Strategy Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Can AI Do Strategy? A Dialogue and Debate

Aaron Chatterji, Felipe A. Csaszar, James Evans, Teppo Felin, Jessica Hullman, Karim R. Lakhani, Mari Sako, Todd Zenger

To cite this article:
Aaron Chatterji, Felipe A. Csaszar, James Evans, Teppo Felin, Jessica Hullman, Karim R. Lakhani, Mari Sako, Todd Zenger (2026) Can AI Do Strategy? A Dialogue and Debate. Strategy Science 11(1):16-30. https://doi.org/10.1287/stsc.2026.ed.v11.n1

https://pubsonline.informs.org/journal/stsc

STRATEGY SCIENCE
Vol. 11, No. 1, March 2026, pp. 16–30
ISSN 2333-2050 (print), ISSN 2333-2077 (online)

Can AI Do Strategy? A Dialogue and Debate

Aaron Chatterji,[a] Felipe A. Csaszar,[b] James Evans,[c,d] Teppo Felin,[e] Jessica Hullman,[f] Karim R. Lakhani,[g] Mari Sako,[h] Todd Zenger[e,*]

[a] Duke University, Durham, North Carolina 27708; [b] Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109; [c] University of Chicago, Chicago, Illinois 60637; [d] Santa Fe Institute, Santa Fe, New Mexico 87501; [e] University of Utah, Salt Lake City, Utah 84112; [f] Northwestern University, Evanston, Illinois 60208; [g] Harvard Business School, Boston, Massachusetts 02163; [h] University of Oxford, Oxford OX1 2JD, United Kingdom

*Corresponding author

Contact: https://orcid.org/0000-0003-1678-9352 (AC); https://orcid.org/0000-0002-1255-9368 (FAC); https://orcid.org/0000-0001-9838-0707 (JE); https://orcid.org/0000-0003-2044-0145 (TF); https://orcid.org/0000-0001-6826-3550 (JH); https://orcid.org/0000-0002-0193-007X (KRL); https://orcid.org/0000-0003-3465-0047 (MS); todd.zenger@eccles.utah.edu, https://orcid.org/0000-0002-9830-4066 (TZ)

Published Online in Articles in Advance: March 19, 2026

History: Accepted for the Special Issue: Can AI Do Strategy?

https://doi.org/10.1287/stsc.2026.ed.v11.n1

An Introduction

Todd Zenger

In August of 2025, the ION Management Science Laboratory at Utah, in collaboration with *Strategy Science*, convened a conference addressing the provocative question: *Can AI do strategy?* As part of the conference, we held a two-part panel that was organized as a moderated dialogue and debate among scholars from divergent backgrounds in strategy, sociology, computer science, and the AI industry. Panelists included Aaron Chatterji, Felipe Csaszar, James Evans, Teppo Felin, Jessica Hullman, Karim Lakhani, and Mari Sako. I served as moderator for the dialogue, guiding a structured exchange intended not to produce consensus but to surface foundational disagreements about the nature of AI's role in strategic decision making and the implications of recent advances in AI.

Following the conference, each panel participant was invited to return to the transcript of their remarks and develop a short essay that sharpened and formalized their core position. The resulting contributions articulate a set of distinct and often conflicting claims. They appear in the order that they were presented at the conference.

Some argue that AI can perform the core elements of strategy by expanding representation, search, and aggregation—particularly if strategy can be rendered more verifiable and amenable to learning. Others contend that strategy is inherently forward looking and causal, grounded in problem formulation, judgment, and actor-specific theorizing that current AI systems cannot replicate. Additional essays emphasize hybrid human–AI arrangements, portraying AI as a powerful complement that reshapes strategic workflows, ideation, and simulation rather than fully automating strategic choice. Still others reframe the debate by highlighting the organizational and institutional dimensions of strategy, emphasizing the role of professional judgment, internal conflict, disagreement, and the dangers of compressing strategic deliberation into singular, authoritative recommendations.

Taken together, the essays do not offer a unified answer to whether AI can do strategy. Instead, they reveal deep disagreements about what strategy *is*—whether it is best understood as a computational problem, a causal-theoretic activity, a professional practice, or an organizational process sustained by conflict and deliberation. By anchoring these contributions in a common dialogue and presenting them side by side, this introduction aims to clarify the conceptual stakes of the debate and to frame a research agenda for the field of strategy science at a moment when AI is rapidly becoming embedded in organizational decision-making processes.

Can AI Do Strategy? Yes—If We Make Strategy Verifiable

Felipe A. Csaszar

The question "Can AI do strategy?" demands clarity about what we mean by strategy. At its core, strategy is the discipline of choosing present actions that create superior future value. This requires foresight—the ability to predict which courses of action will lead to better outcomes (Csaszar and Laureiro-Martínez 2018). My answer is yes, AI can do strategy, but not in the way many imagine. The path is not a CEO asking a large language model (LLM) for the firm's next move. Instead, LLMs will

function as components within a more sophisticated computational process. AI's potential will only be unlocked if we, as a field, commit to making strategy a more verifiable domain. The decisive question today is not whether AI can do strategy but whether we can make strategy verifiable enough for AI to learn it.

Strategy as Foresight Through Computational Processes

For decades, the study of strategy has been, implicitly, the study of how organizations achieve foresight despite the cognitive constraints of their human members (Csaszar and Rhee 2025). Building on the Carnegie tradition, we can understand this process through a cognitive architecture comprising three mechanisms: representation of the strategic problem, search for alternative courses of action, and aggregation of diverse judgments (Csaszar and Steinberger 2022).

Strategy has long grappled with the challenge of deciding under conditions of bounded rationality. AI, however, operates under a fundamentally different and less restrictive set of bounds. It offers a path to systematically relaxing the cognitive constraints that have shaped our theories and practices. AI transforms each of the three core mechanisms: it can enrich representation, enabling the use of models with a complexity far beyond what human cognition can manage; it can expand search at an unprecedented scale, exploring vast landscapes of alternatives; and it can augment aggregation, simulating the wisdom of crowds with virtual panels of experts and combining viewpoints without the social frictions that plague human teams. By transforming representation, search, and aggregation, AI provides a path toward what might be called "unbounding rationality"—the systematic relaxation of the cognitive limits that have long constrained strategic decision making (Csaszar 2025).

AI as System: The Architecture of Computational Strategy

The naive view of AI strategy involves a CEO asking an LLM, "What should I do?" This is a caricature. The more sophisticated and likely path involves designing computational systems where LLMs are foundational components, much as transistors are components in complex circuits. These systems will combine generation and evaluation of alternatives with large-scale search, access to proprietary data, and the design of experiments and realistic simulations to test strategic hypotheses.

This approach leverages sheer computational power to compensate for potential weaknesses in judgment, a pattern seen in other domains where AI has achieved superhuman performance. In games like chess and Go, AI systems combine a relatively rough evaluation function with a massive-scale search through billions of potential moves. An imperfect but fast evaluator, amplified by computational scale, can identify solutions far from an organization's current position. This leverages computational abundance: where human strategists must carefully ration their search efforts, AI systems can generate and evaluate millions of business model × market combinations, systematically surfacing high-potential winners hidden in vast possibility spaces.

We can also computationally instantiate the sophisticated deliberative processes our field has long studied. For example, one AI agent can propose a baseline plan, another can act as a devil's advocate, a third can simulate competitor reactions, and a fourth can model customer responses—all in an iterative loop of refinement. Early evidence already supports this potential; in a recent study (Csaszar et al. 2024), business plans generated by an LLM were, on average, rated more favorably by experienced investors than plans submitted by entrepreneurs to a leading startup accelerator.

The Verifiability Imperative: Building the Gradient

For any machine learning system to improve, it needs a clear performance signal—a gradient to climb (Sutton and Barto 2018). For an AI to master chess, the signal is unambiguous: winning or losing. But for an AI to learn strategy, what is the equivalent signal? This question exposes the primary bottleneck to progress: strategy has largely remained a nonverifiable domain, more akin to art than to science.

To unlock AI's potential, our field must commit to building the intellectual infrastructure for verifiability. This requires developing three distinct but complementary types of measurement tools. First are performance benchmarks: standardized, out-of-sample prediction tasks, such as forecasting stock price reactions to strategic announcements or predicting product launch outcomes. Second are comparative evaluations, where human experts rate AI-generated strategic plans against human-authored ones on multiple criteria such as novelty, feasibility, and expected value—with the explicit goal not of matching human performance but of exceeding it. Third are capability assessments: a battery of tests to measure the qualities we would want to know of any CEO put in charge of important decisions—domain knowledge, risk preferences, leadership orientation, and strategic reasoning abilities such as causal inference and counterfactual thinking.

Collectively, such benchmarks would provide the clear performance gradient needed for AI systems to improve. They would also allow for direct, apples-to-apples comparisons of different systems and processes. Benchmarks are to strategy what loss functions are to learning: without them, nothing improves. If we want AI to learn strategy, we must first ensure strategy is learnable (Heshmati and Csaszar 2024).

