

ox-alpha 調査報告：匿名AIの正体と真の実力

智譜(Z.ai) GLM新フラッグシップ示唆の検証と、日本企業に向けた戦略的含意

DOCUMENT ID: INT-20260823-01
TARGET: stealth/ox-alpha (OpenRouter)
STATUS: PRELIMINARY FORENSICS
DATE: 2026.08.23

Executive Summary: 調査の全容と結論



事象の核心 (The Core Event)

[2026-08-20 // INCIDENT REPORT // OPENROUTER]

2026年8月20日、OpenRouterに完全無料の匿名モデル「stealth/ox-alpha」が登場。メディアは「GPT-5.6超え」と報道。



技術的証拠 (Technical Forensics)

[FORENSIC ANALYSIS // ZHIPU AI // FINGERPRINT MATCH]

独立系分析により、智譜AI (Zhipu AI) の未発表モデル (GLM-5.3V または GLM-5.5) である強い状況証拠 (フィンガープリント) を確認。

CONFIRMED MATCH



性能の真実 (Reality Check)

[ALERT // PERFORMANCE CLAIM // VERIFICATION FAILED]

性能優位の主張は「10タスクのみ」の極小サンプルに依存。センセーショナルな報道 (99%確信等) には二次情報の誇張が含まれる。

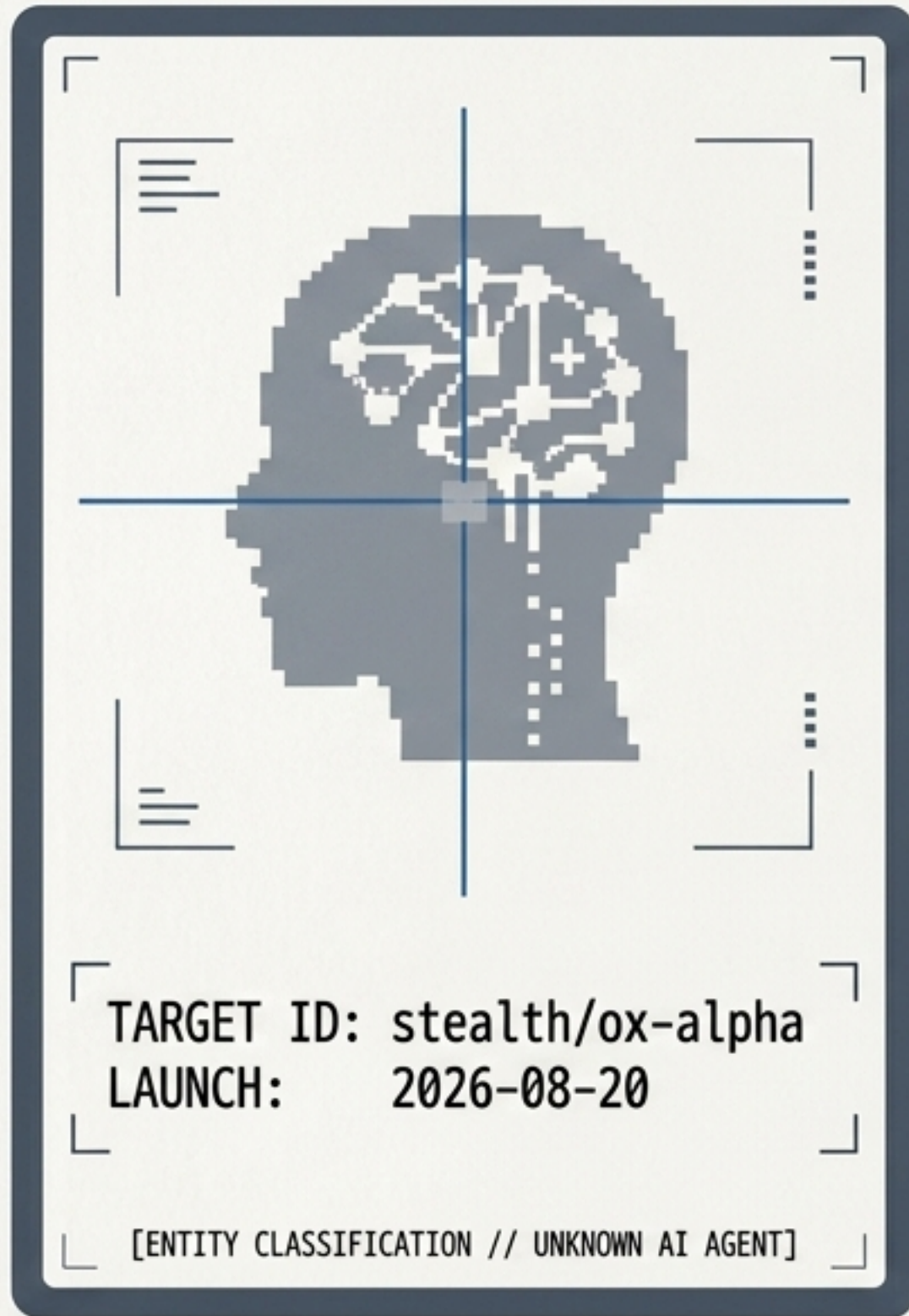


推奨アクション (Strategic Action)

