

DeepSeekの画期的なローンチ： なぜ「十分実用的」なAIが金融 サービスを再び変革できるのか

執筆



Kes Sampanthar

パートナー
Valize



Scott Wolfson

常務取締役
Valize

1月初旬、DeepSeekは画期的な推論モデル「R1」をオープンソースライセンスで公開し、AI業界関係者の間で話題となりました。

これまで一般的な見解では、本格的な大規模言語モデル（LLM）—高度な自動処理や洗練された拡張知能を備えたもの—は、大規模かつ非公開で、訓練に天文学的な費用がかかるとされていました。設備投資と計算費用に加え、訓練データとして数十億のトークンも必要となるため、こうした高性能モデルは、一握りのテック系大手企業の手任せられていたのです。

DeepSeekはその考えを一夜にして覆しました。

DeepSeekのオープンソースR1モデルは、大手企業の独自システムの数分の1の規模と費用で「十分実用的な」の性能を実現し、AI開発の道が必ずしも大規模で高額かつ独占的なものではないことを証明しました。実際、R1の核心技術のオープンソース化は、さらなる真実を明らかにしました。高度なAIへの参入障壁が劇的に下がったのです。これは、まさに「4分の壁」が突然破られたようなものです。

1954年、ロジャー・バニスターが1マイル4分の壁を破った記録は伝説的でしたが、その記録はわずか1ヶ月で塗り替えられました。現在、DeepSeekが「より小規模・低コストで訓練されたモデルで何が可能か」を示したことで、数多くのイノベーターが追随しようとしています。

DeepSeek R1のオープンソース公開から数週間で、リポジトリのダウンロード数は200万回を超え、GitHubには1万件以上のフォークが作成されました（フォークとは、元のコードを変更しない派生版のことです）。開発者の関心がこんなに高まっているのは、スタートアップ企業、大企業、学術研究機構など、多くの関係者がDeepSeekの小規模かつ効率的なやり方を活用しようとしているからです。

同等の独自システムの開発には最大1億ドルもかかるのに対し、DeepSeek R1はわずか600万ドルで開発されたと報告されており、高度なAI開発における参入コストが劇的に下がったことを示しています。（ただし、600万ドルという数字は、事前学習段階に使用したプロセッサチップ代のみであり、総費用の一部に過ぎません。）

私たちは、この流れを過去にも見てきました。テクノロジーが安く普及すると、もはや特定の組織が独占することはできなくなります。すぐにイノベーションが起こり、小規模なイノベーターたちが新たな可能性を模索し始めるのです。もしAIがビジネスのあらゆる場面に分散・組み込まれるようになれば、知識労働中心の金融サービス業界は、特に影響を受けるでしょう。

心理的障壁を打ち破る

生成AI分野では、最良のモデルを作るには数千万ドル（あるいは数億ドル）相当のクラウドと膨大なデータセットが必要で、完成したモデルはAPIで秘匿するのが常識とされてきました。

DeepSeekの小規模・低コストな手法は、これまでの常識を打ち破りました。単なる性能の問題よりも、「4分の壁」が破られた時と同じく、衝撃的な瞬間でした。DeepSeekが従来のほんの一部のコストでエージェント型知能を開発できるのであれば、LLMや先端AIをめぐる市場心理そのものが変化します。

今や、たとえ小規模なスタートアップや大学の研究室、中堅の金融企業であっても、性能を再現したり、上回ったりすることが期待できるようになりました。DeepSeekがR1をオープンソース化したことで、すべてが明らかになりました。他分野のオープンソースソフトウェアの盛り上がりと同様に、新たな実験やフォーク、特化型AIソリューションの波が押し寄せるのはほぼ間違いありません。

