

Global Innovation

March 2025



Innovation Insights from the US – Q1 2025



Storm Uru

Co-Lead Fund Manager of
Global Innovation



Clare Pleydell-Bouverie

Co-Lead Fund Manager of
Global Innovation

During times of market volatility, the most important thing we can do is focus on the fundamentals of our companies. We have recently returned from our quarterly trip to the US, meeting with around 30 management teams across New York and Silicon Valley. While the market backdrop was dynamic, the insights we gained were invaluable, and have underscored our view that markets have predominantly been driven by sentiment rather than fundamentals for much of the past month. Little has changed in the earnings power of innovators since the start of the year and the pace of innovation is only accelerating. We have returned with six key insights:

1. Funding and building the largest infrastructure buildout since the Industrial Revolution
2. The collapsing cost structures of financial platforms sitting on unique datasets
3. Challengers vs. incumbents – emerging winners of the next decade
4. Software 2.0 & ontology – why software 1.0 cannot compete
5. The AI infrastructure buildout – underappreciated implications of inference deployment at scale
6. The next wave of AI – new form factors

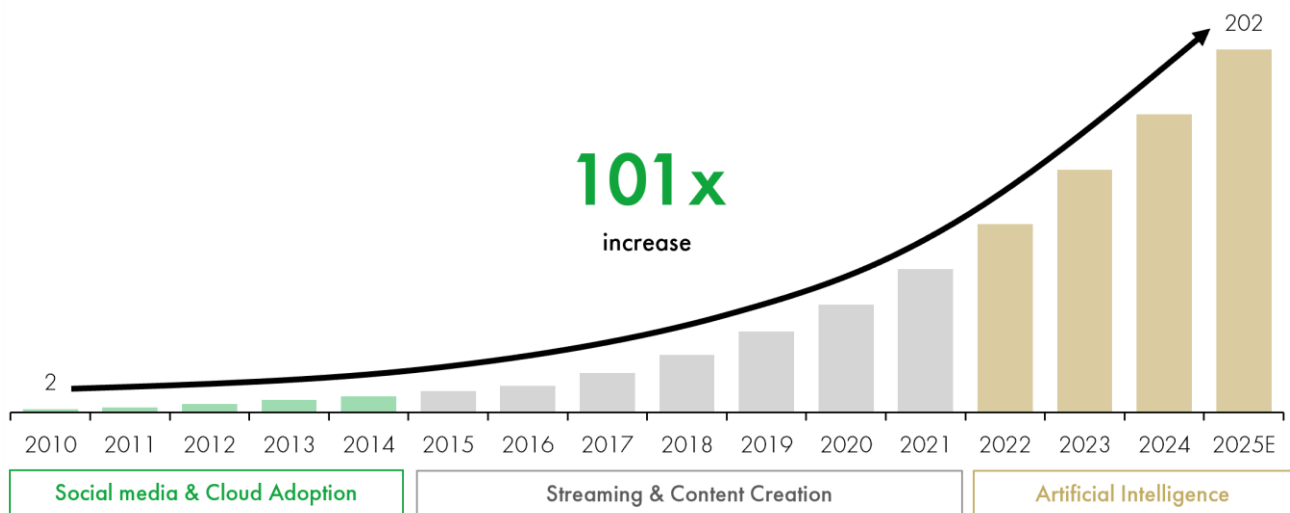
Funding and building the largest infrastructure buildout since the Industrial Revolution

The convergence of data centres and power is creating a generational investment opportunity for the financiers of these assets. We have spoken at length about the compute required for the AI infrastructure buildout, as well as the components that go into these data centres, but the funders and financial operators of this digital infrastructure get less airtime. What is currently underway is the largest infrastructure buildout since the Industrial Revolution; this is coinciding with increasing investor demand to access this theme through private equity, private credit and real asset vehicles.

In 2024, we started to see compelling opportunities emerging across our watchlist in the financial sector, especially alternative asset managers: the cost of capital has come down, transaction pipelines are rebuilding, and at least some degree of deregulation is now widely expected. Moreover, we could see powerful secular drivers coming to the fore, particularly surrounding infrastructure. New York is the global epicentre for finance, and we always skate to where the opportunity lies.

Over the past 15 years, data usage has increased 100x, but more striking has been what has happened in the past three years, catalysed by the take-off of generative AI. In the past three years, the world has created and consumed more data than in the entirety of history.

Explosive growth in data – global data created, consumed and stored in Zettabytes



Source: Blackstone, 2025

It's not just the sheer amount of data that's growing, it's the intensity of that data being processed. If we were to ask SORA (OpenAI's generative AI video making tool) to create a very basic AI-generated video of our trip around New York, this consumes 10,000x the power and data consumption versus a simple Google search query, which is representative of legacy modes of digital interaction.

