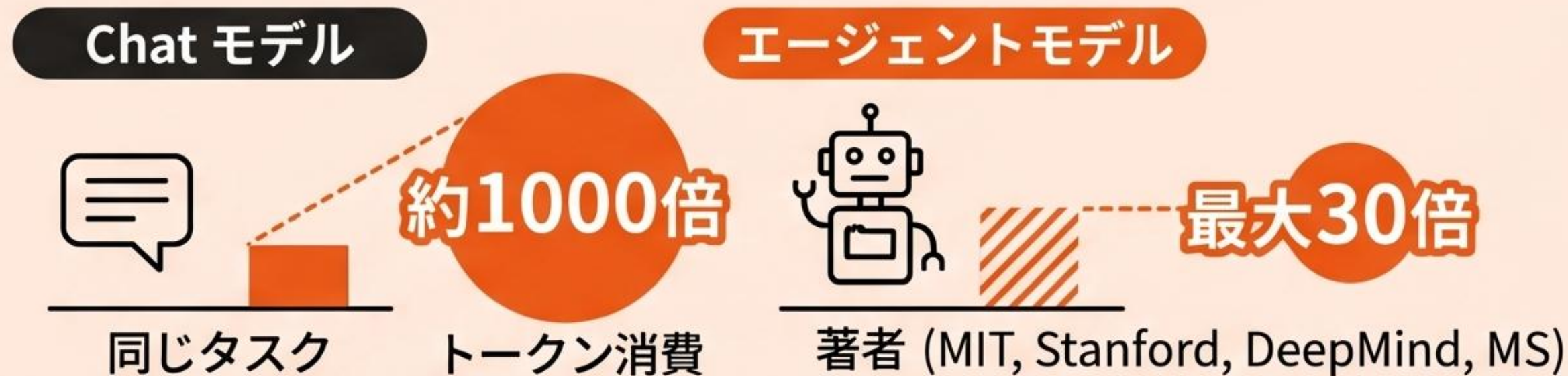
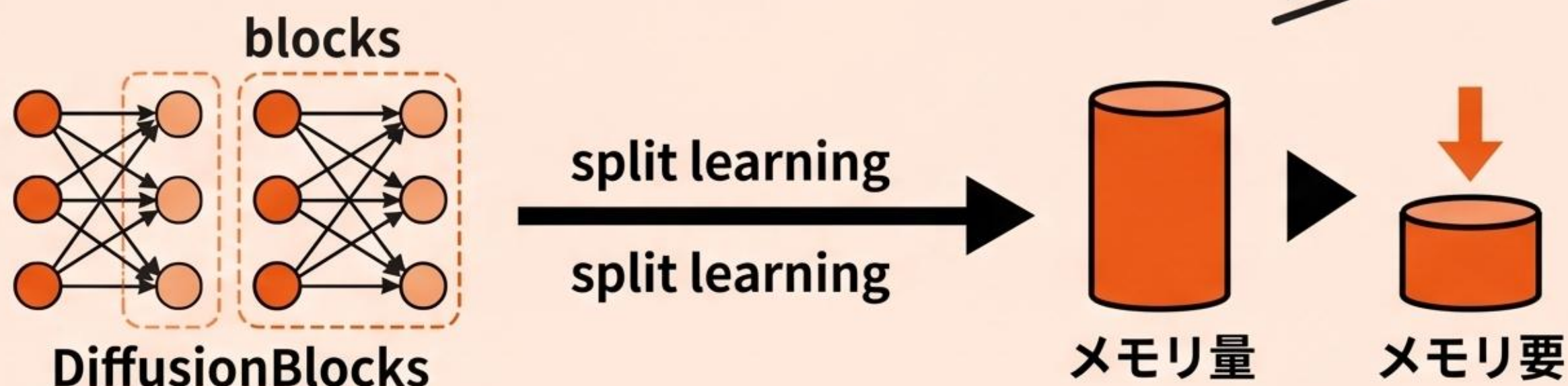
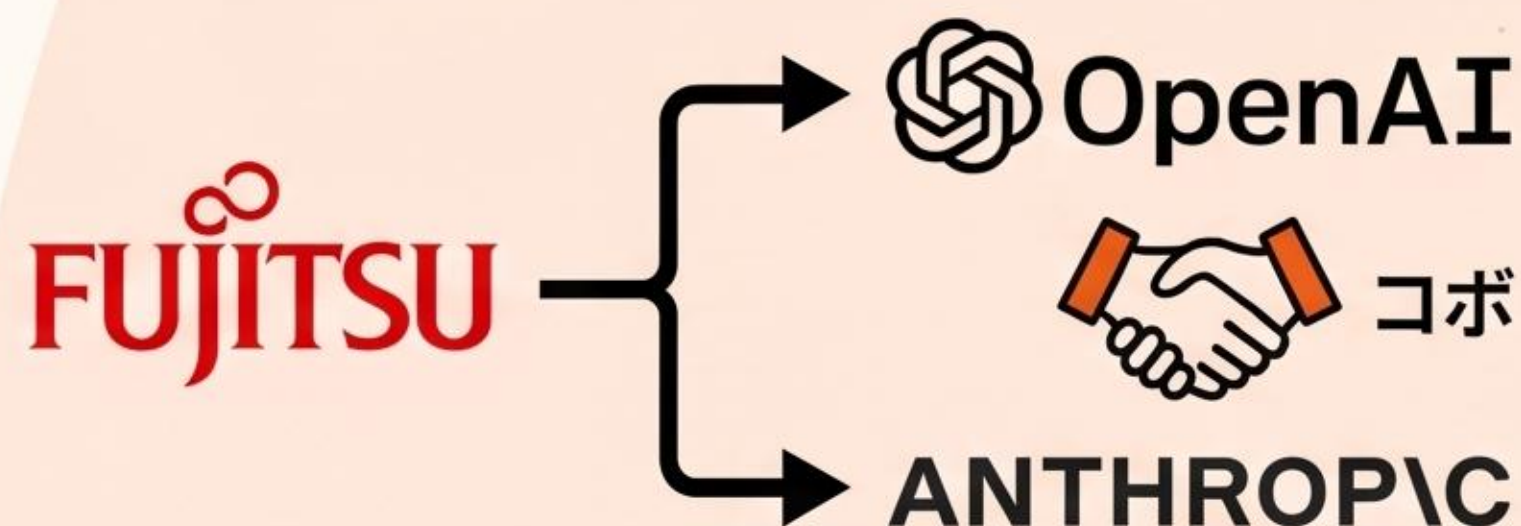




