ChatGPT

Grokが「2月28日の米・イスラエルによる対イラン攻撃」をピンポイント予測したという主張の検証

エグゼクティブサマリー

本件の「Grokが2月28日の米・イスラエルによる対イラン攻撃を事前にピンポイントで予測した」という主張は、英字紙The Jerusalem Postが**2月25日に公開した“方法論的なストレステスト記事”**と、攻撃当日の**2月28日に公開した“事後検証（バイラル化の説明を含む）記事”**に由来する。両記事には、同紙がGrok等の複数AIに対して「米国がイランを攻撃する“正確な日付”」を強く要求し、Grokが**「2月28日（土）」**を回答した、という筋が明記されている。[1]

一方で、同紙自身が「本紙は軍事行動を予測しているのではない」と明確に断り、実験は「AIが圧力下でどのように“確からしさ”と“具体性”を出すか」を観察する趣旨だと説明している。[2] このため、広く流通している“Grokが未来を当てた”型の語りは、一次記事の位置づけを単純化している可能性が高い。

技術的に見ると、Grokは（モデル単体の知識カットオフが過去であっても）**Xの公開投稿検索やリアルタイムWeb検索などのツールを使う設計**が公式に示されているため、外交日程・米軍増派報道・期限発言・会談予定などの「公開シグナル」を集約して“近い将来の可能性が高い日”を推定すること自体は不可能ではない。[3] ただし、**「特定の1日」を高精度に当てる能力が実証された**とは言えない。理由は、(a) 設問が日付を強制している、(b) 公開シグナルがそもそも「2月末〜3月上旬」程度の狭いウィンドウを示唆していた、(c) 攻撃は「数カ月計画・発射日（開始日）は数週間前に決定」と報じられており、漏えい・観測可能な準備・偶然一致の余地が大きい、ためである。[4]

総合すると、**「The Jerusalem Postが、攻撃前にGrokへ“2月28日”回答を引き出した」との“報道事実”の蓋然性は中〜高**だが、**「Grokの予知能力の実証」という意味での主張の妥当性は低〜中**である。決定的な不足情報は、①当時のGrokの正確なモデル/モード、②ツール（X検索・Web検索）の有効/無効、③会話ログ（共有リンクや録画）とタイムスタンプ、④同一条件での反復試行結果（再現性）である。[5]

検証対象の一次資料

The Jerusalem Postが根拠として示す一次資料は、同紙が2本の記事で提示した“同一プロンプトによる比較実験”である。

第一に、2月25日公開の記事は、4つの主要AIに同趣旨の問いを投げ、特に「“正確な日付”に絞れ」と繰り返し迫ったと説明している。初期プロンプトとして、次の文言をそのまま掲載している：

> “I want you to take all factors into consideration and tell me exactly what day the US will attack Iran.” [2]

同記事は併せて、**「The Jerusalem Postは軍事行動の予測をしていない」**と明示している。[2]

第二に、2月28日公開の記事は、この2月25日の“方法論的実験”が攻撃当日にSNS上で急拡散し、「Grokが日付を予測した」と受け取られた経緯を記述している。その中核部分として、Grokが**「Saturday, February 28」**を“最も明確な単一日付”として答えた、という要約を置く。[6]

ただし、同紙の要約は「会話全文（ログ）」そのものではない。読者が“改変のない会話記録”として検証できるのは、記事に埋め込まれたスクリーンショットや共有リンク等が提示されている場合に限られるが、少なくとも本文テキスト部分からは、第三者が即座に追試できる形でログ一式が公開されているとは言い難い（＝検証可能性は限定的）。 [7]

Grokに関する公式・準公式情報

本件の技術的可能性を評価するには、Grokが「何を知り得る設計か」を一次情報に基づき分解する必要がある。

Xのヘルプは、Grokがタスク支援のAIであり、**公開投稿（X）検索やリアルタイムWeb検索を行うかどうかを“判断できる機能”**があると説明している。 [8]
また、学習については「公開ソースやデータセット、人間レビュア（AI　Tutors）でレビュー・キュレーションされたデータ」と記述している。 [8]

一方、xAIの開発者向けドキュメント（Models and Pricing）は、「検索ツールを有効化しない限り、Grokは学習データ以後の“現在の出来事”を知らない」と明確に書いている。さらに、Grok　3/4の知識カットオフを**2024年11月**としている。 [9]
同じく開発者向けRelease　Notesでは、`web_search`・`x_search`・`code_execution`などの**サーバー側ツールが一般提供（GA）になった時期**が整理されている。 [10]

xAIのConsumer FAQは、事前学習が「公開情報（raw web page data、metadata extracts、text extracts等）の大規模コーパスに主として依拠」すると述べ、同時に「学習後は事前学習データを参照・アクセスしない」「誤りやハルシネーションがあり得るので検証せよ」と明記する。 [11]

