

Building on AI's Unfinished Foundation

Generative AI is moving fast, but the surrounding architecture needed for economywide transformation hasn't settled. Companies that understand the difference will know where to invest and where to hold back.

By Kevin J. Boudreau

BY THE ORDINARY MEASURES OF ANY new technology, the current wave of generative AI has moved fast. By some estimates, about 2.4 billion people worldwide use generative AI platforms each month, and coding agents have changed how software is written. Efforts to commercialize the technology

have scaled just as fast. AI coding platform Cursor reportedly passed a \$2 billion revenue run rate by early 2026, and, as of April 2026, Perplexity was reported to have more than 100 million monthly users across its products by challenging one of the internet's most entrenched markets: search. And this growth is not confined to AI-native companies. Salesforce's Agentforce has reached \$1.2 billion in annual recurring revenue, Harvey has spread across large law firms, and Shopify has made AI use a baseline expectation across its operations. By many conventional markers, these developments increasingly resemble the early stages of a platform ecosystem.

Yet AI's larger promise is not to become another successful technology platform. It is to become a true general-purpose technology — like electricity or the internal combustion engine — that reshapes organizations, industries, and, ultimately, the broader economy. Judged against that standard, progress remains shallow.¹ The process of complementary innovation, organizational integration, and economywide transformation expected of a general-purpose technology remains in its infancy.

The obvious culprits — immature models, ordinary adoption friction — are real, but they're not the constraint. Like earlier general-purpose technologies, AI will not become economically transformative simply because it is broadly applicable. It will realize that potential when a surrounding technological, industrial, and institutional architecture enables decentralized organizations to confidently build upon it — in other words, when the technology becomes platformed.

Here, I will explain what platforming entails (the technological, industrial, and institutional architectures a technology needs), why AI remains only partly platformed, and how organizations can innovate and invest effectively while that process is still unfolding.

Platforming a General-Purpose Technology

Scholars have long argued that general-purpose technologies are able to transform economies because they can be applied across many industries while stimulating suc-



cessive waves of complementary innovation — as was the case with electricity, the steam engine, and the internet.² By the same token, the potential of such technologies is unusually hard to realize. Broad transformation requires large numbers of independent organizations to make interdependent investments; redesign products, processes, and business models; develop new capabilities; and coordinate despite deep uncertainty about how the technology and its ecosystem will evolve. Therefore, the central challenge is creating the technological, industrial, and institutional conditions under which decentralized organizations can confidently build upon the technology — a process of platforming. Someone has to build that foundation: It is what makes decentralized downstream integration, complementary innovation, and co-invention possible at all.

Electrification illustrates this. Electricity was technologically proven and commercially viable by 1882, yet widespread electrification did not follow for nearly four decades. Technological architecture stabilized when the Niagara Falls hydroelectric power project (1895-1896) confirmed polyphase alternating current at commercial scale. Industrial architecture matured as a division of labor settled among utilities, equipment makers, and financiers, under the regulated utility model that took hold between 1898 and 1907. Institutional architecture followed, with the first comprehensive state public-utility commissions forming in 1907.

As these technological, industrial, and institutional architectures progressively aligned, organizations gained sufficient confidence to invest and experiment, and electrification accelerated. Platforming emerged through the combined efforts of inventors, manufacturers, utilities, financiers, standards bodies, and regulators. Other general-purpose technologies — notably, personal computing — illustrate alternative pathways: Platform leaders, such as Microsoft and Intel, more directly orchestrated these architectures to support ecosystem growth.³

A technology becomes platformed when a surrounding technological, industrial, and institutional architecture creates conditions stable enough for decentralized organizations to confidently build upon it. Platforming reduces uncertainty by stabilizing expectations about how the technology can be used and how it will evolve. It establishes clear lanes for complementary innovation — where to innovate, where to rely on others, and what can be treated as stable — and the governance, rules, and incentives that enable organizations to capture value from their investments while coordinating with others. Decentralized investment and experimentation can then scale from isolated successes into broad transformation through the alignment of the three architectures.