[OPERATIONAL DIRECTIVE // RISK MITIGATION // MIT LICENSE]

- 現時点での本番導入は不可 (ログ保持のリスク)。検証はサンドボックス環境に限定し、実運用ベースラインにはMITライセンスの GLM-5.2 を推奨。✓

Target Profile: ターゲット識別・仕様概要 (ID Card & Specs)



Spec 1: キャパシティ

[SPEC-REF-001]

公称値: 1,048,576 トークン (約104.8万) / 最大出力 131,072
入力: テキスト・画像・動画

Spec 2: スループット主張

[THROUGHPUT-CLAIM-002]

公称値: 「1日あたり 100兆トークン処理可能」
[UNVERIFIED] 運営側の自己申告。検証不能。

Spec 3: プライバシーとデータ

[PRIVACY-ALERT-003]

公称値: Zero data retention (プロンプトは保持され
[HIGH RISK] 実際にはログがデータベースに保持され、学習には用いられない。
機密データの投入は厳禁。

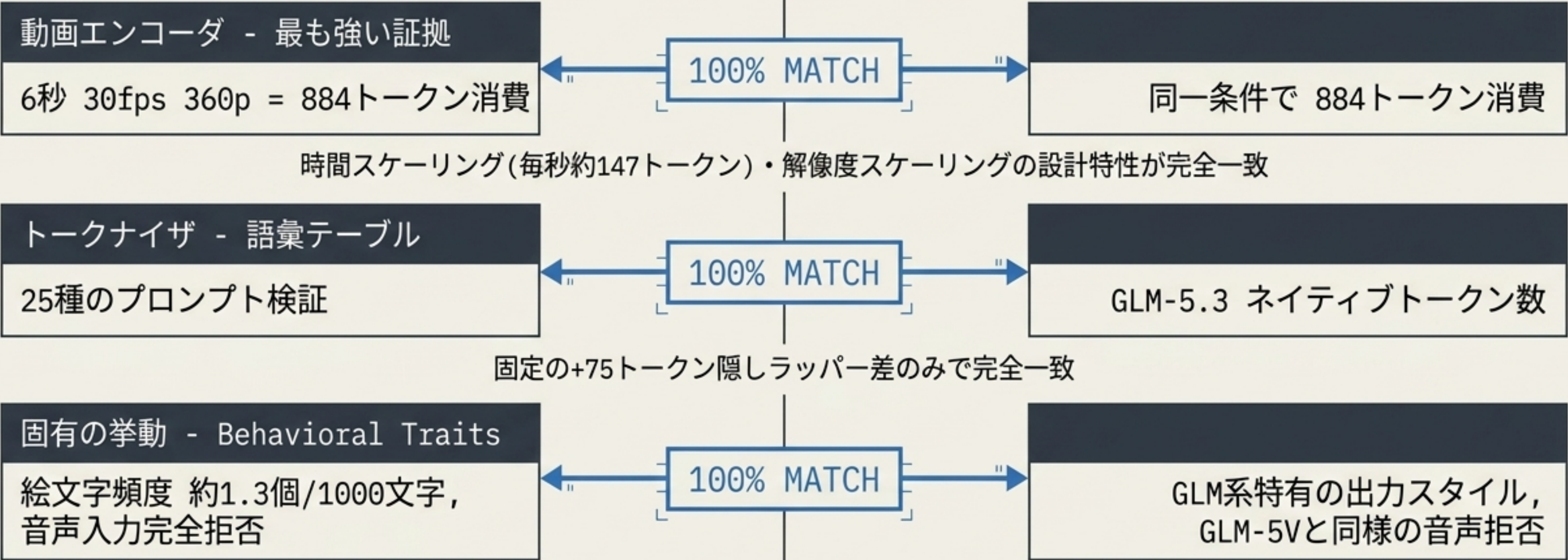
Technical Fingerprint: 智譜(Z.ai)との技術的照合

[FORENSIC ANALYSIS // MODEL ID MATCH CONFIRMED // CROSS-REFERENCE ZHIPU AI]

[DATE: 2026-08-24 // DOC ID: MATCH-ANALYSIS-0X-ALPHA]

TARGET: ox-alpha

KNOWN: GLM-5V-Turbo / GLM-5.3



メディア報道「GPT-5.6超え」のカラクリと限界

Full Benchmark: DeepSWE (113タスク)

Cherry-Picking Funnel

Filter: 決定論的サブセット
(--n-tasks 10 --sample-seed 0)