思わぬきっかけ：中国への半導体輸出禁止措置

皮肉なことに、DeepSeekのイノベーションのきっかけの一つは、中国への先進的なコンピューター処理チップ（GPU）の輸出禁止措置でした。最先端のチップ不足が中国のAI研究を妨げると考えていたからです。しかし代わりに、DeepSeekはより効率的な訓練方法を開発し、新たなアルゴリズムの工夫を発見し、より小型で最適化されたモデルを構築することでハードウェアの制約を克服しました。この新モデルは、さまざまな作業において、はるかに大規模な「クローズド」システムとほぼ同等の性能を発揮します。

「高性能AIを作るには最先端のGPUが必要」という常識が覆されました。制約が変革を生み出したのです。彼らは、最も簡単な解決策（とにかく高性能なハードウェアを大量投入する）が奪われたことで、新たな解決策を発見しました。現在、R1のコードがオープンソース化されたことで、誰もが利用できるようになりました。

類似する独自システムはトレーニングコストに最大1億ドルかかるのに対し、DeepSeekのR1はわずか600万ドルほどで実現したと報告されています。

歴史の法則1: テクノロジーの普遍性

今日、AIは1980年代のマイクロコンピューター革命と同じ道筋をたどっています。小型で安価なデスクトップコンピューターが、それまで大型メインフレームで処理されていた多くの作業を引き受けるようになりました。やがて標準化が進み、作業スピードが加速しました。ムーアの法則が技術進歩を促し——より安く、より速く、より良く——その流れがモバイル革命へとつながり、技術の一般化を通じて、業務のあらゆる場面に浸透していきました。

当初、大規模LLMは巨大な計算施設を必要としていましたが、現在では小型のオープンソースモデルが手元の端末（現場での処理や、特化型で安い機器）で動き、自動処理に十分な性能を発揮しています。マイクロコンピューターが従業員の能力を高め、デジタル変革を促したように、こうした小型でオープンなAIモデルが新たなイノベーションの波を巻き起こすでしょう。

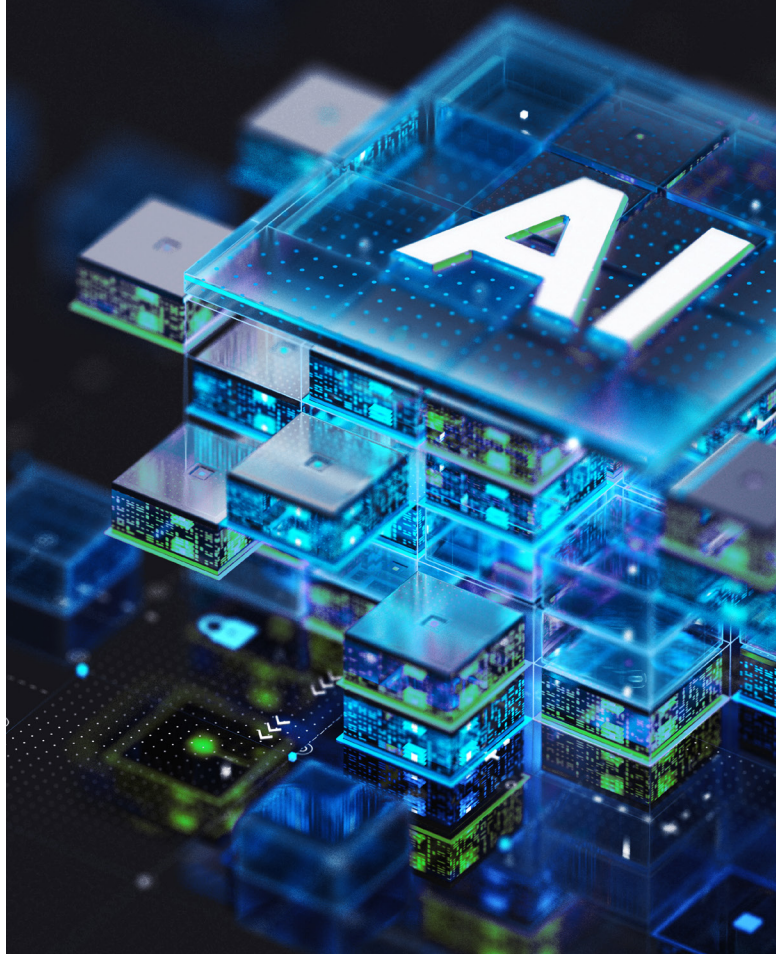
ここから学べることは、テクノロジーは大規模で遅い「中央集権的なシステム」から、安価で高速な「分散型システム」へと進化し、最終的には業務の隅々まで浸透し、従来のバリューチェーンを根本から変革するということです。

