベル	low / medium / high(既定)/ xhigh(新設)の 4 段階[1][3]
知識カットオフ	2026 年 3 月[2]
価格(API 標準)	100 万トークンあたり入力 1.50 ドル、出力 7.50 ドル(思考トークンを含む)。キャッシュ入力 0.075 ドル。1M 全域で単一価格[1][16]
無料枠	Gemini API と AI Studio にレート制限付きの無料枠がある。Gemini 4 で無料枠があるのは Hydrogen が初めて(Fello AI が Google 料金ページを引用)[17]。無料枠のデータは製品改善に使われる
提供面	Gemini API、AI Studio、Android Studio、Gemini for business、Gemini アプリ、Google 検索(全ユーザー・全世界)、Google Sheets、Google Antigravity(有料の Pro/Ultra)[1][2]
無料アプリユーザー	自分では選べない。自動モデル選択で「適切な場合」に使われる。大半のプロンプトは Flash-Lite で処理される[1][18]

なお、無料枠では Google 検索・Google マップのグラウンディングが対象外とする二次情報(OmniaKey)もあるが、当該記事の URL は本調査では確認できていない。

2.2 Gemini 3.8 Flash からの主な変更点(Google の主張)

- 既定設定でタスク完了が 2.5 倍速い[1]。
- Terminal-Bench 4.1 が 18.2%から 57%に向上[1]。
- 事実に反する主張を 5 分の 1 未満に削減[1]。
- ファイル読み込みと検索を並列に実行し、ツール呼び出しが失敗したときの回復力が向上[1]。

2.3 移行時の注意点

Google Cloud の開発者ガイドは、以下の変更点を挙げている[12]。

- 既定の思考レベルが medium から high に変わる。
- thinking_budget と MINIMAL を指定すると HTTP 400 エラーになる。
- temperature、top_p、top_k は無視される。
- 明示的キャッシュはローンチ時点では使えない。暗黙キャッシュは 4,096 トークンから効く。
- モデル名を差し替えるだけでは、コストや挙動が変わる可能性がある。

2.4 Orchestrator(Antigravity)

Carbon と Hydrogen を同じセッション内で動的に切り替える機能である。Google によれば、Terminal-Bench 4.0 で Carbon に近い成功率を出しつつ、クォータ消費を 24%減らし、19%速く完了した[1]。提供はまず Ultra が中心で、Pro 向けは限定的という報道がある(出典媒体は本調査で特定できていない)。

2.5 安全性

- 自動評価では、3.8 Flash と比べて画像からテキストへの安全性が+7.00pp 改善した。一方で不当な拒否は+5.57pp と悪化しており、特にマルチモーダルなプロンプトで目立つ[2]。
- Frontier Safety Framework の評価では、CBRN(生物)とサイバーの両領域で、重大な能力の閾値(CCL)には「達していない」ものの、その手前の「警戒閾値(alert threshold)には到達」と明記されている[2]。
- 詳細な Safety & Security レポートは「近日公開」とされ、11 月 13 日時点では未公開である[2]。

- 公式ブログ側の主張(第三者による検証はない):エージェント作業でのやり過ぎを最大 81% 削減、社内の正直さ評価で 96.7%、Gray Swan の間接プロンプトインジェクション評価で 2.3 倍超の耐性[1]。

3. 性能評価:公式主張と独立検証の区別

3.1 Google の公表値(自社選定の比較)

以下は Google が公表した数値であり、比較対象も Google が選んでいる[1]。ベンダーにとっての最良ケースと見るべきである。

ベンチマーク	Hydrogen	Claude Sonnet 5.5	GPT-6.1 Terra
SWE-sweep	13.5%	3.1%	5.8%
Terminal-Bench 4.1	57.0%	**63.9%**	未報告
Agents' Last Exam v1.1	71.6%	60.4%	69.8%
AutomationBench-AA	76.1%	71.8%	60.2%
OSWorld-2.1(オフライン)	77.6%	**83.9%**	74.2%
MKA	53.4%	52.0%	43.8%
HLE-Diamond	58.8%	53.3%	47.3%