The Platforming of AI: Where Are We Now?

The platforming of AI remains incomplete, but recognizable technological, industrial, and institutional architectures are emerging. Understanding what has stabilized — and what has not — clarifies the opportunities and the frustrations of building on AI before it has been fully platformed. Let's take a look at the current state of AI.

AI's technological architecture is still emerging.

Today's dominant AI architecture rests on a relatively specific trajectory, especially among leading frontier developers: pretrained, predominantly language-based foundation models; specialized hardware; cloud-based training and inference; and API-mediated delivery.⁴ AI is taking shape as a layered stack — chips, cloud infrastructure, foundation models, and the applications built on them. (See "Key Elements of the AI Stack.") The stack's lower three layers are converging on a centralized foundation in which model development and most computation reside with a few cloud-hosted frontier models, with most organizations consuming intelligence remotely through APIs rather than owning it. This departs from the digital services economics that were once taken for granted: Rather than distributing software that runs locally at little additional cost, AI delivers intelligence through continual, cloud-hosted inference, performing heavy computation each time intelligence is used.⁵ Although the prevailing architecture continues to evolve, the likely alternatives — open-weight ecosystems and parallel stacks developed by Chinese companies — are variations on it rather than fundamentally different trajectories. It's likely that to the extent it continues, much of the uncertainty around the lower layers will subside.

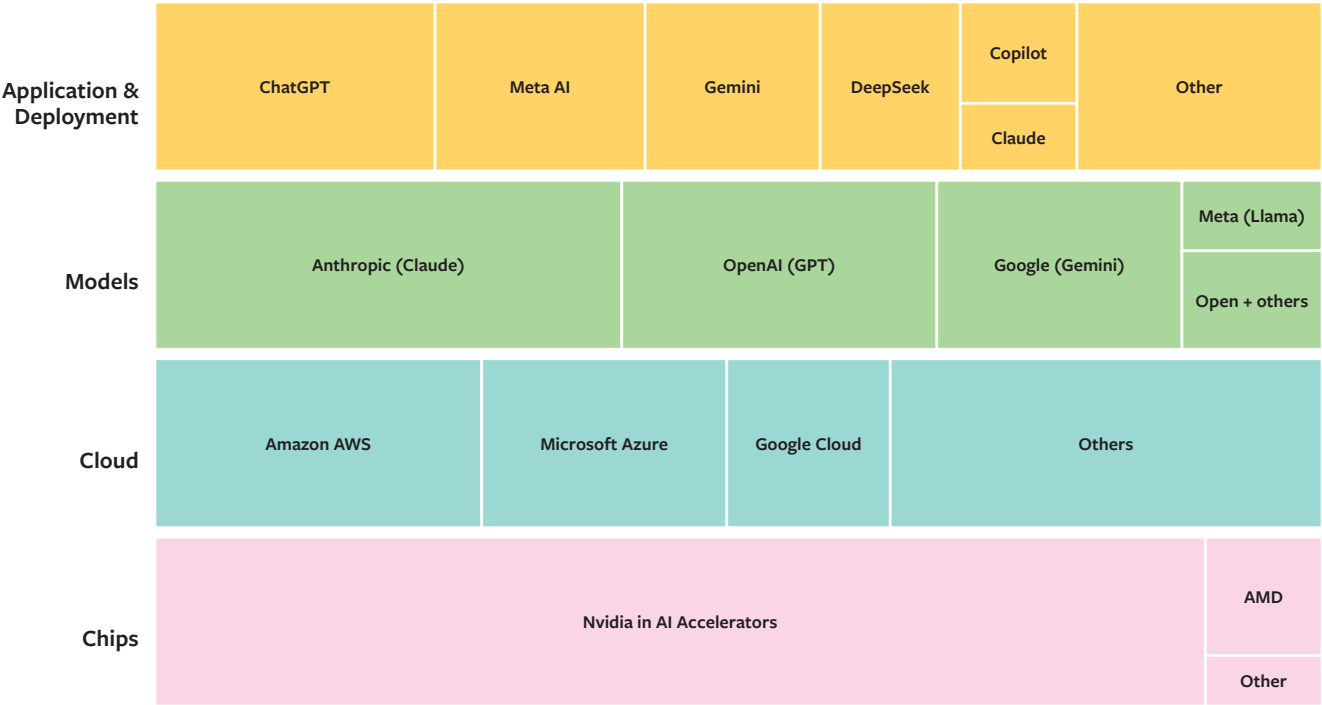
The application and deployment layer, where most organizations hope to build complementary products and services, remains fluid. As emerging orchestration, agent, and middleware layers compete to define how AI should be integrated into products, workflows, and enterprise systems, some companies build around chatbot interfaces and others directly on foundation model APIs. Meanwhile, frontier models keep absorbing capabilities that many people expected to reside elsewhere: Enterprise search, retrieval from knowledge bases, and persistent memory are increasingly being handled by the model rather than by separate software. It remains unclear which abstractions, interfaces, and patterns will become stable enough to support broad complementary innovation.

