

AI Strategic Digest

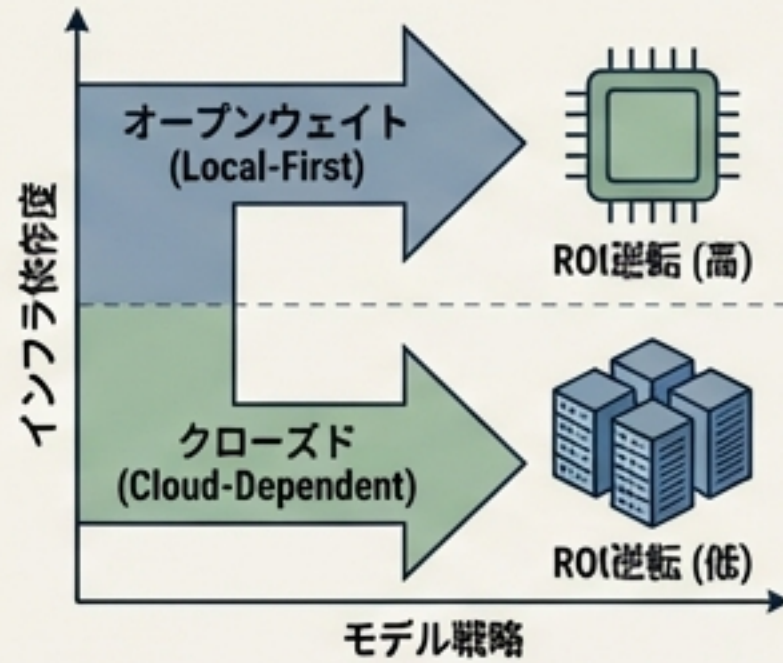
2026年4月の構造変化と実務へのインサイト
ー ハイプからリアリティへ。オープンモデル、
セキュリティ、そしてエコシステムの再編。

DATE:	2026-04-03
TARGET:	CTO / SENIOR ENGINEER / AI STRATEGIST
FORMAT:	VISUAL READING DECK

2026年4月・4つのパラダイムシフト

[Divergence] モデルとインフラの二極化

オープンウェイト (Gemma 4) とクローズド (Qwen3.6-Plus) の戦略的分岐。ローカル回帰によるROIの逆転。



[Paradox] 信頼性の境界 (ZEH)

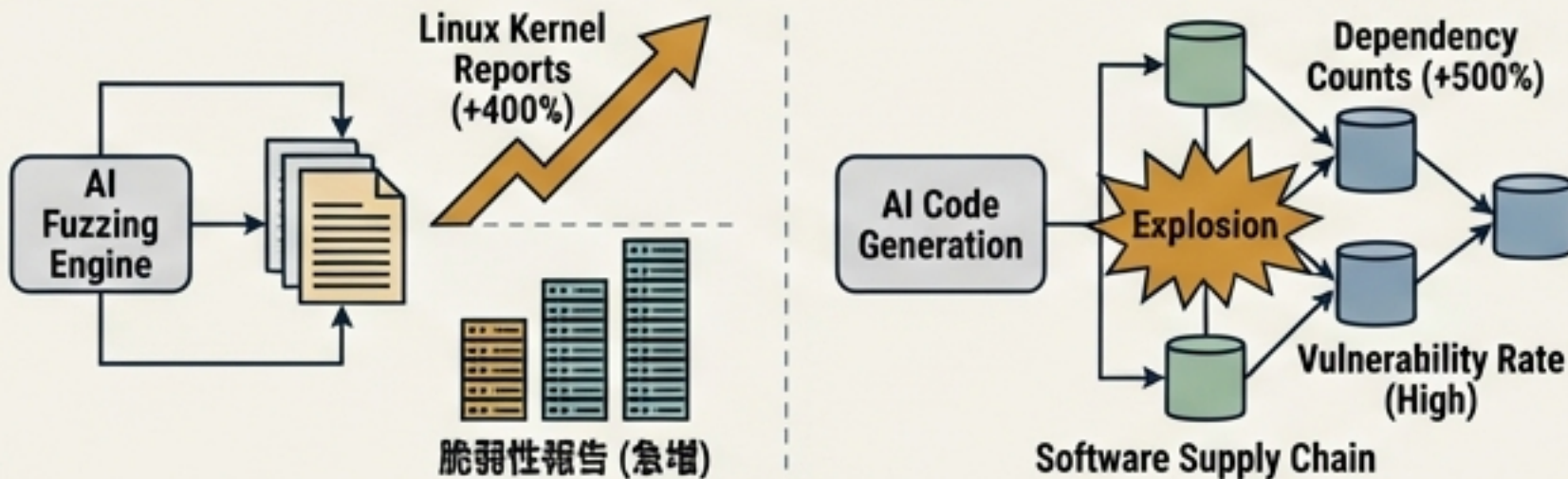
高度化するLLMが抱える System 1思考の限界。「絶対に間違えない入力長」の定量化。