- Google の表(全 10 項目)では、Hydrogen が 7 勝 3 敗である。3 敗はいずれも Sonnet 5.5 が相手で、ターミナル操作とコンピュータ操作の項目である[1]。
- MKA は Google が同時に発表した自社ベンチマークである。

3.2 独立評価(第三者による計測)

評価	結果	解釈
Artificial Analysis Intelligence Index v5.0.1[5]	low 27 / medium 41 / high 51 / xhigh 54。1 タスクあたりコストは 0.29 / 0.80 / 1.87 / 3.16 ドル	xhigh は Sonnet 5.5(max)と GPT-6.1 Sol(max)の各 55 に 1 点差。high は Argon(50)、Terra(51)と同水準
AA:冗長性[5]	xhigh で指数計測に 3 億 400 万出力トークン(中央値は 7,280 万)	非常に冗長。xhigh は費用効率が悪い

評価	結果	解釈
AA:出力速度[5][19]	公式 X は約 200 トークン/秒。個別ページでは xhigh 188、high 236 トークン/秒	Flash 系として高速。引用元により値に幅がある
AA:Terminal-Bench 4.0[5]	54%(high・xhigh とも)	Sonnet 5.5(max)の 64%、Sol(max)の 56%に劣る
AA:AA-Omniscience[5]	作り話の率は約 8%	幻覚の少なさが強み
Arena Agent Arena[20]	総合 5 位、6,500 セッションで +14.8%。上位 10 モデル中最速(中央値 5.3 分/タスク)、中央値 1.6 ドル/タスク	実利用でも速度とコストの優位が見られる
Arena WebDev[20]	10 位(1704)。3.8 Flash は 43 位(1583)	Web 開発では大幅に改善
Arena Text[20]	8 位(1502±9)。3.8 Flash は 10 位(1500±4)	純粋なチャットの人間評価では 3.8 Flash とほぼ同等
Vals.ai Tax Agent Bench[21]	52.30%、85 モデル中 1 位(Carbon は 51.78%)	Google の主張どおり。ただし差は誤差(±3.60)の範囲内
Vals.ai Legal Research Bench[8]	**53.85%、94 モデル中 6 位**(1 位は Carbon の 55.77%)	**Google の「55.8%で同率 1 位」と食い違う**
Vals Index[21]	68.49%、37 モデル中 6 位	上位グループの一角

未計測: METR、Epoch AI、ARC Prize は、11 月 13 日時点で Hydrogen 固有の結果を公表していない。SWE-bench(Verified 等)の独立計測も確認できなかった。

3.3 ソース間の食い違い

1. **Legal Research Bench:** Google は 55.8%で同率 1 位、Vals の公開リーダーボードは 53.85%で 6 位である[1][8]。Google の数値は Carbon の 55.77%とほぼ一致する。ただし上位 6 モデルは約 2 ポイント以内に収まっており、誤差の範囲内でもある。
2. **Carbon の AA 指数:** The Decoder が引用する表では 59、eesel.ai などの記事では 60(high)である[22][23]。指数のバージョン違いの可能性がある(推測)。
3. **出力上限:** 公式ブログは「1M 入力+最大 1M 出力」、モデルカードは「入出力合算で 1M、同期出力は 64K」としている[1][2]。実装上はモデルカードの記述がより正確と見られる(推測)。

4. **xhigh のトークン量:** eesel.ai のまとめでは xhigh は high のほぼ 3 倍、AA 公式 X では「ほぼ 2 倍」とされ、一致しない[23][19]。