AI's industrial architecture is still cohering.

Technological architecture determines how intelligence is built — the trajectory and approach, and the division of the larger problem into components — whereas industrial architecture determines who builds what: how the ecosystem divides problem-solving and commercial activity

Key Elements of the AI Stack

The AI stack today has a settling lower layer — dominated by a handful of chip, cloud, and frontier-model providers — and a still-fluid upper layer where most organizations are trying to build.



Sources: Cloud infrastructure shares: Synergy Research Group (Q4 2025 data, published in February 2026); enterprise large language model spending: Menlo Ventures, 2025; The State of Generative AI in the Enterprise, December 2025, a Western-enterprise sample; GPU shares: Nvidia supplies the large majority of merchant AI accelerators; independent estimates of its revenue share range from roughly 75% to 88% depending on whether hyperscaler custom silicon is counted.

across specialized organizations, and how value creation and capture are distributed among them.⁶ Whether that division of labor can be occupied at all is another matter — one that is dependent on the capabilities organizations build, the skills the labor market supplies, how organizations align with one another and with the wider economy, and the returns that sustain them. Where these misalign, complementary innovation falls far short.

A recognizable division of labor has begun to emerge around the lower layers of the stack: Nvidia in AI accelerators; Amazon Web Services, Microsoft, and Google Cloud in compute; OpenAI, Anthropic, Google DeepMind, and xAI in frontier models; and Meta (Llama), Mistral, DeepSeek, and Alibaba (Qwen) in the open-weight ecosystem. Much of the ecosystem’s measurable investment is concentrated in these foundational layers — chips, compute, power, data centers, cloud, and frontier models.

Some businesses have clearly emerged as complementors on top of the foundation models. But AI has produced nothing like the governed marketplaces and stable inter-

faces earlier developers could build on — the app stores of iOS and Android, or the backward-compatible APIs of Windows — with the clear categories and rules that once gave thousands of them the certainty to commit. OpenAI’s GPT Store has stayed thin; meanwhile, a different candidate layer is forming one level up, around AI app builders such as Replit and Lovable, on which non-developers can generate and ship software. Whether a durable application economy consolidates there, among the model providers, or within incumbent suites is the contest still unresolved.

For now, AI reaches users through a heterogeneous mixture of forms: the frontier providers’ own applications, applications built on frontier models, agents, vertical applications, AI embedded within incumbent software, and proprietary enterprise deployments. Even what counts as the application and deployment layer is unsettled: thin chatbots, AI features embedded in existing tools, orchestration and middleware, wrapper apps, or third-party agents.

