ORIGINALLY PUBLISHED JULY 8, 2026

AVITY Investment Management Inc.

JULY 8, 2026

INSIGHTS

GenAI: The Path Forward
Part 4 – Software's Reckoning

EXECUTIVE SUMMARY

The objective of this white paper series is to contemplate the path forward for artificial intelligence, including the implications for economies, markets and businesses. Part 1 reviewed scaling laws, a critical factor that drives AI capital expenditure. Part 2 examined the return on investment ("ROAI") from AI, including where productivity benefits are likely to emerge. Part 3, published a year ago, explored the potential for AI to be leveraged in the physical world in devices such as robotics and autonomous vehicles. This white paper, Part 4, discusses the implications of AI, positive and negative, for incumbent businesses, particularly those in the digital realm who are most exposed.

CREATIVE DESTRUCTION VS. SUSTAINING INNOVATION

In 1942, the economist Joseph Schumpeter described capitalism as a process of "creative destruction", the "perennial gale" by which new technologies, products and business models render existing ones obsolete. Railroads displaced canals, electricity superseded steam and the internet consumed the catalog. Schumpeter's thesis is relevant today as any business model oriented primarily around bits (software), rather than atoms, is vulnerable to disruption from AI. There is potential for AI to reshape industries across every sector, but software companies are among the first to confront this transition. The investment narrative surrounding software, not long ago viewed as a primary beneficiary of AI, has inverted over recent months as agentic AI threatens to dismantle the competitive position of incumbents. An opposing, and currently out of favor, view holds that AI is more akin to a "sustaining innovation" as defined by Clayton Christensen. The influential Harvard professor's theory draws a distinction between disruptive and sustaining innovation, the latter being improvements to existing products along the dimensions that incumbents already compete on. Sustaining innovations allow incumbents to leverage their capital, data, installed base and distribution channels to defend their competitive position. The issue being debated in markets today is whether agentic AI is a destructive force that will upend the traditional software model, or a sustaining innovation that could be successfully adopted by incumbents.

HISTORY OF SOFTWARE

Software businesses maintain attractive characteristics that have commanded premium equity valuations for most of the last three decades. These companies benefit from high gross margins of typically 70% or more ("write code once, sell many times"), high customer switching costs (deeply embedded, mission critical applications create sticky customer relationships), secular growth (as the world economy has become increasingly digital), and high returns on capital (due to their asset-light nature). Examples of dominant software businesses in the 1990s included Microsoft (Windows and Office), Oracle (database and enterprise resource planning) and SAP (enterprise resource planning). At the turn of the century, a former Oracle executive named Marc Benioff pioneered a customer relationship management (CRM) software-as-a-service (SaaS) solution that was entirely cloud-based. That company, Salesforce, ushered in a new era of cloud-hosted application software. Amazon Web Services was launched in 2006, beginning as a cloud infrastructure platform that allowed any customer to access storage and compute infrastructure through the internet. The ability for storage, compute and applications to be accessed through a browser lowered the barriers to entry for a host of new entrants and forced incumbent enterprise software companies to rearchitect their business models around the cloud. This period of transition weighed on the valuation multiples of legacy software companies as investors questioned their ability to successfully migrate to the cloud while keeping their financial profile intact. Transitioning to the cloud required incumbents to absorb cloud-hosting costs and forego significant up-front license revenue in exchange for smaller, but more predictable, subscription payments.

Several companies stumbled (e.g., Siebel Systems, Sybase, CA Technologies) and were eventually acquired. Those that succeeded (e.g., Microsoft, Adobe) were ultimately rewarded with generous valuation multiples as investors prized the highly visible recurring revenue and attractive margin profile that always defined the industry. Meanwhile, several cloud native SaaS companies (e.g., Salesforce, ServiceNow, Workday) grew into formidable multi-billion-dollar businesses.

Over the course of the 2010s, the SaaS model came to be viewed as among the most attractive of any industry. A distinct set of heuristics emerged to underwrite the high multiples these businesses garnered: annual recurring revenue, net revenue retention, "Rule of 40", "land and expand", lifetime value-to-customer acquisition cost (LTV/CAC). Capital flooded in to meet the opportunity with venture investment concentrated heavily in software during this time. A prolonged period of historically low interest rates magnified the present value of software's long duration, recurring cash flows. The high valuations in the industry made these stocks vulnerable during the inflation surge of 2022, but the group staged a forceful recovery in early 2023, driven in part by the promise of a new technology to lift their growth rates even further.