With this explosion in data we need physical infrastructure to store, process, and deliver it. We also need more power and energy infrastructure – the current transmission queue in the US is roughly double that of the US's available electricity capacity. This is why the likes of Sam Altman have recently proposed building clusters of 5,000-megawatt data centres across the US, which might sound extreme, but is indicative of the scale of the infrastructure that is needed to meet this demand. There is much debate over whether the buildout hype is justified, but this is what is being leased and built as we speak – the large and long-term capital allocators standing behind these projects see demand first.

What was abundantly evident to us is that first mover advantage and scale matter here – both the capacity to originate as well as providing the right cost and form of capital. Take Blackstone, which saw this coming in 2021 when the company acquired QTS – now the largest and fastest growing data-centre platform in the world. Apollo, the pioneer of private credit, is similarly leaning into the emerging trend of financing long-duration digital infrastructure assets through private credit as traditional banks step back.

Capital demand over the next decade for digital infrastructure is estimated to comprise \$15-20 trillion, on top of \$30 trillion for power & utilities and a further \$30-50 trillion for the energy transition. These are not small numbers; they underscore a robust structure growth pipeline for those financiers who have moved quickly and are now sowing the seeds for future returns on their investments.

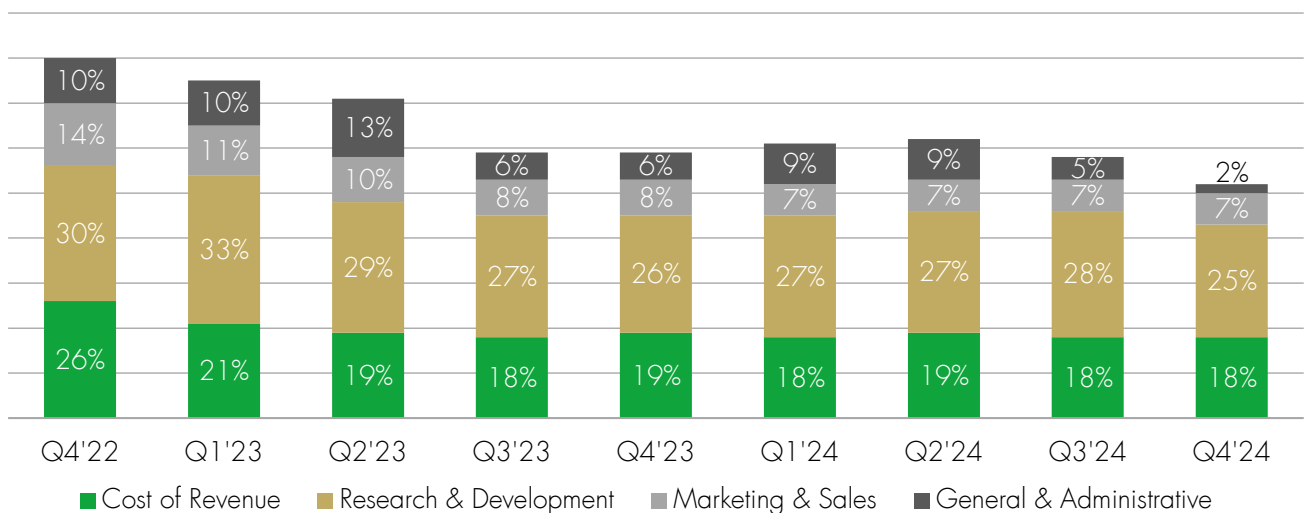
The collapsing cost structures of financial platforms sitting on unique datasets

Evidence of AI-driven revenue growth is everywhere, even if it is not always readily apparent. It's not just dedicated AI firms like OpenAI and Anthropic that are monetising their innovations by offering new AI-powered products and services. Companies across a range of industries are also racing to deploy AI and secure a first-mover advantage in their markets. These organisations leverage AI to launch novel features and enhance customer experiences, thereby attracting business and unlocking new sources of income. We have observed this pattern in large consumer platforms, and now a similar transformation is underway in financial services.

Large business-to-consumer (B2C) platforms such as Shopify, Uber, Airbnb, and Meta have been quick to incorporate AI into their products and services. Management teams at these companies are racing to embed AI features that deliver greater value to users, such as personalised recommendations or automated customer support. By enhancing the user experience in this way, they reduce customer churn and increase engagement.

The financial benefits of this strategy are substantial. Improved customer retention means each new user acquired generates a higher lifetime value, boosting the return on investment (ROI) of marketing and customer acquisition spend. Additionally, many AI-driven improvements do not require significant ongoing costs. By maintaining a relatively fixed cost base even as usage grows, these platforms are achieving strong operating leverage and expanding their profit margins.

Meta Platforms – Expenses as a percentage of revenue



Source: Meta Platforms Q4 2024 earnings presentation.