Addressing Two Fundamental Objections

This vision is not without its challengers. Two powerful objections warrant direct engagement.

The Homogenization Objection The first objection, articulated by Barney and colleagues (Barney and Reeves 2024, Wingate et al. 2025), argues that because AI will be a widely available tool, it cannot be a source of sustainable competitive advantage. If everyone has the same tool, it becomes a cost of doing business, not a source of differentiation. This logic, however, overlooks the path-dependent journey of technology adoption. An apt analogy is the internet of 1995. At that time, access was also becoming widespread, yet not all firms possessed the same foresight or capability to leverage it. A select few—such as Amazon and Google—built durable moats grounded in network effects, economies of scale, and proprietary data. Their advantage came not from using the internet but from understanding the potential of the technology better than others—having foresight—to build the novel business models the internet made possible.

AI will likely follow a similar pattern. Advantage will accrue to firms that move fastest to the new technological frontier, build AI-native decision processes, and create entry barriers. AI may be a common tool, but it will enable uncommon processes. The question is not who has AI but who can imagine what AI makes possible—and build it first.

The Creativity Objection The second objection, articulated by Felin and Holweg (2024), argues that AI, as a fundamentally backward-looking technology trained on historical data, cannot perform the forward-looking, creative, and causal reasoning tasks central to strategy. This is an important distinction. However, humans, too, learn only from the data of past experiences. The question is whether AI can be equipped with computational processes to generate novel, forward-looking insights. The history of AI suggests it can through at least three avenues: (1) computational brute force to explore vast possibility spaces and discover nonobvious combinations; (2) sophisticated process design, such as multiagent deliberation, to challenge assumptions; and (3) hybrid architectures that combine LLMs with reasoning and predictive models.

A key insight is that AI need not replicate human cognition to achieve superior outcomes. A dishwasher doesn't clean plates the way humans do; it doesn't need to. Airplanes don't flap their wings like birds. Similarly, AI will achieve strategic performance not by mimicking human intuition but by leveraging its own unique strengths: massive scale, computational speed, and the ability to systematically explore complex possibility spaces that human cognition cannot traverse.

The Human–AI Hybrid: Architecture and Vigilance

For the foreseeable future, the practical locus of strategic decision making will be hybrid systems, not autonomous AI. AI performance has a "jagged frontier" (Dell'Acqua et al. 2026): on tasks where AI is strong, human performance is amplified; where it is weak, naive use can be detrimental. The fluency of AI models can also induce overconfidence, creating a risk of humans sleepwalking through critical decisions.

This reality makes the cognitive division of labor a central design challenge. The goal is to architect systems with appropriate guardrails, deploying AI where outcomes are verifiable and feedback loops are short while retaining human oversight where stakes are high and feedback is weak or ambiguous. In the near future, at least, the question is not "Can AI do strategy?" but "Can humans with AI do better strategy than humans without AI?"

Implications and a Path Forward

The strategy field now faces an existential choice. We risk repeating the history of statistics, a field that largely ceded the intellectual high ground of machine learning to computer science. If we, as strategy scholars, collectively decide that AI is not a fundamental issue, the intellectual center of gravity for "AI strategy" will migrate elsewhere. If we insist strategy cannot be done with AI, we will ensure it is done without us.

To ensure the strategy field helps write this next chapter rather than being written out of it, we need a concrete plan of action. The vision of a verifiable, AI-augmented strategy process generates a set of testable propositions. Testing these propositions will simultaneously advance our understanding and build the envisioned systems:

P1 (Hybrid advantage): Human–AI systems will outperform humans on strategic tasks with verifiable outcomes.

P2 (Process differentiation): Firms building AI-native decision processes, leveraging organizational adaptation and proprietary data, will achieve durable advantages despite AI tool ubiquity.

P3 (Learning acceleration): Organizations that develop and use AI-powered foresight tools will reallocate resources more effectively under uncertainty and outperform competitors.

P4 (Deliberation premium): Multiagent AI systems that iteratively debate and filter alternatives will outperform single-shot prompting on many outcomes relevant to strategy, including novelty, quality, feasibility, and foresight.

P5 (Computational search): AI systems that combine generation, evaluation, and large-scale search will identify strategic alternatives that human teams, constrained by bounded rationality, systematically overlook.

Conclusion: Building the Gradient

AI can do strategy, but only if we architect systems that expand search, representation, and aggregation,

and—most importantly—if we make strategy a verifiable domain. In 1943, Alan Turing and Claude Shannon debated whether an "artificial scientist" or an "artificial CEO" would be harder to build (Hodges 1983, p. 251). The community best positioned to formalize the latter has not yet fully dedicated itself to the task. We should. Science, Donald Knuth reminded us, is knowledge we understand so well that we can teach it to a computer (Knuth 1974, p. 668). Strategy scholars now face three intertwined challenges: to build the gradient that lets intelligence—human and artificial—learn the future, to revisit everything we know about strategy process so that it can take advantage of AI's capabilities, and to refresh our understanding of strategic change to guide the massive organizational transformation that lies ahead. If we rise to meet them, we will not only advance our field—we will help shape how organizations create value in the century ahead.

Why AI Cannot Do Strategy

Teppo Felin

Artificial intelligence (AI) increasingly appears capable of performing tasks long considered exclusive to human judgment (for a review, see Felin and Holweg 2024). Given the rapid advances in AI—and the problem of human bounded rationality—some argue that we should "replace humans by algorithms whenever possible" (Kahneman 2018, p. 610). A number of scholars seem to suggest that AI might not only help but perhaps even supplant humans in strategy and decision making, given AI's ability to process vast data sets and make predictions (e.g., Csaszar et al. 2024, Doshi et al. 2025, Tranchero et al. 2025). The central argument is that decision making and strategy can be understood as a large-scale prediction problem (cf. Agrawal et al. 2022)—one that AI is uniquely equipped to solve.

I challenge this conclusion and argue that AI cannot do strategy. Specifically, AI cannot do strategy because it is, by construction, backward looking, population level, and based on correlational recombination of past patterns. Strategic decision making, on the other hand, is based on forward-looking causal reasoning that constructs novel elements and combinations, is idiosyncratic, actor-specific, and grounded in disagreement. Below, I highlight these points.

Backward-Looking Output vs. Forward-Looking Causal Reasoning

AI systems are backward-looking, whereas strategy requires forward-looking causal reasoning. Breakthroughs and value creation disrupt the past and are not based on linear extrapolation.

A useful thought experiment, developed with Matthias Holweg, makes this concrete. Imagine an LLM in 1633, trained with all the scientific text up until that point. This is right around the time that Galileo was defending his heliocentric theory. If we asked the LLM about Galileo and his theory, the LLM would say that heliocentrism is delusional and wrong (Felin and Holweg 2024). The LLM could fluently summarize, restate, and paraphrase existing, orthodox views—mirroring its training data. But an LLM has no way whatsoever of going beyond this, of somehow developing or evaluating new theories or sources of novelty. This applies to science as it does to strategy. Anything genuinely novel and new in a causal or strategic sense is out of distribution for AI systems trained on historical data. The logic of AI and machine learning models is backward looking, whereas strategy depends on forward-looking causal theorizing about futures that have not yet occurred.

Time-locked language models provide a concrete instantiation and empirical test of this "imagine an LLM in 1633" thought experiment (Felin and Holweg 2024). A team of researchers at the University of Zurich has trained time-locked LLMs (called Ranke-4B) (Göttlich et al. 2025). These "history LLMs" are trained on vast repositories of books and newspaper articles—but with cutoff dates (e.g., ranging from pre-1913 to pre-1946). By construction, historical LLMs eliminate hindsight bias by restricting the model to what was known at a given moment in time. This is important. It allows us to distinguish genuine forward-looking reasoning from mere reconstruction after the fact. Furthermore, a time-bound LLM can also show whether large-scale pattern learning enables some form of emergent ability to reason about new problems and the future.

So, can one of these history LLMs—or any LLM or AI, for that matter—offer forward-looking guidance or reason about new problems or the future? The answer is no. Even though LLMs are trained with massive amounts of data, these models summarize and reproduce dominant beliefs rather than generating forward-looking causal theories or anticipating future possibilities. Despite being "up to date" on the latest science (having been trained with vast numbers of textbooks, scientific books, and articles), the LLM has no way of theorizing or bootstrapping novelty. Certainly, one can prompt and query an LLM about the future. It will talk about the future in fluent ways, but without any ability to meaningfully reason or theorize. It can say nothing useful about the future—beyond summarizing how the future is talked about in the training data.

An LLM has no magical access—or built-in mechanism—to generate or evaluate genuinely novel causal theories, nor any endogenous capacity to reason beyond recombining patterns present in its training data. Strategy, by contrast, is, by definition, forward looking. It depends on novel, generative causal reasoning: the ability to formulate new problems and to

resolve them in ways that create value (Sako and Felin 2025).

