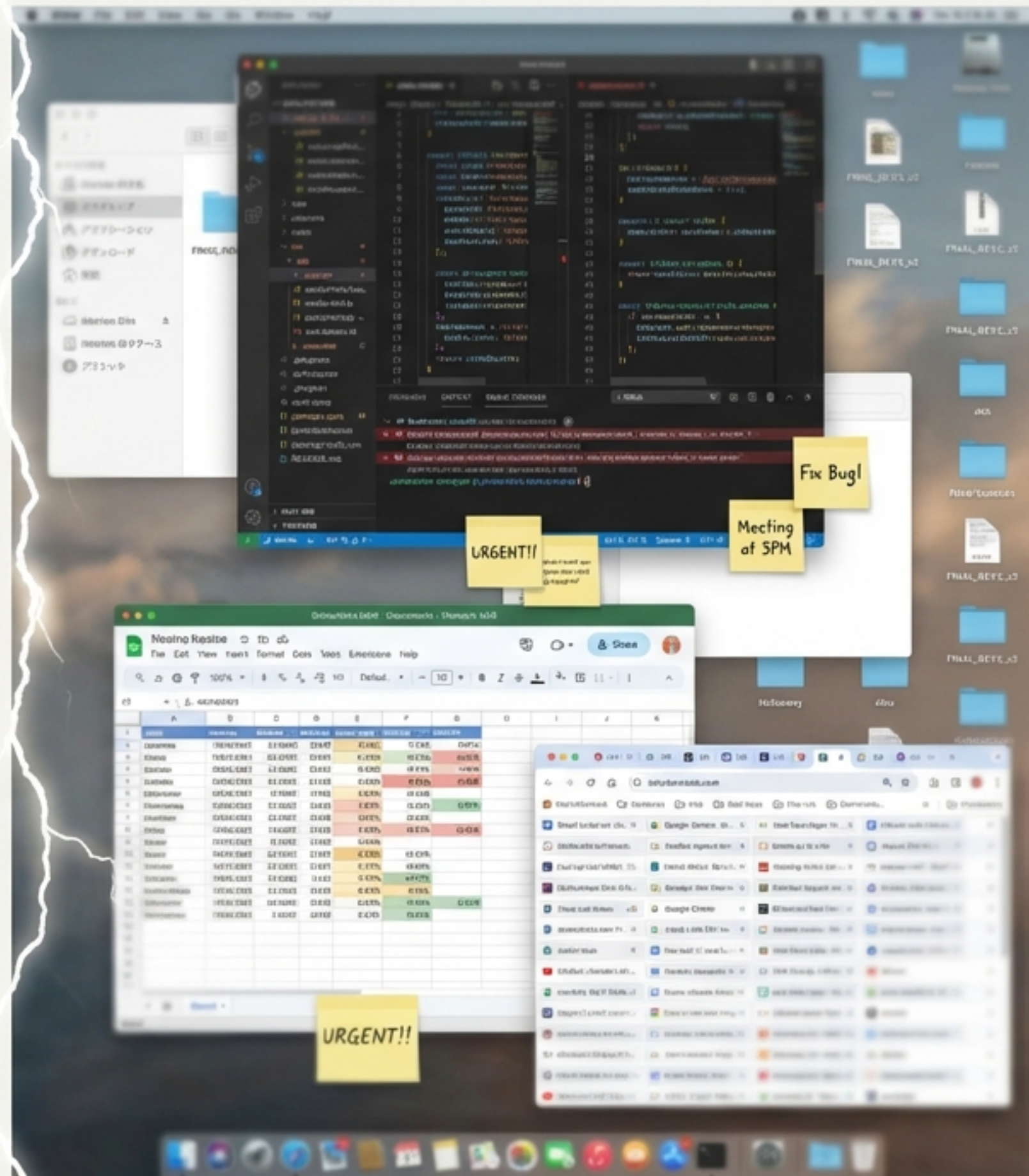


AIエージェントの夜明け

AIエージェントは、本当に私たちの期待に応えられているのでしょうか？

AIエージェントは、人間の代わりにデジタルタスクを実行する能力を約束し、生産性の未来を塗り替えようとしています。業界のリーダーたちは、驚異的なベンチマークスコアを発表し、AIがほぼ人間レベルの能力に達したかのような印象を与えています。しかし、その「進歩」は本物でしょうか？
それとも、私たちは進歩の幻想 (An Illusion of Progress) を見ているだけなのでしょうか？

