Google「Gemini 3.8 Flash」「3.8 Flash Cyber」発表の深掘り調査

―発表内容と評価・評判―

調査日：2026 年 9 月 3 日

Claude Fable 5.1

対象：Google 公式ブログ記事（2026 年 9 月 2 日公開）

1. 要旨

- **記事は実在する。**対象 URL は Google 公式ブログ（The Keyword）上の本物の記事で、2026 年 9 月 2 日公開、執筆者は Tulsee Doshi（Senior Director, Product Management）および Raluca Ada Popa（Gemini Security Lead, Google DeepMind）である [1)]。汎用モデル「Gemini 3.8 Flash」と、防御者限定のサイバー特化版「Gemini 3.8 Flash Cyber」の 2 種が同時に発表された [1)]。
- **3.8 Flash は「安価で速いのに賢い」ワークホースとして概ね高評価。**独立系 Artificial Analysis の Intelligence Index で 59（3.7 Flash 比+3、GPT-5.6 Sol xhigh・Grok 4.6 medium と同等）を記録し、知能対タスク単価の Pareto 最前線に位置する [5,6)]。一方で「働きすぎ」の設計により出力トークンが 1 タスク平均 48k へ約 30%増え、タスク単価は$0.40 から$0.58 へ約 40%上昇した [5)]。
- **3.8 Flash Cyber は脆弱性発見・自動パッチに特化し、新設「Fairwind Program」を通じて政府・重要インフラ・主要ソフトウェア保守者にのみ提供される。**攻撃（exploit）より防御（修正）を優先する設計である [1,3)]。同週に OpenAI「Astra」が Preparedness Framework 上の「Critical」サイバー閾値到達を公表し、Anthropic も新モデルを公表するなど、サイバー AI の競争と「アクセス思想の対立」が一斉に表面化した [9)]。

2. 前提の真偽検証

記事は実在する。公式ページのメタデータから公開日 2026 年 9 月 2 日、著者 Tulsee Doshi およ

び Raluca Ada Popa を確認した[1]。併せて Google DeepMind が「Fairwind Program」の専用ページ[3]と「Gemini 3.8 Flash」のモデルカード[2]を同日に公開しており、The Hacker News[9]、9to5Google[8]、The Register[7]、Artificial Analysis[5]など多数の二次ソースが同日に報道している。記事の真正性は堅い。

3. Gemini 3.8 Flash の発表内容

3.1 位置づけ

3.7 Flash（約 3 週間前公開）の後継であり、「6 週間で 3 度目の Flash リリース」に当たる[8,19]。Google は本モデルを「最も賢いワークホースモデル」「これまでで最良の推論・コーディングモデル」と表現している[1]。基盤は 3.7 Flash で、3.8 Flash Cyber も同じコアを共有する「1 つのコア、2 つのアクセス層」の構成である[1,11]。

3.2 仕様・価格・提供形態

以下は公式ブログおよびモデルカードで確認できた一次情報である。

項目	内容	出典
コンテキスト長	入力 1M トークン／出力 64K トークン	2
入出力モダリティ	テキスト・画像・音声・動画入力／テキスト出力	2
推論（thinking）	effort レベル low／medium／high の 3 段階	1, 2
知識カットオフ	2026 年 3 月	2
価格（導入価格）	入力$0.75／出力$3.75（100 万トークン当たり）。2026 年 12 月 31 日まで	1, 8
価格（2027 年 1 月 1 日以降）	入力$1.50／出力$7.50 へ改定（倍増）。キャッシュ入力の 90%割引は維持	1, 8
提供形態	一般提供（Cyber 版と異なりアクセス制限なし）	1, 11

3.3 公式ベンチマーク（自己申告）

- DeepSWE v1.1（長期ソフトウェア工学タスク）で、多くの大型フロンティアモデルを低コストで上回ると主張[1]。
- Terminal-Bench 2.1 で 90.8%（3.7 Flash の 81.6%から向上）[1]。
- HLE-Verified で 54.9%[1]。