思考モード	System 1 (限界)	安全閾値
定量的指標	ZEH	明確化

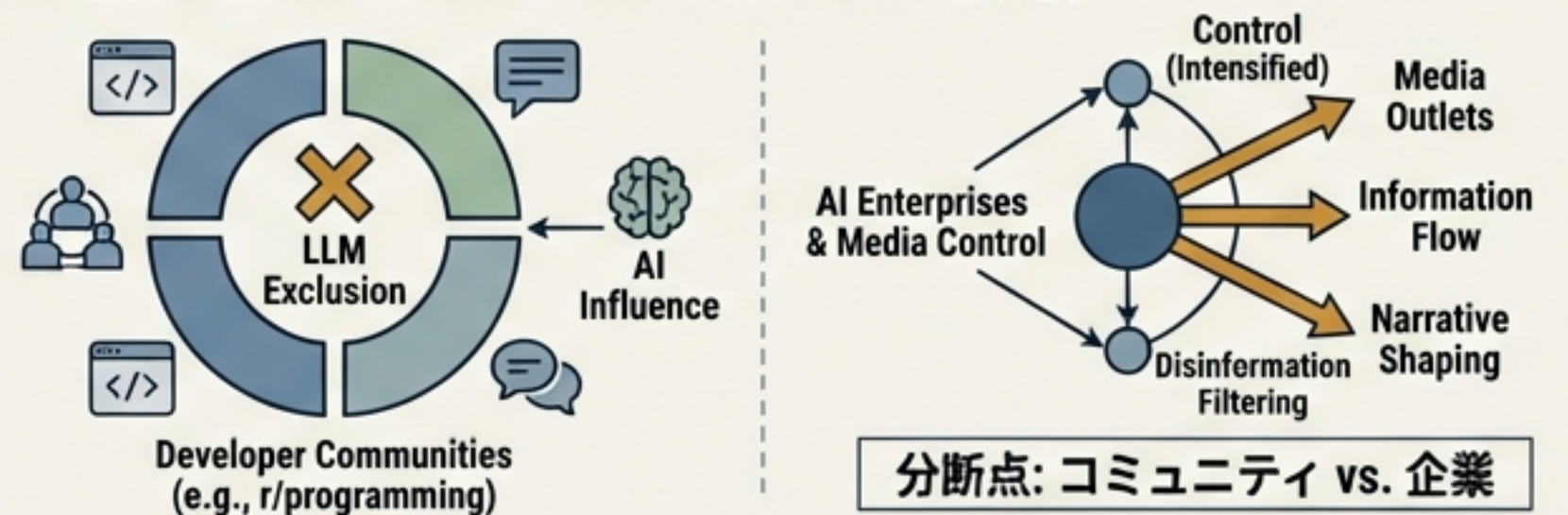
[Friction] 非対称戦と脆弱性

AIファジングによるLinuxカーネル報告急増と、AIコード生成が生むサプライチェーン爆発。



[Synthesis] エコシステムの分断

開発コミュニティ (r/programming) でのLLM排除と、AI企業によるメディアコントロールの激化。



The Open vs. Closed Matrix: Gemma 4 vs Qwen3.6-Plus

Google Gemma 4

戦略 (Strategy)

マルチモーダル・オープン標準 (Apache 2.0)

性能指標 (Performance Metrics)

31B Dense (MMLU-Pro 85.2%, GPQA 84.3%, LiveCodeBench 80.0%) / E2B・E4Bは音声対応。

ユースケース (Use Cases)

ファインチューニング基盤。26B MoEでのローカル画像生成 (ペリカンSVG生成で高評価)。

課題点 (Issues)

ツール呼び出しの失敗 (UnixタイムスタンプのPythonスクリプトを実行せずに誤答)、ローカル推論のバグ (31Bで「---」出力)。

Alibaba Qwen3.6-Plus

戦略 (Strategy)

Bティア・コスト重視 (ホスト専用/クローズド)

性能指標 (Performance Metrics)