Vertical integration further blurs these boundaries as frontier developers move upward into user-facing products while incumbents embed foundation models throughout their suites. Rather than competing within established categories, companies are competing to define them — advancing not merely different products but competing hypotheses about how foundation model capabilities should be organized, accessed, and converted into value.

Meanwhile, returns on complementary investment remain uncertain, most fundamentally with regard to the industry's eventual division of labor, dominant application architecture, and sources of durable advantage. AI also runs into a familiar conundrum: The same foundation models that reduce the cost of innovation also reduce the cost of imitation, making differentiation harder.⁷

Investments in the lower layers also face uncertainty, but for different reasons. Frontier-model developers are pouring vast sums into future scale, ecosystem leadership, and pricing power — expectations that depend on an industrial structure that has not yet emerged. Switching costs remain modest, customers frequently use multiple models, and the mechanisms that historically produced durable platform leadership have yet to develop. Meanwhile, training and inference costs weigh on profitability, and capable open-weight models keep pressuring proprietary ones. Ultimately, these investments are a bet that AI will become platformed enough to generate value sufficient to justify today's spending.

AI's institutional architecture is emerging slowly. Institutional architecture governs how decentralized organizations coordinate, invest, and build on a common foundation. Successful platform ecosystems require more than technology and market forces; they require institutions and governance that let organizations invest independently while remaining collectively coordinated. AI's institutional architecture is emerging through standards bodies, consortia, technology providers, and governments developing protocols, but it remains far less developed than the technological and industrial architectures.

Institutional coordination can arise through several mechanisms. Governments can establish legal frameworks and public standards; industry alliances, standards bodies, and multiparty initiatives can coordinate interoperability and shared conventions. In AI, we might particularly expect the emergence of platform leadership, in which a central company organizes the ecosystem by establishing stable interfaces, governing participation, signaling architectural direction, committing to what it will not absorb, and creating credible incentives for complementary innovation. Such leadership is itself an invest-

ment: It takes a company with enough platform power to set and enforce the terms and enough incentives to bear the cost. Microsoft and Intel exemplify that in personal computing, Apple in the iPhone, and Google in Android.⁸ Today we can see that while many governments have taken a light-touch approach, industry consortia, evaluation frameworks, and private protocols are emerging.

A shared protocol, such as Anthropic's Model Context Protocol, is a useful standard, but platform leadership runs far deeper. It is the active coordination and orchestration of an ecosystem that extends well beyond the platform owner's own boundaries — governing participation, aligning incentives, signaling architectural direction, committing to what it will not absorb, and giving large numbers of independent companies the confidence to build. In that deeper work, today's frontier companies are investing comparatively little. The ecosystem lacks not leadership in technology but leadership in coordination.⁹ Someone has always supplied that coordination — such as a lead company in personal computing, and public authorities and engineering bodies in electricity. In AI, organizations are integrating vertically instead, which is not the same thing.

In the absence of mature platform leadership, many leading companies are instead solving coordination problems through vertical integration, combining frontier models, cloud infrastructure, developer tools, enterprise software, consumer applications, and distribution within integrated ecosystems. But direct control by one company is not ecosystem governance: It forgoes the diversity and decentralized effort of large numbers of independent complementors — the engine of broad transformation — and so may postpone the mature platform ecosystem that it appears to be substituting for.

Building and Innovating Before AI Is Fully Platformed

Most companies will not compete by building frontier models. They will compete by building on them — integrating them into products and operations, and creating the complementary goods and services around them. That places them, awkwardly, where the architecture is least settled: Although the lower layers are converging, the application and integration layer, where most of this building happens, is the part still in flux.

It can seem natural to wait for the costs, risks, and uncertainty of building on AI to fall as the technology becomes platformed. But the real challenge is to keep them in mind and act anyway — for a company to invest in the ways that best strengthen its position while managing the problems of building on a general-purpose technology that has not yet been fully platformed. That

means making investments that will pay off however the architecture settles.

