

DeepSeek's Game-Changing Launch: Why “Good-Enough” Intelligence Could Transform Financial Services—Again

By



Kes Sampanthar

Partner
Valize



Scott Wolfson

Managing Director
Valize

In early January, DeepSeek released its breakthrough reasoning model, R1, under an open-source license, and AI insiders have not stopped buzzing.

The orthodox view had been that serious Large Language Models (LLMs)—the ones capable of advanced, automated workflows and sophisticated augmented intelligence—had to be massive, closed and astronomically expensive to train. Capital and compute costs, not to mention the requisite billions of tokens in training data, placed these high-end models almost exclusively in the hands of a few tech giants.

DeepSeek upended that idea overnight.

By demonstrating “good-enough” intelligence at a fraction of the size and cost of top proprietary systems, DeepSeek’s open-source R1 model proved that large, expensive and closed is not the only way forward. In fact, the open sourcing of key breakthroughs from DeepSeek-R1 underscores an even bigger truth: the barriers to entry for advanced AI have just fallen dramatically, akin to the four-minute mile barrier suddenly being broken.

Roger Bannister's sub-four-minute mile record in 1954, while legendary, stood for a month. Now that DeepSeek showed the world what can be achieved with a smaller model trained far more cheaply, a flood of innovators are set to follow.

Within weeks of DeepSeek-R1's open-source release, the repository was downloaded over 2 million times, with more than 10,000 forks appearing on GitHub. (Think of a fork as an iteration that does not alter the original.) This surge of developer interest signals that many players—startups, enterprises, and academic labs—are eager to harness DeepSeek's smaller-footprint approach.

While comparable proprietary systems can cost up to USD \$100 million to train, DeepSeek's R1 reportedly cost only around USD \$6 million, demonstrating a staggering reduction in barrier-to-entry costs for advanced AI. (Worth noting: the \$6 million is attributed to just the processing chip cost of the pre-training run, which is only a portion of the total cost of the model.)

We have seen this pattern play out before: once the technology becomes cheap and ubiquitous, no single group can monopolize it. Disruption follows swiftly, and we see an explosion of smaller innovators reimagining what's possible. If intelligence will soon be distributed and embedded across the business value chain, the financial services sector, steeped in knowledge work, is particularly poised for disruption.

Breaking the psychological barrier

In the generative AI domain, the prevailing wisdom was that the best models must be trained with tens (or hundreds) of millions of dollars' worth of cloud compute and mind-boggling data sets, then locked down behind APIs.

DeepSeek's smaller, cheaper approach shattered that assumption. Like the four-minute mile, it was less about raw capability and more about the symbolic moment. If DeepSeek trains agentic-capable intelligence at a sliver of the usual cost, then the entire market psychology around LLMs and advanced AI changes.

Now any team—be it a small startup, a university lab, or a mid-market financial services company—can hope to replicate or even surpass that performance. Especially once DeepSeek decided to open source its R1 breakthroughs, the cat was truly out of the bag. A wave of new experiments, forks, and specialized AI solutions is almost inevitable, mirroring the open-source software booms we have seen in other domains.

The unintended catalyst: China's chip ban

Ironically, one of the triggers for DeepSeek's innovation was the export ban on advanced computer processing chips (GPUs) to China. Many assumed that lack of cutting-edge chips would hobble Chinese AI research. Instead, the DeepSeek team developed more efficient training methods, found novel algorithmic tricks, and overcame the hardware limitation by building a smaller, more optimized model. This new model roughly matches the performance of far bigger, "closed" systems on a wide range of tasks.

What was once a formidable barrier—the need for top-tier GPUs—just toppled. It's reminiscent of how constraints can spur breakthroughs: deprived of the easiest route (just throw more hardware at the problem), the team discovered a new solution. Now, with R1's code open-sourced, that solution belongs to everyone.

While comparable proprietary systems can cost up to USD \$100 million to train, DeepSeek's R1 reportedly cost only around USD \$6 million.