パラメータ数非公開。Opus 4.5 / Gemini 3.0レベルを安価に提供。

ユースケース (Use Cases)

コストを重視するAPI利用者 (OpenRouter無料枠)。

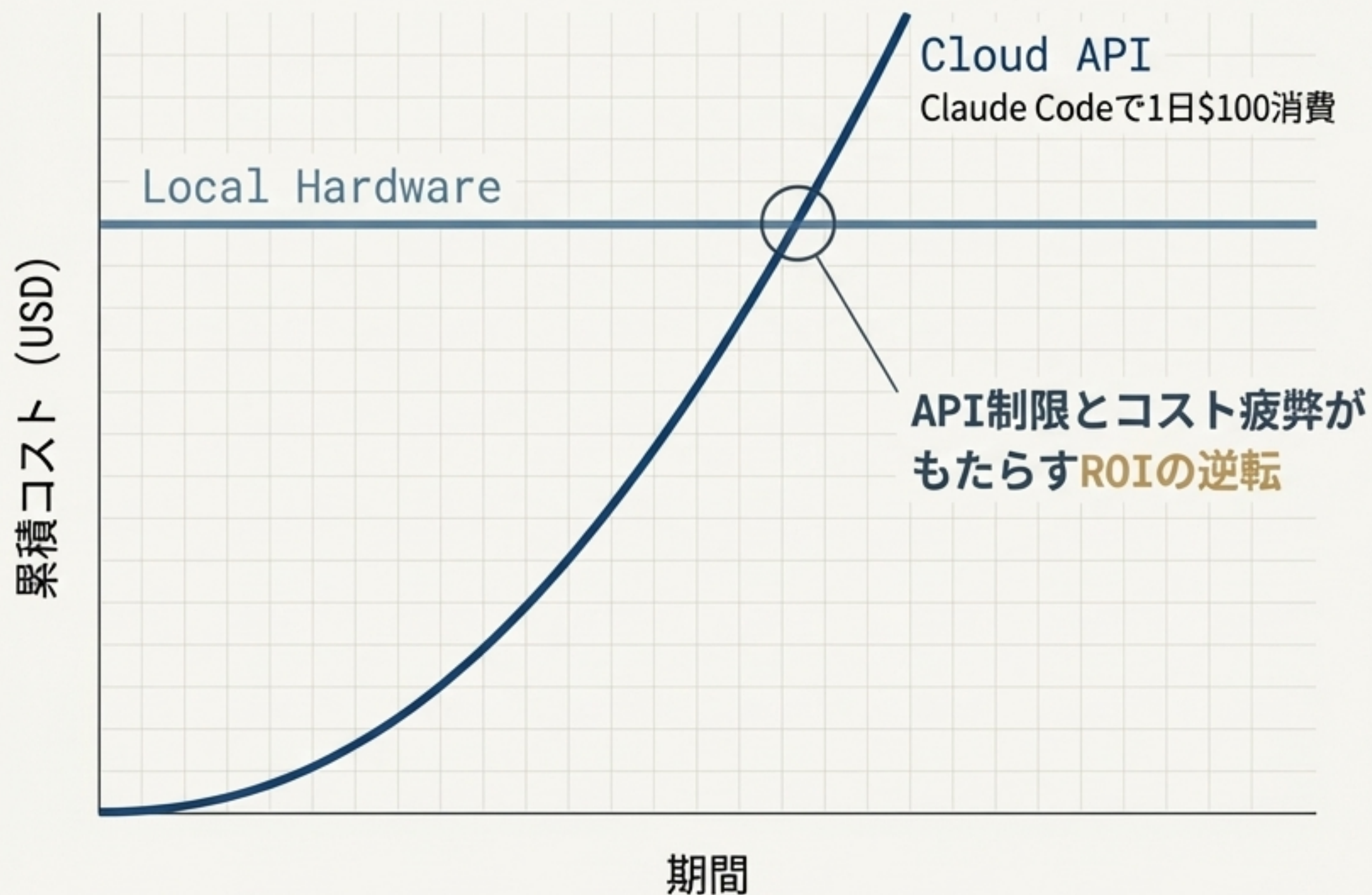
課題点 (Issues)

ベンチマーク比較対象の意図的な選定 (Opus 4.6ではなく4.5との比較)。

The ELO Trap

「ELOスコアの前面表示はミスリーディング (antirez)」。実際にはQwen 3.5 27BがGPQA (85.5%) や HLE-t (48.5%) でGemma 4 Denseを上回り、実コーディングテストでも完成度で勝る。