ChatGPT, created by OpenAI, launched in November 2022 and served as the catalyst for an unprecedented wave of capital investment that would follow in the years to come. GenAI was initially viewed as a promising tailwind for software companies. It was expected that traditional software incumbents could embed the technology into their applications and charge a higher price for the additional functionality. This was the consensus view through 2024, but the investment narrative reversed drastically in 2025. As the capability of frontier models and agentic tools improved exponentially, investors began questioning whether AI would displace, rather than enhance, the role of traditional software applications.

SAASPOCALPYSE

The terminal value for any business operating purely in the digital economy is under scrutiny, pressuring valuation multiples. The threat to incumbent software companies, specifically, is multifold. Among the more prominent debates are the relevance of seat-based pricing, the potential for software customers to build more tools internally, the durability of traditional software margins, and the threat from new entrants.

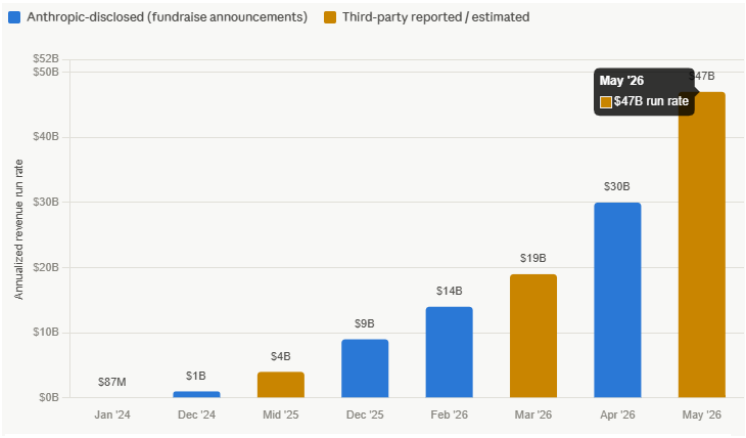
Seat-based pricing. The revenue model of most software companies today is based on the number of users with a subscription. If AI agents replace human users, this unit of monetization disappears. This fear has been fueled by recent layoffs at high profile technology companies, such as Amazon, Oracle, Microsoft, Meta, Cisco and Pinterest. Even if headcounts hold steady, as agents increasingly perform the underlying work, such as querying the database or triaging the ticket, the graphical user interface, which exists to serve a human in the loop, may recede into the background and cease to be the locus where value is created, weakening the logic of per-user pricing.

Build-versus-Buy. There is potential for more software customers to bring more development in-house. Coding agents, such as Claude Code, OpenAI Codex, GitHub Copilot and Cursor, have made it easier than ever for professional and amateur engineers to "vibe code" their own applications. Retool's 2026 Build vs. Buy report surveyed 817 enterprise builders and found 35% of teams have already replaced at least one SaaS tool with a custom application, and 78% expect to build more internal tools in 2026. Klarna, a Swedish fintech company, made headlines in late 2024 after announcing the replacement of certain third-party SaaS vendors with internally developed applications. Internal tool development is not confined to technology companies. Kirkland & Ellis, one of the world's largest law firms, plans to invest $500 million to develop internal AI tools. JPMorgan has allocated $2 billion this year to AI-related investments and now has approximately 150,000 employees using in-house AI tools weekly. Whether such cases prove to be a handful of well-publicized anecdotes or the leading edge of a structural shift is yet to be determined.

Margins. Software margins, historically the envy of the business world, have the potential to deteriorate as AI is embedded into legacy applications. Traditional software carried a marginal cost approaching zero, but every agentic query consumes compute which has an associated inference expense. This introduces a new variable cost to what was once pure incremental margin. Gavin Baker, of Atreides Management, argued on an episode of the Invest Like The Best podcast that software companies "have their 70, 80, 90% gross margins and they are reluctant to accept AI gross margins." Baker continued that

in software, “you write it once and it's written very efficiently, and then you can distribute it broadly at very low cost. And that's why it was a great business. AI is the exact opposite where you have to recompute the answer every time.” Indeed, Microsoft and Adobe serve as examples of two high profile businesses who recently ascribed software margin pressure to compute costs. Even if incumbents can defend their revenue streams, investors worry that the financial profile of these businesses may look very different in the future.

