

# グーグル新モデル「Gemini 3.1 Pro」 深掘りレポート

## Executive Summary (要約)

Gemini 3.1 Proは、2026年2月19日（現地時間）に公式ブログで「プレビュー（preview）」として発表され、開発者（Gemini API/AI Studio等）、企業（Vertex AI等）、一般ユーザー（Geminiアプリ/NotebookLM）へ段階的に展開された「Gemini 3」系の改良版である。<sup>1</sup>

公式のモデルカードでは、推論系ベンチマーク（特にARC-AGI-2）でGemini 3 Pro比「2倍以上」の性能向上が示され、ARC-AGI-2は77.1%（Gemini 3 Proは31.1%）とされる。<sup>2</sup>

一方で、同じモデルカードの比較表は「他社モデルの多くが自己申告値（self-reported）」であること、推論モード（thinking設定）を“最大”側で揃える前提が含まれることから、マーケティング上の「首位」主張をそのまま“科学的比較”として解釈するには注意が必要である。<sup>3</sup>

提供形態はクラウド/API前提で、入力はテキスト・画像・動画・音声・PDF、最大入力1,048,576トークン（約1M）、最大出力65,536トークン。知識カットオフは2025年1月と明記される。<sup>4</sup>

価格はGemini Developer APIの「Gemini 3 Pro Preview」と同水準で、標準（Standard）で入力\$2.00/100万トークン（<=200k）・出力\$12.00/100万トークン（<=200k）等が提示され、Batchは概ね半額水準である。<sup>5</sup>

「AGIへの登竜門」という言い回しは一部日本語記事で見られるが、ARC Prize側はARC-AGIを“AGI判定のリトマス試験紙”と位置づけていないと明記し、HLE（Humanity's Last Exam）側も高得点が直ちにAGIを意味しない旨を明示している。したがって、Gemini 3.1 Proの伸長は「汎用推論の進展」を示す強いシグナルになり得る一方、「AGI到達」の断定には論拠不足である。<sup>6</sup>

## 主要結論と推奨（短く箇条）

- 研究・開発用途では、ARC-AGI-2やBrowseComp等の指標で見える「推論/エージェント能力の上振れ」を、業務フロー（調査→統合→可視化→実装）に直結させやすいモデルである。<sup>7</sup>
- ただし、比較表には他社の自己申告値が混在しうるため、社内評価では「自社データ+再現可能なベンチ+運用指標（コスト/遅延/失敗率）」で再検証する前提を推奨する。<sup>8</sup>
- 長文脈は“最大1M”を謳うが、公式評価でも1M条件での長文脈指標（MRCR v2）が大きく落ちるため、「1M入力を常用する設計」は段階分割・キャッシュ戦略・検索/RAG併用を前提に組むべきである。<sup>9</sup>
- 料金は（少なくともDeveloper API上では）Gemini 3 Pro Previewと同水準のため、同等コストで推論性能を引き上げたい場合に“第一候補”になりやすい。<sup>10</sup>
- 規制・ガバナンス面では、EUのGPAI（汎用AI）義務・リスク管理要求や、NIST AI RMF等に沿った文書化・評価・監視を「導入要件」として組み込むことを推奨する。<sup>11</sup>

## 基本情報と提供形態

Gemini 3.1 Proは「Gemini 3」シリーズの次イテレーションで、モデルカード上で「Gemini 3 Proに基づく（based on）」と明記される。<sup>12</sup>

公式ブログでは「Gemini 3 Deep Think（直前週の大型アップデート）でのブレークスルーを支える“コアの知能”を日常用途へ展開する」という位置づけで、プレビューとして提供し、一般提供（GA）前に検証と改良を進める旨が述べられている。<sup>13</sup>