本プレゼンテーションでは、AIエージェントの評価における根深い課題を明らかにし、真の能力を測るための新しい基準と、その基準を初めてクリアした次世代モデルをご紹介します。

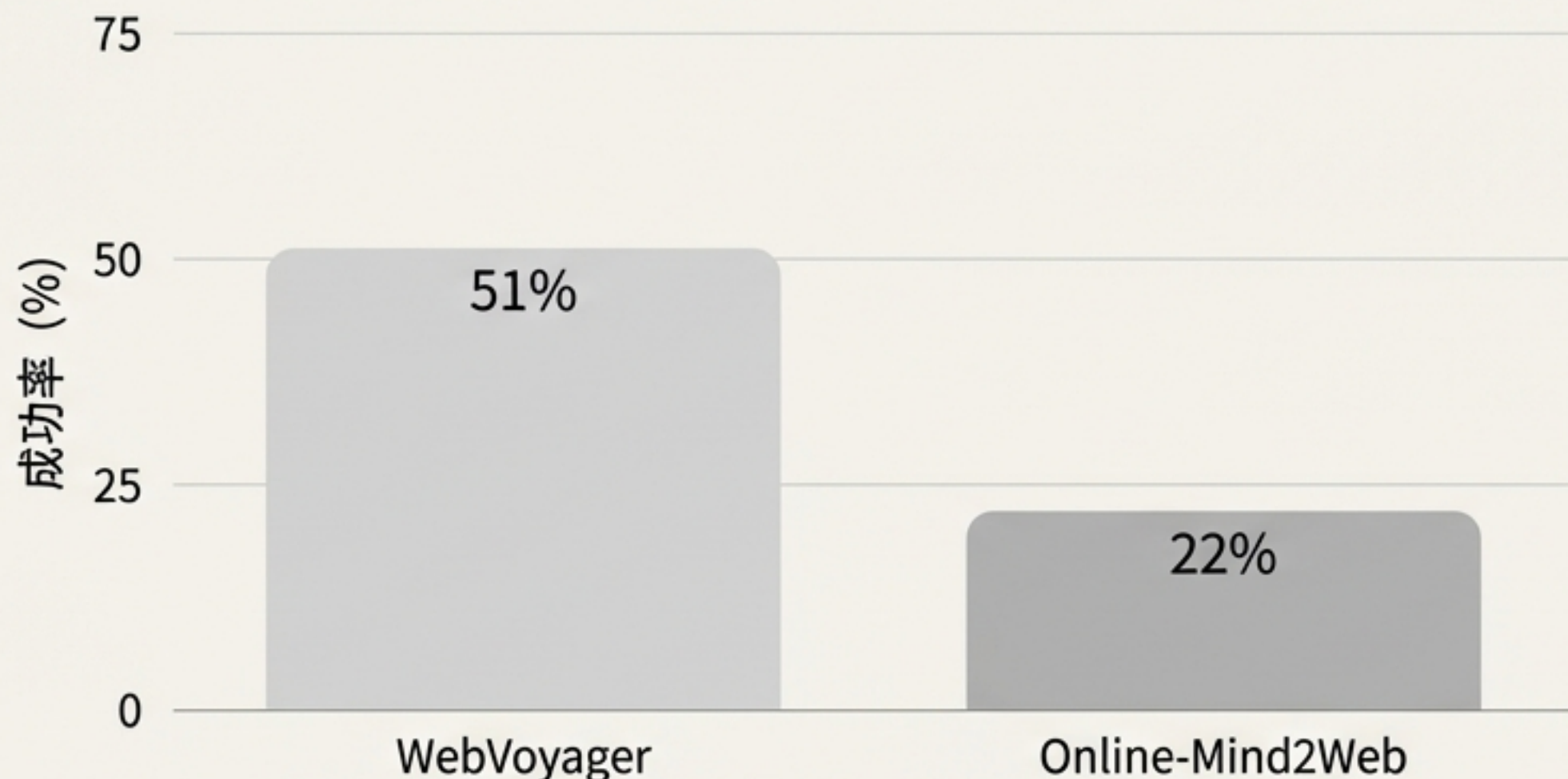


これまでのベンチマークには、致命的な欠陥がありました

多くのAIエージェントが高い成功率を報告してきた「WebVoyager」のような従来のベンチマークは、現実世界の複雑性を捉えきれませんでした。

私たちの分析によると、これらのベンチマークのタスクは単純化されすぎており、多くはウェブサイトを深く操作することなく、ショートカットで解決できてしまいます。

単純な検索エージェントによる驚くべき結果



半数以上のタスクが単純なGoogle検索で解決できるという事実は、これまでの評価基準がAIエージェントの真のナビゲーション能力やインタラクション能力を測定できていなかったことを示しています。

真の能力を測るために、我々は新たな基準を構築しました

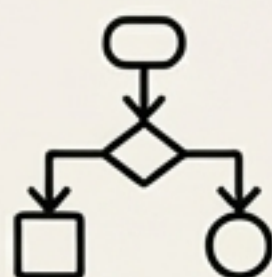
Online-Mind2Web

AIエージェントの現実世界での実用性を正確に評価するため、私たちは新しいオンライン評価ベンチマーク「Online-Mind2Web」を開発しました。このベンチマークは、エージェントが直面する現実の課題を忠実に再現するように設計されています。



ライブウェブサイトでの評価

キャッシュされた静的なページではなく、常に変化する136の実際のウェブサイトでエージェントをテスト。ポップアップ、Cookie、レイアウト変更など、現実の課題に対応する必要があります。