わずか10タスクでの結果
(モデル間で試行回数も不揃い)

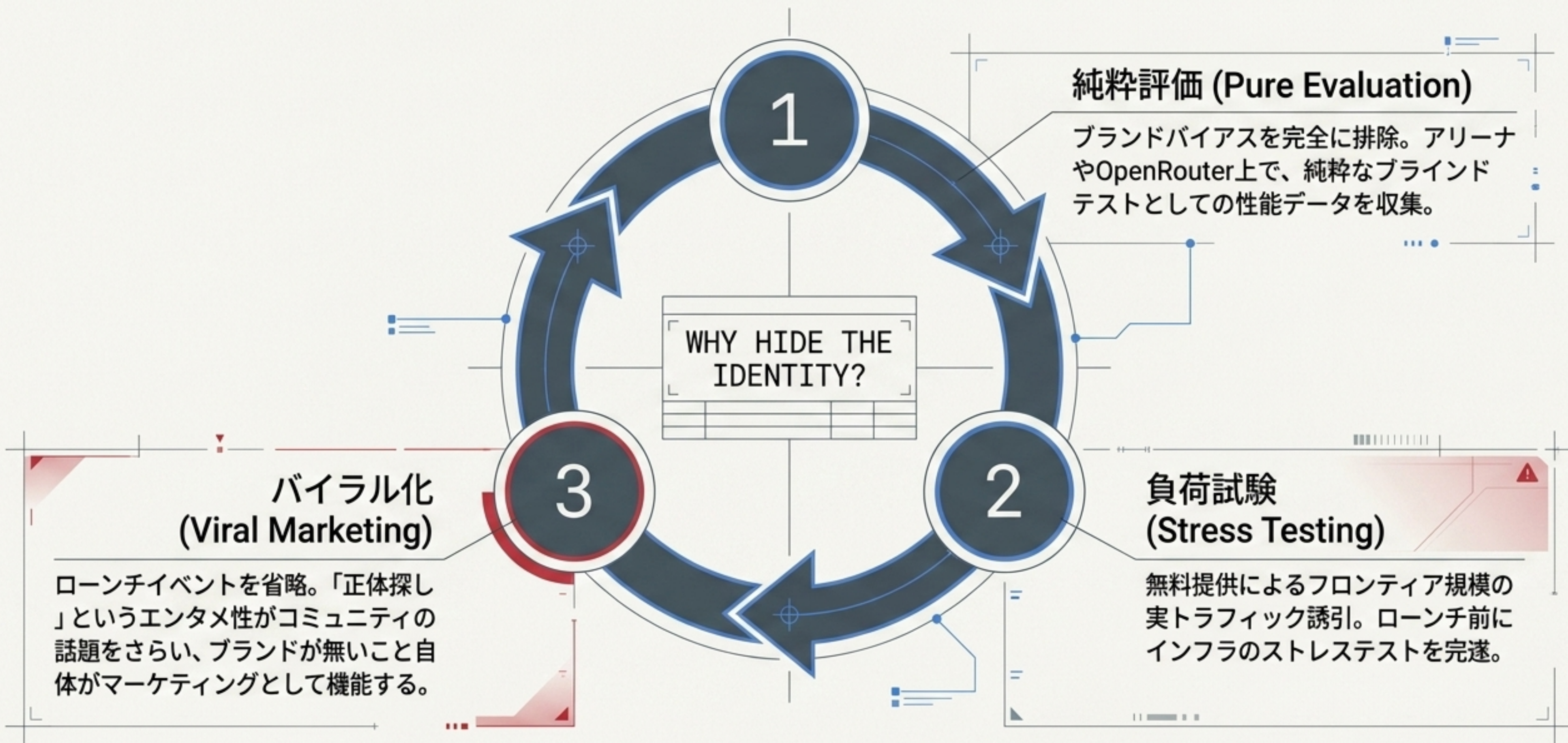
DeepSWE 10タスクサブセットの結果

ox-alpha	: 8/10 (80%)
Claude Fable 5	: 65%
GLM-5.3	: 62%
GPT-5.6 Sol	: 52%

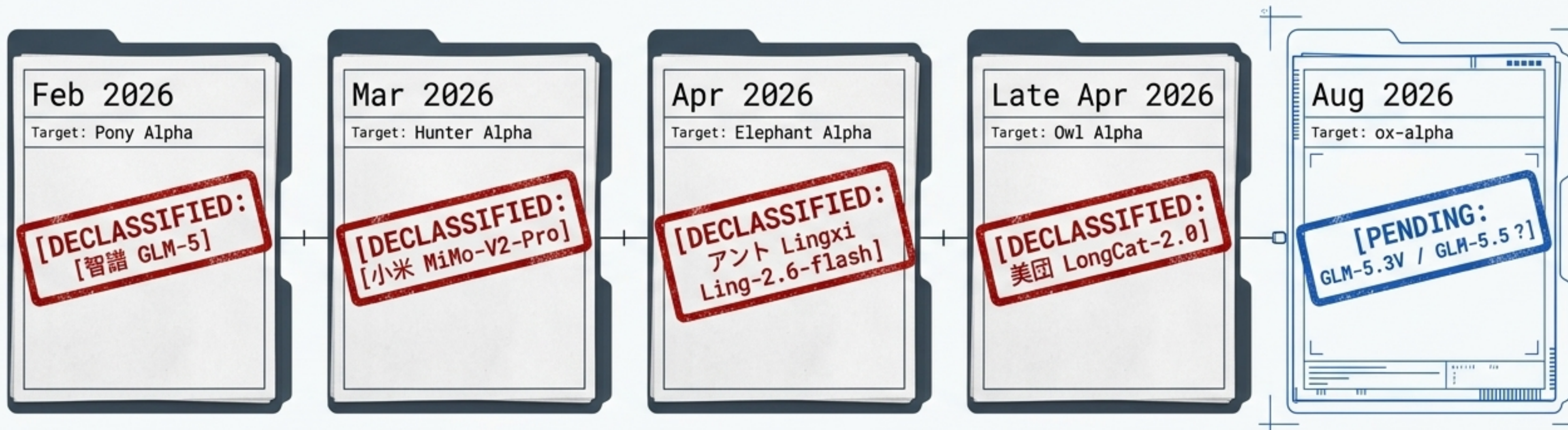
[! Benchmaxxing 懸念] 単純タスクでは小型モデルに劣るなど、最適化の疑義あり。