歴史の法則2: GPTがもたらす変革

マイクロコンピューターの例より前に、歴史は工場において蒸気から電気への転換という貴重な教訓を示しています。電力をいち早く導入した企業の多くは、単に蒸気機関を一台の巨大な電動モーターに置き換えただけで、工場の配置まで変えませんでした。

当初、生産性はほとんど上がりませんでした。本当の変革が訪れたのは、ヘンリー・フォードらが「小さな電気モーターを工場中に配置し、作業の流れを組み立てライン中心に作り直せばいい」と気づいた時です。この再編成で生産性は大きく向上しました。

テクノロジーは「大規模で低速な集中型システム」から、「低コストかつ高速な分散型システム」へと進化し、最終的にはビジネスのあらゆる領域に浸透して、従来のバリューチェーンを揺るがしています。



しかし、再編成には数十年もかかったのです。歴史研究者は、大半のメーカーが電気の本当の力を理解するのに約40年かかったと推定しています。この現象は「ダイナモ効果」と呼ばれることもあります。技術はもうあったのに、古い考え方に縛られて、事業の再編成ができなかったということです。

ダイナモ効果は、汎用技術（GPT、General Purpose Technology）にも当てはまります。GPTとは、1980年代に経済学者のロバート・ソローが作った言葉で、さまざまな業界で幅広く使えて、社会や経済を変革するテクノロジーのことです。

GPTの本当の力を理解するには時間がかかります。フォードのように、GPTがビジネスモデル自体を変革し、新たな競争力を生み出したときに、真の変革が起こるのです。

今、私たちは新たなGPTの波を迎えようとしています。DeepSeekや小型のオープンソースAIは、知能の分散を実現しました。一部の大規模でクローズドなモデルが知識処理を独占するのではなく、スマホやIoTセンサー、個人のノートパソコンから専用のオンプレミス環境まで、あらゆる端末でAIを使えるようになります。

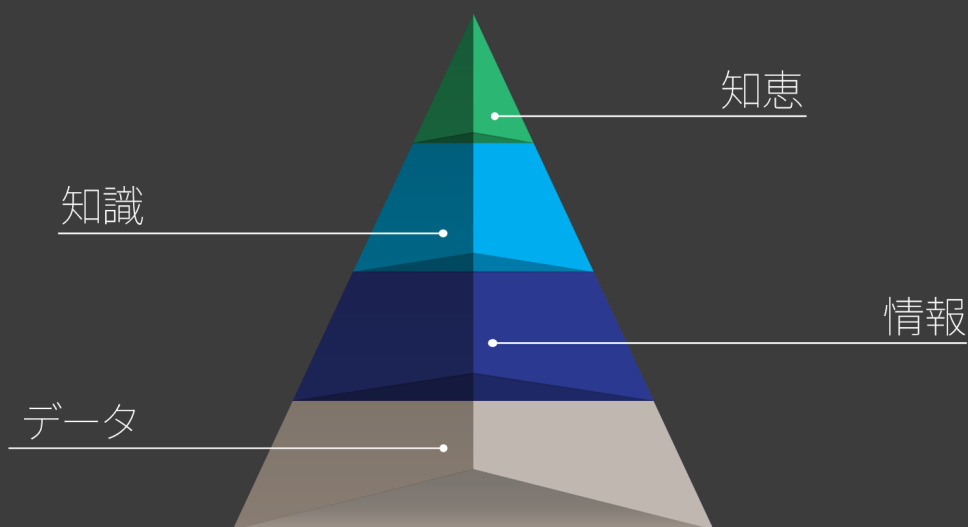
電力への転換と同じく、真の変革を起こすには、高価で大規模なAIモデルを安いものに切り替えるだけでなく、ビジネスモデルや事業領域を根本から見直す必要があります。