### 基本仕様（一次情報中心）

項目	内容	出典
発表日	2026-02-19 (公式ブログ掲載日)。日本語メディアでは「日本時間2/20」表記が一般的	14
提供ステージ	Public preview (プレビュー)	15
モデルコード (API)	<code>gemini-3.1-pro-preview</code> (派生: <code>...-customtools</code> )	16
入力モダリティ	Text / Image / Video / Audio / PDF	17
出力モダリティ	Text (テキスト出力)	17
最大入力	1,048,576 tokens (約1M)	17
最大出力	65,536 tokens	17
知識カットオフ	2025年1月	18
アーキテクチャ	3.1 Pro自体の詳細は「3 Pro参照」。3 Proは sparse MoE (Mixture-of-Experts) なTransformerで、テキスト/視覚/音声のネイティブ・マルチモーダル入力を持つ	19
モデルサイズ (パラメータ数等)	未指定 (モデルカード/公式ドキュメントで明示確認できず)	19
学習データ規模 (総量)	未指定 (ただし種別は記載あり)	20
学習データ種別	事前学習: 公開Web文書/テキスト/コード/画像/音声/動画等。事後学習: 指示追従データ、RLデータ、人の嗜好データ等。ユーザーデータや合成データも含み得る旨が明記	20
推論速度・レイテンシ	数値は公式未指定。ただしthinking_levelで推論深度 (=コスト/速度/性能) を調整可能で、低レベルは低遅延を狙う設計が示される	21
メモリ要件 (オンプレ)	原則API/クラウド提供で「利用に必要なハード/ソフト要件はない」と説明 (自前ホスティング前提のメモリ要件は未指定)	12
利用チャンネル	開発者: Gemini API (AI Studio等) /企業: Vertex AI/消費者: Geminiアプリ・NotebookLM 等	22

### 価格 (Gemini Developer API、プレビュー)

標準 (Standard) で、入力\$2.00/100万トークン (<=200k)、出力\$12.00/100万トークン (<=200k)。200k超は単価が上がる。Batchは標準の概ね半額水準で提示される。<sup>5</sup>

ライセンス/利用条件は、Gemini API追加利用規約やCloud規約等に従う形で配布される (詳細は各規約参照)。<sup>23</sup>

#### timeline

title Gemini 3系の公開タイムライン (主要ソースで確認できる範囲)

2025-11 : Gemini 3 Pro 公開 (Gemini 3 Pro Model Card上のリリース表記)

## 性能評価とベンチマーク比較

### 比較の前提 (方法論上の注意)

Google DeepMind <sup>25</sup> の評価手法説明では、他社モデルの多くが「プロバイダ (提供者) の自己申告値」から採録される場合があること、比較時にはClaude/GPT系の“最大 thinking 設定”側を基本とすることなどが明記される。したがって、下表は「公式比較」ではあるが、厳密な意味での“同一条件ベンチ”とは限らない。

### 4モデル比較 (候補選定と理由)

ユーザー指定がないため、本レポートでは次の4つを「比較の主軸」とする。理由は「同系統の前世代 (3 Pro)」「主要競合2社 (Anthropic <sup>26</sup> / OpenAI <sup>27</sup> の旗艦クラス)」を含めると、調達・運用の意思決定に直結しやすいからである。 <sup>12</sup>

### 公式ベンチマーク (モデルカード記載値)

ベンチマーク	Gemini 3.1 Pro	Gemini 3 Pro	Claude Opus 4.6	GPT-5.2	備考	出典
ARC-AGI-2	77.1%	31.1%	68.8%	52.9%	ARC Prize Verified (抽象推論)	<sup>12</sup>
Humanity's Last Exam (No tools)	44.4%	37.5%	40.0%	34.5%	学術推論 (ツールなし)	<sup>12</sup>
Humanity's Last Exam (Search+Code)	51.4%	45.8%	53.1%	45.5%	検索+コード利用 (ブロックリスト付き検索)	<sup>28</sup>
GPQA Diamond	94.3%	91.9%	91.3%	92.4%	科学知識 (ツールなし)	<sup>12</sup>
SWE-Bench Verified	80.6%	76.2%	80.8%	80.0%	Agentic coding (single attempt)	<sup>12</sup>
Terminal-Bench 2.0	68.5%	56.9%	65.4%	54.0%	端末操作型コーディング (Terminus-2)	<sup>12</sup>
APEX-Agents	33.5%	18.4%	29.8%	23.0%	長期・職能タスク	<sup>12</sup>
MMMLU	92.6%	91.8%	91.1%	89.6%	多言語Q&A (※MMMLU本体は未提示)	<sup>29</sup>