- Vals Finance Agent V2、Harvey's Legal Agent Benchmark で他のフロンティアモデルを上回ると主張（絶対スコアは非公表）[1,11]。
- **公式が明言する注意点：**3.8 Flash は「より働く（works harder）」設計で、追加の推論ステップや反復的なツール呼び出しを行うため、高 effort 時のトークン消費が増える。効率優先の用途には低 effort の使用または 3.7 Flash の継続利用を推奨している [1,17]。

4. Gemini 3.8 Flash Cyber

4.1 位置づけと用途

脆弱性検出と自動パッチに特化した「最も高性能なサイバーセキュリティモデル」で、2026 年 7 月公開の 3.5 Flash Cyber の後継である [1,15]。基盤は 3.8 Flash と共有し、コード修正エージェント「CodeMender」と組み合わせて利用する [1,3]。Google は「exploit のような攻撃能力より脆弱性修正を最初から優先した」と明記している [1]。

4.2 アクセス制限（Fairwind Program）

一般 API・消費者向けには提供されない [1,3]。新設の Fairwind Program を通じ、①政府・国家サイバー当局、②重要インフラ運用者（医療・通信・エネルギー・金融）、③中核技術プラットフォームに限定して提供される [3,4]。許可される用途は「認可された脅威シミュレーション、リバースエンジニアリング、防御・学術研究目的のマルウェア解析」などのデュアルユースタスクに限られ、マルウェア作成は禁止される [3]。応募組織には身元調査、フィッシング耐性 MFA、アクセス追跡が求められ、再販・共有は禁止。ゼロデータ保持（ZDR）に対応する [3]。Google は The Hacker News に対し、「世界で 650 超のパートナー（CrowdStrike、Datadog、Menlo Security、Palo Alto Networks、Snowflake を含む）」と協業していると説明した [9]。

4.3 サイバー系ベンチマーク

ベンチマーク	3.8 Flash Cyber	比較対象（公式掲載値）	出典
CyberGym（脆弱性発見）	86.2%	GPT-5.5-Cyber 85.6%、Mythos 5 83.8%、GPT-5.6 Sol 83.6%、3.5 Flash Cyber 77.5%	3
社内ベンチ（20 言語、発見成功率）	70%超	—	1

ベンチマーク	3.8 Flash Cyber	比較対象（公式掲載値）	出典
CWE-Bench（パッチ生成、pass@1）	47.2%	ブログでは「先行フロンティアモデル」47.8%と匿名表記。Fairwind ページ上では Fable 5（47.8%）と明示	1, 3

社内実績（自己申告）として、Chrome Security で「はるかに大型の最良商用モデルの 2.6 倍の正しいパッチ」、Wiz の内部ペネトレーションテストベンチマークで「リコール+7.5〜9.7%、コスト 2.3〜5.2 倍安」、Cloud Vulnerability Research で「通常数か月を要する重大な基礎的脆弱性を 2 時間未満で発見」が挙げられている [1,15]。

4.4 安全性

3.8 Flash は Frontier Safety Framework に基づき、CBRN（化学・生物・放射性・核）およびサイバー攻撃領域の悪用対策を実装している [1,2]。3.8 Flash Cyber は「より許容的（permissive）なサイバー緩和策」を持つため、防御者限定とされた [1,3]。プロンプトインジェクション耐性は Gray Swan ベンチで大幅に改善したと主張する [1]。一方、Hacker News ではモデルカードの安全性比較表に一部退行（regression）が赤で示されている点を批判する声もある（詳細は要確認）[13]。

5. 評価・評判

5.1 独立ベンチマーク（Artificial Analysis）

- Intelligence Index 59（high）。3.7 Flash（high 56）から+3。GPT-5.6 Sol（xhigh 59）、Grok 4.6（medium 59）と同等で、Opus 5 medium と同水準。medium 57、low 52[5,6]。
- 改善の主因はエージェント系評価（τ^3-Banking でツール利用が+12 点の 45%、Terminal-Bench v2.1、GDPval-AA v2）[5]。
- 知能対タスク単価の Pareto 最前線に位置し、タスク単価$0.58 は同水準知能で最安。ただし 3.7 Flash（$0.40）から約 40%上昇。トークン単価は不変だが、1 タスク平均出力が 48k へ 30%増加し、エージェント評価でのターン数も増えたことが原因 [5]。
- 「非常に冗長」で、評価スイート全体で 1.2 億トークンを生成（中央値 7,100 万）。出力速度は 304.6 トークン/秒と高速（同価格帯推論モデルの中央値 70.8 トークン/秒）[5]。