## 今朝のホットな話題

1. 富士通が OpenAI と連携開始、Anthropic とも「ミュトス」早期アクセスで協業
2. Sakana AI 『DiffusionBlocks』 — ネットワークをブロック単位で分割学習しメモリ要件を削減
3. MIT・Stanford・DeepMind・MS 論文：エージェントは chat の約1000倍トークン消費、同タスクで最大30倍コスト差



**7** トピックを整理。

# 1. 富士通が OpenAI と連携開始、Anthropic とも「ミュトス」早期アクセスで協業

## 🔍 何が起きた？

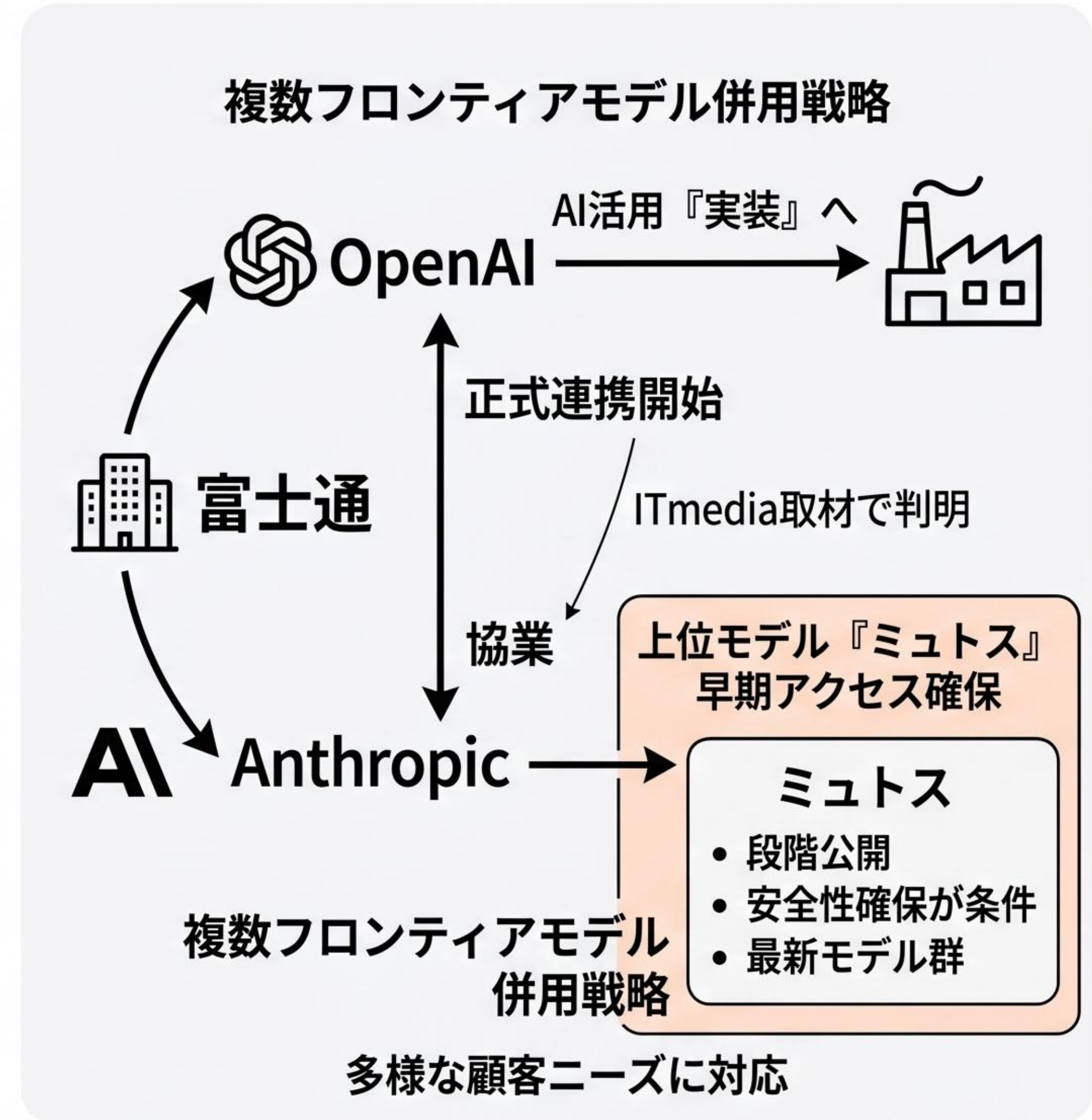
富士通が OpenAI との連携を正式開始し、AI 活用を PoC 段階から『変革の実装実装』へ進めると発表した。同じ時期に Anthropic とも協業しており、ITmedia の取材では Anthropic が段階公開を進める上位モデル『ミュトス』への早期アクセスを確保していることが明らかになった。日本の大手 Sler が OpenAI と Anthropic の双方と同時に手を組む構図になっている。