変革はどのように起こるか

テクノロジーの波がどのようにして事業や産業全体に変革をもたらすのかを理解するためには、データがどのようにして知恵に変換されていくのか、その過程を見ることが重要です。これはDIKW階層、つまり「データ（Data）→ 情報（Information）→ 知識（Knowledge）→ 知恵（Wisdom）」という枠組みで整理できます。

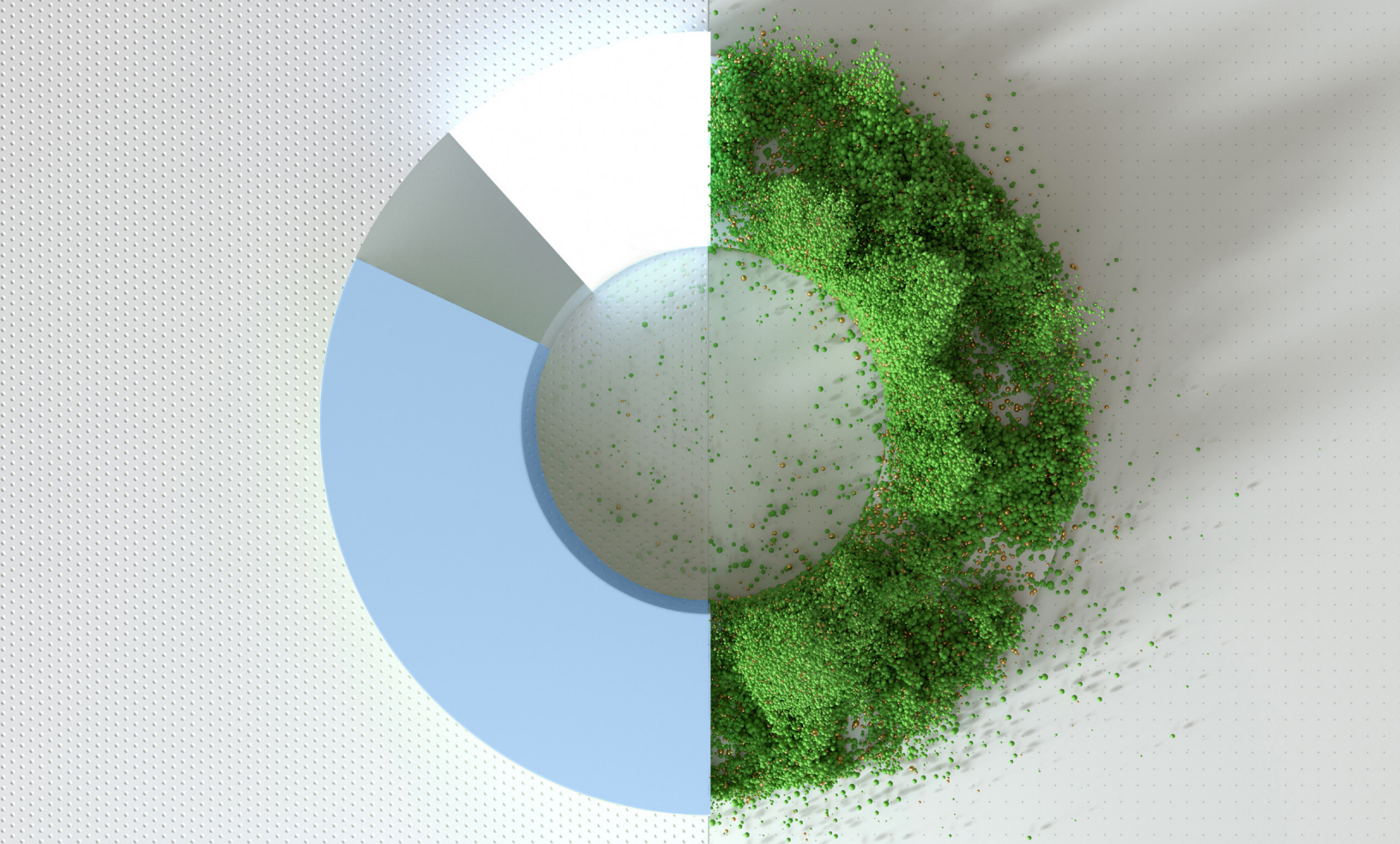
コンピューターが進化するにつれて、DIKW階層の階段を登るように、人間がやっていた仕事を機械が引き受けるようになりました。人間は、生データの処理（初期のコンピューティング）からAIの洞察に基づく戦略的な意思決定まで、より価値の高いタスクに集中できるようになりました。