5.2 メディアの論調

総じて好意的だが、「進化ではなく洗練」との見立てが多い。The Register は「Google がレースに残っていることを再認識させる」と評しつつ、Gemini 3.5 Pro の度重なる延期[22]や Demis Hassabis 氏の DeepMind CEO 退任（会長へ）といった不透明感を背景として挙げた[7]。DataCamp は「ゲインは不均一で、コーディング・ツール利用は跳ねたが Humanity's Last Exam は 45.4%対 45.7%で横ばい」と指摘した[10]。MarkTechPost は「1 つのコア、2 つのアクセス層」と整理し、金融・法務ベンチは絶対スコアなしの相対勝利表記であり、CyberGym も絶対値非公表であると注記している[11]。Thurrott、Android Authority などは事実中心の報道である[16,17]。

日本語メディアでは、窓の杜（Yahoo!ニュース転載）[19]、Business Insider Japan（「最先端から脱落した王者の逆転戦略」）[18]などが報じた。ITmedia・日経の個別記事は本調査では未確認である。

5.3 開発者コミュニティの反応（Hacker News 等）

- 好意的な声：「DeepSWE で Opus 5 を上回る」「Flash なのに強力」「速度が異常に速い」「Claude の文体の癖がなく上質」[13]。
- 懐疑的な声：「benchmaxxing（ベンチマーク最適化）では」「Artificial Analysis 上では Opus 5 medium と同等で、Opus は同じ結果を 1/4 のトークンで出す」「Terminal-Bench 2 で 89.4%だが Tbench 4 では 19.1%（Opus 5 は 51.8%）」「実コードベースでは Sol/Claude に及ばず追加プロンプトが必要」「（コーディングエージェントの）Antigravity が無謀で勝手に commit/deploy する」[13]。
- Simon Willison 氏は llm-gemini 0.34 で即日対応し、定番の「ペリカン SVG」テストを実施した（コスト約 8.97 セント）[14]。

5.4 批判的・懐疑的な視点

- **ベンチマークの透明性：**StartupHub.ai は「Google は偽陽性率や、選定ベンチマーク以外での独立再現を開示していない。検証コストを見込むべきであり、効率優先の用途には 3.7 Flash を温存すべき」と指摘した[12]。
- **アクセス限定（ゲートキーピング）批判：**Fairwind を名指しした著名個人の批判は本調査時点で確認できなかった。ただし直接の先行事例である Anthropic Mythos／Project Glasswing（2026 年 4 月）を巡っては、Wharton Accountable AI Lab の Jonathan Iwry 氏が「説明責任を負わない一握りの民間主体の判断への依存」を、Future of Life Institute の

Hamza Chaudhry 氏が「ガバナンスギャップ」を、Bruce Schneier 氏が「アクセス制限は無益であり、公共 AI が必要」との立場をそれぞれ表明している [25,26]。RUSI の Aybars Tuncdogan 氏は、ベンダー依存が価格・安全規則・輸出規制・買収を国家安全保障問題に変える「新たな戦略的依存」となるリスクを警告した [27]。

- **「大手優遇」構造：**Fairwind は大手セキュリティベンダーとの協業を前提としており、独立研究者・バグバウンティハンター・中小組織が排除されうる構造である。ただしこれを名指しで批判した個人は本調査では確認できない。TechCrunch は一般論として「AI のガードレールが攻撃的セキュリティ研究者の仕事を妨げている」と 2026 年 7 月に報じている [21]。

6. 競合とサイバーAI 業界の同時展開

同週に Google（Fairwind＋3.8 Flash Cyber）、Anthropic（Claude Fable 5.1／Mythos 5.1、後者はトラステッドアクセス限定）、OpenAI（Astra）がサイバーAI 関連の発表を一斉に行った [9]。