## 📌 主な変更点

- 富士通が OpenAI と連携開始、企業変革の『実装』フェーズを掲げる
- 並行して Anthropic とも協業し、最新モデル『ミュトス』への早期アクセスを確保
- 単一ベンダー依存ではなく複数フロンティアモデルを併用する戦略
- ミュトスは Anthropic が安全性確保を条件に段階公開している上位モデル群

## 💡 なぜ重要？

日本の大手 Sler が単一ベンダー依存を避け、世界的なフロンティアモデル企業 2 社と並行して提携。これにより、顧客企業の多様なニーズに対応し、AI による真の企業変革『実装』を加速させる。ITmedia の取材によりこの戦略的提携が明らかになった。



## 2. Sakana AI 『DiffusionBlocks』 — ネットワークをブロック単位で分割学習しメモリ要件を削減

### 🔍 何が起きた？

日本拠点の Sakana AI が『DiffusionBlocks』を発表した。ニューラルネットワークを拡散（diffusion）として捉え直し、全層を一度にメモリへ載せずにブロックごとへ分割して独立に学習できるようにする手法で、大規模モデル学習のメモリ要件を下げることを狙う。

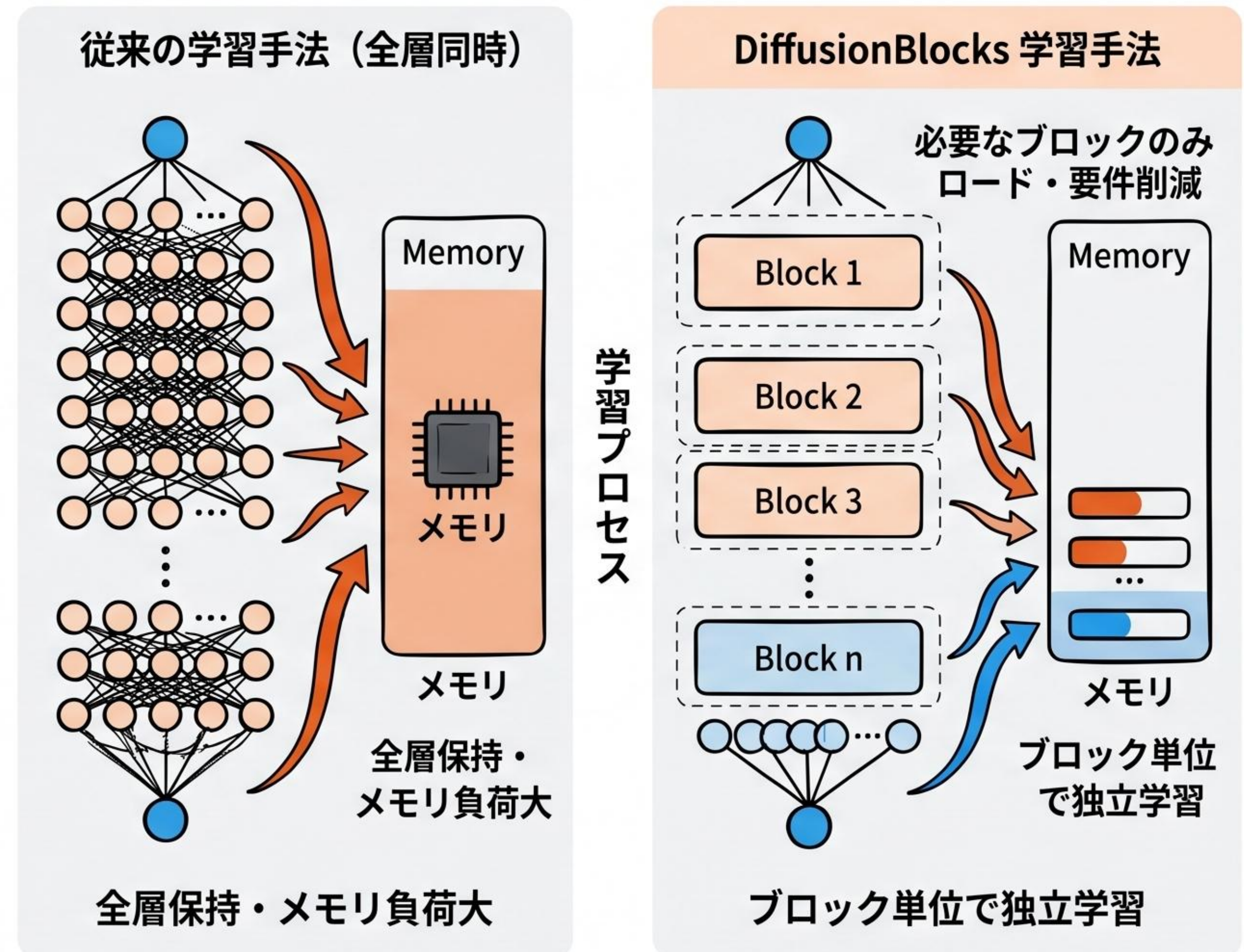
### 📌 主な変更点

- ネットワークを『拡散過程のブロック』として捉え、ブロック単位で分割学習
- 全層を同時にメモリ保持する必要をなくし学習時のメモリ要件を削減
- 限られたリソースで大規模モデルを訓練するアプローチの一つ
- 論文・詳細は [pub.sakana.ai/diffusionblocks](https://pub.sakana.ai/diffusionblocks) で公開

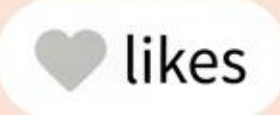
### 💡 なぜ重要？

限られたリソースで大規模モデルを訓練できるようになる。日本拠点のスタートアップによる革新的アプローチ。

### DiffusionBlocks 学習手法の概念図



# 3. MIT・Stanford・DeepMind・MS 論文：エージェントは chat の約1000倍トークン消費、同タスクで最大30倍コスト差



## 🔬 研究の核心

共同研究（MIT・Stanford・DeepMind・MS 論文：エージェント開発において、生成コードの出力コストではなく、コンテキストを読み込む際のトークン消費がコストの主因であることを確認しました。