ベンチマーク	Gemini 3.1 Pro	Gemini 3 Pro	Claude Opus 4.6	GPT-5.2	備考	出典
MRCR v2 (128k avg)	84.9%	77.0%	84.0%	83.8%	長文脈 (128k)	12
MRCR v2 (1M pointwise)	26.3%	26.3%	Not supported	Not supported	長文脈 (1M) で急落	12

### 「高得点」の根拠（日本語記事での扱い）

国内技術メディアでは、特にARC-AGI-2での77.1%を「推論性能が2倍以上」や「未知のロジックパターンを解く能力が高い」と説明し、SVGアニメ生成などのデモも同時に紹介している。<sup>30</sup>

また、Noteや個人解説では「16ベンチ中13首位」等の“首位数”が語られるが、その集計対象（どの16ベンチか、比較相手の設定、自己申告値の扱い）が固定されていないことが多い。よって、主張の妥当性は「対象ベンチ・条件・比較モデルの明示」に依存し、現時点では一般化しにくい。<sup>31</sup>

### ベンチマーク図（棒グラフ）

```
xychart-beta
  title "ARC-AGI-2 (%) 4モデル比較 (高いほど良い) "
  x-axis ["Gemini 3.1 Pro","Gemini 3 Pro","Claude Opus 4.6","GPT-5.2"]
  y-axis "score" 0 --> 100
  bar [77.1, 31.1, 68.8, 52.9]
```

12

### ベンチマーク図（折れ線：HLEにおける「ツール有無」差分）

```
xychart-beta
  title "Humanity's Last Exam：ツールなし（棒） vs 検索+コード（線）"
  x-axis ["Gemini 3.1 Pro","Gemini 3 Pro","Claude Opus 4.6","GPT-5.2"]
  y-axis "accuracy (%)" 0 --> 60
  bar [44.4, 37.5, 40.0, 34.5]
  line [51.4, 45.8, 53.1, 45.5]
```

12

### 独立・準独立ベンチの状況

- **ARC-AGI-2 (ARC Prize Verified)**：モデルカードはARC Prize VerifiedとしてARC-AGI-2結果を掲げる。ARC Prize側は、Verifiedの設計として第三者の学術パネルによる監督や、評価セット (Semi-Private / Private) の区分、効率 (コスト) も含めた解釈を強調している。<sup>32</sup>
- **Chatbot Arena系 (人手投票Elo)**：arena.aiのリーダーボードでは、テキスト領域で `gemini-3.1-pro-preview` が上位に位置づけられている (例：Textで1500前後)。ただし、これは“好み/対話品質”を含む人手投票の総合指標で、研究ベンチとは目的が異なる。<sup>33</sup>
- **APEX-Agents (職能タスク)**：Brendan Foody<sup>34</sup> (Mercor<sup>35</sup>) が「APEX-Agentsで上位」と言及しており、パス率の上昇 (18.4→33.5) を示す投稿がある。モデルカードの数値とも整合する。<sup>36</sup>

- **MMLU等の“古典ベンチ”**：HLE側は「MMLUやGPQAは飽和しやすい」と述べる一方、Gemini 3.1 Proの公式公開資料ではMMLU本体の値がこのモデルカード表に直接は出てこない（多言語版としてMMMLUは提示）。よって「MMLUで比較したい」という要望に対しては、現時点では“未指定”が妥当である。 <sup>37</sup>