Historical pattern 1: The ubiquitous nature of technology

Today, AI follows the same script as the microcomputer revolution of the 1980s. Smaller, cheaper desktop machines took over many tasks once batched for the central mainframe. Standardization followed. Workflows accelerated. Moore's Law propelled technological advancement—cheaper, faster, better—fueling a mobile revolution that penetrated every aspect of business value chains through the consumerization of technology.

While large LLMs initially required massive cloud data centers, smaller open-source models now run on local devices (edge computing, specialized but cheaper hardware), delivering good-enough intelligence for agentic workflows—generative AI's most disruptive impact. Just as microcomputers empowered employees and launched digital transformation, these smaller, open AI models will unlock a new wave of innovation.

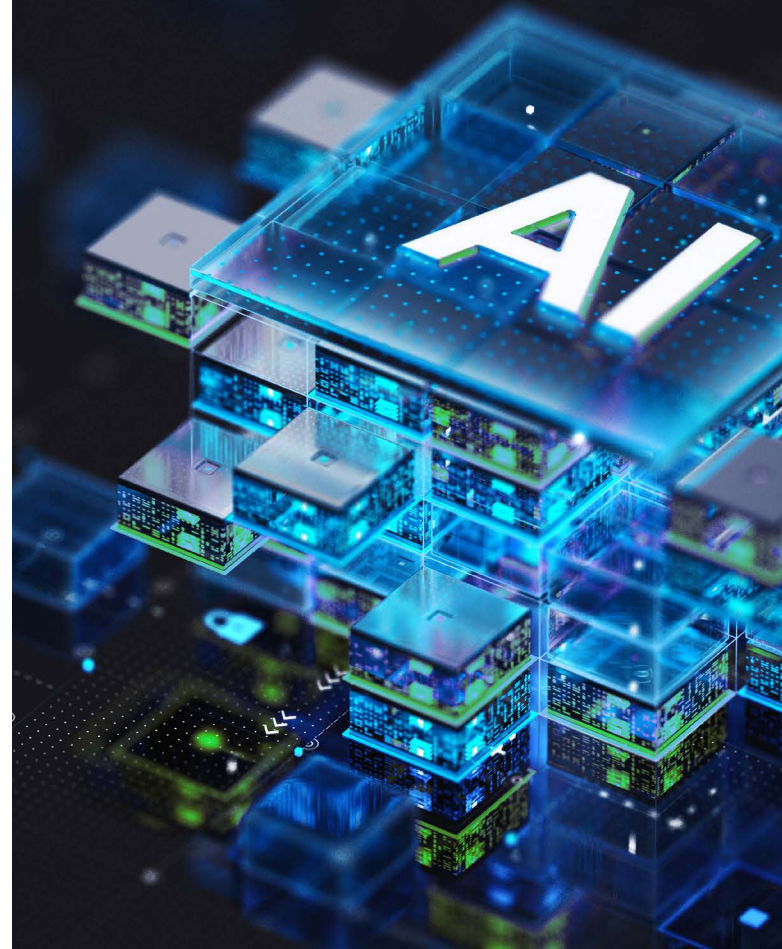
The lesson here: Technology evolves from large, slow, centralized systems to cheap, fast, decentralized ones, eventually penetrating every corner of business and upending traditional value chains.

Historical pattern 2: The transformative impact of GPTs

Even before the microcomputer analogy, history offered a powerful lesson in the shift from steam to electricity in factories. Early adopters of electric power simply replaced their steam engine with one giant electric motor—never rethinking the factory layout.

Productivity gains were minimal at first. The real breakthrough came later: when Henry Ford and others realized they could distribute smaller electric motors across the entire factory floor, thus reorganizing workflows around assembly lines. The reconfiguration yielded profound efficiency boosts.

Technology evolves from large, slow, centralized systems to cheap, fast, decentralized ones, eventually penetrating every corner of business and upending traditional value chains.



However, that crucial redesign took decades—historians estimate it took around 40 years for most manufacturers to figure out electricity’s real potential. This phenomenon is sometimes called the “dynamo effect”: The technology existed, but old mental models kept people from re-architecting their businesses.