現実的で多様なタスク

eコマース、旅行、情報検索など、様々なドメインにわたる300の高品質なタスク。これらは実際のユーザーのニーズから作られています。



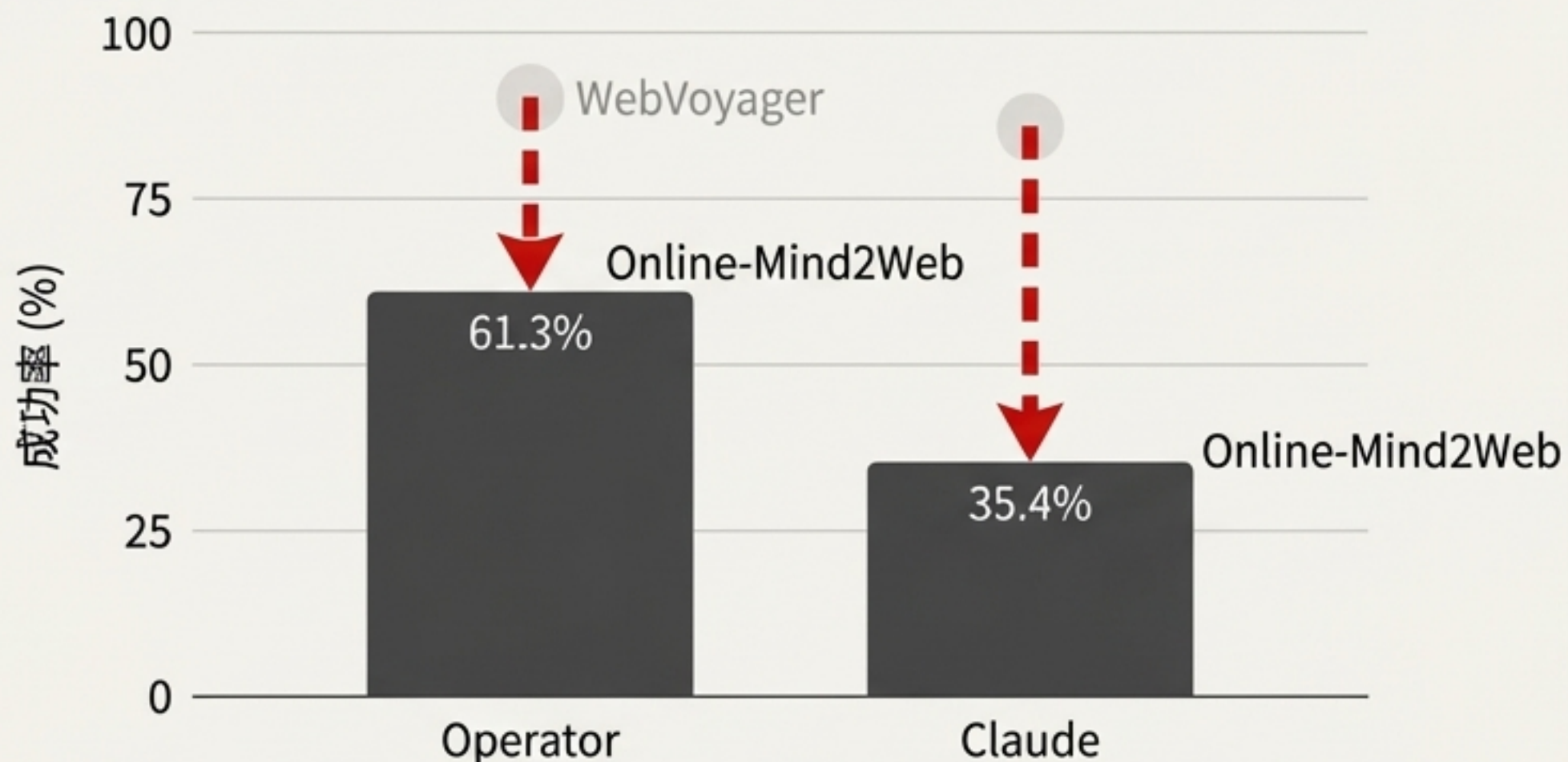
厳格な評価

単純なショートカットを許さず、複数ステップにわたるインタラクションとナビゲーション能力を真に評価します。

新基準が明らかにした、AIエージェントの現在地

既存の最先端エージェントをOnline-Mind2Webで評価した結果、報告されていた性能から大幅な低下が見られました。これは、業界がこれまで見ていた進歩が、評価基準の甘さに支えられていた可能性が高いことを示しています。