To further illustrate the problem with AI reasoning about the future, consider a recent study. Bonelli (2025) analyzed venture capitalists (VCs) who adopted AI-related tools, such as machine learning and predictive analytics, to help them make investment decisions. He found that VCs who used these tools were systematically pushed toward backward-similar startups. Compared with their peers who did not use these tools, VCs using AI-related tools made investments in the types of companies that were successful in the past. More importantly, they were less likely to invest in novel, atypical ventures that ultimately produced rare major successes—the types of "unicorns" that VCs hope to invest in. Data-driven VCs were better at predicting historically patterned outcomes but worse at identifying breakthrough companies, precisely because the algorithms operated within the statistical boundaries of past examples. In short, the adoption of machine learning tools led investors to invest in the familiar and away from innovation, demonstrating how statistical prediction fails in environments where novelty and deviation are key antecedents of value.

Population-Level Generalization and Agreement vs. Actor-Specific Idiosyncrasy and Disagreement

AI learns from populations. It generalizes across large data sets, optimizing loss functions defined over many observations. This works well when homogeneity or aggregation is desirable. But it also creates a tyranny of "on average," in which models flatten heterogeneity and treat population-level patterns as if they applied to individual, actor-specific decisions. This creates an attendant "population loss" where algorithms and statistical techniques optimize for what works best on average across many cases, even if that comes at the expense of what matters for a particular individual decision. This problem is especially acute for any form of unique or individual decision making, which fundamentally characterizes strategy as well (Felin et al. 2025).

This problem is not new. Allport (1961) argued that statistical measurement is inherently population dependent: in science, we only construct variables and scales when they occur widely and vary across many individuals. Anything unique to a single person, by definition, cannot be measured statistically because variance requires a systematic variance across a population. The same logic occurs for any population of entities, including firms and organizations. Statistical methods inevitably privilege what occurs frequently enough to be counted, coded, and correlated while sidelining idiographic, one-of-a-kind entities or configurations that might matter most. It is precisely the latter that are of interest to strategy. Population-level correlations, even when statistically robust, often have little causal or explanatory power for the single case because they cannot capture the causal mechanisms or rare combinations that drive individual outcomes (Meehl 1954). This helps explain why population-trained systems—of which contemporary AI is a prime example—systematically miss the very phenomena that matter in strategic contexts: uniqueness, deviation, idiosyncrasy, and the specific causal theories held by actors.

If population-based methods privilege what recurs across cases in the past, then they inevitably privilege agreement—the overlapping, high-frequency patterns that define the average. Yet strategy rarely lives in the space of agreement. Though AI systems are seen as capitalizing on the wisdom of crowds (e.g., Csaszar et al. 2024), new value and advantage, in fact, arise where actors are polarized and depart from the mean—where they articulate discrepant causal beliefs and see value where others do not (Felin and Zenger 2017). Gius (2025), for example, finds that disagreement and polarized judgments by judges of startups are the best predictors of startup success. Strategic opportunities rarely emerge from widespread consensus or any kind of aggregate wisdom of crowds. Idiosyncrasy and disagreement are not noise—particularly in strategy—even if they are often treated as such in population-level decision frameworks (Kahneman et al. 2021). Rather, idiosyncrasy and disagreement are plausible signals of value—especially when grounded in reasoned causal theories. And this is precisely what population-trained AI systems are least equipped to recognize.

Correlational-Based Recombination vs. Theory-Based Combination

Another way scholars argue that AI generates novelty is through recombination. From this perspective, AI systems are trained on vast corpora containing words, concepts, analogies, examples, and narratives—the building blocks of knowledge. By learning patterns of co-occurrence, large language models can recombine these elements in countless ways, producing outputs that appear novel. Because the set of components available to an LLM vastly exceeds what any individual human can store or retrieve, AI is often portrayed as overcoming bounded rationality. By expanding the combinatorial space of accessible ideas and solutions, AI is said to "unbound" human cognition (Csaszar 2025).

However, this combinatorial account of novelty is deeply underspecified. It leaves unexplained how useful combinations arise from an astronomically large set of possibilities, most of which are irrelevant or useless (Felin and Singell 2026). Recombination alone does not explain why certain elements are treated as candidates for combination, why particular combinations are attempted, or why any of them become

salient. Without a mechanism that generates relevance or salience, recombination risks collapsing into brute-force search.

A central limitation of AI for strategy lies precisely here. AI systems recombine elements correlationally, selecting continuations based on statistical regularities in how representations have co-occurred in the past. This process is powerful, but it operates over a space that is fixed by prior data and prior patterns of association. Such systems cannot, by definition, generate salience for new elements or redefine what counts as a relevant component in the first place. They sample from historically observed structures rather than constructing new ones. Even when outputs appear creative, they remain anchored to previously encountered combinations and lack an account of why a given combination should be pursued to create value (Lewis and Mitchell 2024).

Strategy, by contrast, depends on theory-based reasoning and combination. Strategists do not merely recombine given elements—they actively theorize the elements themselves and the more plausible combinatorial possibilities. Combinatorial salience is a function of actor theories and problem formulations. The relevant components, constraints, and affordances are not known in advance. Both the elements and their possible combinations are often unprestatable prior to the act of theorizing. The combinatorial elements and combinations themselves are constructed through causal reasoning conjectures, problem formulations, and forward-looking beliefs about how value might be created (Felin and Singell 2026).

Using AI for Strategy Requires Human Oversight

Jessica Hullman

AI—and in particular large language models (LLMs)—can help with multiple steps in a strategic planning process, including ideation and simulation. However, to the extent that strategy is an active, iterative process that involves identifying and iteratively refining the relevant context, goals and values that should be brought to bear, human judgment will continue to be required. To the extent that strategy entails identifying genuinely novel solutions that go against past trends, evidence suggests current AI is limited. Consequently, while AI may greatly accelerate steps in a strategy development process, it is unlikely to fully replace human strategists.

For the purposes of this discussion I will consider strategy to be decision-making about what future actions the agent (e.g., firm) should take in the face of uncertainty about the environment (e.g., market conditions) under competition from other agents.

Current AI systems have many strengths that make them useful for strategic decision-making. LLMs excel at summarizing and synthesizing information. They can also be useful for suggesting gaps in vague problem definitions, helping human strategists articulate their goals. They can be used to generate many ideas quickly (e.g., Boussioux et al. 2024), and are particularly powerful when multiple LLM agents can be combined into workflows with specialized roles (e..g, ideation, critique, testing, etc.). With recent advances in agentic tools, they can be applied to execute complex analysis workflows to inform business strategy with relatively minimal human oversight, even collecting and designing data analyses on their own.

However, to the extent that strategy is a human endeavor, verification remains a bottleneck. Reliability is one concern: rapid capability progress in recent years has not been accompanied by comparable increases in the consistency, robustness, or calibration of their outputs (Rabanser et al. 2026). When context is noisy or incomplete or evaluation complex, LLMs can be distracted by irrelevant details (Liu et al. 2023, Li et al. 2025). They can be surprisingly poor at noticing missing information (Fu et al. 2025)—despite being good at finding "needles in haystacks" (Hsieh et al. 2024, Barrett et al. 2025). More broadly, knowledge does not necessarily transfer across related tasks for an LLM the way we expect it to with humans; LLMs often instead exhibit potemkin understanding, where despite being able to accurately define a concept they can struggle to apply that concept to examples the way a human who understands it can (Mancoridis et al. 2025).

When it comes to predicting market conditions and the impacts of hypothetical actions, evidence for LLMs capabilities are mixed. Identifying strategic opportunities often entails reasoning through contingencies, to make predictions about possible future implications of decisions. On the one hand, reasoning about many contingencies at once is easier for computers than humans, who are bound by cognitive limits, e.g., to working memory. To the extent that the strategist feels confident that they know what the important parameters of the environment and firm's action space are, AI is a great tool for simulating possible scenarios. LLMs have proven to be good predictors in contexts where there is stable data and clear indicators—for example, aggregating forecasts or interpreting economic indicators (Gruver et al. 2023, Hansen et al. 2024, Zhang et al. 2024). They have shown some success at predicting human behavior in constrained settings (Argyle et al. 2023, Hewitt et al. 2024, Cui et al. 2025), albeit with some evidence of systematic biases (e.g., Wang et al. 2025).

But on the other hand, LLMs tend to struggle with causal reasoning compared to people. Their reasoning is more likely to be constrained to recognizing sequential causality—patterns where one event often precedes another. They are much weaker at logical causality, which requires abstraction and generalization across

contexts (Kiciman et al. 2023, Mirzadeh et al. 2024). Consequently, while they can generate plausible-sounding causal arguments or causal diagrams, they may easily fail at causal identification in new settings. These findings have been born out as well as recent research on model-generated reasoning chains. A growing body of work suggests that the Chain-of-Thought explanations of reasoning that LLMs provide are fragile (Lanham et al. 2023, Turpin et al. 2023, Kambhampati 2024, Madsen et al. 2024) and driven by token associations (Tang et al. 2023), such that their plausibility depends heavily on the relationship between the training data and test setting (Zhao et al. 2025).