Learn faster than you commit. When the architecture is unsettled, what a company learns is worth more than what it locks in — and the cost of learning is unusually low right now. Much of the experimentation can run on open-weight models — such as Llama, Mistral, or DeepSeek — on a company’s own hardware, where the marginal cost of a query is near zero and proprietary data never leaves the building. Frontier models can be reserved for the work that genuinely needs them. The most valuable thing a company can record — where its experts overrode the model and why — is also at its richest now, while the models are still making enough mistakes to generate corrections. Those corrections also map the jagged frontier of what AI does reliably on a company’s own tasks. But that margin is closing: Once a model reliably beats a company’s experts on a task, the corrections, and the signals they provide, disappear.

Build assets that will survive architectural change. No one yet knows which model, stack, or way of organizing intelligence will win, so businesses should be wary of making bets specific to any one of them. The danger is subtler than it looks. Most companies ask only the technical question “If the model were swapped tomorrow, would the system still run?” and stop there. But even a perfectly swappable model can be locked in commercially, through the way a vendor bills and what its contract permits. Salesforce’s Agentforce meters agent work in “agentic work units” it alone defines, and SAP’s 2026 policy restricts what outside AI is allowed to do with the data inside its software.¹⁰ A company should keep its options open at the applications level, too. Holding back on bets wired to an architecture that is still unsettled is a strategy in itself — not a failure to act.

Invest in complements, not intelligence. Nearly every company will buy rather than build its AI, as will its competitors — who will often purchase the very same models. Intelligence that everyone can rent cannot be anyone’s advantage: The same models lift the floor for a business and its rivals alike, and the work they do well converges toward a common mean.¹¹ Advantage has to come from what the shared model cannot reach, such as proprietary data, domain expertise, trusted customer relationships, distribution, brands, and specialized workflows. This is not a new idea: Rents accrue to the holders of co-specialized complements, not to the freely available input itself.¹²

Where to invest turns on the seam — the interface between layers where value can be captured. A defensible seam rests on something that the layers above cannot easily reproduce: a regulator’s standing trust, an audit trail,

an embedded billing relationship, a proprietary data corpus. This is why many hospitals’ clinical AI runs through Epic: A smarter interface still has to clear the trust bar, which Epic has already done. An illusory seam is a thin wrapper around someone else’s model — useful this quarter but enveloped the next, when the provider folds the same capability into its own product at no extra charge. Before investing in a seam, an organization should ask who will try to take it — the model provider reaching up or the platform incumbent reaching down — and whether what anchors it can be reproduced.

The logic holds even in the case that seems to overturn it: a model capable enough to do the integration, the judgment, and the work itself, leaving little apparent need for a wider ecosystem. Even then, the assets that endure are the ones a model cannot internalize, such as the regulator’s trust, the audit trail, the embedded contract, and the proprietary data. Betting on intelligence pays off only for whoever wins the frontier race; betting on complements pays off regardless of who wins.

The question, then, is not how to own the AI. It is how to own the assets that become more valuable as AI becomes abundant.

Build organizational capability. The most valuable use of AI in this period is the least obvious one: not to do today’s work faster but to become an organization that understands its own workings — where its knowledge sits, how its decisions get made, where its bottlenecks are, and how work moves across its teams. These are the assets the earlier moves depend on. Vendors already have a name for this work — the “forward-deployed engineer” they send to sit inside a customer — and adopters will need the role in-house. The capability a company builds now, such as the memory, the routines, and the judgment about where AI helps and where it does not, is itself among the complements a shared model cannot reach, and it is what keeps producing new ones as the architecture shifts. Learning feeds capability; capability yields the durable complements.

The deep gains from a general-purpose technology always arrive late, and they require two things: The surrounding architecture must settle, and organizations must rebuild themselves around the technology. A company cannot hurry the former. It can begin the latter now, and history suggests that those that do are the ones that pull ahead. Manufacturers that learned to redesign factories around electric power saw gains long before electrification was universal; companies that learned to reorganize around information technology saw benefits long before personal computing matured. The companies that come out ahead in AI may look less like today’s adopters and more like knowledge factories — organizations whose

advantage is not throughput but the rate at which they turn their own operations into validated understanding.