4. 評判

4.1 肯定的な評価

- **Artificial Analysis(公式 X):** high 設定で「Argon 並みの知能を約 19%のコストで」実現したと評価している。high・medium・low はいずれも時間対知能のパレート最適線上にある[19]。
- **Arena:** steerability(指示への追従しやすさ)は 2 位、Praise vs Complaint(称賛と苦情の比率)は 2 位である[20]。
- **Box CEO Aaron Levie:** 同社の評価で 3.8 Flash の 74%から 79%に向上し、所要時間は半分未満になったと報告している(早期アクセスでの結果)[16]。
- **Hacker News:** ローンチスレッドは初日に 380 ポイントを超えた[23]。自社評価を走らせた利用者からは、1 タスクあたりコストは 37%高くなったが所要時間は 40〜60%になったとの報告や、「コード実装のサブエージェント役として優秀」という声があった[6]。
- **日本語圏:** 窓の杜は「Claude Sonnet 5.5、GPT-6.1 Terra のライバル」と位置づけている[24]。GIGAZINE、Impress Watch、ケータイ Watch などが速報した[25](ITmedia も報じたとされるが URL は未確認)。note や Qiita の試用記事では、速度・価格・賢さのバランスや 3.8 Flash からの完成度向上を評価する好意的な内容が目立つ[26]。X 上には日本語能力を高く評価する個人の感想もあるが、あくまで主観である。

4.2 批判・懸念

- **命名の分かりにくさ:** Hacker News で最も多かった論点である。Hydrogen・Argon・Carbon は名前から序列が読み取れず、Haiku/Sonnet/Opus のほうが分かりやすいという意見が見られた[6]。
- **価格:** Simon Willison は、Gemini 4 の価格設定は分かりにくいと指摘している。現時点では Carbon と Argon に対して 25%引きにすぎず(12 月 1 日以降は 75%引きになる)、Claude Haiku 5.5 と GPT-6 Luna(0.10/0.50 ドル)の 15 倍である[7]。3.8 Flash の導入価格(0.75/3.75 ドル)を使っていたユーザーからは、コストが倍になるなら他社の軽量モデルに移るという声も出ている[6]。

- **品質のばらつき:** Willison の定番テスト(自転車に乗るペリカンの SVG)では、Hydrogen の出力は明らかに劣ると評された[7]。Hacker News では、ベンチマーク改善と引き換えに、まれに起きる最悪ケースでのやり過ぎ・作業過多が悪化したという批判もあった[6]。いずれも少数の事例であり、Google 側の主張(やり過ぎを 81%削減)とは食い違う。
- **コーディングの深さ:** The Decoder は、Argon をわずか 4 週間で陳腐化させたと評価する一方、エージェント型コーディングではなお不足があると指摘している[22]。

5. 分析と推測

本章は、上記の事実に基づく分析・推測であり、確認済みの事実ではない。

5.1 競争への影響

- Google は、Gemini 4 世代の水準のモデルを無料の Google 検索と Gemini アプリに一気に投入した。到達ユーザー数の面で他社に対する優位が大きいと考えられる。
- API の価格帯で見ると、Hydrogen は「安価な軽量モデル」(Haiku 5.5、GPT-6 Luna 等。0.10/0.50 ドル)と「中位モデル」(Sonnet 5.5、GPT-6.1 Sol。2/10 ドル)の間の新しい層に位置する。超低価格帯の大量処理(分類やルーティング)には向かず、中位モデルの代替として使うのが合理的である。
- Argon は Hydrogen の登場によって性能でも価格でも上回られ、存在意義が薄れたと見られる。The Decoder も同様の見方を示している[22]。

5.2 企業利用への含意

- 適するのは、高速なエージェント実行、大量文書(PDF・表・動画)の要約や構造化、Workspace 連携である。最上位モデルに計画を任せ、Hydrogen に実行と検証を任せる「分業型」が最も費用対効果が高いと考えられる。
- 既定の high で十分である。xhigh はコストが 69%増えるのに対し、指数の改善は 3 点にとどまる[23]。
- 2027 年 1 月 1 日に 3.8 Flash の単価が Hydrogen と同額に上がる。3.8 Flash を利用中の場合は、それまでに本番ワークロードで Hydrogen の 1 タスクあたりコストを実測することが望ましい。