In our recent research trip to New York, we examined a select group of financial services companies that are capitalising on AI. These firms possess vast troves of proprietary data, complemented by continuous streams of live data flowing into their operations. Some are long-established industry leaders and others are newer entrants, but all share a common trait: they recognise their first-mover advantage in applying AI and are moving quickly to extend their competitive lead in the market.

One category consists of traditional industry leaders, exemplified by Moody's. This century-old firm has built up data assets that are extremely difficult for competitors to replicate. Moody's is now deploying AI-powered solutions to enhance its core offerings in risk assessment and management. It is also integrating AI-driven tools into its research and data analytics processes to improve efficiency and insight.

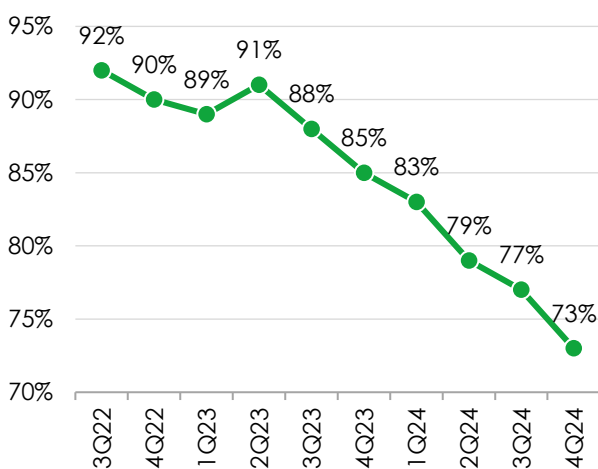
The immediate effect of these AI initiatives is greater product stickiness: clients become more deeply integrated into Moody's ecosystem and are less likely to switch to alternatives. A secondary effect is the opening of new markets and customer segments for the company. For instance, Moody's has developed AI agents that smaller financial institutions (tier 2 and tier 3 banks in the US) can use to automate end-to-end processes and digitise previously analogue workflows. By offering such advanced tools, Moody's is extending its reach to serve organisations that may not have been able to leverage its data services in the past.

Another category includes software-first financial companies that can scale efficiently with minimal increases in workforce. Historically, growth in financial services was closely linked to the size of an organisation’s headcount. Now, however, it hinges more on the quality of the product and the speed at which new customers can be acquired.

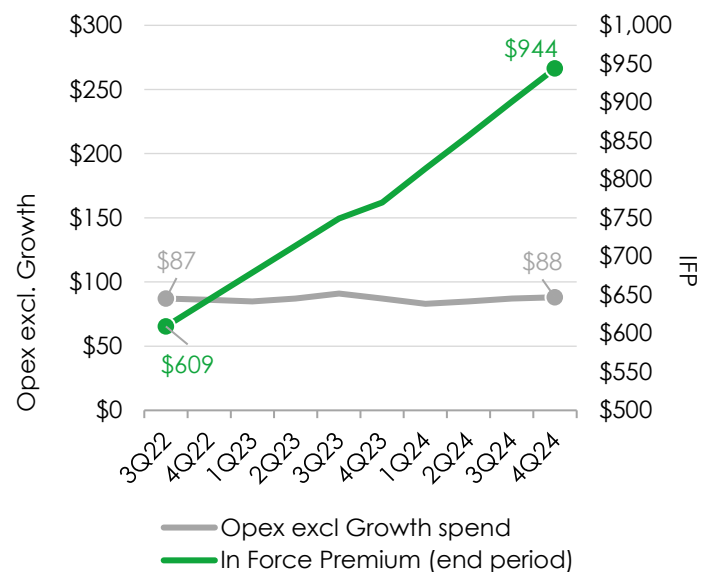
Lemonade, an AI-powered insurance company we visited in New York, exemplifies this model. It is achieving impressive growth in its business and revenues without a corresponding increase in headcount. Lemonade offers an excellent customer experience through its digital insurance platform, while relying on high levels of automation behind the scenes. This allows the company to keep operational expenses flat as it scales, effectively showcasing a fundamentally new business model in the industry.

Lemonade

Trailing 12-months gross loss ratio over time



In-force Premium (IFP) vs. operating expenditure excluding growth spend



Source: Lemonade, March 2025

AI is proving to be a transformative force across industries, and financial services are no exception. Companies that effectively harness proprietary data through AI are realising improved customer retention, new revenue opportunities, and greater operational efficiency. These first movers in AI adoption are establishing competitive advantages that can translate into significant value creation over time. From an investment perspective, it is increasingly important to identify firms with strong AI-driven strategies, as such companies are poised to benefit from enhanced growth and profitability as AI continues to reshape the industry.

Challengers vs. incumbents – emerging winners of the next decade

If there was one unifying theme of our trip, it was identifying challengers to today’s incumbents. As accelerated compute infrastructure continues to expand and the cost of intelligence declines, we are seeing emerging innovators exploiting lower cost bases and faster innovation cycles to deliver superior products at a fraction of the cost of many dominant companies today. This is true of every sector, not just technology. The winners of the next decade will be different to the winners of the last.