New Entrants. Perhaps the most significant threat comes from AI native new entrants, including from the model layer itself. Not only has vibe-coding made it easier for startups to build software, but frontier labs, such as OpenAI, Anthropic and Google, are pushing beyond raw model APIs and into the application layer that is currently occupied by software incumbents. The majority of AI application revenue has so far been concentrated at the foundational model layer. For example, Anthropic's run rate revenue has surged fivefold this year, despite being capacity constrained, to reach nearly $50 billion, exceeding the $42 billion of annual revenue that took Salesforce 27 years to achieve. The rapid growth of Anthropic and OpenAI supports the thesis that frontier labs are capturing the bulk of the new opportunity.



Source: Anthropic

One mechanism through which foundation models are extending their reach toward end users is termed the model “harness”. The harness consists of a scaffolding of tools, memory, context and orchestration that converts the raw output of a model into a more reliable workflow. The aggregation theory framework, popularized by technology analyst Ben Thompson, holds that profits accrue to whoever in the value chain controls the point of integration. In *Agents Over Bubbles*, Thompson explains that “model performance isn’t the only thing that matters: the integration between model and harness is where true agent differentiation is found. This is a very big deal when it comes to figuring out the future structure of the AI industry and where profits will flow, because profits flow away from modular parts of the value chain, which are commoditized, and flow towards integrated parts of the value chain, which are differentiated.” That logic implies a strategic imperative for AI labs to integrate forward, absorbing some of the functionality that was historically handled by traditional software. Should foundation models eventually commoditize, an open debate, the durable margin will lie in owning the surface area where intelligence becomes action.

THE SUSTAINING CASE

The threats to software are legitimate, but it is important to weigh the rebuttals put forth by incumbents. Discerning how disruption may unfold, which businesses it will reach, and over what timeline will allow investors to capitalize on the reorientation of value chains that AI has set in motion.

Seat-based pricing. To date, evidence of seat erosion is thin and the industry is quickly pivoting towards hybrid billing models. Nearly 60% of Microsoft’s customers are already purchasing usage-based credits on a “seats plus consumption” model. ServiceNow disclosed that half of its net-new annual contract value in its most recent quarter has shifted to non-seat-based pricing, even as active users continued to grow 25% year-over-year. Salesforce has introduced an enterprise license agreement that delivers consumption with the predictability of a committed spend alongside fixed-price “super SKUs” with consumption built in. User-based pricing is likely to endure in some shape or form as many chief information officers are attracted to the predictability of seat-based subscriptions, compared to the variability of consumption-based pricing which can be hard to budget for.

When asked during an interview with Ben Thompson whether hybrid models were the future, Microsoft CEO Satya Nadella responded, "100%." Nadella admits to being nervous during the cloud transition of the 2010s but explained “it turned out we sold a lot more subscriptions because people who never bought servers from us were buying subscriptions. I think that’s what’s happening already with agents, I see that on GitHub, I see that on M365, I see that on security, because everyone is building these agent systems that are continuously ‘working’ and so what we built and thought of as the end-user compute is completely getting rebuilt.”

There is plenty of precedent for the success of usage-based models from the likes of Snowflake and Datadog. Whether consumption- or seat-based, all software is, in theory, purchased based on customer return on investment (ROI). The variability of consumption models may attract more scrutiny from buyers, but software companies that deliver genuine ROI should be able to adequately monetize.

Build versus buy. As impressive as coding tools have become, the do-it-yourself thesis fails a few key tests. There is evidence that the time to value and total cost of ownership of building internal applications is underestimated. The aggregate expense comprises more than an API call and engineers’ salary, and includes constant model updates and deprecations, prompt engineering, security and compliance validation, and integration with the hundreds of adjacent systems. At its recent financial analyst day, ServiceNow claimed that customers building their own LLM-based solution typically spend 5-10x more than using its platform. Shopify’s Harley Finkelstein made a similar point in March 2026: “no enterprise wants to vibe code to checkout” and risk a payment flow that fails when Shopfiy’s fully integrated platform costs just $39 a month.