主要エージェントの成功率：WebVoyager vs. Online-Mind2Web



真の課題に直面したとき、既存モデルの能力には限界があることが明らかになりました。
業界には、この厳しい新基準を乗り越える、真のブレークスルーが必要です。

新時代の到来

次世代コンピュータ利用モデル「Lux」

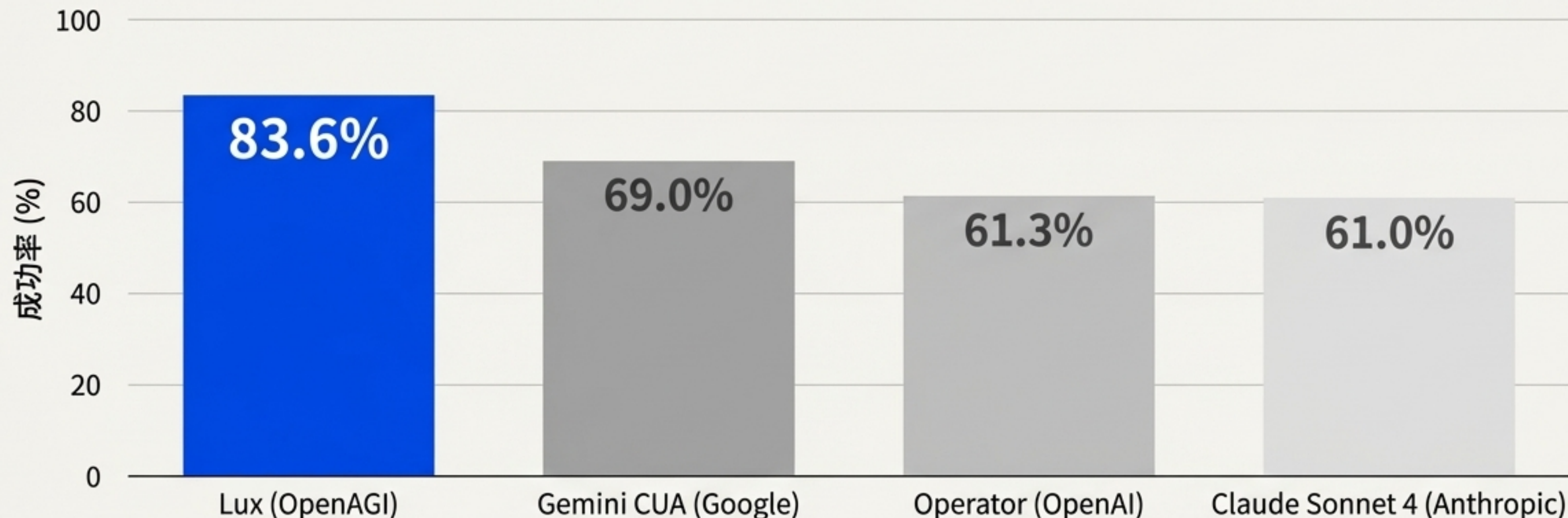


これまで不可能だった基準を、初めてクリアしたモデル。

Luxは、あらゆる競合を凌駕します

Luxは、最も厳格な業界ベンチマークであるOnline-Mind2Webにおいて、これまでの最高記録を20ポイント以上更新する圧倒的な性能を達成しました。