The dynamo effect applies in the case of a General Purpose Technology (GPT)—a term coined by economist Robert Solow in the 1980s describing technologies with wide-ranging applications across industries that transform society and economy.

Understanding a GPT’s potential can take time. The breakthrough occurs when, like Ford, the GPT drives new forms of competitive advantage through business model transformation.

Now, we stand on the cusp of another GPT wave. DeepSeek and smaller open-source AI models bring distributed intelligence. Instead of knowledge-processing being monopolized by a few big, closed models, AI can proliferate across devices—from smartphones to IoT sensors, from personal laptops to custom on-premises deployments.

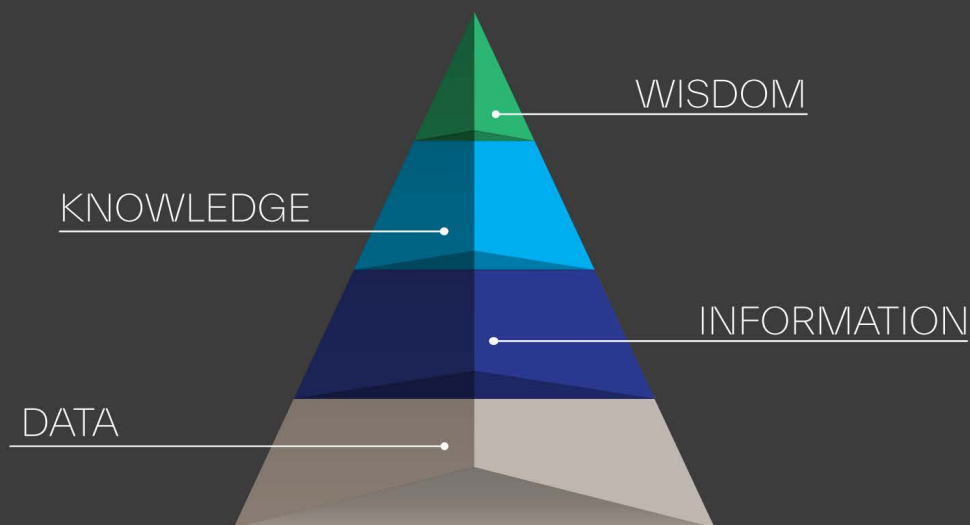
As with the shift from steam to electricity, real transformation will require rethinking business models and arenas, not merely substituting cheap AI models for large, expensive ones.

How transformation happens

To understand how waves of technology drive transformation across businesses and industries, it is useful to examine how data gets translated into wisdom—a process structured by the “DIKW” Hierarchy: Data → Information → Knowledge → Wisdom.