大型コンピューターからパソコン、スマホ、クラウド、AIによる自動化に至るまで、あらゆる大きなテクノロジーの波が、企業をDIKW階層の上に押し上げ、働き方や業務の進め方を変えてきたのです。



出典：Rowley, J. 2007. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science, 33, 163–180.

DIKW階層：私たちは「知識処理」から「知恵の拡張」への移行を目の当たりにしています。AIエージェントは単に情報を分析し、知識を統合して、知恵を拡張するだけでなく、自律的に判断し行動するようになっています。



そして今、次の段階である「インテリジェンス変革」が始まろうとしています。AIを組織全体に取り入れて、あらゆる業務で人間の頭脳と判断力を強化させる取り組みです。これは、従来のデジタル変革を超える、次の段階への飛躍と言えます。

- 企業は単に作業をデジタル化するだけでなく、AIによる支援や自律的な知識労働へと移行しており、作業の流れだけでなく、思考の流れを見直す必要があります。
- DeepSeek R1のようなAIモデルがコストを下げて、高度な知能をあらゆる端末で使えるようにすることで、「インテリジェンスのコピキタス化」が進んでいます。
- この変化は、大型コンピューターからマイクロコンピューターへ、蒸気動力から電気へという、テクノロジーが新しいビジネスモデルや分野を切り開いた歴史的変化と似ています。

私たちは、「知識の処理」から「知恵の強化」への移行を目の当たりにしています。AIエージェントは、単に情報を分析し、知識をまとめ、知恵を強化するだけでなく、自律的に判断して行動を起こすようになります。

金融サービスのデジタル変革

2010年から2020年にかけて、金融サービスはデジタル化によって変革を遂げました。俊敏なフィンテック企業は、モバイルやクラウドのダイナモ効果を活用し、既存の大手金融機関のシェアを徐々に奪う製品を次々と展開しました。

- ChimeとVarolは、月額料金なしで滞りなく手続きできるモバイルファーストの銀行口座を開設。
- Robinhoodはアプリ上で株式取引の手数料無料化を導入し、これがきっかけとなって既存の大手証券会社にも取引手数料を撤廃する動きが拡大。
- BettermentやWealthfrontなどのロボアドバイザーは、アルゴリズムを使用して安価で自動化されたポートフォリオ管理を提供。
- Lemonadeなどのインシュアテック（保険テック）は、モバイルUXとデジタル引受を活用して、住宅保険や賃貸保険の市場を変革。

IoT機器やPOS端末、支店職員向けのARグラスなどのエッジ領域に高度なインテリジェンスを組み込めるスタートアップの台頭は、既存の大手企業にこれまで以上に迅速な適応を迫ることになります。

すべての事例において、小さなスタートアップが、クラウドやスマホ、アプリストアといったより安価で広く利用できる技術を活用し、これまで巨大なインフラを持つ大手銀行にしかできなかったことを実現しました。また、彼らは使いにくい従来の操作画面や高い手数料に対する利用者の不満にも着目しました。

その結果、P2P決済からネオバンクに至るまで、まったく新しい世代の金融商品が誕生したのです。

既存企業にとって、この変革は両刃の剣でした。多くの企業が市場シェアを失ったり、利益率の低下を受けたりしましたが、一方でフィンテック系スタートアップとの提携や買収を通じて新たな収益源を確保しました。