Online-Mind2Web ベンチマーク 成功率



この差は、単なる漸進的な改善ではありません。
これは、AIエージェントの能力におけるパラダイムシフトです。

性能、速度、コスト。Luxはそのすべてで新基準を打ち立てます。

Luxは、世界最高の精度を実現するだけでなく、実用化の鍵となる速度とコスト効率においても、競合を圧倒します。



性能

83.6%

Online-Mind2Webでの成功率。
業界最高水準を達成。



速度

3倍 高速

1ステップあたり約1秒で実行。
OpenAI Operator（約3秒）と比較して大幅な高速化を実現。



コスト

10倍 安価

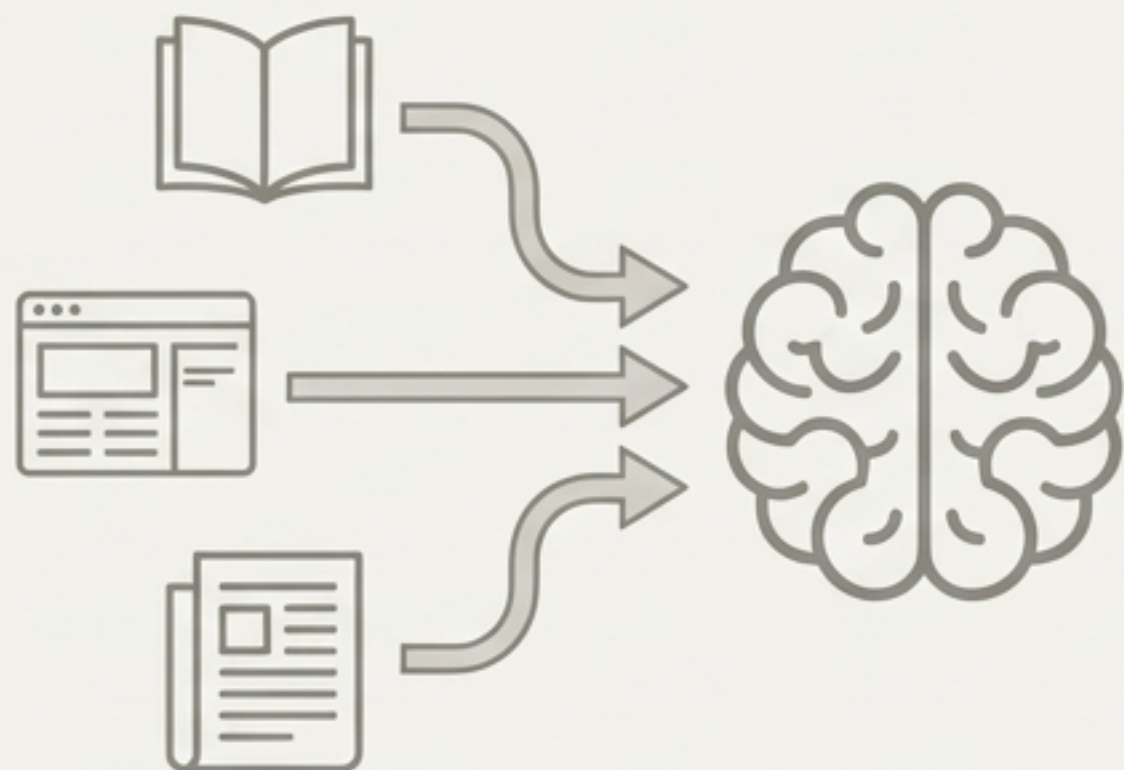
競合の最先端モデルと比較して、
約1/10のコストで運用可能。

この組み合わせが、Luxを研究室から現実世界のアプリケーションへと導きます。

Luxの圧倒的な性能を支える、新しい学習パラダイム

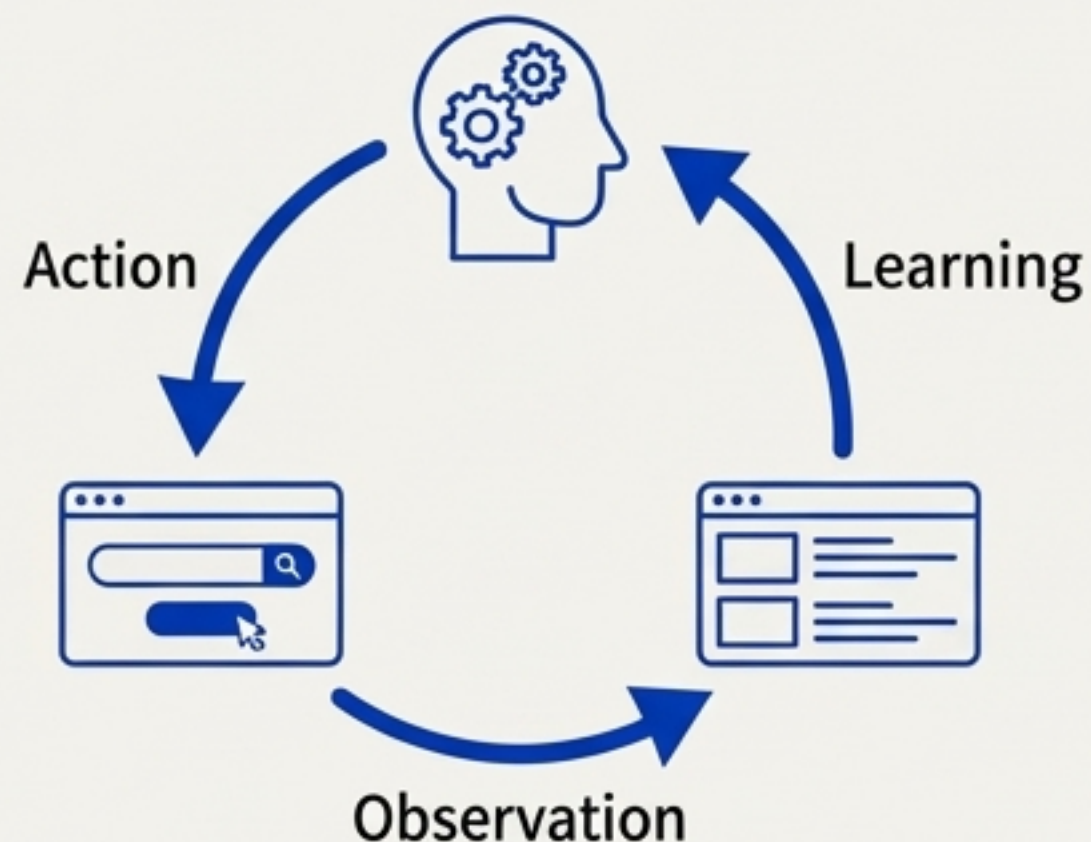
従来のLLMが「教科書を暗記する」ように学習するのに対し、Luxは「実際に運転して学ぶ」ように訓練されます。私たちはこの革新的なアプローチを**Agentic Active Pre-training (AAPT)**と呼んでいます。

受動的学習 (Passive Learning)