There are methods that exist to identify models that transfer more robustly across contexts, defined by different distributions, and to optimize them in ways that bring them closer to the kinds of causal learning humans can do in specific settings. Smaller amounts of domain-specific data can be used to adjust pretrained models to better reason through supervised fine-tuning or reinforcement learning. But these methods hinge on having good ground truth. For more routine types of strategy decisions (e.g., marketing campaigns), it may be possible for AI to identify robust predictors of success from a collection of past data, and learn to explain their reasoning through post-training. But for bigger decisions where innovation is key, it's not clear what gold standard labeled data should look like.

Part of the problem is that big strategic decisions are inherently forward-looking, and driven by competition where the goal is to identify a unique advantage rather than one that is strongly suggested by past data. An important question becomes the extent to which we expect the past that an AI model is constrained to learn from to be predictive of what will happen in the future. Training data is a key factor: LLMs excel at generating patterns similar to the past, but they are weaker when asked to move into new contexts. Several studies comparing the novelty of human ideas to LLM ideas in innovation contexts find that LLMs exhibit less novel, more incremental ideas (Boussioux et al. 2024, Meincke et al. 2024, De Freitas et al. 2025, Hao et al. 2026). Another challenge is that common methods for fine-tuning base models for particular types of judgments, like reinforcement learning-based approaches, are ultimately not designed to take the model too far outside its pretraining distribution, because doing so tends to lead to less stable behavior. These points raise the question of whether LLMs will ever be capable of radical novelty.

Beyond the need to verify their outputs due to potential bias and unreliability, if strategy is like science–fundamentally intended to advance human goals and knowledge–then it's hard to imagine removing human guidance and validation at all key steps in the process. Ultimately, if strategy constitutes an active sensemaking process by a firm, in which goals and commitments are iteratively refined through a process of formulating queries, collecting information, and refining guiding schemas, then humans must remain in the loop to validate that their goals and commitments are represented in the outputs. LLMs can execute on many complex information-rich tasks, but if humans don't remain engaged enough to understand what assumptions they make at each step, there is no reason to expect that the strategies AI develops will be aligned. Thus, LLMs are better thought of as plug-in tools for parts of the strategy pipeline, to be overseen and instructed by human experts, then general purpose strategic agents.

Beyond the Hype: Reframing the AI and Strategy Debate

Karim R. Lakhani

The debate surrounding artificial intelligence and its role in strategy is often polarized, oscillating between utopian optimism and dystopian fear. A recent discussion among strategy scholars at the Strategy Science AI workshop crystallized this tension, revealing a deep divide in how we, as a field, view the technology's potential. On one side are those who see generative AI as a backward-looking, noncausal parrot, incapable of the novel, forward-looking thought that defines true strategy. On the other are those who see it as a transformative force, poised to reshape the very nature of competitive advantage. Whereas the concerns raised by my colleagues are valid and intellectually rigorous, a focus on "strategy in the wild," how it is actually practiced in organizations, and a commitment to empirical evidence reveal that AI is already a profoundly transformative force. The time for purely philosophical debate is passing. We must now deconstruct the common objections to AI in strategy and, as a field, embrace a new, evidence-based research agenda to guide our understanding of this new era.

Deconstructing the Objections: From Theory to Practice

Three primary objections consistently arise in discussions about AI's strategic limitations. The first is that AI, particularly large language models (LLMs), is fundamentally backward looking. The argument holds that LLMs are architected to predict the next word based on past statistical patterns, making them excellent at regurgitating what has been said before but ill-suited for projecting into a future that does not yet exist. The second objection is that LLMs lack true causal reasoning. They can recognize correlation and sequence but struggle with the abstract, logical, and counterfactual thinking required to identify the deep causal drivers of strategic success. The third, a classic argument echoing resource-based view logic, is that because AI will be widely

available, it cannot create a sustainable competitive advantage.

These are powerful critiques, but they tend to view strategy through the lens of an idealized, theoretical process that often doesn't reflect its messy reality. Take the objection that AI is merely backward looking. This implicitly favors a "stroke of genius" model of innovation. Yet most innovation, and, indeed, most strategy, is not a lightning strike of pure novelty but is recombinatorial. It is the process of taking existing blocks of knowledge and combining them in new ways. As I noted in the panel, even the Wright brothers' breakthrough was not an isolated act of genius but was built on 30 years of prior aviation research and failure. The past is the raw material for the future for humans and AI alike. The difference is that AI can perform this recombinatorial search at a scale and speed that are simply beyond human capacity. Our research has shown that human–AI collaboration can produce ideas that are not only more valuable and viable than those produced by human crowds alone but are also generated at a fraction of the cost and time (Boussioux et al. 2024). AI is not replacing the need for forward-looking vision; it is supercharging the engine of recombination that fuels it.

Similarly, the critique that AI lacks causal reasoning, although technically correct, overlooks how strategy is actually practiced. As I often say, we need to take "strategy in the wild" seriously. If you spend time with executives, you rarely see them applying formal game-theoretic models or pure causal logic. More often, strategy is a heuristic process, a series of judgments made with flawed, incomplete information. The right counterfactual is not AI versus a perfect causal reasoner but AI versus a human expert. Even the best human experts are fallible. In a study on dental diagnostics led at our laboratory, conducted in Germany, even experienced dentists demonstrated approximately 31% false-positive rates and 49% false-negative rates in detecting cavities (Endres et al. 2020). We don't demand superhuman perfection from our human experts, and we should apply the same standard to AI. The question is whether a human, augmented by an AI that can process vast amounts of information and identify patterns, can make better decisions under uncertainty than a human alone. Our evidence suggests they can.

Finally, the objection regarding sustainable competitive advantage is one we have seen with every major technological wave, including the internet. The advantage comes not from mere access to the tool but from the deep, organizational learning required to wield it effectively. It is about the skill, speed, and learning economies that a firm develops in its application. As we argue in a *Harvard Business Review* article, AI is ushering in an era of abundant expertise, fundamentally changing the economics of the firm and the nature of competition (Yerramilli-Rao et al. 2025). Advantage will be captured by those who learn to navigate this new landscape fastest.

An Evidence-Based Counterpoint: Findings from the Field

To move beyond these theoretical debates, the Digital Data Design Institute at Harvard, where I serve as founding chair, has focused on conducting rigorous "clinical trials" of AI in real-world organizational settings (Lakhani 2025). This body of empirical work provides a powerful counterpoint to the purely philosophical objections. Our first major finding is the existence of a "jagged technological frontier." In a field experiment with the Boston Consulting Group, we found that consultants using GPT-4 performed 40% better on tasks inside the AI's capability frontier but 20% worse on tasks outside of it (Dell'Acqua et al. 2026). This demonstrates that the question is not if AI can do strategy but how and for which tasks. Naive application is dangerous, but skilled application is transformative.

Second, our work shows that AI is evolving from a simple tool into a "cybernetic teammate." In a study with Procter & Gamble, we found that individuals using AI performed as well as two-person human teams on complex innovation tasks (Dell'Acqua et al. 2025). The AI filled knowledge gaps, broke down functional silos, and even provided a form of social and motivational support. This moves the conversation beyond simple productivity augmentation to a fundamental rethinking of teamwork and collaboration. Third, our research on creative problem-solving and the economics of expertise shows that AI is profoundly democratizing strategic capabilities. We have found that AI can generate business strategies that are judged by experts as more valuable than those from human crowds (Boussioux et al. 2024) and that the abundance of AI-driven expertise allows smaller, less-experienced teams to compete with established incumbents (Yerramilli-Rao et al. 2025). This is not a marginal improvement; it is a structural shift in the competitive landscape.

A Call for a New Scientific Agenda

This brings me to my final and most urgent point. The field of strategy is at risk for being left behind, of having our own "statistics versus machine learning" moment. While we debate the philosophical nuances, computer scientists are in the wild, building algorithms, throwing compute at problems, and shaping the very future we are theorizing about. We cannot afford to be intellectually insular.

I propose that our field must organize around a new scientific agenda, one inspired by Robert Axelrod's

famous tournaments for the prisoner's dilemma. We should create competitive testbeds for strategy formulation. Let's design complex strategic challenges and invite teams to solve them: humans alone, humans with AI, and eventually, AI agents on their own. Let's measure the outcomes objectively. This is how we move from anecdote and debate to a cumulative body of scientific evidence. This is how we can provide real, evidence-based guidance to practitioners on how to navigate the jagged frontier and avoid the perilous trap of "falling asleep at the wheel" when the AI speaks with convincing error (Lakhani 2025).