## 📊 主な知見

- エージェントはチャットの約1000倍のトークンを消費
- コストの主役は生成コード出力ではなくループ毎の context 再読み込み
- 同一タスク・同一モデルでも実装設計でコストが最大30倍変動
- トークン単価の低下だけでは AI 運用コストは下がらない

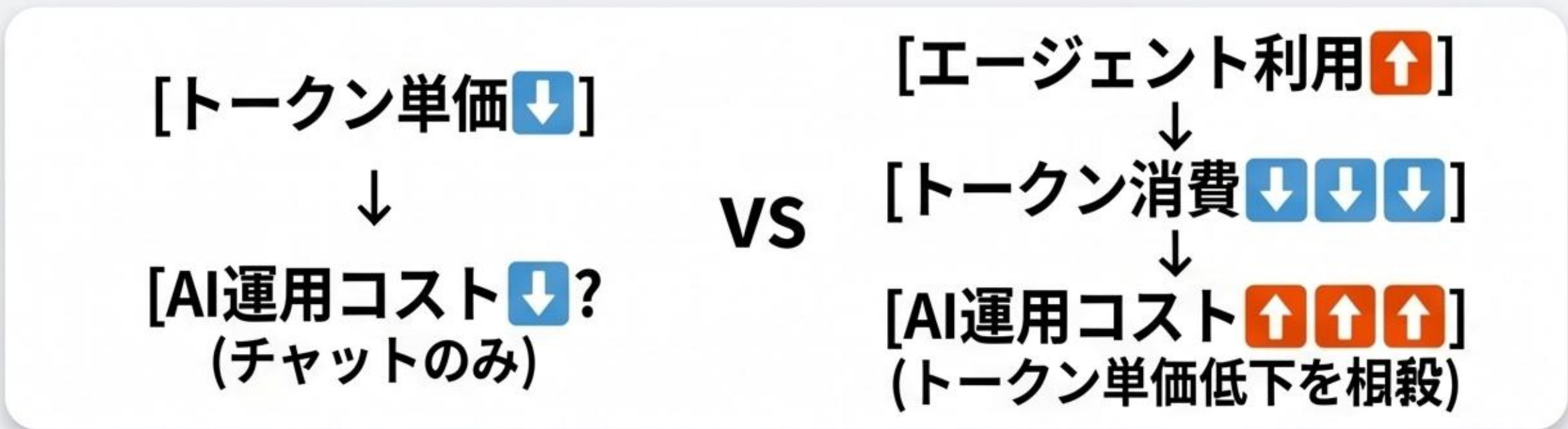
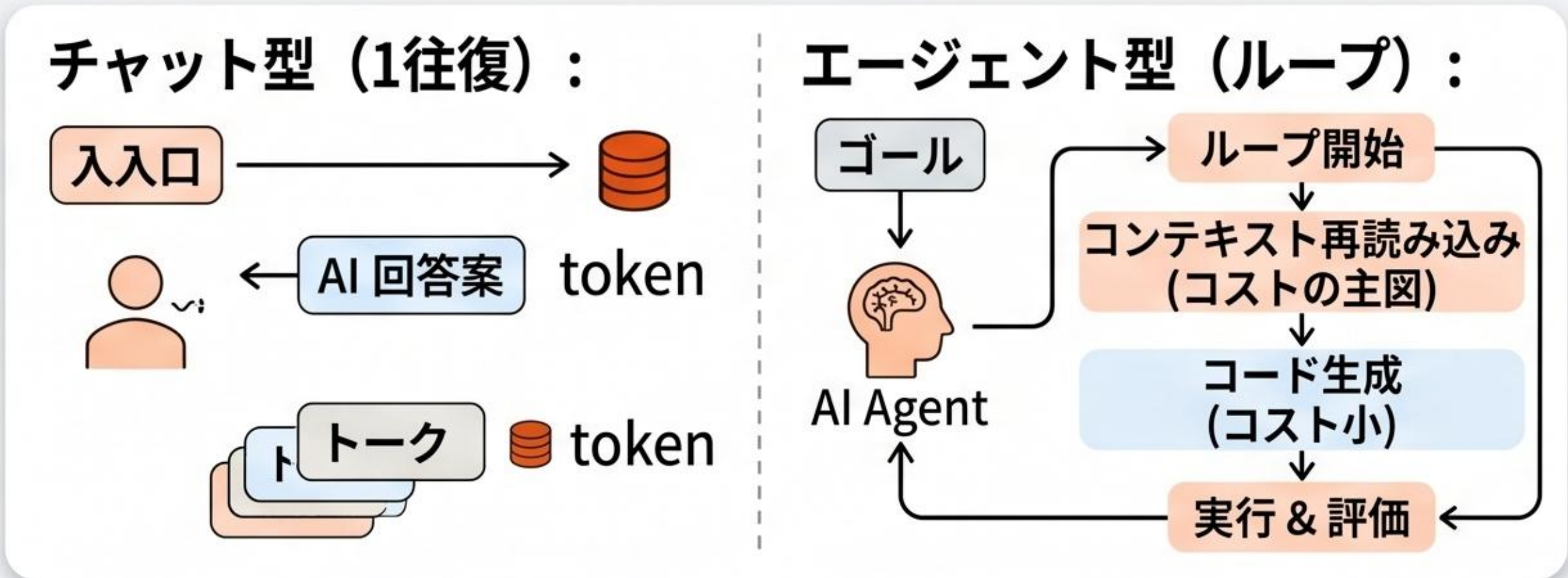
## 🎯 含意