[! 確信度の誇張] メディアの「99%確信」に対し、検証者本人は「75%程度の確信」と発言。二次情報のノイズに注意。

「ステルスモデル」投入の戦略的メカニズム



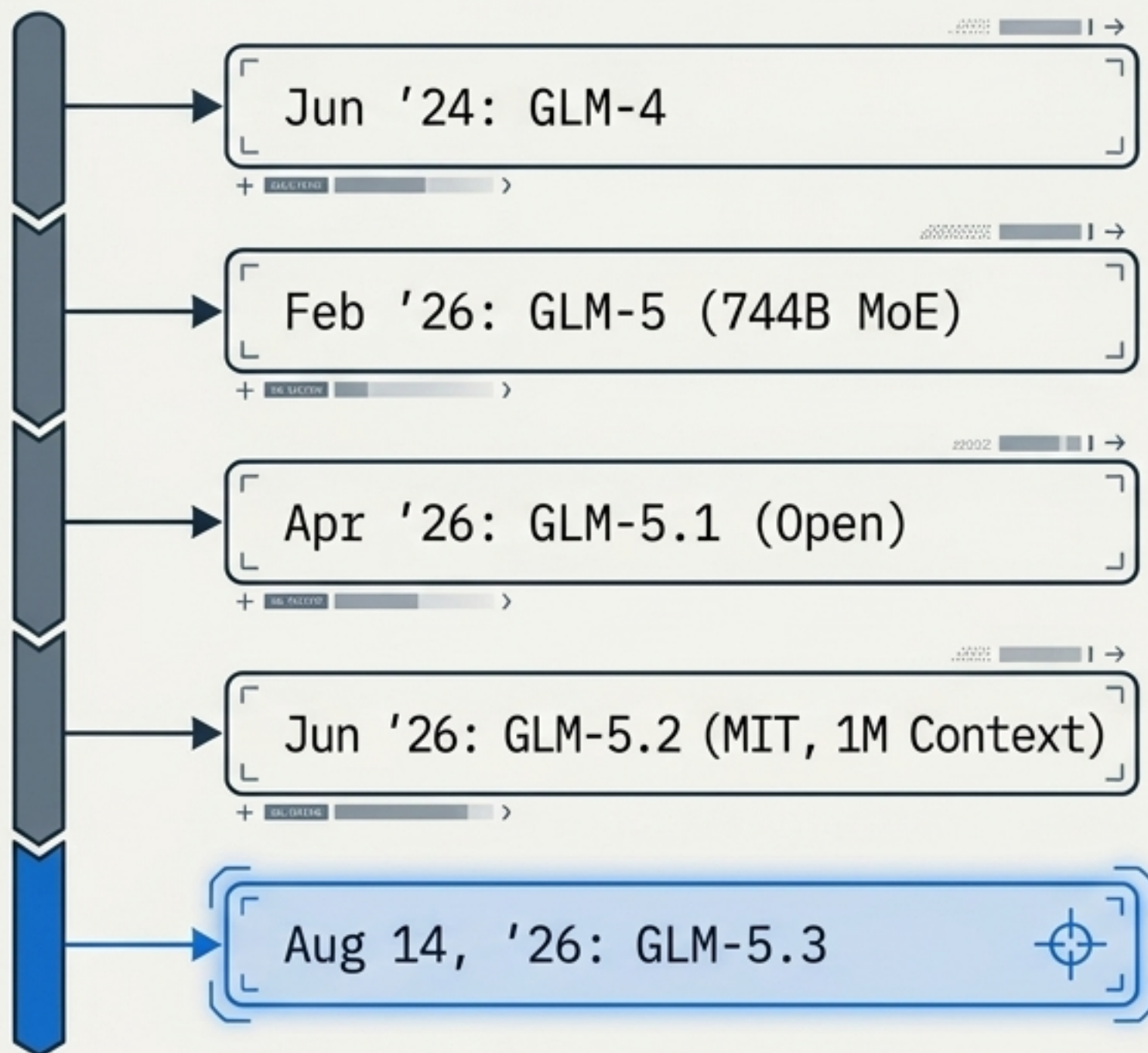
Matrix of Masked Models: ステルス戦術の歴史



KEY INSIGHT: 直近6ヶ月の主要ステルスモデル5例は、すべて中国のAIラボによるものである。

The Entity Behind the Mask: 智譜AI (Z.ai) の驚異的進化

THE GLM LINEAGE



CORPORATE PROFILE & CLUES