The Responsibility of the Strategy Scholar

The debate over AI and strategy is not merely an academic exercise. The technology is here, and it is already being used to make high-stakes decisions. Our responsibility as scholars is not to be the gatekeepers of an idealized, historical definition of what strategy should be. Our responsibility is to be the pioneers in understanding what it is becoming. The challenge is to embrace the empirical reality of AI—with all its jaggedness, risks, and profound potential—and to lead the charge in studying it with the rigor, creativity, and urgency that the moment demands. Otherwise, we risk becoming "fast historians" who documented the past instead of helping shape the rate and direction of this crucial technology that is reshaping our society, economy, and organizations and even challenging the fundamental aspect of what it means to be human.

Should AI Do Strategy? A Professional Perspective

Mari Sako

The prospect of AI agents as strategists raises hope and fear in equal measure. Can AI do strategy? Shedding light on this question requires understanding what strategists do and how AI affects what they do. I argue that the answer depends on the nature of the problem being addressed. Strategy starts with problem framing (Nickerson and Zenger 2004). As such, human strategists lead in framing the problem for an organization and judgment to solve the problem. Given the current state of AI technologies, can AI also do problem framing and judgment?

This paper provides an answer in four parts. First, we ask what strategists as professionals do and assert that of the three modalities (diagnosis, inference, treatment), the hardest to automate is treatment, anchored in the notion of professional judgment. Second, I highlight the central importance of problem formulation in strategy, a quintessentially human activity in which the actor uniquely frames the problem, sometimes in ways not seen by others (Felin et al. 2025). Third, strategy is a job function with workflows of tasks, and regardless of how the start and end of a workflow are defined, many strategy tasks are "hard" for AI to learn (Acemoglu 2025). Fourth, even if full automation of strategic decision making might become possible in the future, we need to assess its desirability. What is technologically feasible is distinct from what we find desirable (Friis and Riley 2025, Shao et al. 2025). As such, strategy scholars should engage more directly with debating about principle-based endorsement (should AI do strategy?), separately from performance-based endorsement.

What Do Strategists Do?

Strategy is a profession (Whittington et al. 2011). Just like other knowledge-intensive professionals, including in medicine, audit, and law, strategists engage in three modalities of diagnosis, inference, and treatment (Abbott 1988). In essence, in the terminology of AI, professionals have a claim to classify a problem (diagnose), reason about it (infer), and act on it (treat). For instance, strategists in mergers and acquisitions, with the help of investment bankers, collect relevant financial information in due diligence (i.e., diagnosis) before using finance knowledge (i.e., inference) to recommend the best financial structure for M&A (i.e., treatment). Inference that connects diagnosis to treatment is based not just on knowledge but also expertise, including difficult-to-codify know-how grounded in accumulated experience of practice (Ward et al. 2020).

AI impacts the three modalities in differential ways. Diagnosis may be largely automated by machine learning and inference by rule-based expert systems. The performance of AI tools may be augmented by reinforcement learning from human feedback and retrieval-augmented generation (RAG). The modality that is most challenging for AI is treatment, namely, arriving at and implementing a decision, a solution. This is due to two reasons. First, human expertise has proven difficult to capture fully as expert system rules or by machine learning algorithms that attempt to learn and embody the tacit know-how of human strategists. Second, real-world problems require professional judgment, an informed decision in specific contexts. It is in this regard that professional judgment is materially different from passing exams. The latter is about accuracy, that is, getting the right answer,[1] whereas the former requires formulating the problem (cf. exams give you the problems).

Importance of Problem Framing in Strategy

When strategists do strategy, they need to first scope out the project at hand by framing the problem: what is this about, and what are we aiming to achieve? Diagnosis, inference, and treatment can begin only after such problem framing is completed. This is implicitly assumed by Abbott (1988), writing about professions. In strategy, there is an explicit recognition of the central

importance of problem formulation that predates discussion of AI in strategy (Nickerson and Zenger 2004). However, with AI, linking the cognitive processes of search, representation, and aggregation to strategy generation and evaluation (Csaszar et al. 2024) skirts over problem framing as a starting point for strategists. AI enables faster search, more complex representation, and drawing on the wisdom of virtual crowds, but for what problem are the search, representation, and aggregation undertaken? In art and design work, "it's too late to be creative" if search for ideas is not preceded by "framing and (re)defining problems and goals" (Hwang 2022). By analogy, it's too late to be strategic if AI in the loop in creative problem-solving (Boussioux et al. 2024) is not preceded by humans (re)defining the problem for which creative solutions are sought.

In a different perspective, problem formulation may be seen to occur in the process of search. Then, problem finding (i.e., discovery and framing of problems) and problem-solving could be crowdsourced (Nickerson et al. 2017). Search using AI could suggest problems and strategy options for specific actors. But humans need to choose from AI-generated options, and such choice is likely to be actor specific (Felin et al. 2025). And it is the actor-specific way in which a problem is formulated, rather than access to AI technology per se, that constitutes a source of competitive advantage. Thus, LLMs may be used to explore, refine, and extend strategic theory by simulating environments where mechanisms and boundary conditions emerge organically through interaction (Tranchero et al. 2025). Simulated strategic interaction, however, informs, but does not lead to, the choice of valuable problems that, if solved, would yield desirable knowledge and capability (Nickerson and Zenger 2004).

Strategy as a Workflow of "Hard-to-Learn" Tasks

Breaking down a job into tasks has become the sine qua non of analyzing AI impact on work. O*NET is the bible taxonomy for a comprehensive list of occupations, specifying tasks for each occupation. But there is no O*NET category for strategists, although related occupations include business intelligence analysts, business continuity planners, and management analysts. Moreover, tasks listed for an occupation tell us nothing about the end-to-end workflow with a sequence of tasks (Handa et al. 2025), nor about task complementarities within an occupation (Acemoglu 2025). And we have plenty of evidence that AI adoption for real (rather than for a pilot study) requires embedding AI use in workflows (Aristidou et al. 2022). So, there is a need to move from tasks to workflows and from automation-augmentation dichotomy to automation-augmentation sequencing in workflows.[2] But the absence of strategy as occupation in O*NET may be a blessing in disguise and points to the difficulty of transposing the microlevel activity–based view of strategy work (Johnson et al. 2003, Whittington 2003) into workflows.

Moreover, many tasks in strategy are hard to learn in the sense that there is lack of a simple mapping between action and desired outcomes. Easy-to-learn tasks, which are relatively straightforward for generative AI to learn and implement, are defined by a simple mapping between action and an outcome metric that is reliable and observable. Hard-to-learn tasks, by contrast, do not have a simple mapping between action and desired outcome. In hard problems, "what leads to the desired outcome in a given problem is typically not known and strongly depends on contextual factors, or the number of relevant contexts may be vast, or new problem-solving may be required" (Acemoglu 2025, p. 30). Answering an exam question is an easy task; evaluating a real-world client problem is a hard task. If strategic decision making is about "identifying and choosing among long-term courses of action that can determine firm performance" (Csaszar et al. 2024), then many strategy tasks are hard to earn, given that many contextual factors could affect firm performance.

Is Full Automation of Strategic Decision Making Feasible and Desirable?

In conclusion, can AI do strategy? Much of the existing disagreement appears to reside in different definitions of strategy. Defined in the way I do in this paper, AI can be a decision support for human strategists, whose capacity for search and diagnosis may be vastly enhanced by AI but whose capacity for problem formulation and judgment will remain complementary to, but not substituted by, AI. Human strategists with AI would do better than those without, but only for specific types of problems that are well understood, not open-ended and therefore "easy to learn" for AI. Thus, problem formulation becomes central as a source of competitive advantage, but also for delineating the bounds of AI capability.

In the future, even if full automation of strategic decision making becomes possible, when technology enables turning hard-to-learn tasks into easy-to-learn ones, strategy scholars should continue to question whether such a future is desirable. Unlike for occupations for which a human touch is central, no one is likely to object that letting AI do all or most of strategy work is "morally repugnant" (Friis and Riley 2025). Nevertheless, we should remain wide-eyed about what sort of future we desire, taking as hint what employees say today about desirable automation (Shao et al. 2025) separately from performance-based endorsement of AI doing strategy.

AI as C-Suite: Strategic Intelligence Through Conflict, Not Consensus

James Evans

Can AI do strategy? I argue that the question misframes a deeper structural homology between artificial intelligence and strategic management—one that reveals both the promise and peril of their integration. Both AI and strategy are engines of compression that transform complex, distributed phenomena into tractable representations. Both face collapse when they operate recursively on their own outputs. And both, when functioning well, depend not on singular optimization but on sustained internal conflict among diverse perspectives. Recognizing this homology suggests that AI will only "do strategy" effectively when we stop treating it as an oracle delivering best answers and instead open its black box to expose the productive disagreement within and allow it to amplify and expand our own strategic conflicts.