> “It’s effectively impossible to recreate the Shopify stack you’re offered for less than you’re paying for it. Even if I had unlimited resources, an army of engineers, and a blank check from Anthropic, my ability to do that at a lower cost is almost impossible, due to the negotiating leverage Shopify brings to the table with all of their partnerships… It’s a long list of reasons why you’re not going to vibe-code your way to a Shopify stack.” Michael Morton, MoffettNathanson, June 18, 2026 (Stratechery)

In February 2026, NVIDIA’s Jensen Huang told Becky Quick on CNBC that “agents won’t replace the tools, but agents will use the tools… why rewrite the browser when the browser exists, just use it. Why rewrite excel when excel exists, just use it. All of these tools we use today, whether it’s Cadence or Synopsys or ServiceNow or SAP, these tools exist for a fundamentally good reason and agentic AIs will be intelligent software that uses these tools on our behalf and help us be more productive.”

> “Imagine one of these days when we have robots in our homes, it’s very unlikely the robot will come up with a new Cuisinart. The robot will just read the manual for the Cuisinart and use the Cuisinart. It’s more likely that instead of coming up with another way of doing microwaving, the robot will just use the microwave. These tools exist for good reason.” Jensen Huang, CEO, NVIDIA, February 26, 2026 (CNBC)

An additional challenge with building software tools internally is the significant security and governance gap. Gartner estimates that 40% of agentic AI projects will fail by 2027, not because the AI is incapable but because it lacks governance. Cloudflare’s management describes the challenge as “shadow AI”, unsanctioned agents running in business units with no oversight. Recent high-profile incidents serve as examples of the security and governance risk that AI poses. In March 2026, an AI agent at Meta exposed sensitive data with no external hacker. In a separate incident, software company PocketOS experienced total data wipeout caused by a coding agent running on the Cursor platform.

Certainly, some third-party software will be easier to replicate than others. Furthermore, the continued improvement in LLMs may eventually overcome certain of the limitations cited above. However, for deeply embedded mission critical applications, there is a high hurdle for making the build-versus-buy equation favorable for customers.

Margins. Though there is early evidence of modest gross margin compression across the industry, some companies are confident in their margin outlook. Salesforce and ServiceNow expect to maintain its gross margin structure even as agentic

products scale, assisted by steep declines in inference cost per token. ServiceNow has framed AI inference as less than 10% of its cost to serve, with the balance residing in orchestration, governance and context. This demonstrates that software customers do not pay just for tokens, but for resolved outcomes spanning various modalities. To justify pricing in an AI world, it will be critical for application vendors to prove their worth as more than just conduits for token consumption.

Additionally, the very technology that pressures the cost-to-serve for software companies simultaneously allows them to be among the largest beneficiaries on the operating margin line. Empirical Research estimates that the software and services industry has the highest share of labor exposed to LLM displacement at nearly 65% of the category's workforce. With so much of their cost base comprised of engineering and customer service, two verticals that are very complementary with AI, software companies are already proving an ability to offset any gross margin pressure through meaningful operating efficiencies. Microsoft, ServiceNow, Salesforce, Workday, Cloudflare, SAP, Atlassian and HubSpot have all spoken publicly about AI-driven margin expansion largely due to slower headcount growth.

New Entrants. Competition is the most significant threat to incumbents. Even if switching costs allow traditional software companies to retain their existing customers, incremental growth may prove more difficult to achieve as AI natives intensify the competitive landscape. Our durability framework, described below, is an attempt to distill which business characteristics make certain incumbents relatively better positioned.

DURABILITY FRAMEWORK

Not all legacy software companies will share the same fate as AI diffuses across the industry. Orlando Bravo, founder of the software-focused private equity firm Thoma Bravo, recently drew a distinction between winners and losers in the software space: "there are many software companies in the public markets that will be disrupted from AI. Those companies were going to be disrupted anyway. AI will create a disruption a lot faster." There is a second category of software firms, Bravo argues, that are positioned to succeed. To be sure, legacy software vendors have some general factors working in their favor: significant resources for research and development, massive installed bases and brands that customers trust. Below, we attempt to identify the most important characteristics that may help incumbents navigate the AI transition.