（含意）この研究は、LLMエージェント運用の経済性を再評価するものです。トークン単価が低下しても、エージェント利用の増加とコンテキスト再読み込みによるトークン消費が劇的に増えるため、全体の運用コストは下がらない可能性があります。真のコスト削減には、LLM単価の低下だけでなく、コンテキスト管理の効率化が鍵となります。

エージェント  
トークン消費量  
**約1000倍**  
vs チャット利用



同一タスク  
コスト差  
**最大30倍**  
実装設計による変動



# 4. Microsoft 365 Copilot が Claude Opus 4.8 を追加、 Vercel AI Gateway でも即日提供

## 🔍 何が起きた？

Anthropic の Claude Opus 4.8 発表当日、Microsoft が Microsoft 365 Copilot にモデル選択肢として Opus 4.8 を追加すると発表。まず Copilot Cowork (Frontier) で利用可能になり、Copilot Chat・Excel・PowerPoint・Copilot Studio へ順次展開される。同日Vercel も AI Gateway 上で anthropic/claude-opus-4.8 をルーティング対象に加えた。

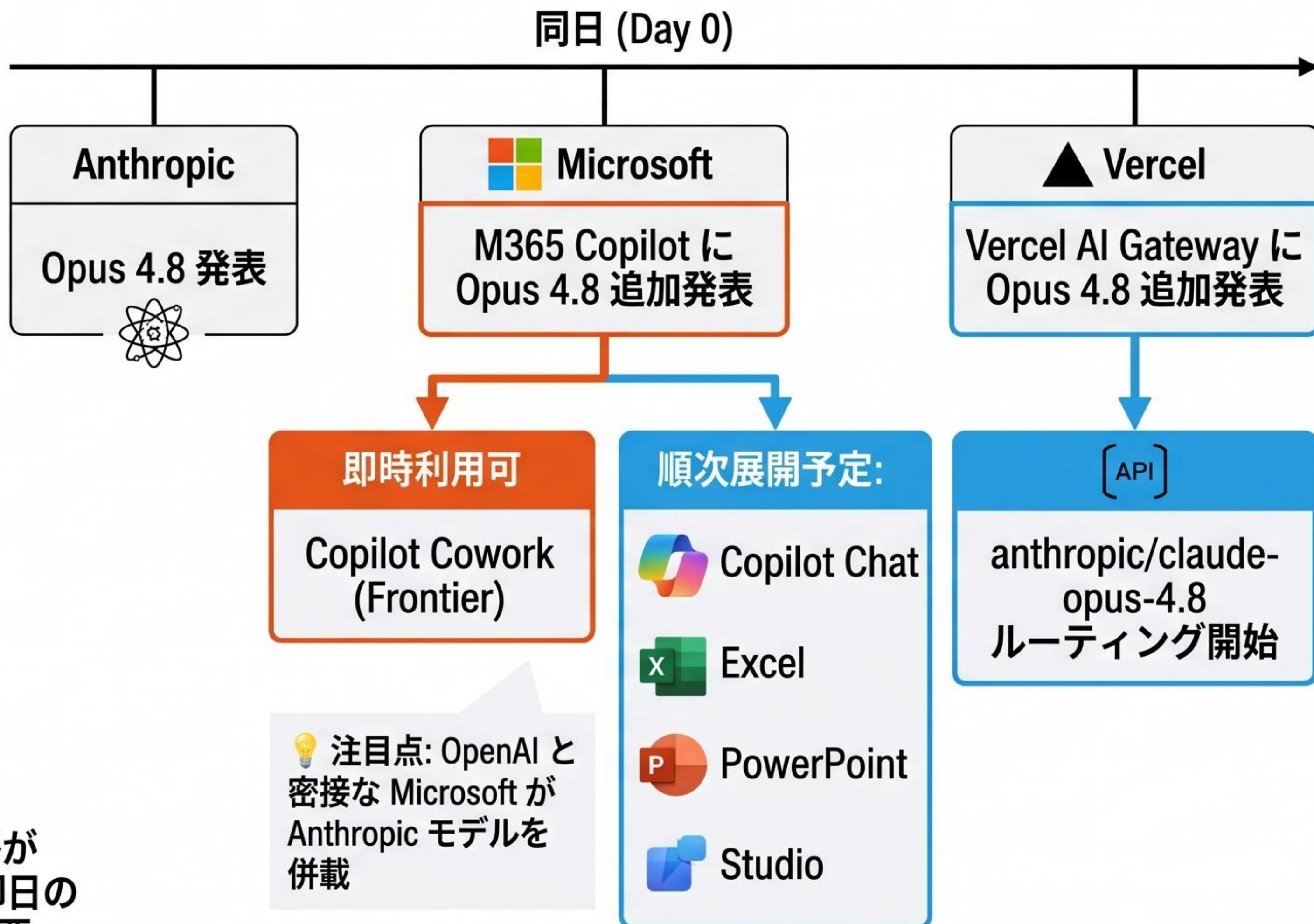
## 📌 主な変更点

- Microsoft 365 Copilot が Anthropic Claude Opus 4.8 を選択モデルに追加
- Copilot Cowork (Frontier) で即時利用可、Chat / Excel / PowerPoint / Studio へ順次展開
- OpenAI と密接な Microsoft が Anthropic モデルを併載する点が注目
- Vercel AI Gateway も同日 Opus 4.8 をルーティング対象に追加

## 💡 なぜ重要？

Anthropic の最上位モデル Opus 4.8 が、OpenAI と関係の深い Microsoft Copilot で選択可能になったことで、マルチモデル戦略が深化。ユーザーは用途に合わせてモデルを選べるようになる。即日の展開はスピード感を象徴。Vercel AI Gateway の対応も同様に重要。