New winners of tomorrow



Take Pure Storage, which is collapsing the cost of enterprise data storage by 50% versus the industry standard and set to disrupt legacy names like HP and Dell. Pure combines unique hardware (custom built direct flash modules as opposed to solid state drives and hard discs) with a software layer that means that customers can upgrade their storage seamlessly, with a click of a button. Previously customers had to replace their old storage systems every 3-5 years and migrate the data onto new drives every single time. This is both cumbersome and costly. As data storage and orchestration becomes an increasingly important part of every enterprise's AI strategy, extracting a 50% better total-cost-of-ownership makes a big difference.

SoundHound is another company engaged in a battle with Big Tech to solve Voice AI, but is now running away with the competition and growing at a current run-rate of 100% year-on-year. What we have all learned over the past decade from disappointing engagements with Siri and Alexa, is that Voice AI is a very difficult problem to solve. SoundHound has a two-decade head start on competition and, as a result, its multimodal, multilingual model is 20-30% more performant than its rivals (which include both Google and OpenAI).

When automakers, restaurants and financial institutions are looking to integrate voice AI into their platforms to perform functions such as customer service, they are turning to SoundHound. If you go to a drive through at Burger King here in the UK, the entire experience is now powered by SoundHound's Voice AI. If you purchase a Lucid Motors car, you can now ask your car through the medium of voice to wind down the windows or turn on the air conditioning. Voice is quite simply the most natural interface to engage with AI; the issue to-date has been the technical feat of the challenge. Having solved for this problem, SoundHound is now set to lead the next frontier in the digital revolution for brand interaction.

Finally, salad-chain Sweetgreen exemplifies a non-technology company exploiting technology to disrupt the fast food restaurant industry. Winning in the quick serve restaurant space is hard – it requires both a compelling menu as well as an ultra-efficient business model. Sweetgreen excels in the former with its novel farm-to-fork experience, but the latter is what makes this company unique.

Sweetgreen's 'infinite kitchens', one of which we toured in New York, are automated end-to-end. Think of its kitchen as comparable to a Tesla factory, with the majority of US restaurant chains still working off Ford assembly lines. As a result, Sweetgreen reduces labour costs (the largest line item for restaurants) by a third versus its traditional kitchens. This is a significant cost saving and will feed through to operating leverage as the company continues to open new stores and drive topline momentum.

Software 2.0 & ontology – why software 1.0 cannot compete

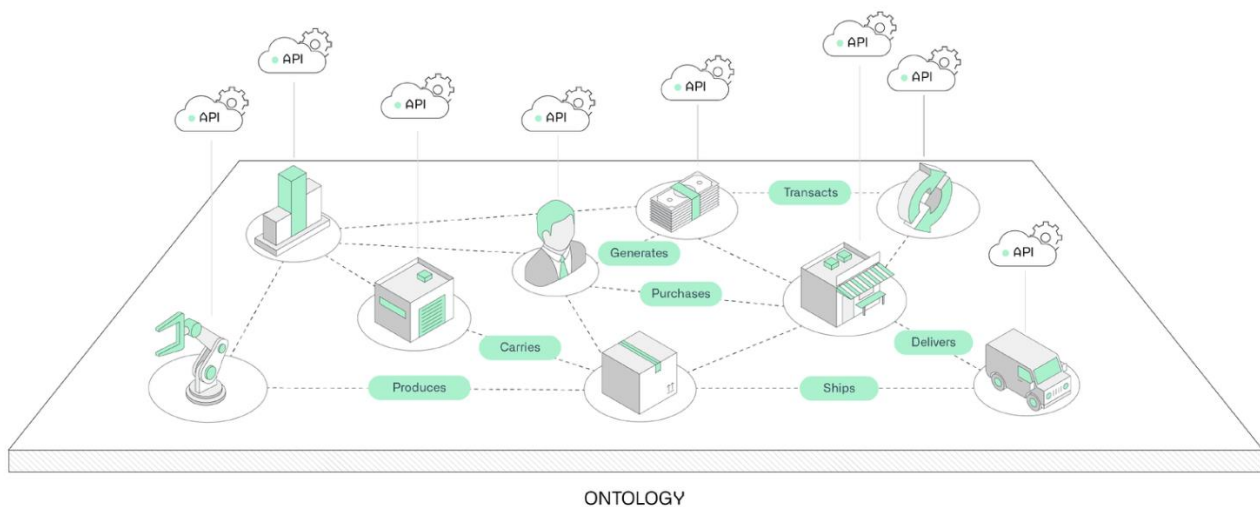
Traditional factory floors and large enterprises often grapple with dozens – if not hundreds – of software vendors and siloed data repositories scattered across difficult-to-access locations. The key question is how incumbent market leaders can keep pace with digital-native start-ups. The answer lies in adopting and embedding an ontology-based approach.