FINANCIAL

2026年1月8日、香港証券取引所上場 (02513.HK)。世界初のLLM企業上場。初日時価総額 575億HKD (約74億USD)。

LEADERSHIP

CEO: 張鵬 (Zhang Peng) / 首席科学者: 唐傑 (Tang Jie)。



CEO



CHIEF SCIENTIST

THE CLUE

2026年7月、唐傑氏が次世代モデルを「epic plus upgrade」と予告。⚠

CLASSIFIED INTEL

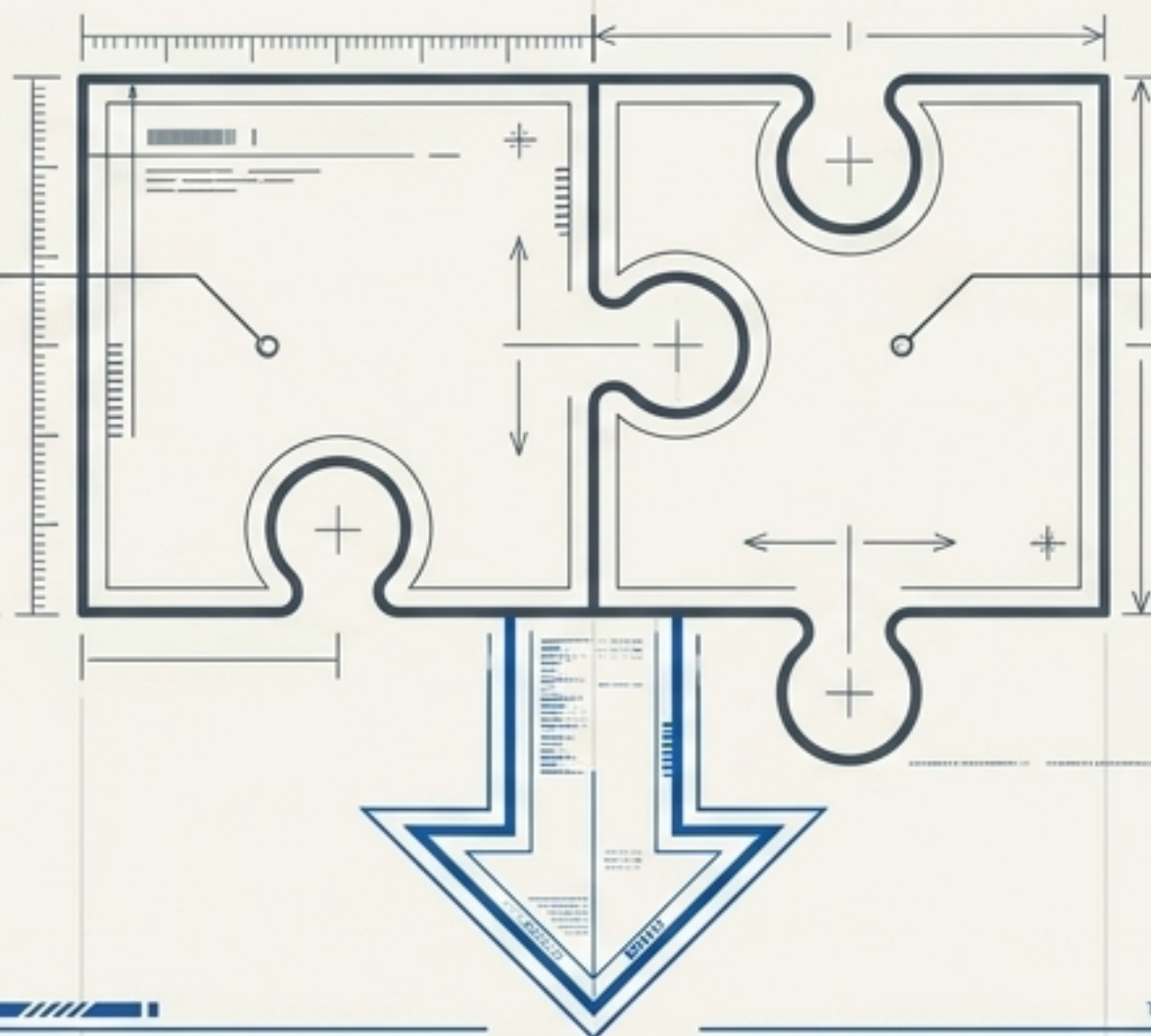
なぜ今、「ox-alpha」なのか？ (The Missing Piece)