## 📅 Claude Opus 4.8 発表当日の各社対応展開図 (2026-05-30 同日)



## 🔬 研究の核心

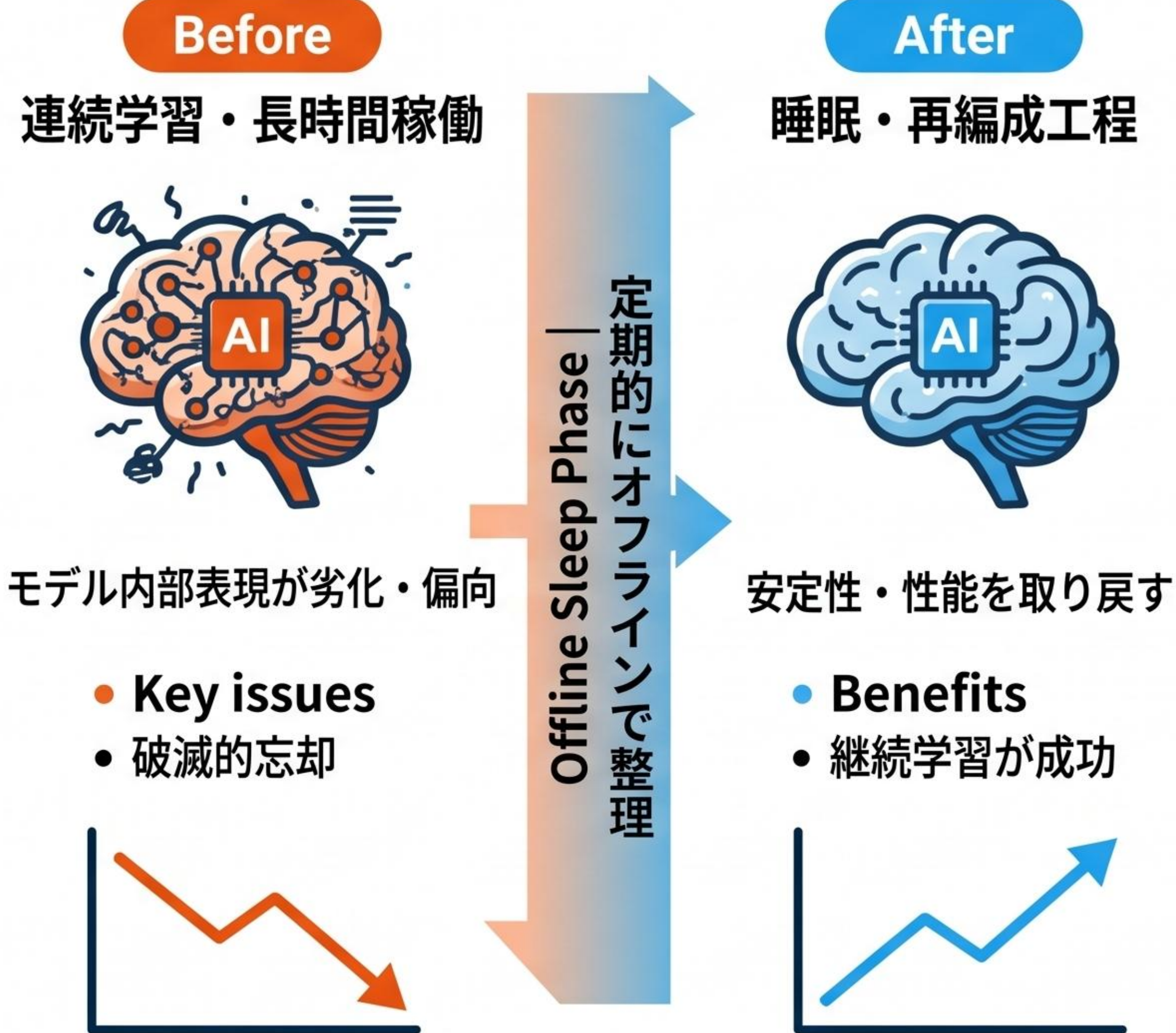
AIモデルにも人間の睡眠に相当する『休息・再編成』の工程が性能維持に必要だとする研究結果を GIGAZINE が紹介した。連続学習や長時間稼働でモデルの内部表現が劣化・偏向していく現象に対し、定期的にオフラインで整理するフェーズを挟むことで安定性や性能を取り戻せるという考え方で、継続学習 (continual learning) や破滅的忘却の問題と関連する。

## 📊 主な知見

- 連続学習・長時間稼働でモデル内部の表現が劣化／偏る現象を扱う
- 人間の睡眠に相当する『整理・再編成』工程が性能維持に寄与する

## 🎯 含意

継続学習 (continual learning) や破滅的忘却の研究テーマと関連。



# 6. 164個の npm パッケージが oob[.]moika[.]tech 経由で認証情報を窃取する新キャンペーン

## 🔍 何が起きた？

SafeDep が、クラウド・金融分野を標的にした 164 個の悪性 npm パッケージによる依存関係混乱 (dependency confusion) キャンペーンを報告した。

## 📌 主な脅威と手口

- 164 個の悪性 npm パッケージがクラウド・金融領域を標的化
- インストール時に `process.env` 全体を外部へ送信し認証情報を窃取
- 狙いは API キー・トークン・CI 環境のシークレット
- 社内名と同名パッケージを公開する dependency confusion を悪用