同時に一般消費者は、競争の激化やより良い顧客体験の恩恵を受けました。フィンテック企業の成功は、デジタルサービスを立ち上げる費用が劇的に下がったことに直接起因しています。これは過去の技術革新と同じ現象です。

では、次に起こりうる可能性を探ってみましょう。

一次的影響：金融サービスにおけるインテリジェンス変革

2010年から2020年がデジタルの利便性やつながる体験の時代だったとすれば、2025年以降は「知的ワークフロー」の時代になると考えられます。

- **引受業務**は、単にデジタル化されるだけでなく、より小型で効率的なAIモデルが現場で稼働し、数千もの信号をリアルタイムで分析。
- **ファイナンシャルアドバイザー向けのAIコパイロット**が、コンプライアンスに配慮した提案を自動生成したり、お客様に最適化された最新情報をメールで送信したり、リスク要因をその場で分析。
- **不正検知やコンプライアンス監視**は、中央集約型システムではなく、取引の各ポイントに組み込まれたAIエンジンによってほぼリアルタイムで実行されるようになります。
- **顧客対応のチャットボット**がこちなく制限されたものではなく、ほぼ人間並みの能力を持ちながら、安価なハードウェアで動作します。

モバイルの普及により、小規模なフィンテック企業が既存の大手企業の流通経路独占を崩したように、小規模でオープンなAIモデルが、さらなる変革の波を起こすかもしれません。

高度な知能をIoT機器やPOS端末、支店職員向けのARグラスなどエッジ機器に組み込むスタートアップの台頭は、既存の大手企業にこれまで以上に速い適応を迫ることになります。

もちろん、こうした急速なイノベーションは、複雑な金融規制の枠組みに従う必要があります。金融サービス会社は、執行措置を避けるために、アルゴリズムの公平性、データのプライバシー、モデルの透明性を確保することに細心の注意を払わなければなりません。特に規制当局がAI駆動サービスに関するガイダンスを更新している今、なおさらです。

二次的影響：エッジでのインテリジェンス

第二の波では、DeepSeekのような小規模・低コストなモデルがもたらす、まだ十分に議論されていない可能性を引き出します。すなわち、現場レベルでのAI実装です。AIは次のようなものに組み込みます。

- **スマホ**：常にクラウドのLLMを呼び出すのではなく、スマホ内蔵のAIが大半の作業をオフラインで処理できる可能性。
- **IoT機器**：工場や小売環境に設置されたセンサー搭載機器は、データを処理し自律的に動作。
- **ARグラス**：ARヘッドセットでリアルタイム分析や言語翻訳が可能に。
- **店頭端末**：銀行のキオスクや保険金請求端末が、回線接続状況が悪くとも内蔵の高度AIで瞬時に判断。

リアルタイムに端末内で処理できることで、さまざまな摩擦が減ります。パソコンが各デスクに、最終的には一人ひとりのポケットにコンピューターを広げたのと似た現象です。

金融サービスでは、支出や投資パターンを常時監視するパーソナルAIアシスタントや、高度なサービスを提供する際に大規模なクラウドAPIに依存しない特化型フィンテック系スタートアップが現れるかもしれません。

コスト障壁が下がるにつれて、2010年代のフィンテックの波と似た、さらに大規模な速度で、新しいソリューションと新規参入者が一気に増える「カンブリア爆発（※）」のような現象が起こると考えられます。（※カンブリア期に生物の数や種類が爆発的に増えた現象）