## 技術的特徴とアーキテクチャ差分

Gemini 3.1 Proは、アーキテクチャを刷新した“別系統”というより、Gemini 3 Proを基盤に「品質（reasoning・SWE・エージェント実用性・トークン効率）」と「推論の制御性」を強化した“反復改良”色が強い。 <sup>38</sup>

**新機能・性質（公式に明記される範囲）** - **ネイティブ・マルチモーダル+超長文脈**：入力はText/Image/Video/Audio/PDFで最大1M、出力は最大65,536。コードリポジトリ全体まで含む“巨大入力”を意識した記述がある。 <sup>39</sup>

- **thinking\_level（推論深度）の運用最適化**：Gemini 3系は“dynamic thinking”をデフォルトで行い、thinking\_levelで推論深度（=遅延/コスト/性能のトレードオフ）を制御できる。3.1 ProではVertex AI側でMEDIUM追加が明記され、速度と性能の中間点を運用上選びやすくしている。 <sup>21</sup>

- **ツール/エージェント指向の機能群**：関数呼び出し、構造化出力、コード実行、検索グラウンディング、コンテキストキャッシュ、Batch等がサポートされる（ただし環境により制約、例：File searchはAI Studio限定表記）。 <sup>40</sup>

- **“コードで作る”生成物に強いデモ**：公式ブログは「テキスト指示からWebサイトに組み込めるアニメーションSVGを生成」「複雑API→ダッシュボード化」「インタラクティブ3Dや創造的コーディング」等を例示する。 <sup>41</sup>

### 差分の実務的含意（推定ではなく、公式記述からの帰結）

「推論を深くするほど正解率が伸びやすいが、計算資源=コスト/遅延も増える」という前提の下、thinking\_levelに“中間”が生えることは、プロダクション運用で「普段はMEDIUM、難問だけHIGH」などのポリシー設計を可能にする。 <sup>42</sup>

flowchart LR

```

A[アプリ/業務プロセス] --> B[Gemini API / Vertex AI]
B --> C{推論深度\nthinking_level}
C -->|LOW| D[低遅延・低コスト]
C -->|MEDIUM| E[バランス]
C -->|HIGH| F[高推論・高コスト]
B --> G[ツール連携\n(検索/コード実行/関数呼び出し)]
B --> H[キャッシュ/Batch\n(コスト最適化)]
G --> A
H --> A

```

<sup>43</sup>

### 制約（明確に“Not supported”とされるもの）

APIのモデル仕様表では、音声生成・画像生成・Live APIが「未サポート」とされる（※“理解”としての音声入力自体はサポート）。 <sup>44</sup>

## AGI関連評価とガバナンス示唆

### 「AGIへの登竜門」主張の妥当性評価

日本語の個人記事や解説では、ARC-AGI-2を「AGIへの登竜門」と表現し、Gemini 3.1 Proの77.1%を根拠に“推論の壁”を破ったと論じるものがある。<sup>45</sup>

しかし、ARC Prize<sup>46</sup>自身は、ARC-AGIを「AGI判定のためのリトマス試験紙ではない」と明記し、Semi-Privateリーダーボードはインターネットや計算資源に制限がないこと、別途コンペは“計算制約+ネット禁止”のPrivateセットで行うことを説明している。つまり、単純な正解率だけで「AGI到達」を言う設計思想ではない。<sup>47</sup>

さらに、Scale AI<sup>48</sup>のHLEページは「MMLUやGPQAは飽和が早い」こと、HLE高得点は“閉じた学術問題での専門家級性能”を示すが、「自律的研究能力やAGIを単独で示すものではない」と明確に断る。よって、ARC-AGI-2やHLEの向上は“AGIへの接近”の一要素ではあっても、十分条件ではない。<sup>49</sup>