Building before AI is platformed is necessarily about investing in what the technology cannot supply — learning, complements, and organizational capability that compound while the architecture is still in motion. It is not about adopting the most tools or predicting the final form first. ■

Kevin J. Boudreau is a professor of strategy, entrepreneurship, and innovation at Northeastern University's D'Amore-McKim School of Business, with appointments in the Khoury College of Computer Sciences and the College of Social Sciences and Humanities. He is a research associate in the Productivity, Innovation, and Entrepreneurship program at the National Bureau of Economic Research.

REFERENCES

1. On the speed of AI adoption — faster than it was for the personal computer or the internet at a comparable stage — see A. Bick, A. Blandin, and D.J. Deming, “The Rapid Adoption of Generative AI,” *Management Science*, Articles in Advance, published online Jan. 20, 2026, <https://doi.org/10.1287/mnsc.2025.02523>. On its still-limited and uneven use in production, see the U.S. Census Bureau's Business Trends and Outlook Survey for 2023-2026; and on pre-generative AI technologies, see K. McElheran, J.F. Li, E. Brynjolfsson, et al., “AI Adoption in America: Who, What, and Where,” *Journal of Economics and Management Strategy* 33, no. 2 (summer 2024): 375-415, <https://doi.org/10.1111/jems.12576>. On the jagged, task-by-task pattern of where AI is reliable, see F. Dell'Acqua, E. McFowland III, E. Mollick, et al., “Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality,” working paper 24-013, Harvard Business School, September 2023. On still-modest measured productivity and why the deep gains of a general-purpose technology arrive only with a lag, compare these three articles: D. Acemoglu, “The Simple Macroeconomics of AI,” *Economic Policy* 40, no. 121 (January 2025): 13-58, <https://doi.org/10.1093/epolic/eiae055>; E. Brynjolfsson, D. Rock, and C. Syverson, “The Productivity J-Curve: How Intangibles Complement General Purpose Technologies,” *American Economic Journal: Macroeconomics* 13, no. 1 (January 2021): 333-372, <https://doi.org/10.1257/mac.20180386>; and P.A. David, “The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox,” *The American Economic Review* 80, no. 2 (May 1990): 355-361. On the electrification sequence, see also T.P. Hughes, “*Networks of Power*” (Johns Hopkins University Press, 1983).
2. T.F. Bresnahan and M. Trajtenberg, “General Purpose Technologies ‘Engines of Growth?’” *Journal of Econometrics* 65, no. 1 (January 1995): 83-108, [https://doi.org/10.1016/0304-4076\(94\)01598-T](https://doi.org/10.1016/0304-4076(94)01598-T); and R.G. Lipsey, K.I. Carlaw, and C.T. Bekar, “Economic Transformations: General Purpose Technologies and Long-Term Economic Growth” (Oxford University Press, 2006). On platforming as the building of technological, industrial, and institutional infrastructure, see K.J. Boudreau, L.B. Jeppesen, and M. Miric, “Artificial Intelligence as a Platform Technology: Strategic Implications of Competing on Top of an AI Platform,” chap. 4 in “*Handbook of Artificial Intelligence and Strategy*,” eds. F.A. Csaszar and N. Jia (Edward Elgar, 2026).
3. On platform leadership and ecosystem coordination, see A. Gawer and M.A. Cusumano, “How Companies Become Platform Leaders,” *MIT Sloan Management Review* 49, no. 2 (winter 2008): 28-35.
4. On the decomposition of a complex system into nearly independent components and the interfaces that join them, see H.A. Simon, “The Architecture of Complexity,” *Proceedings of the American Philosophical Society* 106, no. 6 (Dec. 12, 1962): 467-482; for platforms specifically, see C.Y. Baldwin and C.J. Woodard, “The

Architecture of Platforms: A Unified View,” in “*Platforms, Markets and Innovation*,” ed. A. Gawer (Edward Elgar, 2009), 19-44; and, for an integrative framework, see A. Gawer, “Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework,” *Research Policy* 43, no. 7 (September 2014): 1239-1249, <https://doi.org/10.1016/j.respol.2014.03.006>.