逆転するROI：ローカルLLM回帰の波



MBP M5 Max 128GB
+ MLX 8-bit

Qwen3-Coder-Next
(Sonnet 4.5相当) がローカルで実用可能に

NVIDIA GB10

デュアルGPU構成の1/5の消費電力で120B MoEを稼働
(システムメモリ制約には注意)

Orchestration

AMD Lemonade (統合ランタイム /
Open WebUI・VSCode連携)

Capabilities

Text Generation, TTS, STT, Vision

Hardware (ROCm/Vulkan)

CPU

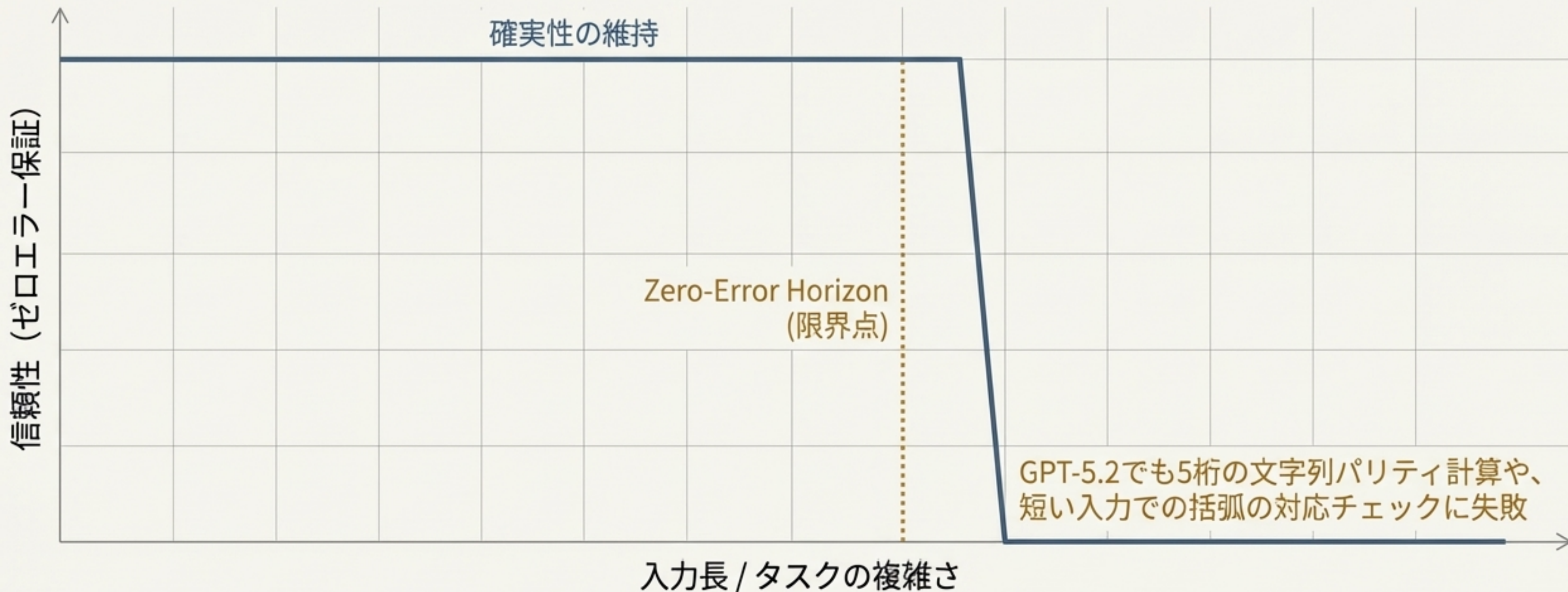
GPU
(7900 XTX /
120Bモデルで50t/s)

NPU
(Ryzen AI / 小型
モデル常時起動)

Snap/PPA/RPM提供: 依存関係
(ROCm等) を吸収し、「イン
ストールしたら動く」体験を実
現。

■ **トレードオフ (Trade-offs):**
Ollamaほどのポータビリティは
なく、NPUカーネルはプロプラ
イエタリ。Docker非対応がコ
ンテナ派の障壁に。

■ 「統合サーバーが重い場合は
llamacpp-rocm の単体利用も
選択肢。」



System 1 (LLMの直接処理)

「速い思考」。确实な計算には不適合。

System 2 (ツール・コード実行)