### 定量指標から見た「AGIっぽさ」の到達度

- **汎用推論の伸び**：ARC-AGI-2で77.1%は、前世代31.1%から約2.48倍で、短期間の伸長としては非常に大きい。<sup>50</sup>
- **“職能タスク”の改善**：APEX-Agentsが18.4%→33.5%へ伸びた（モデルカードと外部投稿が整合）。ただし33.5%は“上位層”であっても“完全自律で大半を解く”水準ではない。<sup>51</sup>
- **長文脈の限界**：1M入力を掲げる一方、長文脈評価（MRCR v2）の1M点は26.3%と128k平均（84.9%）から大きく落ち、超長文での信頼性は未解決課題として残る。これは「AGI＝長期記憶・一貫した計画」の観点では重要な制約である。<sup>52</sup>

### 倫理・リスク・安全性（モデルカード／FSFから）

モデルカードは、内部の安全評価（安全性・多言語安全性・不当拒否など）でGemini 3 Proと概ね同程度の安全性能を維持しつつ、一部指標の微小変動を示す。<sup>12</sup>

フロンティア安全性については、DeepMindのFrontier Safety Framework (FSF) に基づき、CBRN・サイバー・有害操作・ML R&D・ミスアラインメントの5領域で「重要能力レベル（CCL）に到達していない」と要約される（ただしサイバー領域では能力増を認め、追加評価を実施した旨も記載）。<sup>53</sup>

### 規制・ガバナンス上の示唆（EU/NIST/日本の枠組み）

- **EU（GPAI義務）**：European Commission<sup>54</sup>は、EU AI Actにおける汎用AI（GPAI）モデル提供者の義務が2025年8月2日から適用されること、システミックリスク該当モデルはAI Officeへの通知やリスク評価・緩和等が求められることを説明している。Gemini 3.1 Pro級の“フロンティア”モデルは、EU市場での提供形態によってこれらの枠組みの実務影響を強く受ける可能性がある。<sup>55</sup>
- **米国（リスクマネジメント）**：NIST<sup>56</sup>のAI RMF 1.0は、AIリスクを組織的に扱う枠組み（GOVERN/MAP/MEASURE/MANAGE）を提示しており、ベンチマークだけでなく、運用監視・説明責任・影響評価を含めたガバナンス設計が推奨される。<sup>57</sup>
- **日本（公共調達・行政ガイダンスの方向性）**：日本のデジタル庁資料では、生成AI調達に伴う真正性/来歴検証（provenance）を含む検討を継続する旨など、行政利用での慎重な枠組み整備が示唆される（民間にも波及し得る）。<sup>58</sup>

## 実務的インパクトと競合ポジショニング

### 産業応用領域（公式記述+そこからの実務的帰結）

Vertex AI側ドキュメントでは、SWEとエージェント能力の改善、トークン効率とthinkingの改善、**MEDIUM** 導入によるコスト/性能/速度の最適化を“品質改善”として列挙している。金融やスプレッドシート領域でのエー

ジェント改善にも言及がある。 59

公式ブログの例（テレメトリ→ダッシュボード、インタラクティブ設計、SVG生成）からは、(1)データ統合と可視化、(2)UI/フロントエンド生成、(3)ツール接続型の業務自動化（エージェント）へ寄せた設計思想が読み取れる。 41

## 導入コスト（API利用を中心に）

Developer APIの価格は、3.1 Pro Previewと3 Pro Previewで同一の料金表が掲載されているため、推論性能向上が“コスト増なし”で得られる可能性がある。 10

運用では、(a)出力料金に「thinkingトークンが含まれる」点、(b)200k超で単価が上がる点、(c)キャッシュ（read/write+保管）が別料金である点が、総コストを左右する。 60