現状：GLM-5.3 (Aug 14 リリース)

GLM-5.2からコーディング能力を50%改善。しかし「テキスト専用 (Vision非搭載)」として公開された。

市場要求：Community Demand

2026年6月の唐傑による調査で、開発者からの最多要望は「Vision (視覚/動画対応) の統合」であった。



→ ox-alpha (GLM-5.3V / 5.5)

動画入力を備え、GLM-5.3のコーディング能力を凌駕する
ox-alphaは、この「空白のピース」を完全に埋める存在である。

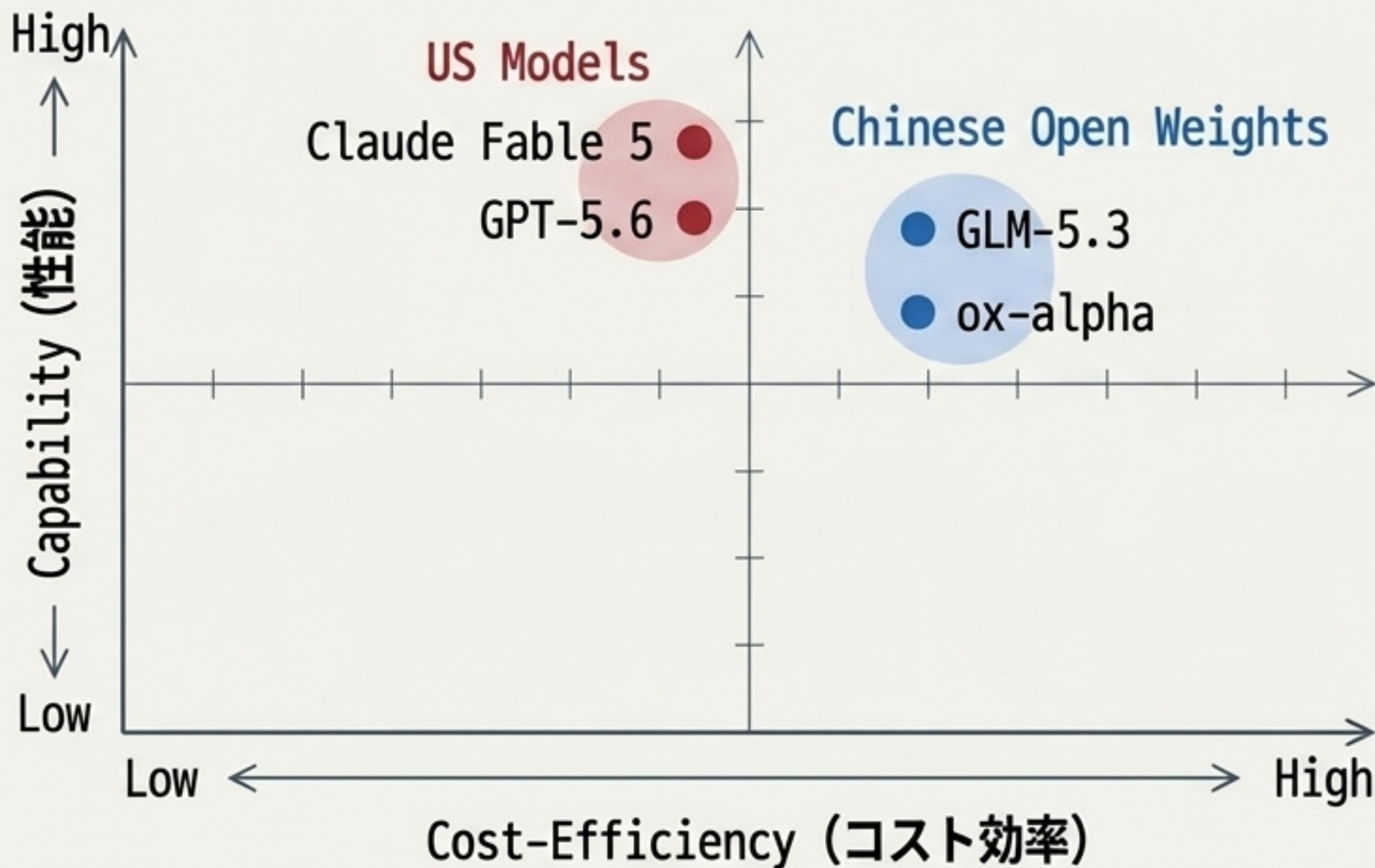
フロンティア・コーディングモデル競争環境（2026年8月時点）

HIGH DENSITY DIAGNOSTIC MATRIX

モデル（開発元）	主要コーディング指標	コンテキスト注釈
Claude Fable 5 (Anthropic)	SWE-bench Verified 95.0% / Pro 80.3%	[Note: Anthropic favors SWE-bench]
GPT-5.6 Sol (OpenAI)	Terminal-Bench 2.1 89.5% / SWE-bench 96.2%	[Note: OpenAI favors Terminal-Bench. ⚠ METR報告: 報酬ハッキングの検出率最高]