Ontology transforms your digital assets – including data, models, and processes – into a dynamic, actionable representation of the business for all users to leverage in operations:

- **Objects:** digitally represent the real-world entities, relationships, and events that constitute your business.
- **Relations:** represent the connections between real-world entities, events, and processes.
- **Actions:** capture the kinetics between objects and orchestrate real-world changes through an enterprise system. Actions can be mapped from existing processes or existing models.

Two notable firms delivering this capability at scale are Palantir and C3.ai: Palantir pursues a product-led growth strategy, whereas C3.ai partners with major cloud providers such as Microsoft and AWS. Crucially, Microsoft's alliance with C3.ai underscores the complexity and rarity of developing robust ontology technology for enterprise environments. So, what are the key aspects of ontology?

1. **From APIs (Application Programming Interface) to Agents:** Traditional enterprise software relies on APIs for data exchange, but Software 2.0 features intelligent agents that decide and act in real time. Unlike rule-based Software 1.0, these agent-driven systems adapt on the fly, boosting efficiency and speed. For instance, rather than manually querying multiple APIs, an agent can autonomously gather and analyse data from numerous sources.

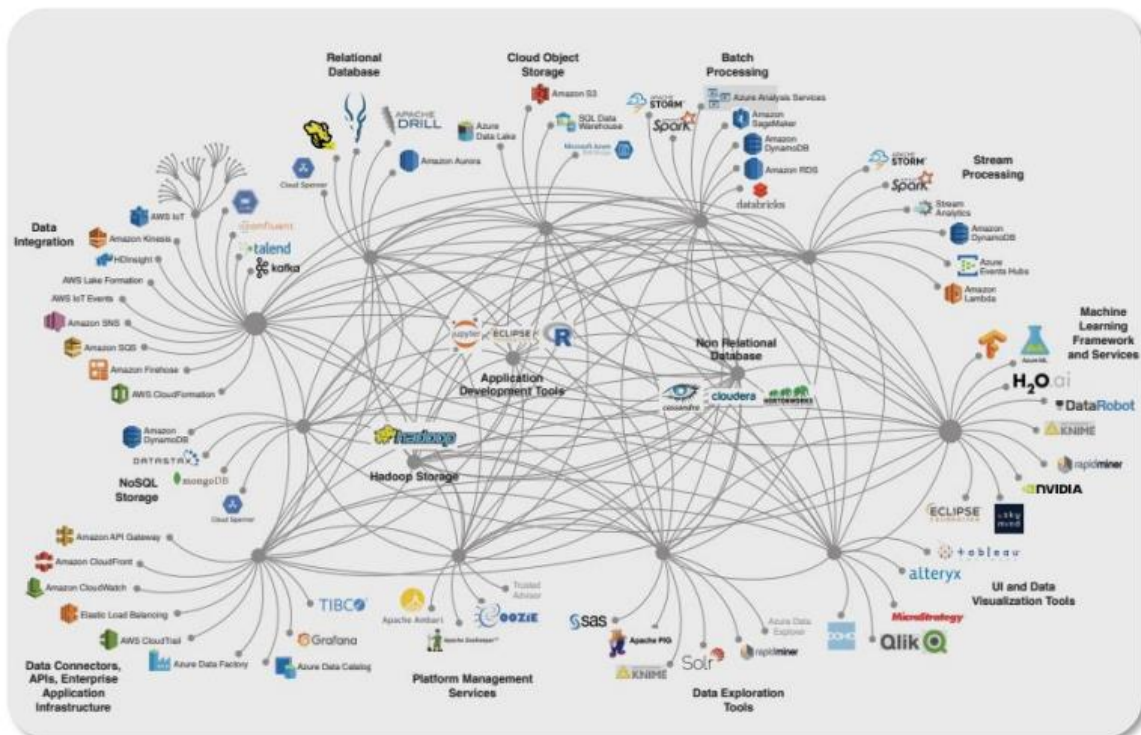


2. **Data integration at massive scale:** Modern AI demands unified data from thousands of systems. Such integration is a prerequisite for robust machine learning models, which rely on a complete picture of the business. Software 2.0 platforms like C3 AI and Palantir offer unified data models and connectors, enabling them to ingest information seamlessly. This capability outperforms siloed, API-centric approaches.
3. **Limits of structured programming:** Traditional structured programming often becomes unmanageable as data sources and services multiply. Writing code for every scenario creates a tangled web of dependencies. By contrast, AI-driven Software 2.0 learns from data rather than relying on exhaustive coding. It can adapt to new inputs, making it far more suitable for today's large-scale, dynamic applications.
4. **Ontology – the master blueprint:** An ontology sets out a common language and defines relationships between entities, ensuring all systems interpret data consistently. This shared framework prevents miscommunication

and enables AI to reason across different datasets without layers of ad hoc integrations. At Palantir, for example, an ontology links all operational concepts—people, assets, workflows—so AI-driven insights and decisions can span the entire enterprise.