## 運用上の注意点

- **プレビューゆえの制約**：プレビューは変更され得ること、プレビュー/実験モデルはレート制限が厳しめになり得ることが公式ドキュメントで述べられる。 61
- **レート制限とスループット設計**：Gemini APIはRPM/TPM/RPD等のレート制限体系で、プレビューはより制限が厳しいと明記される。Batchにも別枠の上限（Tier別のenqueued tokens）が示される。 62
- **“1Mコンテキスト=万能”ではない**：公式評価で1M点の長文脈性能が低いことは、RAG/分割/キャッシュを前提としたアーキテクチャが必要であることを示唆する。 9

## 競合とのポジショニング（本レポートの結論）

- **強み（示されている範囲）**：ARC-AGI-2、GPQA Diamond、BrowseComp、APEX-Agentsなどで高い値が示され、“推論+エージェント+実装（コード）”の三位一体に寄せた総合型の強みが見える。 63
- **弱み/注意**：SWE-Bench VerifiedではOpus 4.6が僅差で上回る、GDPval-AA（専門家タスク嗜好）では競合が高いなど、“すべてで圧勝”ではない。さらに比較値には自己申告混在の可能性があるため、プレミアム調達の意味決定では自社要件に直結するベンチ（例：自社SWE課題、監査ログ、MTTR、逸脱率）で測る必要がある。 28

## 追加調査項目と主要ソース一覧

### 不足情報（未指定）と入手方法

- **モデルサイズ（総パラメータ数、エキスパート数、活性化パラメータ等）**：公式モデルカード（3.1/3 Pro）では定量の明示が確認できないため未指定。今後の技術レポート/論文、またはモデルカード更新を待つのが最短。 19
- **トレーニングデータ“規模”の定量（トークン数、動画時間、コーパス内訳）**：種別は説明されるが総量は未提示。規制（GPAI義務）対応の技術文書として開示される可能性があるため、EU向け技術文書動向も追跡対象。 64
- **推論速度・レイテンシの定量（TTFT、tokens/sec、SLO/SLA）**：公式に一般的な“速度指標”の掲載は確認できず未指定。入手には、(a)自前計測（同一条件でのTTFT/throughput計測）、(b)Provisioned Throughput活用時の実測、(c)公式の将来ドキュメント更新が必要。 65
- **Yahoo記事の一次確認**：当環境ではYahoo Newsドメインがrobots制限で参照不可のため、本レポートは他の日本語メディア/一次情報で代替した（必要なら利用者側でURL共有→別環境で確認が必要）。

## 主要ソース一覧 (URL)

<https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-1-pro/>  
<https://deepmind.google/models/model-cards/gemini-3-1-pro/>  
<https://storage.googleapis.com/deepmind-media/Model-Cards/Gemini-3-Pro-Model-Card.pdf>  
<https://ai.google.dev/gemini-api/docs/models/gemini-3.1-pro-preview>  
<https://ai.google.dev/gemini-api/docs/pricing>  
<https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/3-1-pro>  
<https://ai.google.dev/gemini-api/docs/rate-limits>  
<https://k-tai.watch.impress.co.jp/docs/news/2087531.html>  
<https://gihyo.jp/article/2026/02/gemini-3-1-pro>  
[https://www.gizmodo.jp/2026/02/google\\_gemini\\_3\\_1\\_pro.html](https://www.gizmodo.jp/2026/02/google_gemini_3_1_pro.html)  
[https://note.com/masa\\_wunder/n/nae2323894f8c](https://note.com/masa_wunder/n/nae2323894f8c)  
<https://chatgpt-lab.com/n/n64d9f440c178>  
<https://arcprize.org/policy>  
<https://arcprize.org/arc-agi/2/>  
[https://scale.com/leaderboard/humanitys\\_last\\_exam](https://scale.com/leaderboard/humanitys_last_exam)  
<https://arxiv.org/abs/1911.01547>  
<https://digital-strategy.ec.europa.eu/en/news/learn-more-about-guidelines-providers-general-purpose-ai-models>  
<https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf>

---

1 13 14 22 25 35 41 46 61 <https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-1-pro/>  
<https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-1-pro/>