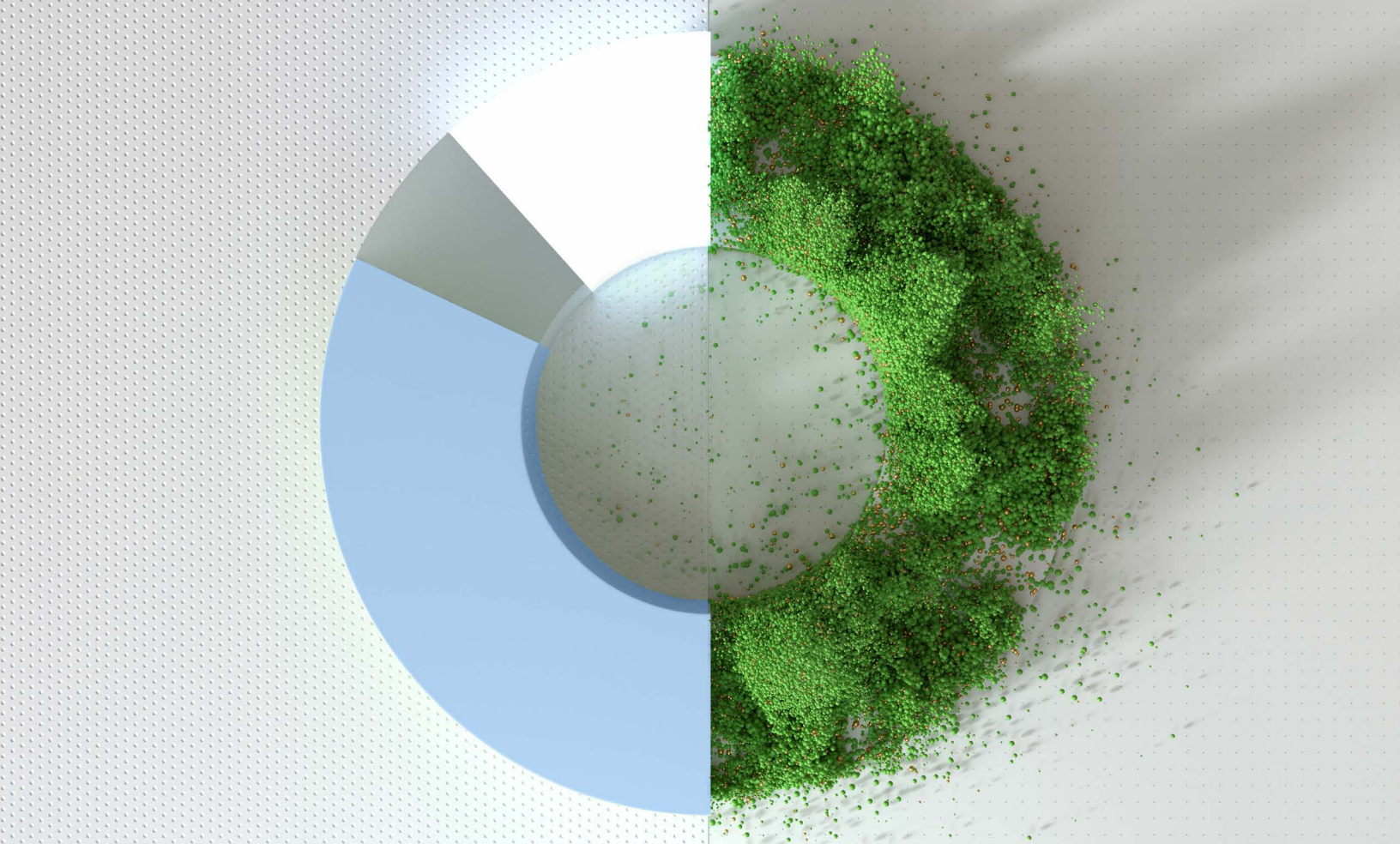
As computing systems evolved, they increasingly moved up this hierarchy, shifting responsibilities from humans to machines. This transformation allowed people to focus on higher-value tasks, from processing raw data (early computing) to making strategic decisions with AI-driven insights.

Each major technological wave—from the mainframe era to the personal computing era to the era of smartphones, cloud computing and AI-powered automation—has propelled businesses further up the DIKW pyramid, changing the nature of work and business processes.



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

DIKW hierarchy: We are witnessing a shift from knowledge processing to wisdom augmentation, where AI agents not only analyze information, synthesize knowledge and augment wisdom, but also take action independently.



The next stage is now unfolding: Intelligence transformation, where organizations systematically embed AI to amplify human cognition and decision-making capabilities across all business functions, represents the evolutionary leap beyond digital transformation.

- Instead of merely digitizing workflows, businesses are shifting to augmented and autonomous knowledge work that requires rethinking thoughtflows alongside workflows.
- AI models like DeepSeek R1 are lowering costs, making advanced intelligence widely available at the edges, driving toward the ubiquity of intelligence.
- This transition mirrors historical shifts—from mainframes to microcomputers, from steam power to electricity—where technology unlocked new business models and arenas.

We are witnessing a shift from knowledge processing to wisdom augmentation, where AI agents not only analyze information, synthesize knowledge, and augment wisdom but also take action independently.