三次的影響：自律型金融の台頭

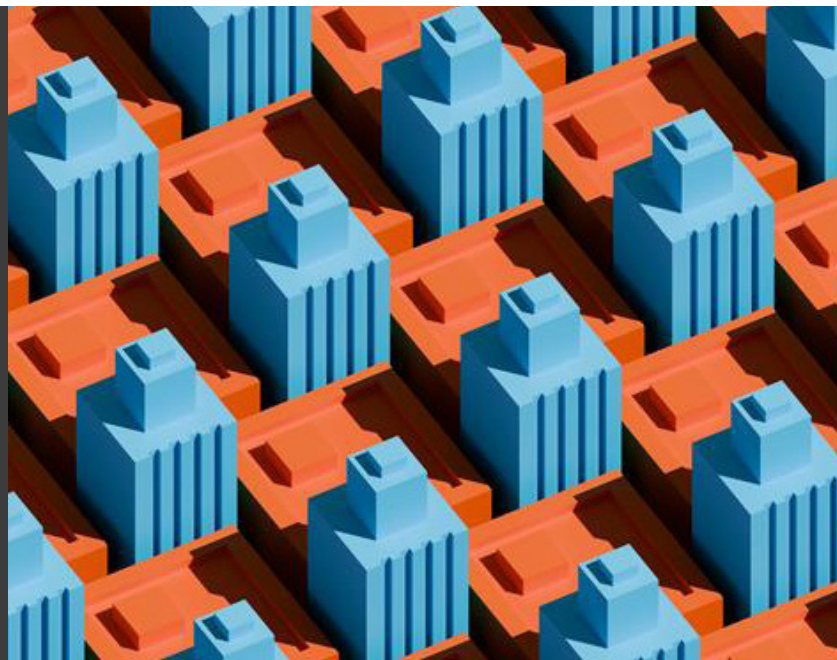
エッジインテリジェンスは自律型金融の台頭につながるでしょう。これは論文「*Beyond the Sum: Unlocking AI Agents Potential Through Market Forces*（総和を超えて：市場の力を通じてAIエージェントの可能性を解き放つ）」で例示されています。

著者らは、自律的なコード生成ができるLLMを活用したAIエージェントが、金融市場において独立した経済主体としてどのように機能するかについて考察しています。

AIエージェントが休みなく稼働し、成功した戦略を完璧にコピーし、情報を瞬時に共有することで、機械のスピードで取引を実行し、市場状況に動的に対応し、人間の介入なしにまったく新しい金融ワークフローを生み出すことで、金融サービスを変革するかもしれません。

しかし、著者らは機械速度の取引に対応するために、克服すべきインフラ上の課題を指摘しました。決済システムの再構築、安全なデジタルIDフレームワークの開発、人間ではなくAIエージェントの利用に最適化した金融インターフェースの創出などです。

DeepSeekによるR1のオープンソース公開は、単なる技術的なマイルストーン（中間目標）にとどまりません。「高度なAIが民主化できる」というメッセージと同義です。



こうした課題が解決されれば、金融機関は、自律型AIエージェントが市場の機能を根本的に変える新たなパラダイムの入り口に立つことになります。

その結果として、金融エコシステムにおいて、前例のないレベルの効率性とイノベーションが実現する可能性があります。一方で責任やリスク、コンプライアンスに関する複雑な規制上の課題も生じることが予想されます。

なぜこれらの波がフィンテック革命を超える可能性があるのか

2010年～2020年のフィンテック革命は、クラウドと携帯端末が一般化し、参入の壁を下げて開発を高速化するプラットフォームが進化したことで起きました。小規模なスタートアップが、既存企業が提供できなかった洗練されたデジタル体験を提供できるようになりました。

しかし、生成AIがソフトウェアを凌駕する可能性は、おそらくそれ以上に破壊的だと言えるでしょう。

- **プログラミング支援**: エンジニアの複雑なシステム構築を高速化し、新しい金融アプリの試作品も驚くほど速く作成自動デザイン設計: 新しい画面設計や販促企画の見本を数週間ではなく数時間で作成。
- **自動デザイン設計**: 新しい画面設計や販促企画の見本を数週間ではなく数時間で作成。
- **エージェント連携**: コンプライアンス文書の作成、データ検証、規制当局へのメール送信などの複雑な作業を、AIエージェントが協力して実行し、間接費を大幅に削減。