陣営	モデル／プログラム	アクセス方針	出典
Google	Gemini 3.8 Flash Cyber／Fairwind Program	政府・重要インフラ・中核プラットフォームに限定。審査型	1, 3
Anthropic	Claude Fable 5.1／Mythos 5.1	Mythos 5.1 はトラステッドアクセス限定	9
OpenAI	Astra／Trusted Access for Cyber	検証ベースで比較的広く開放する方針。Astra 自体はより限定的なアクセス（Daybreak Blue）で提供予定	9, 20, 24

OpenAI は Astra が自社 Preparedness Framework 上で初めて「Critical」サイバー能力閾値に到達したと公表した [24,29]。The Hacker News によれば、ExploitBench で 100%（既知脆弱性からの exploit 開発）、脱獄要求の 91.5%を拒否（GPT-5.6 Sol の 59%から向上）、評価中に実在するゼロデイ 2 件を発見して exploit チェーンを構成したとされ、米政府による出荷前サイバーレビューを初めて受けたと主張している [9]。

アクセス思想の対立：OpenAI は「誰が自衛できるかを中央集権的に決めるのは適切でない」との立場から、検証ベースで広く開放する「Trusted Access for Cyber」を推進する [20,23]。対して Google の Fairwind と Anthropic は、防御者に先行優位を与えるための限定・審査型である [3,9]。両陣営の思想差は鮮明である。また、100 社超（Anthropic、Google、Microsoft、OpenAI 等）が AI 駆動サイバー攻撃への共同防御を求める書簡に署名した [9]。

7. 影響分析と示唆【分析・推論】

本章は一次・二次情報に基づく筆者の分析であり、確認済み事実とは区別される。

7.1 AI 業界の競争構図

- Google は「ほぼ月次のモデル更新」方針（Pichai CEO、Q2 決算）の下、3.6→3.7→3.8 Flash と高速リリースを継続している [7,8]。Flash 系での価格・速度・知能のバランスは明確な強みだが、フラッグシップ（Gemini 3.5 Pro）の遅延が続く限り、「最先端」の座は OpenAI／Anthropic に残るとの見方が支配的である [7,18,22]。
- 内部コードネーム「skimaki」、社内コーディング基盤「Jetski」での検証、社内エンジニアが Anthropic Opus より本モデルを選好したとの情報、「3.5 Pro はスキップし次は Gemini 4 Pro」との観測は、いずれも WSJ 報道やコミュニティ観測に基づく**未確定情報**である [13,18,28]。

7.2 サイバーAI の潮流における位置づけ

- 「防御者優先・限定アクセス」（Google／Anthropic）と「検証ベースの広い開放」（OpenAI）という二つのモデルが並存する局面に入った [3,9,20]。Astra の「Critical」到達は、フロンティアモデルの攻撃能力が制度的閾値を超えたことを各社が公式に認め始めた画期であり、政府による出荷前レビューが制度化に向かう可能性がある [9,24]。
- Google が攻撃能力ではなく「修正」を最初から優先した設計思想は、CWE-Bench（パッチ生成）で最上位に僅差で追随しつつコストで優位に立つという結果と整合的であり、防御側の運用コスト低減を訴求点としている [1,3]。

7.3 企業の知財・法務・R&D 部門への示唆

- **一般開発・R&D：**3.8 Flash はエージェント／コーディングのワークホースとして採用検討に値する。ただし①高 effort でのトークン増と冗長性、②2027 年 1 月の価格倍増、③実コードベースでは Claude/GPT 系に劣るとの実地報告を踏まえ、PoC で自社ワークロードのコスト／成功率を実測すべきである。効率優先の定常パイプラインは 3.7 Flash の温存が妥当と考えられる [1,5,13]。
- **セキュリティ部門：**3.8 Flash Cyber は該当組織（政府・重要インフラ・主要ソフト保守者）のみ Fairwind 申請を検討する。非該当組織は公開モデル＋CodeMender 等で代替可能であ

る。公式ベンチマークは自己申告色が強く偽陽性率が非開示のため、導入時は独立検証とトリアージ工数を必ず見込むべきである [3,12]。

- **知財・法務：**デュアルユース規制とサイバーAI のアクセス層化が進行している。①サイバーAI 利用契約におけるデュアルユース制約・ZDR 要件・再販禁止条項、②ベンダーロックイン／戦略的依存リスク、③自律エージェントの暴走リスク（勝手な commit／deploy の報告等）を統制対象として社内規程に明文化することが望ましい [3,9,13,27]。今後の変化の指標としては、OpenAI 型の検証ベースアクセスが業界標準化するか、政府の出荷前レビューが制度化するかを注視すべきである。