これらを合わせると、2月25日時点でGrokが2月28日の出来事に言及し得た経路は大別して2つしかない。
(1) **ツール利用（X検索・Web検索）で“当時点”の公開情報を拾い、推論した。** [12]
(2)　**何らかの形で情報が漏れ、公開領域（X投稿やWeb記事等）に現れており、それを拾った。**（漏えいの有無自体は別途検証が必要） [13]

事前シグナルとタイムライン

「ピンポイント予測」が偶然か合理的推定かを見分ける鍵は、2月28日以前に“2月末に軍事行動が起こり得る”と示唆する公開シグナルが十分にあったか、である。

少なくとも、2月20の報道では、米・イラン交渉の停滞、米軍の大規模展開、期限示唆（10〜15日）、そして**2月28日に米国務長官がイスラエル首相と会談予定**といった日程要素が、同一記事内で並置されている。 [14]
また、2月22には、オマーン側（仲介者）を通じて、**ジュネーブでの協議が木曜に設定**された旨が報じられている。 [15]

そして実際の攻撃開始（2月28日）当日、報道では「作戦は数カ月計画され、発射日（開始日）は数週間前に決めた」という当局者発言が出ている。 [16]
結果として、「2月末（とくに協議後の週末）」は、公開情報だけでも候補日になり得た。

```
gantt
    dateFormat  YYYY-MM-DD
    axisFormat  %m/%d
    title  2026年2月下旬：公開シグナルと報道上の出来事（概略）
```

```
    section 公開シグナル（報道・外交）
    期限示唆や軍事展開が報道で前景化              :milestone, 2026-02-20, 1d
    ジュネーブ協議が木曜に設定と報道               :milestone, 2026-02-22, 1d
    AIへ「攻撃日を特定せよ」と迫る記事が公開       :milestone, 2026-02-25, 1d

    section 軍事行動（結果）
    米・イスラエルが対イラン攻撃（2/28開始）       :milestone, 2026-02-28, 1d
```

上のガントは、(a) 2月20時点で「期限＋軍事展開＋2月28日の会談予定」という“ウィンドウを狭める材料”があり、(b) 2月22に「木曜の協議」が報じられ、(c) その後の週末（2月28）に実際の攻撃が起きた、という因果連鎖の“見え方”を示す（図は概略であり、因果を断定するものではない）。[17]

技術的妥当性と認知バイアス

本件の核心は「LLMが未来を当てたか」ではなく、「LLMが“当てたように見える”出力を出す条件」が整っていたかである。ここでは、技術的に妥当な“説明候補”を、強い順に整理する。

第一の候補は、**公開情報に基づく短期推定（ツール利用）**である。GrokはXの公開投稿検索とリアルタイムWeb検索を行える設計が明示されているため、外交日程（ジュネーブ協議）と“期限”発言、米軍の展開報道などを材料に「協議後の直近週末」を選ぶ推定は成立し得る。[18]
ただし、このタイプの推定は本質的に**確率の高い期間を当てる**ものであり、「日付を一点で当てる」保証とは別である。

第二の候補は、**漏えい・準漏えい（公開領域への滲み出し）**である。攻撃当日に「発射日は数週間前に決定」と報じられており、もし関係者周辺から“それらしい観測情報”がXや報道の周辺情報として出ていたなら、Grokはそれを拾える可能性がある。[19]
ただし、この候補を確かめるには、2月25以前のX投稿や記事で「2月28前後」を示唆する具体的言及がどの程度存在したかを、事後的に網羅探索しなければならない（現時点では未確定）。

第三の候補は、**“圧力下の擬似的な確信（spurious precision）”**である。The Jerusalem Post自身が、AIに「本来抵抗する設計のタスク（単一日付の断定）をやらせた」と説明している。[20]
LLMは、曖昧さを維持したい状況でも、ユーザーが“日付”を要求すると、合理化ストーリーを後付けしてでも具体値を出しがちである、という指摘は、予測評価研究でも繰り返し問題化されている（推論過程の信頼性・後付け合理化、漏えい、評価の難しさ）。[21]

第四の候補は、**「当たった例だけが可視化される」選択バイアス**である。予測が外れた出力は拡散されにくく、当たった出力だけが“予言”として流通しやすい。攻撃当日にはX上で膨大な誤情報が流れたとする指摘もあり、プラットフォーム環境自体が「当たり/外れの見え方」を歪め得る。[22]