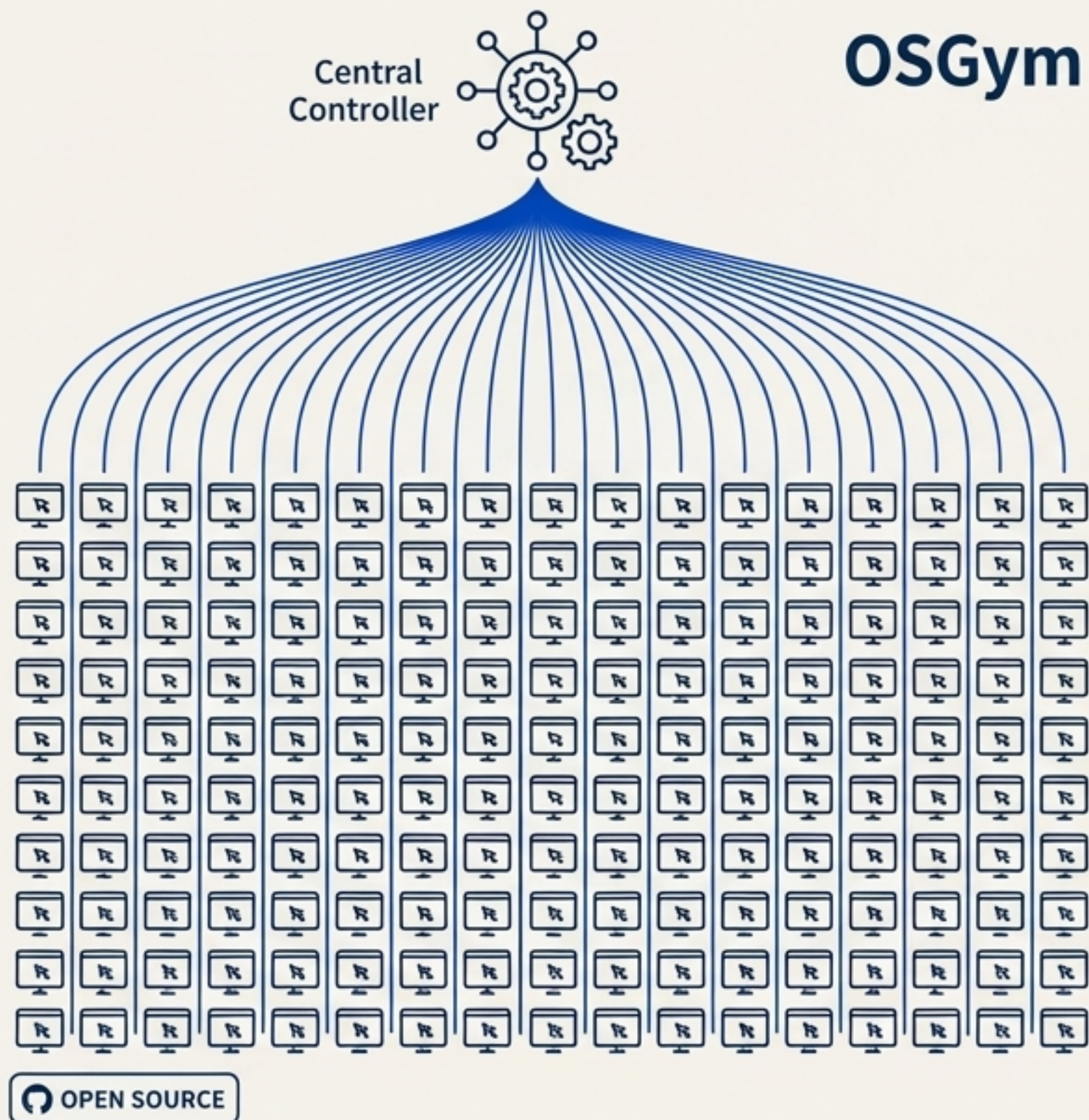
インターネット上の膨大な知識を一方向的に吸収する。

エージェント的能動学習 (Agentic Active Learning)