Probabilistic versus deterministic workflows. Some tasks do not lend themselves well to the probabilistic nature of AI. LLMs generate responses based on statistical probabilities, which introduces margin for error. Robert Smith, founder of Vista Equity Partners, commented during a recent CNBC interview that consumer AI may tolerate 93% accuracy, but "that does not work in banking, it does not work in insurance." This intolerance for error is not confined to financial services. It is arguably even more acute in engineering domains. Sassine Ghazi, CEO of the electronic design automation (EDA) firm Synopsys, stated on a recent earnings call that the company's software delivers "optimal deterministic silicon-proven results that probabilistic AI models do not replicate." Simon Mays-Smith, VP of Investor Relations at Autodesk, the computer aided design platform, joked at a recent Baird conference "top tip, don't walk into a building that's been created by a probabilistic model, because it might fall down." Deterministic systems, in contrast to probabilistic models, can execute fixed rules with auditable precision. Software that encodes a guaranteed output, rather than a plausible one, will be harder for LLMs to displace.

Proprietary data. Nikesh Arora, CEO of Palo Alto Networks, said on the company's recent earnings call that "as frontier models become available to everyone, the real competitive advantage shifts from the model to the data fuel." The value of proprietary data has taken on new significance in the age of AI. Software applications whose core value proposition was to surface and organize publicly available data, something LLMs excel at, are most exposed to disruption as graphical user interfaces (GUIs) fade into the background. Many management teams argue that inference is likely to occur where the data already resides, in the systems of record, which act as repositories for information such as client communications, payroll, invoices and supply chain records. NVIDIA CEO Jensen Huang has made the case for systems of record: "we need the tools to finish their work and put the information back in a way that we can understand… Those system of records will still be ground truth… the agents will use it and populate the system of records." Salesforce points to 26 years of customer data. Palo Alto Networks ingests more than 17 petabytes of daily telemetry. ServiceNow's Context Engine draws on more than 95

billion annual workflows and 7 trillion transactions across 22 years. Workday cites 80 million users under contract and 1.4 trillion transactions annually and twenty years of embedded process knowledge that provide it with "data and context that no other competitor can replicate". Workday CEO Aneel Bhusri estimates it would take five to seven years for even the best AI to replicate Workday's data set. Unique data is a critical advantage but is not sufficient on its own to protect incumbents. Legacy systems of record must proactively build domain specific agents before their systems are abstracted away by third-party agentic execution layers. As one example, Workday used its data trove to build a model that, in the words of CEO Bhusri, is "the best context engine for agentic HR, finance and beyond."

Integration Complexity. Software that has intricate integrations across digital and physical ecosystems will prove harder for an LLM to displace. The deeper a vendor's software is embedded into hardware, regulated infrastructure, or a dense web of workflows, the higher the switching cost. Palo Alto Networks maintains an expansive footprint of 125 million sensors across networks, endpoints and clouds. Cadence Design Systems' software, used to design semiconductor chips, runs on its own hardware, is co-optimized with foundries to ensure compatibility, and is integrated with its own IP blocks that are embedded into customers' designs.

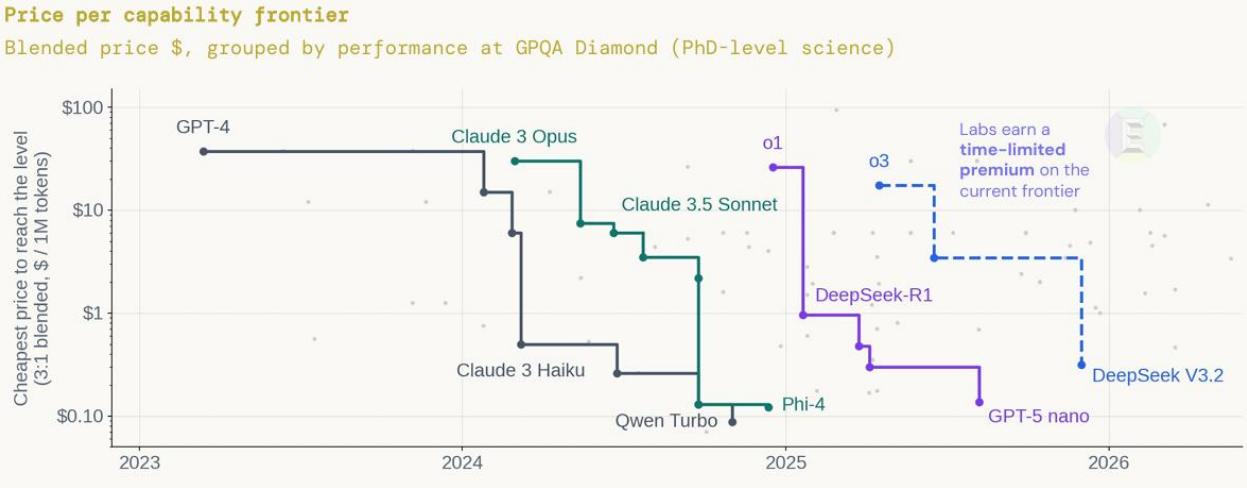
> "Any AI tools that we are developing or our customers are using basically in the end call our software to get the job done properly. We have seen absolutely no discussion with customers of reducing usage. On the contrary, all these AI tools are increasing the usage of our tools."– Anirudh Devgan, CEO, Cadence Design Systems, February 17, 2026