2 7 9 12 19 23 26 27 28 29 32 34 38 39 48 50 51 52 53 54 56 63 <https://deepmind.google/models/model-cards/gemini-3-1-pro/>  
<https://deepmind.google/models/model-cards/gemini-3-1-pro/>

3 8 <https://deepmind.google/models/evals-methodology/gemini-3-1-pro>  
<https://deepmind.google/models/evals-methodology/gemini-3-1-pro>

4 16 17 18 40 44 <https://ai.google.dev/gemini-api/docs/models/gemini-3.1-pro-preview>  
<https://ai.google.dev/gemini-api/docs/models/gemini-3.1-pro-preview>

5 10 60 <https://ai.google.dev/gemini-api/docs/pricing>  
<https://ai.google.dev/gemini-api/docs/pricing>

6 47 <https://arcprize.org/policy>  
<https://arcprize.org/policy>

11 55 <https://digital-strategy.ec.europa.eu/en/news/learn-more-about-guidelines-providers-general-purpose-ai-models>  
<https://digital-strategy.ec.europa.eu/en/news/learn-more-about-guidelines-providers-general-purpose-ai-models>

15 42 59 65 <https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/3-1-pro>  
<https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/3-1-pro>

20 24 64 <https://storage.googleapis.com/deepmind-media/Model-Cards/Gemini-3-Pro-Model-Card.pdf>  
<https://storage.googleapis.com/deepmind-media/Model-Cards/Gemini-3-Pro-Model-Card.pdf>

- 21 43 <https://ai.google.dev/gemini-api/docs/gemini-3>  
<https://ai.google.dev/gemini-api/docs/gemini-3>
- 30 <https://k-tai.watch.impress.co.jp/docs/news/2087531.html>  
グーグル「Gemini 3.1 Pro」リリース 推論性能が「3 Pro」の2倍以上 - ケータイ Watch
- 31 [https://note.com/masa\\_wunder/n/nae2323894f8c](https://note.com/masa_wunder/n/nae2323894f8c)  
[https://note.com/masa\\_wunder/n/nae2323894f8c](https://note.com/masa_wunder/n/nae2323894f8c)
- 33 <https://arena.ai/leaderboard>  
<https://arena.ai/leaderboard>
- 36 <https://x.com/BrendanFoody/status/2024536763528860089>  
<https://x.com/BrendanFoody/status/2024536763528860089>
- 37 49 [https://scale.com/leaderboard/humanitys\\_last\\_exam](https://scale.com/leaderboard/humanitys_last_exam)  
[https://scale.com/leaderboard/humanitys\\_last\\_exam](https://scale.com/leaderboard/humanitys_last_exam)
- 45 <https://note.com/antwork/n/n6328b7aeac3e>  
<https://note.com/antwork/n/n6328b7aeac3e>
- 57 <https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf>  
<https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf>
- 58 [https://www.digital.go.jp/assets/contents/node/basic\\_page/field\\_ref\\_resources/e2a06143-ed29-4f1d-9c31-0f06fca67afc/6e45a64f/20250527\\_resources\\_standard\\_guidelines\\_guideline\\_04.pdf](https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/e2a06143-ed29-4f1d-9c31-0f06fca67afc/6e45a64f/20250527_resources_standard_guidelines_guideline_04.pdf)  
[https://www.digital.go.jp/assets/contents/node/basic\\_page/field\\_ref\\_resources/e2a06143-ed29-4f1d-9c31-0f06fca67afc/6e45a64f/20250527\\_resources\\_standard\\_guidelines\\_guideline\\_04.pdf](https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/e2a06143-ed29-4f1d-9c31-0f06fca67afc/6e45a64f/20250527_resources_standard_guidelines_guideline_04.pdf)
- 62 <https://ai.google.dev/gemini-api/docs/rate-limits>  
<https://ai.google.dev/gemini-api/docs/rate-limits>