5.3 知財実務への含意

- 1M トークンのコンテキストと多数文献の並列読み込みは、先行技術文献の一括読解、クレームチャートの下書き、図面を含む公報の要約に有利である。幻覚の少なさ(AA-Omniscience で約 8%)もプラスに働く[5]。
- ただし、Hydrogen を特許・知財の用途で直接評価した公開データは確認できなかった。Vals Legal Research Bench は米国法の 8 分野を対象としており、特許・知財分野を含まない。しかも Hydrogen は同ベンチで 6 位である[8]。
- 知識カットオフは 2026 年 3 月である[2]。最新の公報や審査基準を扱うには、検索グラウンディングや RAG が前提になる。
- 機密の発明情報を扱う場合、無料枠(データが製品改善に使われる)は避け、有料枠または Gemini for business(Google Cloud 規約)を使うべきである。
- 過去案件で正解が分かっている先行技術調査とクレーム対比を用いて、Hydrogen(high)、Carbon、他社の中位モデルを比較検証することが望ましい。明細書作成支援は、人が最終確認する前提の下書き用途に限定すべきである。

6. 推奨事項

5. **3.8 Flash からの移行は前提とし、まず medium/high で検証する。** パラメータ変更(thinking_level、キャッシュ、temperature)を確認したうえで、コストと時間を実測する。
6. **難しい判断は Carbon、実行は Hydrogen** という分業設計を検討する。
7. **コスト最優先のバッチ処理**では、GPT-6.1 Terra(AA 指数で同点、1 タスク 0.96 ドル)や超低価格モデルも比較対象にする。
8. **法務・知財用途では Google のベンチマーク主張をそのまま採用しない。** 自社データで評価し、無料枠に機密データを入れない。
9. **Safety & Security レポートの公開と、METR・Epoch AI・ARC Prize による独立評価を追跡する。**

7. 本調査の限界

- 発表から約 1 日の時点での調査である。ベンチマーク、価格(Batch・Flex の適用、無料枠のレート制限値)、提供範囲は今後変わる可能性がある。無料枠のレート制限値や Batch・Flex の価格は公表が確認できていない。
- 公式ブログに掲載されたベンチマークの多くは自社選定の比較で、第三者による検証が済んでいるのは一部(Artificial Analysis、Arena、Vals)だけである。
- Arena のリーダーボードのスナップショットは日付表示が 11 月 11 日となっており、発表前のテスト段階の数値を含む可能性がある。
- Hacker News、X、note などの評判は少数の初期ユーザーによるもので、代表性は限られる。プロモーション目的のブログも多く、数値の引用元には注意が必要である。

参考文献

URL はいずれも 2026 年 11 月 13 日時点の調査で取得したものである。本文と文献の対応は、調査レポートが本文中で明示した情報源に基づいて付した。

[1] Google 公式ブログ「Introducing Gemini 4 Hydrogen」(Koray Kavukcuoglu、2026 年 11 月 12 日)
https://blog.google/innovation-and-ai/models-and-research/gemini-models/introducing-gemini-4-hydrogen/

[2] Google DeepMind「Gemini 4 Hydrogen Model Card」(HTML 版/PDF 版、2026 年 11 月 12 日更新)
https://deepmind.google/models/model-cards/gemini-4-hydrogen/
https://storage.googleapis.com/deepmind-media/Model-Cards/Gemini-4-Hydrogen-Model-Card.pdf

[3] Google Cloud ドキュメント「Gemini 4 Hydrogen」モデルページ(Gemini Enterprise Agent Platform)
https://docs.cloud.google.com/gemini-enterprise-agent-platform/models/gemini/4-hydrogen

[4] orcarouter.ai ブログ(Gemini 4 Flash/Hydrogen のリーク経緯の整理)
https://www.orcarouter.ai/blog/gemini-4-flash-hydrogen-leak

[5] Artificial Analysis「Gemini 4 Hydrogen」モデルページ
https://artificialanalysis.ai/models/gemini-4-hydrogen

[6] Hacker News ローンチスレッドほか
https://news.ycombinator.com/item?id=50471487
https://news.ycombinator.com/item?id=50471431

[7] Simon Willison's Weblog(2026 年 11 月 12 日)および X 投稿
https://simonwillison.net/2026/Nov/12/llm-gemini/
https://x.com/simonw/status/2121039423441981444

[8] Vals.ai「Legal Research Bench」リーダーボード
https://www.vals.ai/benchmarks/legal_research