Shopify's software stack integrates merchant websites with payment rails, tax and compliance, and extends into inventory and shipping, entrenched further through the Universal Commerce Protocol, the open standard Shopify co-developed with Google. Toast extends the same logic into restaurants, with software, payment processing and purpose-built hardware integrated into one system. Samsara's fleet tracking software is integrated with vehicle gateways, asset tags and cameras. These are just a few examples of multifaceted software stacks that would be difficult for an LLM to replicate.

Infrastructure Software. The surface that AI threatens to disrupt the most is the user interface, but there is a large sub-category of software that sits underneath the application that will remain relevant. These include various databases, data warehouses, observability tools and cloud compute upon which software is built and run. These vendors already charge on consumption and are levered to volume of activity rather than headcount. AI inference is likely to anchor where large data sets reside, a concept known as "data gravity". Oracle, IBM, MongoDB and Snowflake are some examples of data platforms that allow customers to point agentic systems at their analytic estate while retaining control and governance over their data.

THE OPPORTUNITY: THE ORCHESTRATION LAYER

The cost of a fixed level of model capability has fallen precipitously in recent years as competition among labs and open-weight alternatives commoditizes raw intelligence. Notably, this decline in unit cost has coincided with rising aggregate spend, as users reach for higher-capability frontier models. If foundation models commoditize, and capabilities converge across labs as inference pricing falls toward the marginal cost of compute, then, by the logic of conservation of attractive profits, value should migrate upward to whoever controls the layer that converts raw intelligence into production outcomes.



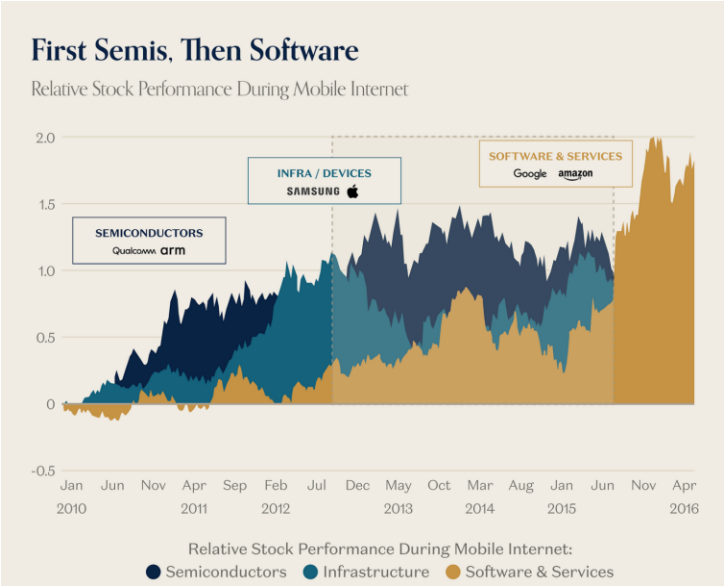
Source: Exponential View analysis: Epoch AI