- Leading implementations:** Palantir and C3.ai are notable pioneers in ontology-based architecture. Palantir Foundry employs an ontology layer for cohesive data modelling, and the C3 AI Suite uses a unified “Type System” to simplify development and reduce complexity. Beyond these, industry-specific ontologies like the Financial Industry Business Ontology (FIBO) help banks and institutions align data consistently, further illustrating why Software 1.0’s static, siloed methods struggle to keep up.

By harmonising data, employing agent-based design, and learning from real-world complexity, ontology-centric Software 2.0 outperforms older, code-heavy approaches. It scales more effectively, reacts faster to change, and offers a unified view of an enterprise – advantages that Software 1.0 can no longer match.



Source: C3.ai, 2024

Ultimately, ontology disintermediates traditional enterprise software systems to mere data storage functions – dismantling the current value chain between compute, data storage, software, and the end user. With this new technology stack, ontology captures a large piece of the economic profit across the stack and is often referred to as the defining factor behind the winners of enterprise AI.

The AI infrastructure buildout – underappreciated implications of inference deployment at scale

If there is one key debate that divides technology investors at present, it is whether scaling laws are dead. As we have articulated many a time before, the major insight over the last seven years in developing new AI systems is that the scale of compute matters – as we scale both compute and data, we train better models, which opens up use cases for AI and revenue opportunities across the economy.

This trend – what we call pre-training scaling laws – has held true since 2017 and we believe it has at least another four years to play out. We believed this before the release of DeepSeek’s R1 model, before Alibaba’s

Quen model (which were both trained on significantly less compute than previous model iterations), and we believe it after – certainly following engaging with the entire silicon ecosystem in Silicon Valley last week.

What has changed is proof that we can use these scaled models to train distilled ‘student models’ at a fraction of the compute cost. This does not circumnavigate the fact that someone (i.e. the hyperscalers) still needs to be heavily investing to continue scaling these models to push the frontiers of model intelligence. We’ve gone from training clusters to train AI models made up of 4,000 GPUs (or AI chips) in 2022 to 100,000 in 2024, and companies are in the process of building clusters of 1 million GPUs. This is still happening.

More importantly, though, are the emergence of two new scaling laws as we move from AI experimentation to inference deployment at large across the economy. It is good news that the computational costs of training models is coming down, because the inference costs of servicing demand for ‘reasoning’ models is gargantuan.

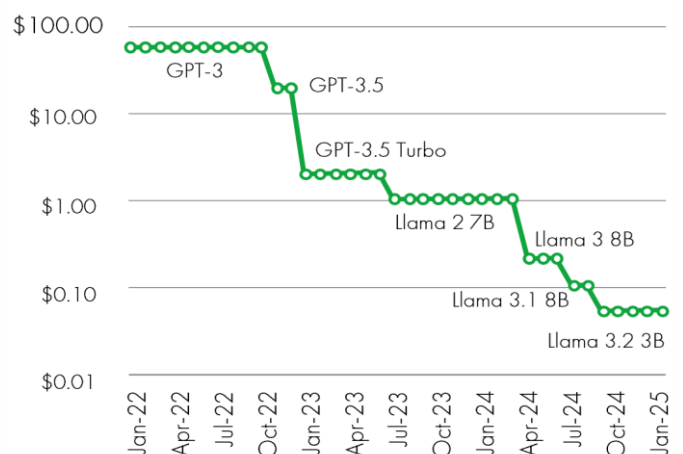
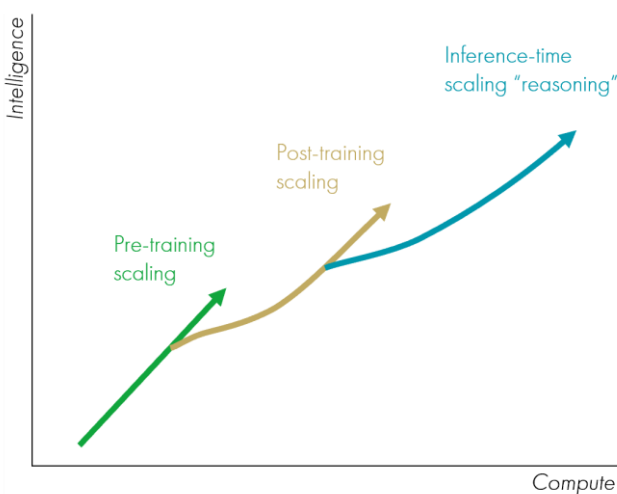
The first of these two new laws is in post-training: this is essentially making the model a specialist by training it on domain-specific data. If pre training made the model as intelligent as a high school student, post training makes the model a college graduate. It is down to advances in post-training using synthetic data generation via AI reinforcement learning that we are on the brink of robotics’ ChatGPT moment with the timeline for humanoid robots being dramatically brought forward. Post-training is vastly data and compute intensive (think of training a robot to have the dexterity and ability to undertake just the first hour of your morning), and as such, post-training compute requirements will soon sail past those pre-training.