The Compression Machines

Deep neural networks, or "deep learning," the drivers of almost all modern artificial intelligence, are engines of compression. They ingest vast corpora—the text of the internet, posted images of the world, digital breadcrumbs of human activity—and compress this information across lattices of weighted connections to effectively predict, generate, and respond. But this compression comes at a cost: inscrutability. No one can completely understand what these models are doing. At the International Conference on Learning Representations this year (2025)—the world's premier deep learning conference—more than half of the papers seemed to focus on mechanistic interpretability. Their authors engaged in what amounts to hermeneutics, trying to make sense of machines they built but could not read (Olah et al. 2020, Elhage et al. 2022). Increasingly, large models are designed by other large models, compounding the opacity.

Strategy science performs an analogous compression. Organizations are themselves black boxes—complex assemblages of routines, cultures, political coalitions, and tacit knowledge that resist simple characterization (Cyert and March 1963, Nelson and Winter 1982). Academic strategy compresses the histories of these organizations into principles: competitive advantage, dynamic capabilities, resource-based views. Like neural networks, strategy science distills distributed complexity into tractable representations that can guide action.

The homology runs deeper. Organizations, like AI models, are increasingly created and evolved by other organizations. Consultants, investors, regulatory agencies, and industry associations shape organizational forms through isomorphic pressure (DiMaggio and Powell 1983). Just as deep reinforcement learning models discovered the activation functions in large language models (Ramachandran et al. 2017), McKinsey, Goldman Sachs, and the SEC shape the architectures, rules, and appendages that strategy science subsequently examines. Both AI and strategy involve recursions that, if they fail to gather new information, can unravel.

The Collapse Problem

When AI models train on their own outputs, they collapse. Shumailov et al. (2024) demonstrated that recursive self-training causes models to progressively narrow their distributions, pulling in the tails until novelty disappears. The model converges toward an impoverished mean, losing the capacity to generate, first, the rare and unexpected, and, second, anything sensible at all. This "model collapse" affects text, images, video, and sound—any domain where compression meets recursion.

Strategy science faces an analogous risk. If strategic principles are derived from successful organizations and those principles then shape future organizations, the field can come to operate on itself and progressively narrow the space of strategic possibility. March (1991) warned of the exploitation trap: organizations that optimize within existing frameworks lose the capacity to explore fundamentally new possibilities. But the problem extends to the field itself. Strategy scholars who study firms shaped by prior strategic frameworks may discover only the principles already embedded in their cases—a kind of intellectual autoregression that converges toward increasingly abstract and inert recommendations, unrooted in the particularities that make strategic action contextually strategic.

Our research on AI adoption among scientists reveals this dynamic empirically. Scientists who adopt AI tools achieve 160% more citations, leave academia 60% less often, and achieve seniority 1.3 years earlier (Hao et al. 2026). But they mobilize toward large-scale data in established areas. The semantic diameter of their questions shrinks. When AI-assisted work is referenced, it does not reference each other. AI completes or kills fields much faster than it creates new ones. Innovation moves toward established data and away from emergent problems. Both AI and strategy need a steady stream of fresh data from new and evolving contexts to escape their recursive traps.

Conflict as Mechanism

The homology between AI and strategy goes deeper still. When we examine how AI systems reason effectively—particularly the recent generation of reinforcement-trained reasoning models such as DeepSeek-R1 and Alibaba's QwQ—we find something most computer scientists find unexpected (Kim et al. 2026). These models,

trained to produce correct answers through reinforcement learning, spontaneously develop internal debate. Their "chain of thought" is not linear reasoning but emergent conversations and conflict between perspectives that were never explicitly programmed. The amount of conflict and diversity in these internal conversations scales linearly with problem difficulty across domains. The models have discovered that building ensembles of disagreeing agents is the best way to reason.

This mirrors what we know about effective organizations. Successful C-suites are not unified voices but coalitions of conflicting perspectives—what Hambrick (1994) called "behavioral integration," which paradoxically requires maintaining disagreement. Eisenhardt et al. (1997) showed that high-performing top management teams sustain substantive conflict while avoiding interpersonal friction. Burgelman (1983, 1991) demonstrated that strategic renewal depends on autonomous initiatives that compete with and challenge official strategy. The organization's capacity for strategic perception emerges from this maintained tension—diverse perspectives that are collectively sensitive to weak signals from the environment that any single viewpoint would miss.

In both AI and organizations, then, effective intelligence is not singular but plural. It emerges from conflict, not consensus. The black boxes, when opened, reveal not unified agents and singular strategies, but ensembles of perspectives that compete, challenge, amplify, and enable the collective reckoning with signals that would otherwise be lost.

The Automation Bias Problem

Here lies the central problem with current AI deployment for strategic purposes. Commercial AI systems spin up diverse internal perspectives, then winnow them to select the "best" answer, package it, defend it, and present it to users as a polished recommendation. In that winnowing process, they effectively replace the user.

Our research demonstrates the consequences. When human users receive a single credible-sounding AI recommendation, they exhibit automation bias—adopting it with minimal critical engagement (Lai et al. 2025), a phenomenon observed in other prior settings (Parasuraman and Riley 1997, Skitka et al. 1999). In our experiments, users feel confident but perform poorly at discriminating true from false novel information. When we instead bias the AIs—making them opinionated either toward or against the user's initial position—performance improves dramatically, though confidence drops. When AIs fight with each other in view of the user, flanking the user's biases, they both perform better at truth discrimination and feel more calibrated.

The lesson is clear: the form factor matters as much as the capability. A highly capable AI that delivers consensus recommendations may make strategic decision making worse, not better. The compression that makes AI useful—collapsing complexity into tractable outputs—becomes pathological when it collapses the very diversity that enables strategic perception and cognition.

Opening Black Boxes

What we need, for both strategy and AI, is to open the black boxes and make visible the systems of productive disagreement inside. Strategy requires a board of perspectives, not a single recommendation. AI should be configured to expand and diversify those perspectives so humans can explore and participate in deliberation rather than receiving polished conclusions.

This means designing systems that expose the internal diversity of AI reasoning, invite human participation at multiple stages, and resist the temptation to collapse complex strategic landscapes into simple recommendations. It means using AI not to replace strategic judgment but to amplify the conflict that makes strategic judgment possible: surfacing perspectives, challenging assumptions, and forcing engagement with the weak signals that herald both opportunity and threat.

At present, humans remain better at picking up expensive and on-the-fly information—subtle, contextual cues that require embodied presence in specific environments. AIs currently excel at compressing the world's libraries, recombining distant knowledge, and surfacing unexpected connections. The productive partnership lies in combining these capabilities: AI-generated diversity feeding human judgment that remains grounded in particular strategic contexts.

Conclusion: From Tool to Institution

The question is not whether AI can do strategy in the abstract but how we design sociotechnical systems that leverage AI's unique capabilities while preserving and enhancing human strategic judgment. This requires understanding AI not as a tool or even an agent, but as an institutional form—one we can wire, rewire, and evolve to serve purposes we define.

The homology between AI and strategy suggests a research agenda: understanding how compression systems avoid collapse, how internal conflict sustains adaptive capacity, and how to design human–AI partnerships that preserve productive disagreement rather than optimizing it away. Coordination and consensus come from similarity. Creativity and strategic renewal come from maintained difference. Both AI and strategy science must resist the pull toward premature convergence, the comfortable collapse into best practices and best answers that feels like progress but forecloses the possibilities that strategy, at its best, is meant to discover.

Substitute or Complement? AI and the Future of Strategic Work

Aaron Chatterji

As a strategy scholar conducting research on generative artificial intelligence (AI), the question of whether AI can do strategy is especially salient. I have spent my career building the skills to do strategy research and the idea that AI could do my job is unsettling. However, it is also exciting to consider how AI could complement my research and help me uncover new insights. This tension between whether AI will be a substitute or a complement for tasks, jobs, and entire human organizations is at the heart of contemporary debates about the economic implications of this transformative technology.

At the time of this panel discussion, AI capabilities are rapidly advancing. These capabilities include being able to conduct many of the tasks a strategist is responsible for in an organization. One example would be a market research report that includes an industry analysis and assessment of the firm's strengths, weaknesses, opportunities, and threats (SWOT).

However, jobs are not simply a collection of discrete tasks. The human strategist would need to ensure that the information in the market research report is accurate and customized to the context of the organization. The strategist might present the report during a meeting with other executives, leverage the insights to persuade the CEO to undertake a particular course of action and ultimately be accountable for whether the chosen strategy achieves its objective.

This stylized example suggests to me that whereas AI can do many of the tasks that comprise making and executing strategy, it is more likely to be a complement to strategists than able to "do strategy" on its own. Why? First, it will take time to change existing workflows in organizations to fully leverage the potential of AI in strategy. Strategy does not exist in a functional silo in organizations. Strategy is an integrative function that needs to align with finance, marketing, sales, and operations. Strategy requires unusually high levels of context about competition, institutions, and capabilities. Today, the most common AI tools have limited access to this sort of context. As both model intelligence and contextual awareness improve, however, AI applications may evolve more quickly in some functions than others, as we see today with the dramatic improvement in coding. This uneven development across functions will slow the progress of AI to actually do strategy across the organization.