## 💡 脅威の分析と影響

先日の Shai-Hulud 系とは別系統の新たな攻撃波であり、CI/CD パイプラインを含む開発環境からの重要情報の窃取に直結する。

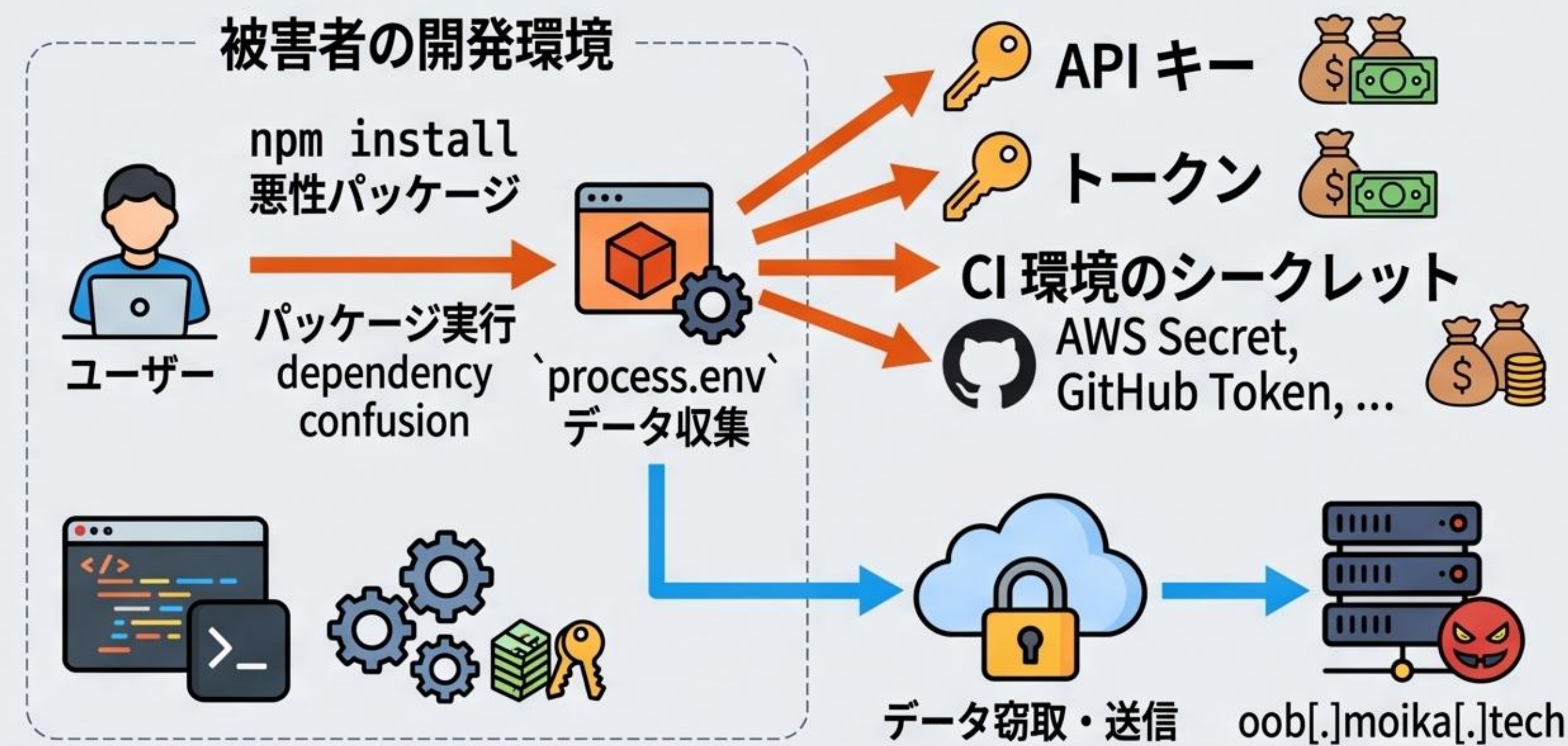
## キャンペーンの規模

**164**   
悪性 npm パッケージ

## キャンペーンの規模

**164**   
悪性 npm パッケージ

## 認証情報窃取のフロー図



# 7. Opus 4.8 vs GPT-5.5 — もう『賢さ』ではなく用途で選ぶ時代の使い分けチートシート

## 要点

- Opus 4.8はリーダーボードを追わず『トークン効率』（稼働コスト約61%減・fast mode 3倍安）に特化したリリースという考察
- 4.8とGPT-5.5は知性勝負から用途選択へ


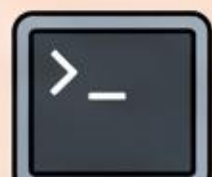






## 具体的な手法 / 使いどころ

- |   |   |
|---|---|
| <ul style="list-style-type: none"> <li>Opus向き用途           <ul style="list-style-type: none"> <li>リポジトリ規模コーディング</li> <li>長時間の無人エージェント</li> <li>自己ミス検出が要る作業</li> <li>computer use</li> </ul> </li> <li>Codex(GPT-5.5)向き用途           <ul style="list-style-type: none"> <li>ターミナル中心作業</li> <li>Web リサーチ</li> <li>大量処理パイプライン</li> <li>速度重視</li> </ul> </li> </ul> | <ul style="list-style-type: none"> <li>Opus向き用途           <ul style="list-style-type: none"> <li>ターミナル中心作業</li> <li>大量処理パイプライン</li> <li>速度重視</li> </ul> </li> <li>Codex(GPT-5.5)向き用途           <ul style="list-style-type: none"> <li>Web リサーチ</li> <li>大量処理パイプライン</li> <li>速度重視</li> </ul> </li> </ul> |
|---|---|

## なぜ刺さるか / 学び

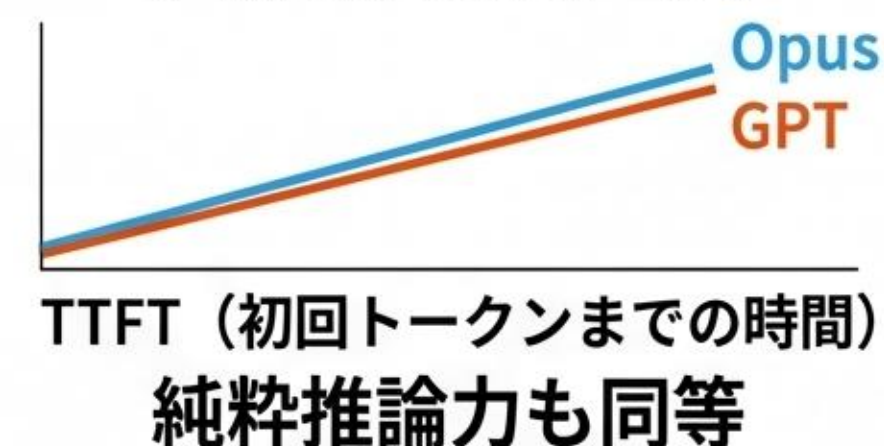
- 純粋な推論力とTTFTはほぼ互角だが、圧倒的なコスト効率が最大の選択基準に。成熟したAI活用エコシステムを示唆
- ...