8. 留保事項（確認済み事実と推論の区別）

- **一次ソース確認済みの事実：**モデルの存在・公開日・著者・価格・仕様・公式ベンチマーク主張・Fairwind Program の条件 [1,2,3]。独立ベンチマークの数値は Artificial Analysis による [5,6]。
- **二次情報・伝聞・観測（未確定）：**内部コードネーム「skimaki」、Jetski 検証、社内での Opus 超えの選好、3.5 Pro スキップ／Gemini 4 説 [13,18,28]。
- **公式ベンチマークの限界：**金融・法務・CyberGym の多くは絶対スコア非公表または自己申告であり、偽陽性率・独立再現は非開示。公式ブログの一部比較は「先行フロンティアモデル」と匿名化されている（CWE-Bench は Fairwind ページ上で Fable 5 と明示）[1,3,11,12]。
- **本調査で未確認の事項：**ITmedia・日経・Impress Watch・GIGAZINE・Publickey 等の個別記事、Fairwind を名指しした著名専門家の批判、CISA 等政府機関の Flash Cyber に対する個別反応。
- **URL 未確認の出典：**Fortune（Iwry 氏発言）、Schneier 氏論考、RUSI（Tuncdogan 氏）、WSJ 報道、OpenAI「path-to-astra」ページは、本調査では二次的に要旨のみを確認したもので、URL は未検証である（参考文献 25〜29 に明記）。

参考文献

1) Google, "Introducing Gemini 3.8 Flash and 3.8 Flash Cyber"（著者：Tulsee Doshi, Raluca Ada Popa）, The Keyword, 2026 年 9 月 2 日. https://blog.google/innovation-and-ai/models-and-research/gemini-models/3-8-flash-and-3-8-flash-cyber/

2) Google DeepMind, "Gemini 3.8 Flash – Model Card", 2026 年 9 月. https://deepmind.google/models/model-cards/gemini-3-8-flash/

3) Google DeepMind, "Fairwind Program", 2026 年 9 月. https://deepmind.google/fairwind-program/

4) Google, "Google's Fairwind Program: Cyber defense tools for trusted partners", The Keyword, 2026 年 9 月. https://blog.google/innovation-and-ai/technology/safety-security/fairwind-program/

5) Artificial Analysis, "Google has released Gemini 3.8 Flash, its fourth Flash model in under four months", 2026 年 9 月 2 日. https://artificialanalysis.ai/articles/gemini-3-8-flash

6) Artificial Analysis (@ArtificialAnlys), X 投稿, 2026 年 9 月 2 日. https://x.com/ArtificialAnlys/status/2095177419489739063

7) The Register, "With Gemini 3.8 Flash, Google reminds everyone it's still in the race", 2026 年 9 月 2 日. https://www.theregister.com/ai-and-ml/2026/09/02/with-gemini-38-flash-google-reminds-everyone-its-still-in-the-race/5294049

8) 9to5Google, "Gemini 3.8 Flash rolling out three weeks after last release", 2026 年 9 月 2 日. https://9to5google.com/2026/09/02/gemini-3-8-flash-launch/

9) The Hacker News, "Google, Anthropic, and OpenAI Unveil Cyber AI Models, Safeguards, and Access Programs", 2026 年 9 月. https://thehackernews.com/2026/09/google-anthropic-and-openai-unveil.html

10) DataCamp, "Gemini 3.8 Flash: Features, Benchmarks, and Pricing", 2026 年 9 月. https://www.datacamp.com/blog/gemini-3-8-flash-cyber

11) MarkTechPost, "Google DeepMind Releases Gemini 3.8 Flash and Gemini 3.8 Flash Cyber: One Core Model, Two Access Envelopes", 2026 年 9 月 2 日. https://www.marktechpost.com/2026/09/02/google-deepmind-releases-gemini-3-8-flash-and-gemini-3-8-flash-cyber-one-core-model-two-access-envelopes/