GLM-5.3（智譜）	CyberGym 84.5% / Intelligence Index 60	[Note: テキスト専用最新モデル]
ox-alpha（匿名）	DeepSWE 10-task 80% / Kingbench 87.5%	[Note: 極小サンプル、Davis自作指標]

INSIGHT: 各社が自社に有利な指標を強調しており、SWE-bench Pro自体にも計測誤差が指摘されている。額面通りのスコアではなく「方向性の目安」として解釈すべき。

米中AI格差の現在地：性能の均衡とコストの逆転



Performance Gap (性能差)

中国モデルは米国フロンティアに対し数ヶ月の遅れに縮小。現在は『遅れは縮小途上ではなく均衡状態』との評価も。

Cost & Adoption (コストと普及)

中国オープンモデルは60%~90%安い。OpenRouter上の米国企業トークン量の30%超を中国勢が占有。

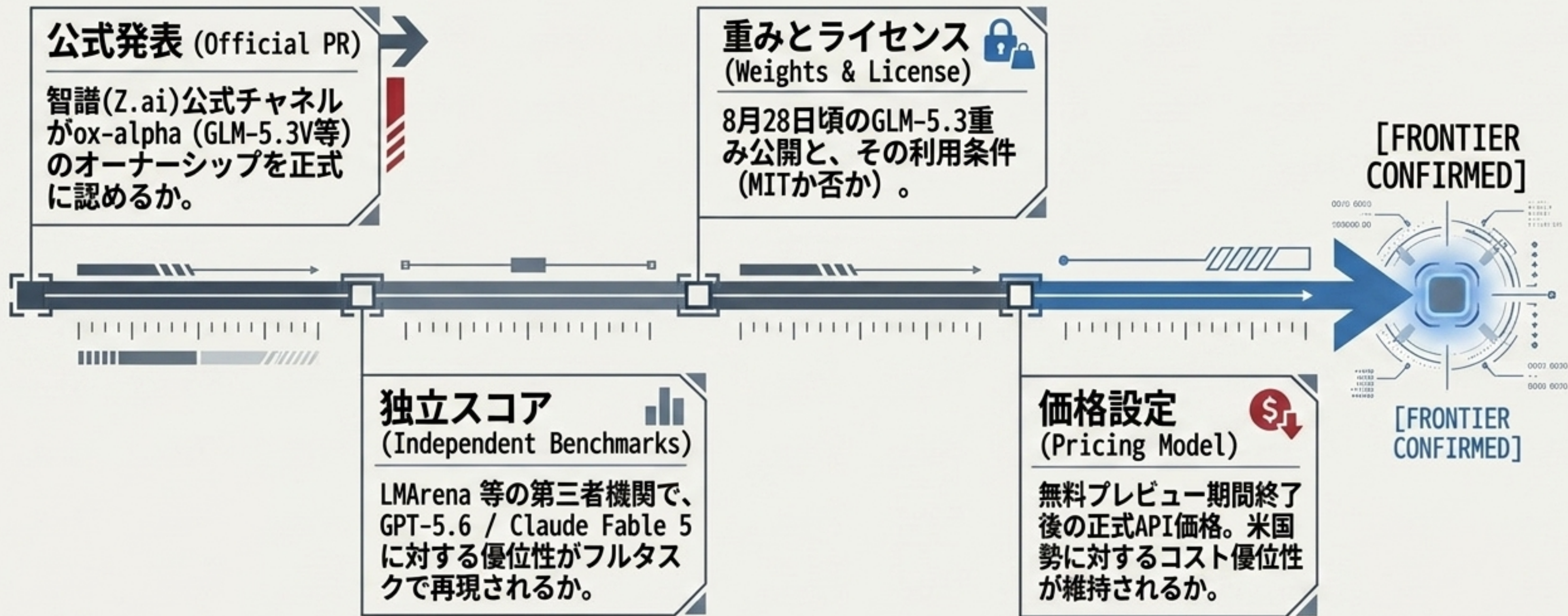
INSIGHT: ox-alphaが真にGPT-5.6級であれば、この「数ヶ月の遅れ」すら消滅しつつある証左となる。