「遅い思考」。「LLMにパリティ計算をさせるな。パリティ計算するスクリプトを書かせろ」。

AIによる非対称戦：脆弱性発見とサプライチェーン

The Creation Threat: サプライチェーン爆発



- LiteLLM侵害（40分で50万台に影響）、axios npm攻撃。
- AIコーディングアシスタントが安易に外部パッケージをインポート。SOC2では防げず、推移的依存関係が爆発。

The Discovery Threat: メンテナンス崩壊



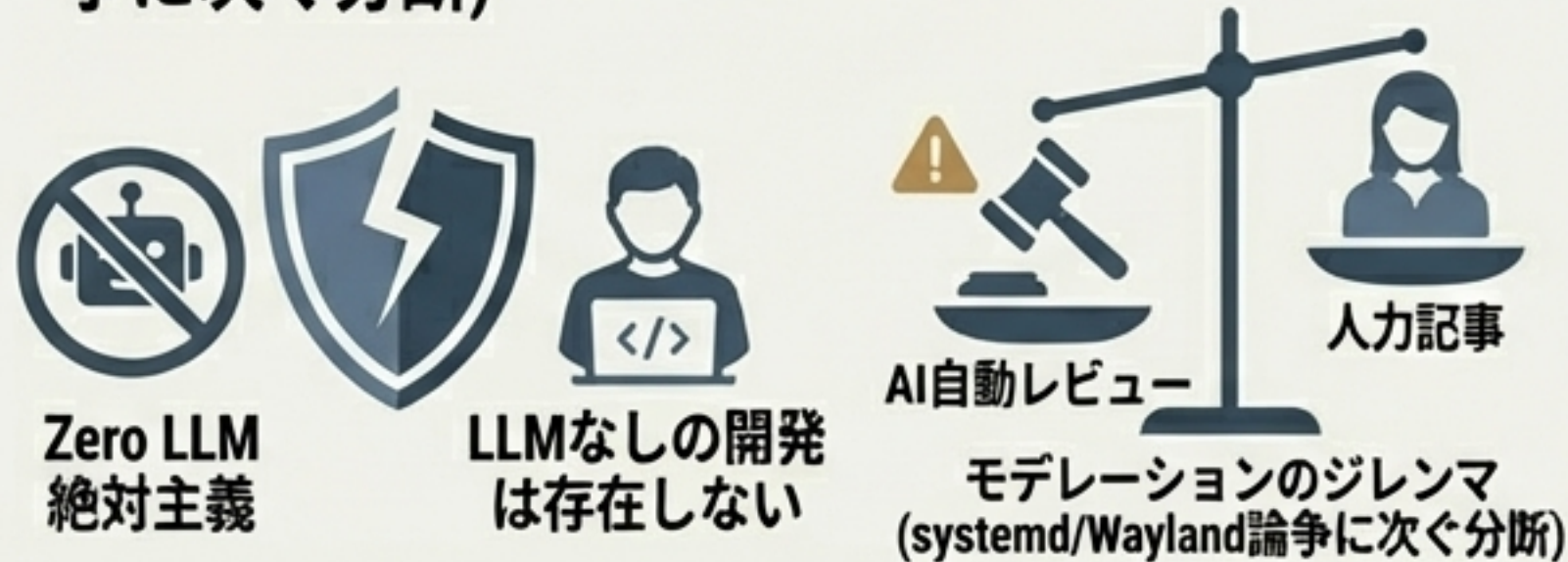
- Linuxカーネルの脆弱性レポート急増（LWN.net）。
- AIファジングが長年の未知バグを発掘。「バグ発見速度が修正速度を上回る」。