[9] Google 公式ブログ(開発者向け)「Build with Gemini 4 Hydrogen」
https://blog.google/innovation-and-ai/technology/developers-tools/build-with-gemini-4-hydrogen/

[10] Google 公式ブログ(Google 検索向け)Gemini 4 Hydrogen による検索アップグレード
https://blog.google/products-and-platforms/products/search/gemini-4-hydrogen-search-upgrade/

[11] Google DeepMind「Model Cards」一覧
https://deepmind.google/models/model-cards/

[12] Google Cloud ドキュメント「Gemini 4 Hydrogen」開発者ガイド(移行ガイド)
https://docs.cloud.google.com/gemini-enterprise-agent-platform/models/guides/gemini-4-hydrogen

[13] X 公式アカウント投稿(News from Google、Google AI Studio)
https://x.com/NewsFromGoogle/status/2120943020803801198
https://x.com/GoogleAIStudio/status/2120976872729174329

[14] Business Insider 報道の転載・要約(MLQ.ai、Investing.com)
https://mlq.ai/news/google-is-testing-a-fast-new-gemini-model-called-hydrogen-with-staff-business-insider-reports/
https://www.investing.com/news/stock-market-news/google-tests-new-gemini-4-hydrogen-model-for-speed-report-says-93CH-5019557

[15] 発表前の解説記事(kie.ai、explainx.ai、ArgoLink)※発表前の内容であり現在は古い
https://kie.ai/blog/what-is-gemini-4-hydrogen
https://explainx.ai/blog/gemini-4-hydrogen-flash-internal-test-claimed-vs-verified-november-2026
https://argolink.io/en/blog/what-is-gemini-4-hydrogen

[16] VentureBeat「Google launches Gemini 4 Hydrogen: a faster, smarter, pricier Flash successor」
https://venturebeat.com/technology/google-launches-gemini-4-hydrogen-a-faster-smarter-pricier-flash-successor

[17] Fello AI「Gemini 4 Hydrogen」解説記事
https://felloai.com/gemini-4-hydrogen/

[18] Notebookcheck「Gemini 4 Hydrogen: What free users and Google AI Plus get」
https://www.notebookcheck.net/Gemini-4-Hydrogen-What-free-users-and-Google-AI-Plus-get.1449493.0.html

[19] Artificial Analysis 公式 X 投稿
https://x.com/ArtificialAnlys/status/2120941511223185722

[20] Arena(旧 LMArena)リーダーボード(Agent/WebDev/Text)および公式 X 投稿
https://arena.ai/leaderboard/agent/chat

https://arena.ai/leaderboard/code/webdev
https://arena.ai/leaderboard/chat/text?scoring=raw
https://x.com/arena/status/2120940869889175840

[21] Vals.ai「Gemini 4 Hydrogen」モデルページ
https://www.vals.ai/models/google_gemini-4-hydrogen

[22] The Decoder「Google's Gemini 4 Hydrogen outscores the bigger Argon just four weeks after its launch」
https://the-decoder.com/googles-gemini-4-hydrogen-outscores-the-bigger-argon-just-four-weeks-after-its-launch/

[23] eesel.ai「Gemini 4 Hydrogen review」
https://www.eesel.ai/blog/gemini-4-hydrogen-review

[24] 窓の杜(Impress)Gemini 4 Hydrogen 記事
https://forest.watch.impress.co.jp/docs/news/2156117.html

[25] 日本語速報(GIGAZINE、Impress Watch、ケータイ Watch)
https://gigazine.net/news/20261113-gemini-4-hydrogen/
https://www.watch.impress.co.jp/docs/news/2156136.html
https://k-tai.watch.impress.co.jp/docs/news/2156225.html

[26] 日本語の試用記事(note、Qiita)
https://note.com/masa_wunder/n/nf56ecfd6826e
https://note.com/humble_bobcat51/n/nd6d5ac8fb54b?hl=en
https://qiita.com/kentaro_kawamura/items/2d1a2f8196d0b89c0cbe
https://qiita.com/AIHubMix/items/140251a8b323c8e78d97