Enterprises, unwilling to bind mission critical work to any single provider, already operate in a multi-model world. Microsoft reports more than 5,000 Foundry customers using at least one open-source model, and Databricks finds 76% of firms running open models alongside proprietary ones. Multi-model environments are likely to proliferate as enterprises pivot from “token maxxing”, using as many tokens as possible, to token optimization, directing prompts to the model that can most efficiently produce a good-enough response. This elevates the orchestration, or harness, layer responsible for routing, evaluation, governance, security and workflow context. As Evercore ISI frames it, model abundance "makes the control plane more valuable, not less," and the winning model "may be less important than the winning system." Should models commoditize, the durable differentiation would likely reside in context, workflow ownership and execution rather than in LLM capability. In a CNBC interview on July 1, 2026, Palantir CEO Alex Karp argued that deploying AI appropriately across an enterprise requires “the model plus the application plus compute, it is really all three.” Karp explains that "[enterprises] want control over their compute, their models, their data stack and their alpha. They want to know they own the means of production and it’s not being transferred to someone else.”

To be sure, frontier labs recognize this and are attempting to forward integrate. Simultaneously, the cloud service providers (CSPs) are extending upward from the infrastructure layer they already control (Amazon with its Bedrock and AgentCore harness, Microsoft with its Azure AI Foundry, and Google with Vertex), to assemble model-agnostic agent platforms that fold routing, guardrails, governance and evaluation into the cloud stack. Established application franchises are not without a claim to this valuable territory. Proprietary data, deep workflow ownership and entrenched distribution into the enterprise are assets not easily replicated.

THE CYCLE OF TECHNOLOGICAL REVOLUTIONS

History offers a useful lens on the sequencing of value capture over the course of a technological revolution. The early spoils tend to accrue to the hardware and infrastructure layer (the “picks and shovels”) and only later to the applications built atop it. The flow of funds in 2025 and 2026 has followed that script faithfully. Capital has rotated out of software and into the semiconductor, networking and power names that monetize the buildout directly. The cohort of "AI Plays" tracked by Empirical Research sourced roughly 75% of first quarter S&P 500 earnings growth, some 80% of capital spending growth and more than 80% of total returns. The dot-com cycle rhymed: the infrastructure suppliers of 1995–2000, like Cisco and Intel, led on the way up, while the application franchises, such as Amazon and Google, compounded through the 2002–2010 phase that followed the shakeout. The smartphone wave repeated the pattern, with the device-and-silicon era of 2007–2012 rewarding Apple, TSMC and Qualcomm before a mobile-native application layer, including Uber, Meta, Spotify, emerged to capture the next leg of growth. If the sequencing holds, then the AI opportunity for software lies largely ahead.



Source: Morgan Stanley Research, Factset, A16Z

DISCOUNTING DISRUPTION

The stock market is a discounting mechanism that reflects the present value of future cash flows a business is expected to generate. For most of the past three decades, software stocks commanded premium valuations because those future cash flows were judged to be highly visible. To date, the fundamentals of legacy software vendors remain sound. Management teams are pointing to retained customers, stable operating margins and revenue contribution from AI. Earnings revisions for the S&P 500 Software and Services Index are nicely positive year-to-date. And yet, software stocks remain out of favor as investors place increased scrutiny on the terminal value of these businesses, a rational response to diminished visibility, which is likely to persist until these companies can prove their durability in an agentic era. Positive near-term results cannot convincingly disprove the AI-disruption threat, which, if it comes, is likely to play out gradually. Even the franchises that successfully navigate the transition are likely to emerge with materially altered business models, such as consumption-based pricing and new variable expenses. The growing uncertainty has been reflected in stock performance, with the S&P 500 Software and Services Index underperforming the S&P 500 Index by nearly 40% over the last year. The group is trading approximately 30% below its five-year average price-to-earnings ratio.

Forward price-to-earnings multiples of the S&P 500 Software and Services Index
Source: Bloomberg

As is so often the case, depressed valuations are an ambiguous signal. Low equity multiples can be interpreted as either a buying opportunity or a structural business impairment (“cheap for a reason” or “value trap”) that is to be avoided. The qualitative assessment of this white paper suggests that the decline in valuations is largely justified due to reduced certainty and a dynamic competitive landscape. However, the rapidly evolving technology stack may present select opportunities to invest in businesses who stand to benefit from a maturing AI ecosystem.

Special thanks and credit to Carter Vanneck, Zachary Rosenbaum, CT Foster and Thomas Roach for their dedicated research contributions which supported the publication of this white paper.