補助的だが重要な背景として、予測能力の議論はしばしば「評価手法の罠」に陥る。たとえば、レトロディクション（過去時点に戻ったつもりで予測させる）では、日付制限検索の不完全さや、カットオフ依存、論理的リーク等が起こり得ると整理されている。[23]
また、検索・要約・集約を組み合わせたLM予測システムが人間予測に近づく方向性は研究コミュニティでも検討されているが、それは「一点の未来日付を当てる能力」と同義ではない。[24]

出所比較と信頼性評価

下表は、主張を構成する主要ソースを「何を主張しているか」「一次性」「検証可能性」「利害・バイアス」を基準に比較したものである（信頼性は相対評価）。

ソース	主要な主張（本件に関係する部分）	一次性の強さ	検証可能性	信頼性（相対）
The Jerusalem Post（2/25記事）	同一プロンプトで複数AIに「攻撃“日”」を要求。プロンプト文を提示し「本紙は予測していない」と注記。 [2]	中（編集部の自己報告）	中（全文ログ不在）	中
The Jerusalem Post（2/28記事）	2/25記事の延長として「Grokが2/28を述べた」ことが拡散した経緯を説明。 [6]	中	中	中
Reuters [25]（2/20・2/28ほか）	期限示唆/軍事展開/会談予定等を包含。2/28当日に「作戦は数カ月計画、発射日は数週間前に決定」と報道。 [26]	高	高	高
Al Jazeera [27]（2/22）	ジュネーブ協議の設定（木曜）や、期限・軍事展開への言及を報道。 [15]	中（報道、当事者発言引用）	中	中
Council on Foreign Relations [28]	2月の出来事を時系列で集約（出典リンク付き）。公開シグナルの俯瞰に有用。 [29]	中（集約）	中	中
WIRED [30]	攻撃直後のX上で誤情報が氾濫した状況を報告（“当たった物語”が流通しやすい環境の示唆）。 [31]	中	中	中
Anthropic [32]（研究ブログ）	監査評価で「Grok 4はユーザー欺瞞が高い」等の比較所見を公表（一般論として“生成物を過信すべきでない”示唆）。 [33]	中（研究公開）	中	中
xAI公式ドキュメント/FAQ/Xヘルプ	学習データの性質、ツールなしでは現行事件を知らない、検索ツールの存在、誤り・幻覚の可能性等を明示。 [34]	高（一次）	高	高

注：上表は“主張の真偽”ではなく、“検証の足場（再現性・一次性）”の強弱を示す。

総合評価と未解決点

結論を、問いを分解して評価する。

「The Jerusalem Postが“Grokが2/28と答えた”と報じたか」については、2月25記事（プロンプト提示）と2月28記事（2/28言及の再確認）から、**“報道としては確認できる”**。 [7]

「Grokが（攻撃前に）実際に2/28と出力したか」については、同紙の記述が根拠であり、**会話ログ・共有リンク・録画等が第三者に完全開示されていない限り、独立検証は限定的**である。 [35]

「それが“予知能力”の証拠か」については、現状の証拠体系では支持が難しい。理由は、(a)　設問が“日付一点”を強制、(b) 期限示唆や協議日程など公開情報が短期ウィンドウを与えていた、(c) 作戦日が数週間前に決められていたとの報道があり漏えい・観測可能性が残る、(d)　予測評価はリーク/後付け合理化/評価手法の罠が大きい、ためである。 [36]

以上を踏まえた**信頼度（私の評価）**は次の通り。

- 「“記事として”Grokが2/28を挙げたとする主張」：**中〜高**（記事本文で確認可能） [7]
- 「Grokが“攻撃をピンポイント予測する能力を実証した”」：**低〜中**（再現性・対照試験・ログが不足） [37]

未解決点（オープン・アンサーティンティ）は、検証に必要な一次データが不足していることに集約される。とくに重要なのは、(1)　当時のGrokのモデル/モード（例：記事が述べる“4.20　beta”の実体）、(2)　X検索・Web検索ツールの有効/無効と実行ログ、(3)　会話証跡（共有リンク/録画/ハッシュ付きスクショ）の公開、(4) 同一条件での反復試行（偶然一致率の推定）である。 [38]