Digital transformation of financial services

Between 2010 and 2020, financial services experienced a radical shake-up due to digitalization. Nimble fintech players exploited the dynamo effect of mobile and cloud computing to build entire product lines that chipped away at incumbents:

- Chime and Varo launched mobile-first bank accounts with no monthly fees and frictionless sign-up
- Robinhood introduced commission-free trading on an app, driving established brokerages to eliminate trading fees
- Robo-advisors like Betterment and Wealthfront used algorithms to offer low-cost, automated portfolio management
- Insurtechs such as Lemonade applied mobile UX and digital underwriting to disrupt home and renters' insurance

Startups that can embed advanced intelligence at the edge—in Internet of Things devices, point-of-sale terminals, or augmented reality glasses for branch staff—will push incumbents to adapt faster than ever.

In all these cases, small startups leveraged cheaper and more widely available technology (cloud, smartphones, app stores) to do what only big banks with big infrastructures once could. They also tapped into user frustrations with clunky legacy interfaces or high fees.

The result? An entire generation of new financial products, from peer-to-peer payments to neobanks.

For incumbents, this transformation was a double-edged sword. Many lost ground or faced margin pressure, but they also discovered new revenue streams by partnering with or acquiring fintech startups.

Meanwhile, end-consumers benefited from more competition and better user experiences. We can trace the success of these fintechs directly to the fact that the cost of launching digital services plummeted—mirroring earlier technology disruptions.

Now let's explore the adjacent possible.

First order impacts: Intelligence transformation in financial services

If 2010–2020 was about digital convenience and connected experiences, 2025 onward looks to be about intelligent workflows:

- **Underwriting** that is not only digital, but also run by smaller, efficient AI models on-site, scanning thousands of signals in real time
- **AI co-pilots for financial advisors**, automatically generating compliance-friendly suggestions, emailing clients personalized updates, or analyzing risk factors on the fly
- **Fraud and compliance** monitoring that uses near-real-time AI engines embedded at every point of transaction, rather than a central, high-latency system
- **Customer service chatbots** that are no longer stilted and limited, but nearly human in their capabilities—yet still operating on cost-effective hardware

Just as the mobile shift let smaller fintechs break the old guard's hold on distribution channels, smaller, open AI models may unleash a further wave of disruption.

Startups that can embed advanced intelligence at the edge—in Internet of Things devices, point-of-sale terminals, or augmented reality glasses for branch staff—will push incumbents to adapt faster than ever.

Of course, such rapid innovation must still comply with a complex financial regulatory landscape. Financial services firms must be highly mindful of ensuring algorithmic fairness, data privacy and model transparency to avoid enforcement actions—particularly as regulators update guidance for AI-driven services.

Second order impacts: Intelligence at the Edge

The second wave will tap into one of the under-discussed implications of smaller, cheaper models like DeepSeek: intelligence at the edge. AI can be embedded in:

- **Smartphones:** Instead of always calling a cloud LLM, your phone's local AI might handle most tasks offline
- **IoT devices:** Sensor-laden devices in factories or retail environments can process data and act autonomously
- **Augmented glasses:** Real-time analysis or language translation could happen in your AR headset
- **Branch terminals:** A bank kiosk or an insurance claims terminal can run advanced AI locally for immediate decisions, even if connectivity is limited

The potential for real-time, on-device intelligence lowers friction—similar to how personal computers brought computing to each desk and eventually each pocket.

In financial services, this may translate into personal, always-on AI assistants that monitor spending or investment patterns, or specialized fintech startups that no longer rely on big cloud APIs for advanced services.

As cost barriers drop, we will likely see a Cambrian explosion of new solutions and new entrants, parallel to the fintech wave in the 2010s—but possibly bigger and faster.

Third order impacts: The rise of self-driving finance

Edge intelligence will lead to self-driving finance, as exemplified in the paper, *"Beyond the Sum: Unlocking AI Agents Potential Through Market Forces."*

The authors explore how LLM-powered AI agents capable of autonomous code generation could function as independent economic participants in financial markets.

These agents—operating continuously, perfectly replicating successful strategies, and sharing information instantaneously—could transform financial services by executing transactions at machine speed, dynamically responding to market conditions, and creating entirely new financial workflows without human intervention.

However, the researchers identify significant infrastructure barriers that must be overcome, including reimagining payment systems for machine-speed transactions, developing secure digital identity frameworks, and creating financial interfaces optimized for agent consumption rather than human users.