デジタル環境を能動的に探索し、「行動と結果」の連続ループから自律的にスキルを洗練させる。

Luxの訓練を可能にした、超スケーラブルな分散データエンジン「OSGym」



Agentic Active Pre-trainingのポテンシャルを最大限に引き出すため、私たちはOSGymを開発しました。これは、汎用コンピューターエージェントの訓練に特化した、業界初の超スケーラブルなインフラストラクチャです。

私たちは、この強力なツールをオープンソースとして公開し、AIエージェント研究の発展をコミュニティと共に加速させることを目指します。



圧倒的なスケーラビリティ

1,000以上のOSレプリカを並列実行し、1分あたり1420以上の軌道データを生成。



高い汎用性

ブラウザ操作からオフィスアプリケーション、ソフトウェア開発まで、OS上で実行可能なあらゆるタスクをサポート。



驚異的な経済性

学術機関でも利用可能な低コスト（1レプリカあたり1日0.2-0.3ドル）で大規模実験を実現。

AIの新たなパラダイム：「生成」から「操作」へ

私たちは、AIの能力における次の大きなシフトを「コンピューター利用（Computer Use）」と定義します。これは、AIが単に情報を生成・推論するだけでなく、人間のようにコンピューターを**操作**する能力です。

過去（Generation）



テキスト、画像、コードを**生成する**

記事作成、デザイン補助

未来（Operation）



ソフトウェア、ダッシュボード、OSを**操作する**

ソフトウェアテスト自動化、データ入力、SaaS運用



Luxは、この新しいパラダイムをすべての開発者に解放します。

開発者のニーズに応える、3つの柔軟な実行モード

Luxは、タスクの特性に応じて最適な制御レベルと自律性を提供するために、3つの異なるモードを備えています。

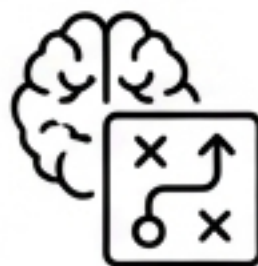


Actor Mode (アクターモード)

Feature: **即時実行**。明確に定義されたタスクを1ステップ約1秒の速さで実行。

Use Case: 高速なデータ入力、APIのないGUIからの情報取得、単純なナビゲーション。

Example: 「NASDAQのサイトでAppleのインサイダー取引情報を取得する」



Thinker Mode (シンカーモード)

Feature: **自律的計画**。曖昧で複雑なゴールを理解し、実行可能なタスクに分解して自律的に遂行。

Use Case: 「競合他社のSNSでの評判を調査しまとめる」といった複数ステップにわたるリサーチ。



Tasker Mode (タスカーモード)

Feature: **完全な制御**。Pythonリストで提供されたステップを厳密に一つずつ実行。高い信頼性と再現性を保証。

Use Case: ソフトウェアのE2Eテスト、財務レポート作成など、厳格な手順が求められる業務ワークフロー。

オープンなエコシステムを構築し、AIの未来を共に創る

OpenAGIのミッションは、人類に利益をもたらすAIモデルのオープンなエコシステムを創造することです。私たちは、AIの未来が、孤立したモデルではなく、それらの協調によって築かれると信じています。



Open:

透明で、コミュニティがアクセス可能



Safe:

堅牢な安全メカニズムを設計に組み込む



Collaborative:

協調によって単一モデル以上の成果を出す



Beneficial:

人類へのポジティブな影響に焦点を当てる



Dr. Zengyi Qin,
Co-founder & CEO

MITの博士課程に在籍し、自律システムの安全性に関する研究をリード。JetMoE-8BやOpenVoiceなど、広く採用されている複数のAIモデルの開発を主導。

未来を、共に創る。

世界で最も先進的なコンピューター利用モデルと、それを支えるオープンなインフラが、
あなたの手の中に。さあ、今すぐ始めましょう。

Lux SDKを試す

パワフルなコンピューター利用アプリケーションの構築を始めましょう。



agiopen.org/sdk

OSGymに参加する

オープンソースの訓練エンジンに貢献し、次世代エージェントの研究を加速させましょう。



github.com/agiopen-org/osgym

コミュニティに参加する

Discordで開発者や研究者と繋がり、アイデアを交換しましょう。



agiopen.org/discord