5. On the economics of information goods — high fixed costs and near-zero marginal costs — see C. Shapiro and H.R. Varian, “*Information Rules: A Strategic Guide to the Network Economy*” (Harvard Business School Press, 1998).
6. This is consistent with the industry architecture of the sectorwide template governing the division of labor and where value accrues. See M.G. Jacobides, T. Knudsen, and M. Augier, “Benefiting From Innovation: Value Creation, Value Appropriation and the Role of Industry Architectures,” *Research Policy* 35, no. 8 (October 2006): 1200-1221, <https://doi.org/10.1016/j.respol.2006.09.005>. The emphasis here falls on the correspondence between the problem space set by the technological architecture and the capabilities, skills, and organization required to work on it: the industry-level counterpart of the firm-level clustering of technology, workplace organization, and skill documented in T. Bresnahan, E. Brynjolfsson, and L.M. Hitt, “Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence,” *Quarterly Journal of Economics* 117, no. 1 (February 2002): 339-376, <https://doi.org/10.1162/003355302753399526>. On the returns that sustain such investment, see D.J. Teece, “Profiting From Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy,” *Research Policy* 15, no. 6 (December 1986): 285-305, [https://doi.org/10.1016/0048-7333\(86\)90027-2](https://doi.org/10.1016/0048-7333(86)90027-2).
7. K. Boudreau, L.B. Jeppesen, and M. Miric, “The Paradox of Platform-Based Entrepreneurship: Competing While Sharing Resources,” *SSRN*, Aug. 4, 2021, <http://dx.doi.org/10.2139/ssrn.3897435>; and Boudreau, Jeppesen, and Miric, “Artificial Intelligence as a Platform Technology.”
8. On platform leadership, see M.A. Cusumano and A. Gawer, “The Elements of Platform Leadership,” *MIT Sloan Management Review* 43, no. 3 (spring 2002): 51-58; and K. Boudreau and A. Hagiu, “Platform Rules: Multi-Sided Platforms as Regulators,” in “*Platforms, Markets and Innovation*,” ed. A. Gawer (Edward Elgar, 2009), 163-191. The latter treats a platform's authority to write and enforce the rules of participation as the basis for its incentive to invest in goods the ecosystem shares.
9. On platforms as bases that others build on, supplied by a lead vendor responsible for the architecture, and on co-invention by adopters, see T.F. Bresnahan and S. Greenstein, “Technical Progress and Co-invention in Computing and in the Uses of Computers,” *Brookings Papers on Economic Activity: Microeconomics* (1996): 1-83, <https://doi.org/10.2307/2534746>. A lead vendor is one mode of provision; a public authority and a private collective are others.
10. Salesforce's Agentforce run rate and “agentic work unit” were sourced from its quarterly earnings releases for 2025-2026; SAP's third-party AI API restrictions were sourced from its published terms in “SAP API Policy v.4.2026a,” PDF file, published on the SAP Help Portal, late April 2026, <https://help.sap.com>.
11. On AI raising the mean while compressing the distribution among individual workers in a large customer-support setting, see E. Brynjolfsson, D. Li, and L. Raymond, “Generative AI at Work,” *Quarterly Journal of Economics* 140, no. 2 (May 2025): 889-942, <https://doi.org/10.1093/qje/qjae044>.
12. The complementary-assets logic follows Teece, “Profiting From Technological Innovation,” 285-305.

Reprint 68129. For ordering information, see page 6. Copyright © Massachusetts Institute of Technology, 2026. All rights reserved.