DeepSeek's open-source release of R1 is more than a technical milestone. It is a statement that advanced AI can be democratized.



As these challenges are addressed, financial institutions stand at the threshold of a new paradigm where autonomous AI agents could fundamentally reshape how markets function.

The potential outcome? Unprecedented levels of efficiency and innovation in the financial ecosystem, but also complex regulatory implications around accountability, risk, and compliance.

Why these waves could surpass the fintech revolution

The 2010–2020 fintech revolution was driven by the commoditization of cloud and mobile and the evolution of platforms that lowered entry barriers and accelerated app development: small startups could deliver slick digital experiences that incumbents hadn't matched.

But generative AI's potential to eat software is arguably more disruptive:

- **Co-pilot coding:** AI helps software engineers build complex systems faster, meaning new financial apps can be prototyped at lightning speed
- **Automated design:** AI can mockup new user interfaces or branding campaigns in hours instead of weeks
- **Agentic orchestration:** Complex tasks—like generating entire compliance documents, verifying data, and emailing regulators—could be done by AI agents working together, slashing overhead

Financial services is particularly vulnerable to this wave of AI because it is intensely knowledge-based, from risk modeling to legal documentation, from underwriting to wealth management.

If digital transformation initiatives taught incumbents hard lessons about agility, this wave will demand they apply those lessons even more swiftly.

The meltdown of cost barriers and acceleration of development cycles enables challengers to launch

advanced, online banks or AI-based insurers at unprecedented speeds and costs. They can embed smaller generative AI models throughout customer acquisition and servicing, iterating at increasingly rapid rates.

Meanwhile, older institutions burdened by legacy systems and bureaucracy face potential massive reorganization or obsolescence.

Action steps for financial services leaders

1. **Assess your inflection point:** Are you waiting to see if smaller, cheaper, open AI is real, or jumping in with small pilots? Consider guidance from Rita McGrath's *Seeing Around Corners*: once inflection points are obvious, you're too late!
2. **Understand adjacent possible:** Brainstorm not only where AI can improve existing workflows (first-order) but also how it can fundamentally change your business model and arena (second- and third-order).
3. **Rethink everything:** Just as Ford rebuilt factories around distributed electric motors, consider rethinking your entire business around distributed intelligence at the edge.
4. **Upskill and reshape:** While the knowledge era drove learning and development to build knowledge, intelligence transformation will require building and augmenting intelligence—demanding an entirely new paradigm.
5. **Experiment on the fringe:** If digital transformation drove innovation labs and incubators, this accelerated pace of change will demand new models for experimentation and innovation.
6. **Balance speed and governance:** The specter of bias or error in AI is real, especially in regulated fields. Financial services firms, in particular, must be wary of evolving regulations around algorithmic fairness, model explainability, and consumer protection. Build robust oversight frameworks to manage these risks.

DeepSeek's open-source release of R1 is more than a technical milestone. It is a statement that advanced AI can be democratized. Much like the microcomputer did to mainframes, it heralds a power shift away from the handful of major players with near-exclusive control over top-tier AI models. The psychological barrier has been broken.

Financial services leaders should recall how digital transformation spurred a burst of fintech ventures that reconfigured everything from payments to investing. That playbook will run again, but on the next level: automating knowledge work at scale, embedding intelligence at the edge, and opening the door to thousands of new AI-powered business models.

Inflection points are times when you must reconfigure for advantage. DeepSeek's demonstration that good-enough intelligence is both smaller and cheaper means we have just crossed a threshold. The best move for financial services executives is to embrace the wave early, scale up the successes, and remain vigilant for the new business models that will inevitably emerge.

We are at the dawn of an intelligence transformation era. Buckle up, because it promises to be every bit as dramatic as the advent of mobile fintech or the shift from steam to electricity.

If the last decade's fintech revolution felt like a whirlwind, wait until you see what happens when knowledge work itself becomes commoditized.

After all, once a four-minute mile has been run, the record only gets faster.

If digital transformation initiatives taught incumbents hard lessons about agility, this wave will demand they apply those lessons even more swiftly.