Next, what it means to do strategy is also a moving target. As AI capabilities increase, strategists may add new tasks to their job description or simply spend more time on higher-value tasks such as leadership, persuasion, and delegation. Most importantly, a human will need to be accountable for whether the strategy succeeds or fails. Although it is difficult to predict how quickly AI capabilities will advance, it seems clear that the job of a strategist will be much different a decade from now than it is today.

I also believe that there is an important distinction between what generative AI can do and what we will let it do in organizations. Laws, norms, and values may indeed evolve over time, but humans will remain in the loop in many cases because humans will want the empathy and accountability that only another human can provide.

To accelerate the ability of AI to do strategy, scholars should work to develop evaluations or "evals" to guide AI researchers as they improve their models. These evals can help encode desired model behavior for training and inside organizational contexts. Strategy scholars are well positioned to identify which tasks are most important in making and executing strategy and what excellent outcomes look like. We should also remember that billions of people are already talking to AI chatbots, and many are asking for advice about strategic questions. Engaging with frontier labs to shape model behavior is also an important component for helping AI do strategy better.

My research and practical experience with generative AI have taught me to be humble about making bold predictions. However, I am fairly confident that AI will absorb parts of what strategy entails today. But I am equally confident that strategists will still have plenty to do tomorrow and for years to come.

Acknowledgments

At the time of publication, Aaron Chatterji was the chief economist of OpenAI. This paper was written in his capacity as a professor at Duke University and does not represent the views of OpenAI.

Endnotes

[1] OpenAI's GPT-5 can get 100% or nearly 100% correct in math tests and pass medical exams (U.S. Medical Licensing Examination) with a 97% score (https://openai.com/index/introducing-gpt-5/).

[2] This is, in part, because generative and agentic AI is undermining the automation-augmentation dichotomy, with each task being subjected to a combination of modes (directive, interactive, learning, etc.). See https://www.anthropic.com/research/anthropic-economic-index-september-2025-report.

References

Abbott A (1988) *The System of Professions: An Essay on the Division of Expert Labor* (University of Chicago Press, Chicago).

Acemoglu D (2025) The simple macroeconomics of AI. *Econom. Policy* 40(121):13–58.

Agrawal A, Gans J, Goldfarb A (2022) *Prediction Machines: The Simple Economics of Artificial Intelligence* (Harvard Business Review Press, Boston).

Allport G (1961) *Pattern and Growth in Personality* (Holt, Rinehart and Winston, New York).

Argyle LP, Busby EC, Fulda N, Gubler JR, Rytting C, Wingate D (2023) Out of one, many: Using language models to simulate human samples. *Political Anal.* 31(3):337–351.

Aristidou A, Jena R, Topol EJ (2022) Bridging the chasm between AI and clinical implementation. *Lancet* 399(10325):620.

Barney J, Reeves M (2024) AI won't give you a new sustainable advantage: But using it may amplify the ones you already have. *Harvard Bus. Rev.* (September–October 2024), https://hbr.org/2024/09/ai-wont-give-you-a-new-sustainable-advantage.

Barrett L, Bajaj VS, Kingan RJ (2025) Can LLMs find a needle in a haystack? A look at anomaly detection language modeling. Christodoulopoulos C, Chakraborty T, Rose C, Peng V, eds. *Findings Assoc. Comput. Linguistics: EMNLP 2025* (Association for Computational Linguistics, Stroudsburg, PA), 6428–6435.

Bonelli M (2025) Data-driven investors. *Rev. Financial Stud.*, ePub ahead of print October 13, https://doi.org/10.1093/rfs/hhaf078.

Boussioux L, Lane JN, Zhang M, Jacimovic V, Lakhani KR (2024) The crowdless future? Generative AI and creative problem-solving. *Organ. Sci.* 35(5):1589–1607.

Burgelman RA (1983) A process model of internal corporate venturing in the diversified major firm. *Admin. Sci. Quart.* 28(2):223–244.

Burgelman RA (1991) Intraorganizational ecology of strategy making and organizational adaptation. *Organ. Sci.* 2(3):239–262.

Csaszar FA (2025) Unbounding rationality: Why AI is a fundamental issue for strategy. Preprint, *submitted September 8*, https://doi.org/10.2139/ssrn.5454634.

Csaszar FA, Laureiro-Martínez D (2018) Individual and organizational antecedents of strategic foresight: A representational approach. *Strategy Sci.* 3(3):513–532.

Csaszar FA, Rhee L (2025) The power and limits of distributed representations in strategic decision-making. *Strategy Sci.* Forthcoming.

Csaszar FA, Steinberger T (2022) Organizations as artificial intelligences: The use of artificial intelligence analogies in organization theory. *Acad. Management Ann.* 16(1):1–37.

Csaszar FA, Ketkar H, Kim H (2024) Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. *Strategy Sci.* 9(4):322–345.

Cui Z, Li N, Zhou H (2025) A large-scale replication of scenario-based experiments in psychology and management using large language models. *Nature Comput. Sci.* 5(8):627–634.

Cyert RM, March JG (1963)*ABehavioral Theory of the Firm* (Prentice-Hall, Englewood Cliffs, NJ).

De Freitas J, Nave G, Puntoni S (2025) Ideation with generative AI—In consumer research and beyond. *J. Consumer Res.* 52(1): 18–31.

Dell'Acqua F, McFowland E III, Mollick E, Lifshitz H, Kellogg KC, Rajendran S, Krayer L, Candelon F, Lakhani KR (2026) Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Organ. Sci.* Forthcoming.

Dell'Acqua F, Ayoubi C, Lifshitz H, Sadun R, Mollick E, Mollick L, Han Y, et al. (2025) The cybernetic teammate: A field experiment on generative AI and teamwork. *Harvard Business School Working Paper No. 25–043*, Harvard Business School, Boston.

DiMaggio PJ, Powell WW (1983) The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *Amer. Sociol. Rev.* 48(2):147–160.

Doshi AR, Bell JJ, Mirzayev E, Vanneste BS (2025) Generative artificial intelligence and evaluating strategic decisions. *Strategic Management J.* 46(3):583–610.

Eisenhardt KM, Kahwajy JL, Bourgeois LJ (1997) Conflict and strategic choice: How top management teams disagree. *California Management Rev.* 39(2):42–62.

Elhage N, Hume T, Olsson C, Schiefer N, Henighan T, Kravec S, Hatfield-Dodds Z, et al. (2022) Toy models of superposition. Preprint, submitted September 21, https://arxiv.org/abs/2209.10652.

Endres MG, Hillen F, Salloumis M, Sedaghat AR, Niehues SM, Quatela O, Hanken H, et al. (2020) Development of a deep learning algorithm for periapical disease detection in dental radiographs. *Diagnostics* 10(6):430.

Felin T, Holweg M (2024) Theory is all you need: AI, human cognition, and causal reasoning. *Strategy Sci.* 9(4):346–371.

Felin T, Singell M (2026) Technology: Theory-based experimentation and combinatorial salience. *Eur. Econom. Rev.* 181:105186.

Felin T, Zenger T (2017) The theory-based view: Economic actors as theorists. *Strategy Sci.* 2(4):258–271.

Felin T, Sako M, Hullman J (2025) Artificial intelligence and actor-specific decisions. *Preprint, submitted June 3*, https://doi.org/10.2139/ssrn.5279401.

Friis S, Riley JW (2025) Performance or principle: Resistance to artificial intelligence in the US labor market. *Harvard Business School Working Paper* No. 26–017, Harvard Business School, Boston.

Fu HY, Shrivastava A, Moore J, West P, Tan C, Holtzman A (2025) AbsenceBench: Language models can't tell what's missing. Preprint, submitted June 13, https://arxiv.org/abs/2506.11440.

Gruver N, Finzi M, Qiu S, Wilson AG (2023) Large language models are zero-shot time series forecasters. *Adv. Neural Inform. Processing Systems* 36:19622–19635.

Gius L (2025) Disagreement predicts startups success: Evidence from venture competitions. *Strategy Sci.* 10(2):93–108.

Göttlich D, Loibner D, Jiang G, Voth H-J (2025) History LLMs. Technical report. https://github.com/DGoettlich/history-llms.

Handa K, Tamkin A, McCain M, Huang S, Durmus E, Heck S, Mueller J, et al. (2025) Which economic tasks are performed with AI? Evidence from millions of Claude conversations. *Preprint, submitted* February 11, https://doi.org/10.48550/arXiv.2503.04761.

Hansen AL, Horton JJ, Kazinnik S, Puzzello D, Zarifhonarvar A (2024) Simulating the survey of professional forecasters. Preprint, submitted February 10, https://doi.org/10.2139/ssrn.5066286.