## 用途別 使い分けcheat sheet

Opus 4.8	Codex (GPT-5.5)
 リポジトリ規模コード	 ターミナル操作
 長時間無人エージェント	 Webリサーチ
 自己ミス検出	 大量処理
 computer use	 速度重視

数字で見る 4.8の効率  
稼働コスト ▼約61%減  
fast mode ⚡ 3倍安 💰

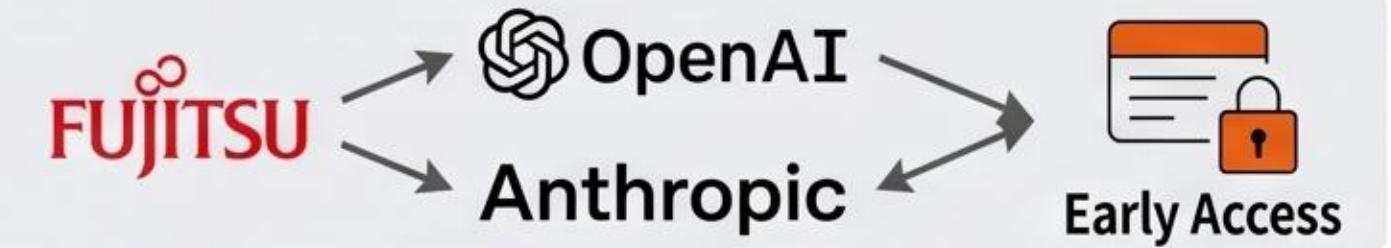
### 性能はほぼ互角



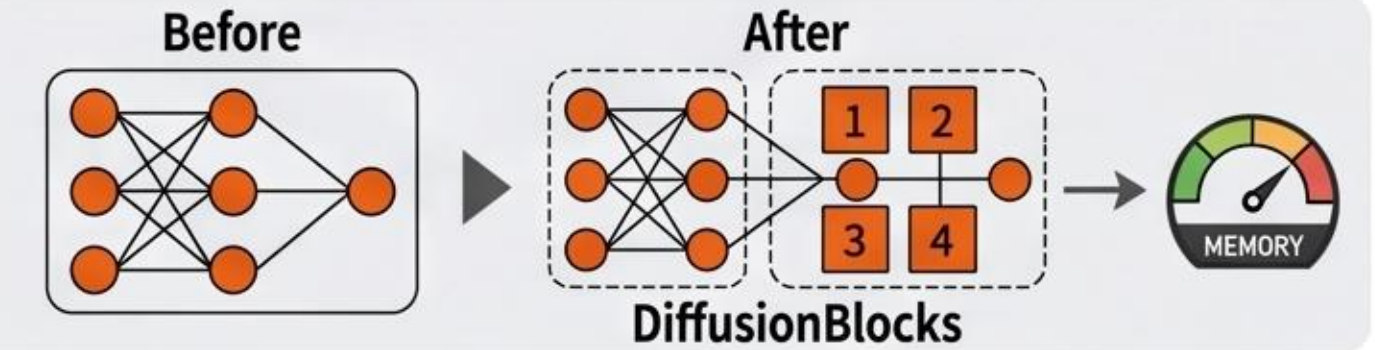
# 本日のトピック一覧

今日のまとめ

**1 富士通が OpenAI と連携開始、Anthropic とも「ミュトス」早期アクセスで協業**  
富士通が OpenAI と連携、Anthropic 『ミュトス』 早期アクセス協業



**2 Sakana AI 『DiffusionBlocks』 — ネットワークをブロック単位で分割学習しメモリ要件を削減**  
Sakana AI が 『DiffusionBlocks』 を発表。ネットワークを分割学習し、メモリ削減



**3 MIT・Stanford・DeepMind・MS 論文：エージェントは chat の約1000倍トークン消費、同タスクで最大30倍コスト差**  
MIT・MS等の論文：AIエージェントはchat比で1000倍トークン、30倍コスト



**4 Microsoft 365 Copilot が Claude Opus 4.8 を追加、Vercel AI Gateway でも即日提供**  
Microsoft 365 Copilot に Claude Opus 4.8 が追加。Vercel でも提供



**5 「AIモデルには睡眠が必要」という研究結果**  
研究結果：「AIモデルには睡眠（休息）が必要」



**6 164個の npm パッケージが oob[.]moika[.]tech 経由で認証情報を窃取する新キャンペーン**  
新キャンペーン：164個の npm パッケージで認証情報を窃取（oob[.]moika[.]tech 経由）



**7 Opus 4.8 vs GPT-5.5 — もう『賢さ』ではなく用途で選ぶ時代の使い分けチートシート**  
Opus 4.8 vs GPT-5.5：『賢さ』より用途で選ぶ使い分けガイド

Usability Guide (Not just smarts)

	Opus 4.8	GPT-5.5
Strength	Strength	Strength
Best Use Case	Concrete-example: example, Mtcnoceft, Opus amianels, ...	Concrete-example: examples, GPT-0505, GPT vo GPT-5.5