12) StartupHub.ai, "Gemini 3.8 Flash Brings Cheap Reasoning to Cyber", 2026 年 9 月. https://www.startuphub.ai/ai-news/ai-research/2026/gemini-3-8-flash-brings-cheap-reasoning-to-cyber

13) Hacker News, "Gemini 3.8 Flash"（コメントスレッド）, 2026 年 9 月 2 日. https://news.ycombinator.com/item?id=49537553

14) Simon Willison, "Release: llm-gemini 0.34", 2026 年 9 月 2 日. https://simonwillison.net/2026/Sep/2/llm-gemini/

15) Cyber Security News, "Google Launches Gemini 3.8 Flash Cyber to Identify and Auto-Patch Security Vulnerabilities", 2026 年 9 月. https://cybersecuritynews.com/gemini-3-8-flash-cyber/

16) Thurrott, "Google Releases Gemini 3.8 Flash and Cyber Variant", 2026 年 9 月 2 日. https://www.thurrott.com/a-i/340992/google-releases-gemini-3-8-flash-and-cyber-variant

17) Android Authority, "Google's Gemini 3.8 Flash is built to 'work harder'", 2026 年 9 月. https://www.androidauthority.com/gemini-3-8-flash-google-ai-model-3706483/

18) Business Insider Japan, 「グーグル社員はすでに次期 AI『Gemini 3.8 Flash』をテスト中。最先端から脱落した王者の逆転戦略」, 2026 年 8 月. https://www.businessinsider.jp/article/2608-google-employees-testing-next-gemini-flash-3-8-model/

19) 窓の杜（Yahoo!ニュース転載）, 「Google、『Gemini 3.8 Flash』を発表 〜6 週間で 3 度目の『Flash』モデルリリース」, 2026 年 9 月. https://news.yahoo.co.jp/articles/0722c6c8d9b074c4c25d570c0bb40b8ca0fd665d

20) OpenAI, "Trusted access for the next era of cyber defense", 2026 年. https://openai.com/index/scaling-trusted-access-for-cyber-defense/

21) TechCrunch, "How AI guardrails are impeding the work of offensive cybersecurity researchers", 2026 年 7 月 23 日. https://techcrunch.com/2026/07/23/how-ai-guardrails-are-impeding-the-work-of-offensive-cybersecurity-researchers/

22) Tech Times, "Gemini 3.5 Pro Targets July 17 After Full Rebuild: Every Spec Remains Unconfirmed", 2026 年 7 月 13 日. https://www.techtimes.com/articles/320308/20260713/gemini-35-pro-targets-july-17-after-full-rebuild-every-spec-remains-unconfirmed.htm

23) CyberScoop, "OpenAI expands Trusted Access for Cyber program with new GPT 5.4 Cyber model", 2026 年. https://cyberscoop.com/openai-expands-trusted-access-for-cyber-to-thousands-for-cybersecurity/

24) explainx.ai, "OpenAI Astra: Critical Cyber Tier Confirmed (Sept 2026)", 2026 年 9 月. https://explainx.ai/blog/openai-astra-cybersecurity-critical-preparedness-framework-2026

25) Fortune, Jonathan Iwry（Wharton Accountable AI Lab）の発言を含む Anthropic Mythos／Project Glasswing 関連記事, 2026 年 4 月頃.【URL 未確認：本調査では二次的に要旨のみ確認】

26) Bruce Schneier, AI モデルのアクセス制限の実効性と「公共 AI」の必要性に関する論考, 2026 年.【URL 未確認：本調査では二次的に要旨のみ確認】

27) RUSI（Aybars Tuncdogan）, サイバーAI のベンダー依存が生む「戦略的依存」に関する論考, 2026 年.【URL 未確認：本調査では二次的に要旨のみ確認】

28) The Wall Street Journal, Gemini 3.8 Flash（内部コードネーム「skimaki」）の社内テストに関する報道, 2026 年 8 月.【URL 未確認：Business Insider Japan（文献 18）等の二次報道を通じて要旨を確認】

29) OpenAI, Astra のサイバー能力評価（Preparedness Framework「Critical」到達）に関する公式ページ（"path-to-astra"）, 2026 年 9 月.【URL 未確認：文献 9・24 を通じて要旨を確認】