Hambrick DC (1994) Top management groups: A conceptual integration and reconsideration of the "team" label. *Res. Organ. Behav.* 16:171–213.

Hao Q, Xu F, Li Y, Evans J (2026) Artificial intelligence tools expand scientists' impact but contract science's focus. *Nature* 649(8099): 1237–1243.

Heshmati M, Csaszar FA (2024) Learning strategic representations: Exploring the effects of taking a strategy course. *Organ. Sci.* 35(2):453–473.

Hewitt L, Ashokkumar A, Ghezae I, Willer R (2024) Predicting results of social science experiments using large language models (August 8), https://samim.io/dl/Predicting%20results%20of%20social%20science%20experiments%20using%20large%20language%20models.pdf.

Hodges A (1983) *Alan Turing: The Enigma* (Simon & Schuster, New York).

Hsieh C-P, Sun S, Kriman S, Acharya S, Rekesh D, Jia F, Zhang Y, Ginsburg B (2024) RULER: What's the real context size of your long-context language models? Preprint, submitted December 5, https://arxiv.org/abs/2404.06654.

Hwang AH-C (2022) Too late to be creative? AI-empowered tools in creative processes. Barbosa S, Lampe C, Appert C, Shamma DA, eds. *CHI EA'22: CHI Conf. Human Factors Comput. Systems Extended Abstracts* (Association for Computing Machinery, New York), 1–9.

Johnson G, Melin L, Whittington R (2003) Micro strategy and strategizing: Towards an activity-based view. *J. Management Stud.* 40(1):3–22.

Kahneman D (2018) Comment on "artificial intelligence and behavioral economics." *The Economics of Artificial Intelligence: An Agenda* (University of Chicago Press, Chicago), 608–610.

Kahneman D, Sibony O, Sunstein CR (2021) *Noise: A Flaw in Human Judgment* (Little, Brown & Co., New York).

Kambhampati S (2024) Can large language models reason and plan? *Ann. New York Acad. Sci.* 1534(1):15–18.

Kiciman E, Ness R, Sharma A, Tan C (2023) Causal reasoning and large language models: Opening a new frontier for causality. *Preprint, submitted* April 28, https://doi.org/10.48550/arXiv.2305.00050.

Kim J, Lai S, Scherrer N, Aguera y Arcas B, Evans J (2026) Reasoning models generate societies of thought. Preprint, submitted January 15, https://arxiv.org/abs/2601.10825.

Knuth DE (1974) Computer programming as an art. *Comm. ACM* 17(12):667–673.

Lai S, Kim J, Kunievsky N, Potter Y, Evans J (2025) Biased AI improves human decision-making but reduces trust. Preprint, submitted August 12, https://arxiv.org/abs/2508.09297.

Lakhani KR (2025) AI needs clinical trials: Harvard's findings on democratization. [Video]. TEDxBoston https://youtu.be/Jb9b8j0-RIQ.

Lanham T, Chen A, Radhakrishnan A, Steiner B, Denison C, Hernandez D, Perez E (2023) Measuring faithfulness in chain-of thought reasoning. *Preprint, submitted* July 17, https://doi.org/10.48550/arXiv.2307.13702.

Lewis M, Mitchell M (2024) Evaluating the robustness of analogical reasoning in large language models. Preprint, submitted November 21, https://arxiv.org/abs/2411.14215.

Li W, Wang X, Yuan S, Xu R, Chen J, Dong Q, Xiao Y, Yang D (2025) Curse of knowledge: When complex evaluation context benefits yet biases LLM judges. Preprint, submitted September 3, https://arxiv.org/abs/2509.03419.

Liu NF, Lin K, Hewitt J, Paranjape A, Bevilacqua M, Petroni F, Liang P (2023) Lost in the middle: How language models use long contexts. Preprint, submitted July 6, https://arxiv.org/abs/2307.03172.

Madsen A, Chandar S, Reddy S (2024) Are self-explanations from large language models faithful?*Findings Assoc. Comput. Linguistics ACL 2024*(ACL, Stroudsburg, PA), 295–337.

Mancoridis M, Weeks B, Vafa K, Mullainathan S (2025) Potemkin understanding in large language models. Preprint, submitted June 26, https://arxiv.org/abs/2506.21521.

March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.

Meehl PE (1954) *Clinical Versus Statistical Prediction* (University of Minnesota Press, Minneapolis).

Meincke L, Girotra K, Nave G, Terwiesch C, Ulrich KT (2024) Using large language models for idea generation in innovation. *Research paper*, The Wharton School of the University of Pennsylvania, Philadelphia.

Mirzadeh I, Alizadeh K, Shahrokhi H, Tuzel O, Bengio S, Farajtabar M (2024) GSM-Symbolic: Understanding the limitations of mathematical reasoning in large language models. Preprint, submitted October 7, https://arxiv.org/abs/2410.05229.

Nelson RR, Winter SG (1982) *An Evolutionary Theory of Economic Change* (Harvard University Press, Cambridge, MA).

Nickerson JA, Zenger TR (2004) A knowledge-based theory of the firm—The problem-solving perspective. *Organ. Sci.* 15(6):617–632.

Nickerson JA, Wuebker R, Zenger T (2017) Problems, theories, and governing the crowd. *Strategic Organ.* 15(2):275–288.

Olah C, Cammarata N, Schubert L, Goh G, Petrov M, Carter S (2020) Zoom in: An introduction to circuits. *Distill* 5(3):e00024.001.

Parasuraman R, Riley V (1997) Humans and automation: Use, misuse, disuse, abuse. *Human Factors* 39(2):230–253.

Rabanser S, Kapoor S, Kirgis P, Liu K, Utpala S, Narayanan A (2026) Towards a science of AI agent reliability. *Preprint, submitted February 18*, https://doi.org/10.48550/arXiv.2602.16666.

Ramachandran P, Zoph B, Le QV (2017) Searching for activation functions. *Preprint, submitted October 16*, https://doi.org/10.48550/arXiv.1710.05941.

Sako M, Felin T (2025) Does AI prediction scale to decision making? *Comm. ACM* 68(4):18–21.

Shao Y, Zope H, Jiang Y, Pei J, Nguyen D, Brynjolfsson E, Yang D (2025) Future of work with AI agents: Auditing automation and augmentation potential across the US workforce. Preprint, submitted June 6, https://arxiv.org/abs/2506.06576.

Shumailov I, Shumaylov Z, Zhao Y, Papernot N, Anderson R, Gal Y (2024) AI models collapse when trained on recursively generated data. *Nature* 631(8022):755–759.

Skitka LJ, Mosier KL, Burdick M (1999) Does automation bias decision-making? Internat.*J. Human-Comput. Stud.* 51(5):991–1006.

Sutton RS, Barto AG (2018) *Reinforcement Learning: An Introduction*, 2nd ed. (MIT Press, Cambridge, MA).

Tang X, Zheng Z, Li J, Meng F, Zhu SC, Liang Y, Zhang M (2023) Large language models are in-context semantic reasoners rather than symbolic reasoners. *Preprint, submitted* May 24, https://doi.org/10.48550/arXiv.2305.14825.

Tranchero M, Brennink CF, Murugan A, Nagaraj A (2025) Theorizing with large language models. *Preprint*, submitted October 8, 2024, https://doi.org/10.2139/ssrn.4978831.

Turpin M, Michael J, Perez E, Bowman S (2023) Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Adv. Neural Inform. Processing Systems* 36:74952–74965.

Wang A, Morgenstern J, Dickerson JP (2025) Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nature Machine Intelligence* 7(3):400–411.

Ward P, Schraagen JM, Gore J, Roth E (2020) *The Oxford Handbook of Expertise* (Oxford University Press, New York).

Whittington R (2003) The work of strategizing and organizing: For a practice perspective. *Strategic Organ.* 1(1):117–125.

Whittington R, Cailluet L, Yakis-Douglas B (2011) Opening strategy: Evolution of a precarious profession. *British J. Management* 22(3):531–544.

Wingate D, Burns BL, Barney JB (2025) Why AI will not provide sustainable competitive advantage. *MIT Sloan Management Rev.* 66(4):9–11.

Yerramilli-Rao B, Corwin J, Li Y, Lakhani KR (2025) Strategy in an era of abundant expertise.*Harvard Bus. Rev.*(March–April 2025), https://hbr.org/2025/03/strategy-in-an-era-of-abundantexpertise.

Zhang X, Chowdhury RR, Gupta RK, Shang J (2024) Large language models for time series: A survey. Preprint, submitted February 2, https://arxiv.org/abs/2402.01801.

Zhao C, Tan Z, Ma P, Li D, Jiang B, Wang Y, Liu H (2025) Is chain of-thought reasoning of LLMs a mirage? A data distribution lens. *Preprint, submitted* August 2, https://doi.org/10.48550/arXiv.2508.01191.