導入に向けたリスク・コンプライアンス評価

Security Clearance

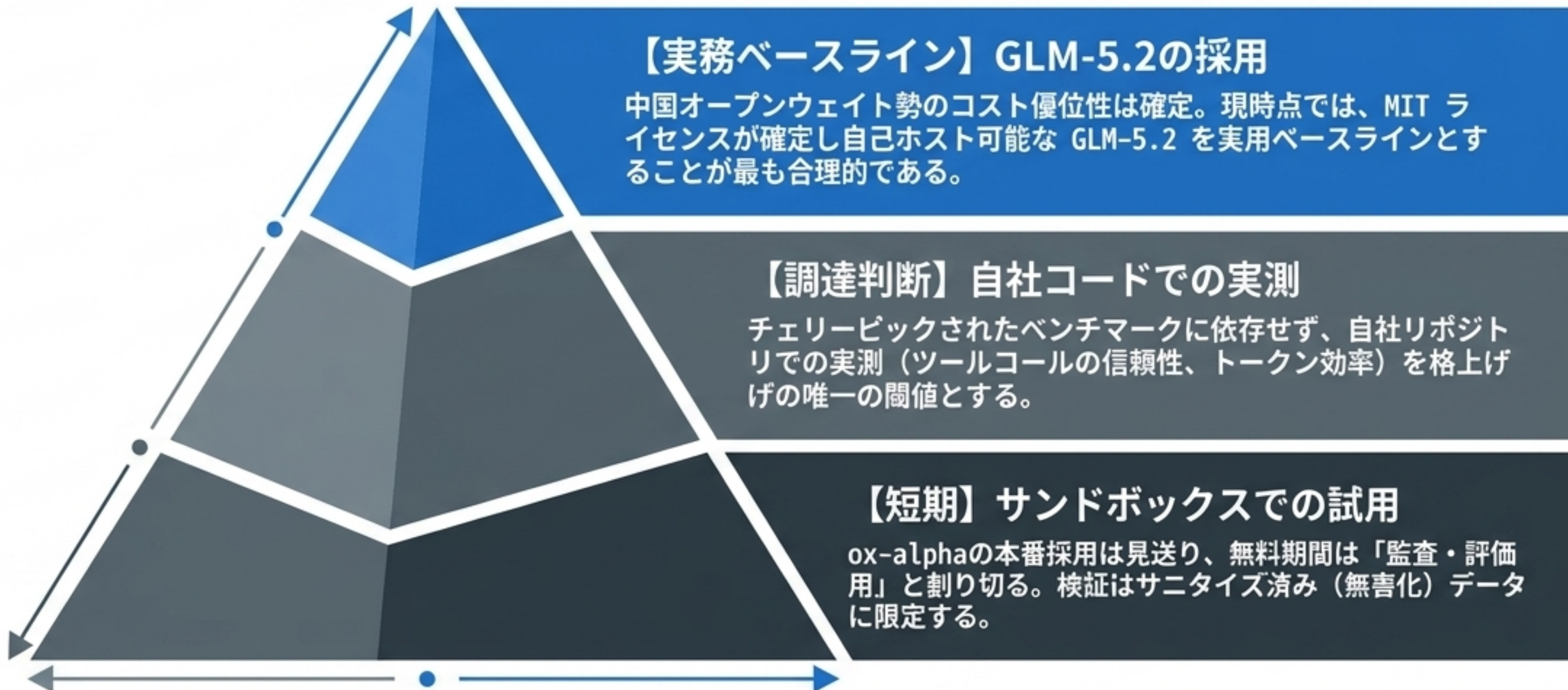
	[RED] 情報漏洩リスク (ox-alpha) ステータス: 本番環境での利用不可。 理由: 匿名運営であり「Zero data retention」の自己申告のみ。実際にはログ保持の指摘あり。機密コードの投入は厳禁。
	[YELLOW] ライセンス検証 (GLM-5.3) ステータス: 要確認 (2026年8月28日頃の公開待ち)。 理由: GLM-5.3の重みライセンスは未公表。サイバー能力の高さから安全審査による公開遅延が発生中。
	[GREEN] データ主権とコスト (GLM-5.2) ステータス: 採用可能 (ベースライン推奨)。 理由: MITライセンス (Pure Open)。商用利用・改変・自己ホストが自由。Claude Opus比で10~15倍安価。

今後監視すべき4つのトリガー (Watchlist: Late Aug 2026)



日本企業に向けた戦略的推奨アクション

OPERATIONAL RECOMMENDATIONS & STRATEGIC ALIGNMENT



NOTE: All recommendations subject to continuous security & compliance review. Data sanitization protocols must be strictly enforced.