

SPECIAL ISSUE ARTICLE

SMS | Strategic Entrepreneurship Journal



WILEY

AI and Entrepreneurship

New wine, old bottles? Effective strategies of new GenAI firms

Kathleen M. Eisenhardt | Xilan Zhang

Department of Management Science and Engineering, Stanford University, Stanford, California, USA

Correspondence

Kathleen M. Eisenhardt, Department of Management Science and Engineering, Stanford University, Huang Engineering Center, 475 Via Ortega, Stanford, CA 94305, USA.

Email: kme@stanford.edu

Abstract

Research Summary: Blending insights from GenAI, strategic reasoning, and empirical grounding in five illustrative ventures, we contribute a theoretical framework for effective strategies of new GenAI firms. They face *bifurcated uncertainty* (i.e., high uncertainty of a nascent market, moderate uncertainty of a somewhat predictable technology trajectory) such that two *complementary strategic logics*—Process Perspective (PP) and Resource-based View (RBV)—are relevant. Second, we identify *process-based* (i.e., well-designed business model and product innovation process) and *resource-based* (i.e., superior data, efficient orchestration, complementary resources) advantages. Some relate to GenAI, some generalize. Third, we address what is unique about GenAI entrepreneurship, including *strategic implications* of GenAI, *technology strategy* when there is upstream technical dominance and many downstream firms, and a *cross-technology comparison* with technologies like software and predictive AI.

Managerial Summary: The rapid trajectory of GenAI is stunning. Yet many new GenAI firms are simply thin wrappers around well-known LLMs. Drawing on five illustrative ventures, we contribute a framework for effective strategies for new GenAI firms. These firms face *bifurcated uncertainty*

(i.e., high uncertainty of a nascent market, moderate uncertainty of a somewhat predictable technology trajectory). This calls for creating advantage by navigating high uncertainty markets to accomplish key strategic tasks and relying on foresight to build strategically valuable resources. Specifically, we identify *process-based* (well-designed business model and product innovation process) and *resource-based* (i.e., superior data, efficient orchestration, complementary resources) advantages. We also address what is unique about GenAI entrepreneurship, including *strategic implications* of GenAI and a *cross-technology comparison* with technologies like software, predictive AI, and quantum.

KEYWORDS

bifurcated uncertainty, entrepreneurial strategy, generative AI, process perspective, resource-based view

1 | INTRODUCTION

The rapid trajectory of increasing artificial intelligence (AI) capabilities, especially generative AI (GenAI), is stunning. Since the large-scale introduction of ChatGPT in fall 2022, GenAI progress has maintained a torrid pace. Large language models (LLMs) that form the current foundation of GenAI are getting bigger, more diverse, and more useful. The scale of financial investment in these models and related infrastructure has been massive.

Many argue that GenAI's ability to produce, manipulate, and analyze content in unprecedented ways has the potential to redefine corporate boundaries, reshape industries, alter competitive advantage, and change the nature of work (e.g., Allen & McDonald, 2026; Dell'Acqua et al., 2026; Hatmaker, 2026; Noy & Zhang, 2023). Others argue that, while it is a significant technology, its impact may be overestimated for some individuals and some large, incumbent firms (e.g., Allen et al., 2026; Brynjolfsson et al., 2025; Jacobides & Ma, 2026; Kilminster et al., 2026; McCracken, 2026; Yotzov et al., 2026).

Regardless of which scenario (or another one) is accurate, GenAI is clearly driving a tsunami of new firms. In fact, GenAI ventures accounted for 50% of Y-combinator firms in 2024 (Kim & Koning, 2026) and captured over 60% of U.S. venture capital in 2025 (Pitchbook & National Venture Capital Association, 2026). Moreover, since new firms are often less inertial than incumbent firms, they are likely to exploit GenAI opportunities more rapidly (Ozcan et al., 2026). Yet, new GenAI firms often seem to resemble those of the Internet bubble with many offering little strategic differentiation beyond thin wrappers around well-known LLMs.

Our purpose is to unpack the strategies that are likely to be effective for new GenAI firms. By *new GenAI firms*, we mean recently founded firms that produce GenAI-powered products or what Kim & Koning (2026) term the “product channel.” They sit in the Applications layer of the GenAI stack where entrepreneurial opportunities are particularly abundant (Appendix A). They serve a broad range of customers including individual consumers and professionals (e.g., Gamma, OpenEvidence), small business (e.g., Fazeshift), government (e.g., WebAI), and enterprise (e.g., Sierra AI). We exclude ventures that adopted GenAI solely for their internal workflows (what Kim & Koning (2026) term the “process channel”).