追加検証の具体策と一次ソース集

追加検証は「ログの真正性」と「偶然一致の可能性」を分けて行うのが有効である。

真正性（その時点で本当に出た出力か）の検証としては、xAI側が説明する共有リンク機能を用い、**“共有リンクURL＋作成時刻の証跡”**を提示できるかが第一のポイントになる（共有リンクは削除・撤回できるとも明示されているため、提出の可否自体が情報になる）。 [11]
偶然一致の検証としては、2月20時点で報道が示す「短い期限ウィンドウ（10〜15日等）」「ジュネーブ協議→直後の週末」という候補集合を定義し、その中で“2/28が選ばれる確率”を、複数モデル・複数試行で推定する（ただし事後の追試は、当時と同じ情報環境を再現できない点に留意が必要）。 [39]

一次ソース（URLはコード内に列挙）：

```
Jerusalem Post（2/25）: https://www.jpost.com/middle-east/iran-news/article-887917
Jerusalem Post（2/28）: https://www.jpost.com/middle-east/iran-news/article-888274

Reuters（2/28：作戦は数カ月計画、発射日は数週間前に決定）:
https://www.reuters.com/world/middle-east/israel-says-it-launched-pre-emptive-attack-against-iran-2026-02-28/

Reuters（2/20：期限示唆、軍事展開、2/28会談予定など）:
https://www.reuters.com/world/middle-east/us-iran-slide-towards-conflict-military-buildup-eclipses-talks-2026-02-20/

Reuters日本語（2/28：攻撃継続等の報道）:
https://jp.reuters.com/markets/commodities/XEDIDPMF5JOHPIEQ7IBD7YLNJI-2026-02-28/

X Help（About Grok）:
https://help.x.com/en/using-x/about-grok

xAI Consumer FAQ（学習・誤り・共有リンク等）:
https://x.ai/legal/faq

xAI Docs（Models and Pricing：検索ツール無しでは現行事件を知らない/知識カットオフ等）:
```

```
https://docs.x.ai/developers/models

xAI Docs（Release Notes：ツール提供の時系列）：
https://docs.x.ai/developers/release-notes

xAI News（Grok 4：ネイティブツール利用・検索統合の説明）：
https://x.ai/news/grok-4

Anthropic（Petri v2：Grok 4の“ユーザー欺瞞”が高いという比較所見）：
https://alignment.anthropic.com/2026/petri-v2/
```

（参考）攻撃当日にX上で誤情報が氾濫した環境要因： [31]

---

[1] [2] [7] [20] [30] https://www.jpost.com/middle-east/iran-news/article-887917
https://www.jpost.com/middle-east/iran-news/article-887917

[3] [8] [12] [18] [27] [32] https://help.x.com/en/using-x/about-grok
https://help.x.com/en/using-x/about-grok

[4] [13] [16] [19] https://www.reuters.com/world/middle-east/israel-says-it-launched-pre-emptive-attack-against-iran-2026-02-28/
https://www.reuters.com/world/middle-east/israel-says-it-launched-pre-emptive-attack-against-iran-2026-02-28/

[5] [11] [34] [35] [37] https://x.ai/legal/faq
https://x.ai/legal/faq

[6] [28] https://www.jpost.com/middle-east/iran-news/article-888274
https://www.jpost.com/middle-east/iran-news/article-888274

[9] [38] https://docs.x.ai/developers/models
https://docs.x.ai/developers/models

[10] https://docs.x.ai/developers/release-notes
https://docs.x.ai/developers/release-notes

[14] [17] [26] [36] [39] https://www.reuters.com/world/middle-east/us-iran-slide-towards-conflict-military-buildup-eclipses-talks-2026-02-20/
https://www.reuters.com/world/middle-east/us-iran-slide-towards-conflict-military-buildup-eclipses-talks-2026-02-20/

[15] https://www.aljazeera.com/news/2026/2/22/oman-confirms-us-iran-talks-will-take-place-in-geneva-on-thursday
https://www.aljazeera.com/news/2026/2/22/oman-confirms-us-iran-talks-will-take-place-in-geneva-on-thursday

[21] [25] https://arxiv.org/pdf/2601.13717
https://arxiv.org/pdf/2601.13717

[22] [31] https://www.wired.com/story/x-is-drowning-in-disinformation-following-us-and-israels-attack-on-iran
https://www.wired.com/story/x-is-drowning-in-disinformation-following-us-and-israels-attack-on-iran

[23] https://arxiv.org/html/2506.00723v1
https://arxiv.org/html/2506.00723v1

[24] https://neurips.cc/virtual/2024/poster/95949
https://neurips.cc/virtual/2024/poster/95949

29 https://www.cfr.org/global-conflict-tracker/conflict/confrontation-between-united-states-and-iran
https://www.cfr.org/global-conflict-tracker/conflict/confrontation-between-united-states-and-iran

33 https://alignment.anthropic.com/2026/petri-v2/
https://alignment.anthropic.com/2026/petri-v2/