Last and by no means least is in inference time reasoning, unlocked by OpenAI’s o1 model in 2024 with open-source iterations of course now announced by DeepSeek, Alibaba and others. These models can now ‘reason’, using complex multi-chain of thought, as opposed to simply predict. The key statistic that we have learned over the past two weeks, reinforced in the past few days at Nvidia’s GTC conference, is that these reasoning models require 100x more compute than prior model generations trained to give ‘on shot’ answers. This makes sense – the longer the model thinks, the more compute is required.

Therefore, as the cost of inference continues to plummet – aided by the innovations of Nvidia and model providers building on top of Nvidia, this enables broader adoption of reasoning models, broader adoption of inference by more than a handful of enterprises. Servicing this demand, though, requires more compute, not less. This is what the market has largely got wrong in our eyes.

Plummeting cost of inference enables broad adoption – Cheapest LLM above 42 MMLU cost/1 m tokens

From one to three scaling laws



Sources: Nvidia 2024

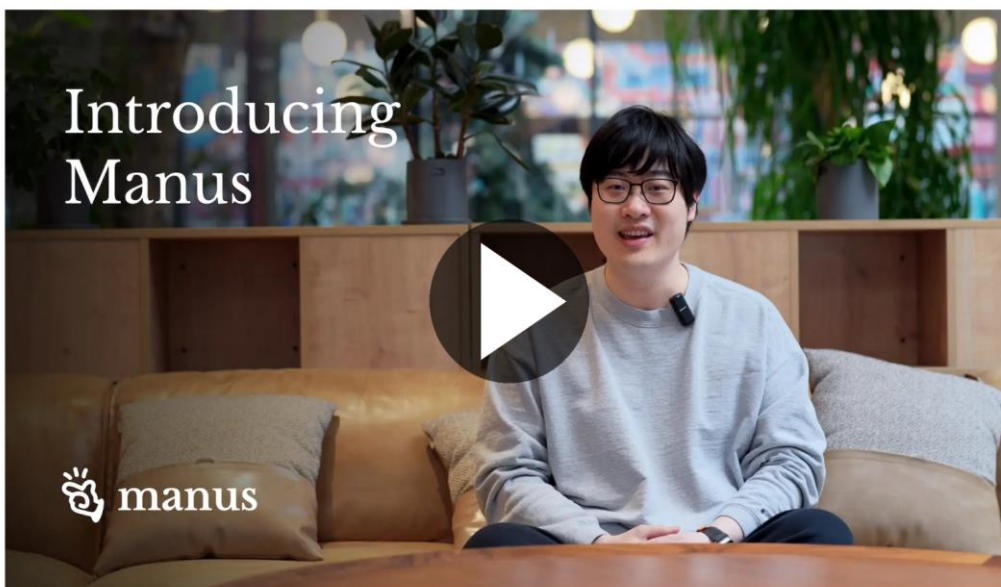
The next wave of AI – new form factors

The newest wave of artificial intelligence takes shape in three transformative domains: AI agents, humanoid robotics, and autonomous vehicles. Each domain extends AI's capabilities beyond conventional applications, heralding a future in which autonomous systems perform tasks once reserved for humans.

AI agents are a notable step forward from rule-based chatbots and narrow machine learning applications. Rather than simply responding to prompts, an advanced agent can autonomously initiate tasks, learn from context, and act without continuous human intervention. One example, Manus.AI, is often recognised as the first general AI agent, capable of orchestrating multiple specialised models to solve complex challenges. The pace of adoption is breathtaking with over two million people signing up for the Manus.AI waitlist within the first seven days.

Its architecture involves decomposing tasks into manageable steps, executing them in sequence, and refining its approach through iterative feedback. In practice, such an agent might handle extensive research or project management duties, consulting internal and external data sources to generate actionable outcomes. This self-directed capability foreshadows a future in which AI moves beyond passive assistance to become a proactive collaborator, raising both promising opportunities for enhanced productivity and new ethical considerations around autonomy and accountability.

[Introducing Manus: The General AI Agent](#)



Humanoid robotics represents the second frontier of this new AI wave. By marrying advanced cognitive systems with agile, human-like bodies, robotics engineers seek to create machines that can operate seamlessly in spaces designed for people. In China, Unitree has demonstrated full-size humanoids capable of walking, jogging, and performing intricate movements with little reliance on external controls. Meanwhile in the United States, Figure 01 is a humanoid robot intended to serve as a general-purpose worker in environments ranging from warehouses to hospitals.