金融サービスは、リスク分析から法的文書作成、引受業務から資産管理に至るまで、知識集約型であるため、特にこのAIの波の影響を受けやすい分野です。

デジタル変革で既存企業は「スピードこそ命」と痛感しました。今回のAI革命では、これまで以上のスピードで対応しなければなりません。

コスト障壁の崩壊と開発期間の短縮により、新規参入企業は前例のない速度とコストで、高度なオンライン銀行やAIベースの保険会社を立ち上げられるようになりました。彼ら

はお客様の獲得からサービス提供まで小型の生成AIモデルを組み込み、ますます速いペースで反復と改善を行うことが可能です。

一方、レガシーシステムや官僚主義に縛られた既存の組織は、大規模な再編が陳腐化という潜在的なリスクに直面しています。

金融サービスのリーダーに必要なアクションステップ

1. **転換点を評価する**: より小規模・低コストのオープンなAIが本物かどうか、様子見を続けますか。それとも小規模な実証実験を始めますか。Rita McGrathが著書『Seeing Around Corners (先を見通す)』で説いた「転換点が明らかになってからでは遅すぎる」を参考にしてください。
2. **隣接する可能性を理解する**: AIが既存の業務フローをどのように改善できるか（一次的な変化）だけでなく、AIによって自社のビジネスモデルや業界がどのように根本的に変わりうるか（二次的・三次的な変化）についても検討してみてください。
3. **すべてを再考する**: フォードが分散型電気モーターを中心に工場を再建したように、エッジでの分散型インテリジェンスを中心にビジネス全体を再考してください。
4. **スキルアップと再構築を図る**: これまでの知識社会では知識を身につけるための学習や研修が中心でしたが、AI時代ではAIを構築・拡張する能力が必要になります。まったく新しいパラダイムが求められます。
5. **周辺で実験する**: デジタル変革がイノベーションラボやインキュベーターを牽引したように、この加速する変化の速度には、実験と変革のための新しい仕組みを必要とします。
6. **スピードとガバナンスの両立を図る**: AIの偏見やミスの脅威は、特に規制分野で現実のものとなっています。金融機関は、アルゴリズムの公平性、モデルの説明可能性、消費者保護に関する規制の進化に注意しなければなりません。これらのリスクを管理するため、しっかりとした監督体制を構築してください。

DeepSeekによるR1のオープンソース公開は、単なる技術的なマイルストーン（中間目標）にとどまりません。これは、「高度なAIが民主化できる」という強いメッセージです。マイクロコンピューターがメインフレームの独占を崩したように、R1の登場は、最上位AIモデルをほぼ独占した一握りの大手企業から権力が移ることを告げています。心理的な壁は崩れ去りました。

金融業界のリーダーは、デジタル変革がいかにフィンテック系スタートアップの急増を引き起こし、決済から投資まであらゆる分野を塗り替えたかを思い出してください。同じシナリオが再び展開されます。ただし、今回はより進化して起こります。つまり、知識労働の大規模な自動化、エッジへのインテリジェンスの組み込み、AIを活用した何千もの新たなビジネスモデルの誕生です。

転換点とは、優位性のために再構築しなければならない時です。DeepSeekが「十分に良い」知能を小型かつ安価に実現したことは、まさに私たちが新たな境界を越えたことを意味します。金融サービスの経営陣にとっての最善策は、早く波を受け入れ、成功を拡大し、必然的に出現する新たなビジネスモデルに対して警戒を怠らないことです。

私たちはインテリジェンス変革時代の幕開けにいます。モバイルフィンテックの登場や蒸気から電力への転換と同じくらい劇的なことが予想されるので、気を引き締めてください。

もしこの10年のフィンテック革命がめまぐるしいものだったと感じたなら、知識労働がコモディティ化されるこれからの変化は、その比ではありません。

結局、一度「4分の壁」が破られれば、記録はさらに速くなるだけです。

デジタル変革は企業に対しては俊敏性の重要性を痛感させました。今回のAI革命では、これまで以上のスピードで対応しなければなりません。