「リリースしたら洞窟に戻るモデルの終焉」－定期アップデート自動化（unattended-upgrades / kexec）と標準ライブラリ活用が必須に。

分断されるエコシステムと企業のメディア戦略

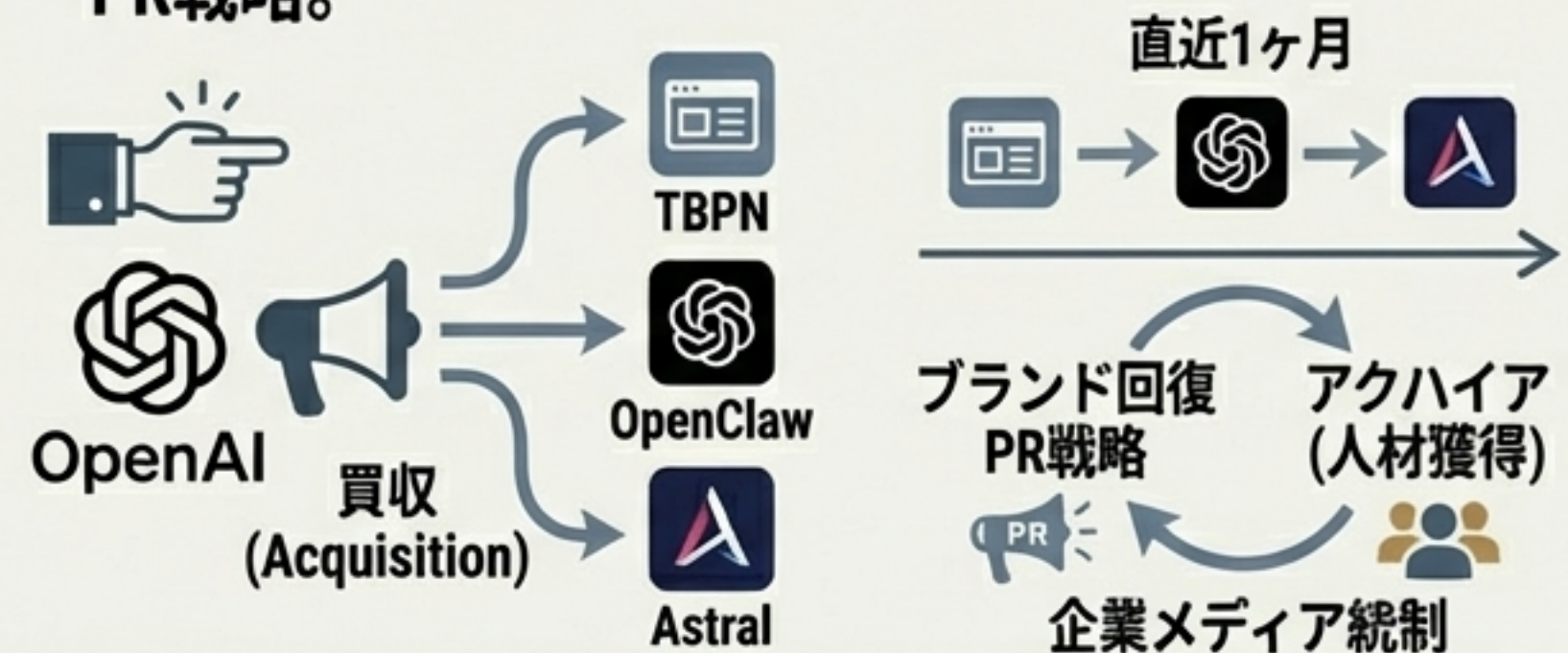
Community Rejection: コミュニティの反発

- 4月2日、Reddit最大級のr/programmingがLLM関連議論・生成コードを全面禁止。
- 「ゼロLLM絶対主義」vs「LLMなしの開発は存在しない」。AI自動レビューが人力記事を誤判定するモデレーションのジレンマ。(systemd/Wayland論争に次ぐ分断)



Corporate Media Control: 情報の統制

- OpenAIがTBPN（登録者5.8万人）を買収。直近1ヶ月でOpenClaw, Astralに続く3件目。
- 再生数3,000程度のメディア買収は「アクハイア（人材獲得）」および毀損したブランド回復のPR戦略。



Risk Warning: モデル比較記事やレビューを読む際、背後にある「資本関係（メディア中立性の喪失）」の確認が必須スキル化。

Actionable Insights : 次の一手

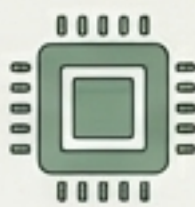
Adopt (採用)



- 標準ライブラリの徹底活用 (AIコード生成による推移的依存関係の削減)



- Linuxカーネル等への定期アップデート自動化 (CVEベースのトリアージ構築)



- Denseモデル (24B~27Bクラス) を用いたローカル推論のPoC (Ollama / LM Studio)

Monitor (監視)



- オープンモデル (Gemma 4等) のツール呼び出し精度の進化



- AIメディア・ベンチマークの資本背景 (誰が資金を出しているか)

Avoid (回避)



- 表面的なベンチマーク (ELOスコア) のみでのモデル選定



- LLMへの直接的な計算タスクの丸投げ (ZEHの教訓に基づくSystem 1利用の禁止)



- クラウドAPIのみに依存した無自覚なスケーリング (TCOの定期的な再評価)