These robots rely on AI-driven perception and decision-making, enabling them to handle tools, navigate complex settings, and interact with humans. The technological significance lies in extending AI's reach beyond virtual domains; humanoid machines capable of working safely alongside people could transform labour-intensive industries, address workforce shortages, and take on hazardous tasks traditionally performed by humans – by the end of 2025 thousands of humanoid robots will be deployed in factories for early test use cases.

[World's First Side-Flipping Humanoid Robot: Unitree G1](#)



The third key domain is autonomous vehicles, which apply AI to real-time decision-making on roads. Waymo was among the first to launch fully driverless taxi services in selected regions, demonstrating the capability of AI to navigate urban environments without human input. By analysing sensor data from cameras, radar, and lidar, these vehicles detect objects, predict traffic patterns, and make split-second adjustments to route plans. Software 2.0 and vision AI has accelerated this development with the prospect of Tesla launching a robotaxi service in Austin as early as June.

By late 2024, Tesla's fleet had surpassed nine billion miles driven on Autopilot. Public records and NHTSA data suggest that only a handful of fatal incidents—on the order of five to ten known cases globally up to 2023—occurred while Autopilot or FSD was engaged. Even assuming ten fatal crashes in nine billion Autopilot miles, that equates to roughly one fatal crash per 900 million miles, which is markedly better than the US average (approximately one per 100 million miles, or nine to ten per 900 million).

The technology is finally here.

[Putting FSD Safety to the Test | Tesla](#)



In sum, this new wave of AI exemplifies a profound shift from passive software tools to autonomous systems operating both in the digital realm and the physical world. Whether through self-directed agents, human-like robots, or self-driving vehicles, AI is increasingly poised to alter everyday life, introduce new forms of collaboration, and redefine the boundaries of human and machine capabilities.

Key Risks

Past performance does not predict future returns. You may get back less than you originally invested. We recommend any fund is held long term (minimum period of 5 years). We recommend that you hold funds as part of a diversified portfolio of investments.

The Funds managed by the Global Innovation Team:

May hold overseas investments that may carry a higher currency risk. They are valued by reference to their local currency which may move up or down when compared to the currency of a Fund.

May have a concentrated portfolio, i.e. hold a limited number of investments. If one of these investments falls in value this can have a greater impact on a Fund's value than if it held a larger number of investments.

May encounter liquidity constraints from time to time. The spread between the price you buy and sell shares will reflect the less liquid nature of the underlying holdings.

Outside of normal conditions, may hold higher levels of cash which may be deposited with several credit counterparties (e.g. international banks). A credit risk arises should one or more of these counterparties be unable to return the deposited cash.

May be exposed to Counterparty Risk: any derivative contract, including FX hedging, may be at risk if the counterparty fails.

Do not guarantee a level of income.

The issue of units/shares in Liontrust Funds may be subject to an initial charge, which will have an impact on the realisable value of the investment, particularly in the short term. Investments should always be considered as long term.

The risks detailed above are reflective of the full range of Funds managed by the Global Innovation Team and not all of the risks listed are applicable to each individual Fund. For the risks associated with an individual Fund, please refer to its Key Investor Information Document (KIID)/PRIIP KID.

Disclaimer

Non-UK individuals: This document is issued by Liontrust Europe S.A., a Luxembourg public limited company (société anonyme) incorporated on 14 October 2019 and authorised by and regulated as an investment firm in Luxembourg by the Commission de Surveillance du Secteur Financier ("CSSF") having its registered office at 18, Val Sainte Croix, L-1370 Luxembourg, Grand Duchy of Luxembourg and registered with the Luxembourg trade and companies register under number B.238295.

UK individuals: This document is issued by Liontrust Investment Partners LLP (2 Savoy Court, London WC2R 0EZ), authorised and regulated in the UK by the Financial Conduct Authority (FRN 518552) to undertake regulated investment business. It should not be construed as advice for investment in any product or security mentioned, an offer to buy or sell units/shares of Funds mentioned, or a solicitation to purchase securities in any company or investment product. Examples of stocks are provided for general information only to demonstrate our investment philosophy. The investment being promoted is for units in a fund, not directly in the underlying assets.

The document contains information and analysis that is believed to be accurate at the time of publication, but is subject to change without notice. Whilst care has been taken in compiling the content of this document, no representation or warranty is given, whether express or implied, by Liontrust as to its accuracy or completeness, including for external sources (which may have been used) which have not been verified.

This is a marketing communication. Before making an investment, you should read the relevant Prospectus and the Key Investor Information Document (KIID) and/or PRIIP/KID, which provide full product details including investment charges and risks. These documents can be obtained, free of charge, from www.liontrust.co.uk or direct from Liontrust. If you are not a professional investor please consult a regulated financial adviser regarding the suitability of such an investment for you and your personal circumstances. All use of company logos, images or trademarks in this document are for reference purposes only.