Our focus on new GenAI firms fills the gap between *individual-level* GenAI usage (e.g., Dell'Acqua et al., 2026) and *firm-level* GenAI adoption by incumbents (e.g., Jacobides & Ma, 2026; Kilminster et al., 2026). While valuable,



these streams leave open new GenAI firms. Our strategy lens complements organizing new GenAI firms (e.g., DeSantola et al., 2023; Kim & Koning, 2026). New GenAI firms face a challenging strategic situation. They anchor their products on powerful upstream models that they do not own and that rivals can use. Thus, differentiation is essential but difficult (also von Krogh et al., 2026). They also face challenging unit economics, elusive learning economies, rapid technological change, and hallucination. These features of GenAI have strategic implications that shape the effective strategies of new GenAI firms (next section) and differentiate them from technologies like software and predictive AI.

GenAI also offers an intriguing theoretical setting to explore two strategic logics. New GenAI firms typically face nascent or disrupted markets where products, rivals, business models, and customers are highly uncertain (Zuzul & Tripsas, 2020). Here, the *Process Perspective* by which entrepreneurs succeed in high-uncertainty markets by accomplishing strategic tasks like designing an advantageous business model or forming an effective ecosystem is a relevant strategic logic (Bingham et al., 2019; Eisenhardt et al., 2026; Ozcan & Eisenhardt, 2009; Volmar & Eisenhardt, 2025). Yet at the same time, AI is progressing along a somewhat predictable trajectory. GenAI was a surprise for many, but not most computer scientists. Here, the *Resource-based View* by which executives have sufficient foresight to develop strategically valuable resources is also a relevant strategic logic (Barney, 1991; Wernerfelt, 1984; Wilde & Packard, 2026). As Furr and Eisenhardt (2021) argue, there is also a sweet spot of uncertainty where the strategic logics of the Process Perspective (PP) and Resource-based View (RBV) overlap (also Ross & Lieberman, 2026). GenAI offers an opportunity to explore this sweet spot.

We use an abductive approach (He & Puranam, 2026). We began by developing a general understanding of GenAI technology. We then chose five new GenAI firms as illustrations. These firms were founded since 2022 with varied founding teams, markets, customers, and products that increase generalizability. While these firms are more successful than most, their success increases the richness of available information and points to well-defined strategies that have proven at least initially effective across multiple contexts. At the same time, we used the PP and RBV strategic logics to explain (if possible) why the strategies of these new GenAI firms were effective. We also remained open to other existing and emergent strategic logics. Finally, we iterated among our general insights into GenAI, strategic logics, and illustrative firms to develop a theoretical framework of effective strategies for new GenAI firms.

We have several contributions at the nexus of entrepreneurship, strategy, and organization theory. Our primary contribution is a theoretical framework for how GenAI entrepreneurs create effective strategies including major sources of advantage. These entrepreneurs follow a temporal sequence of (1) initially addressing high market uncertainty and process-based advantages (i.e., well-designed business models and product innovation processes), (2) adding resource-based advantages (i.e., superior data, efficient orchestration, complementary resources) to keep pace with a moderately predictable technology trajectory, and (3) relying on both logics as the technology advances. Thus, the Process Perspective and Resource-Based View are *complementary* (not competing) strategic logics that together form complete strategies. Figure 1 summarizes this theoretical framework.

A novel contribution is the *varied generalizability* of these advantages—i.e., some relate to unique GenAI features (e.g., well-designed product innovation process, efficient orchestration), while others generalize widely across many new firms (e.g., well-designed business model, complementary resources). We also deepen understanding of these advantages with illustrative cases of new GenAI firms that achieved early success. While these firms may illustrate one advantage especially well, they often pursue most or even all of them. Overall, our framework is both theoretically reasoned and empirically grounded in actual firms.

We also introduce the novel concept of *bifurcated uncertainty*—i.e., high uncertainty of a nascent or disrupted market and moderate uncertainty of a somewhat predictable technology trajectory. In contrast, much entrepreneurship research focuses on navigating high uncertainty (e.g., Camuffo et al., 2024; McDonald & Eisenhardt, 2020; Sarasvathy, 2024; Snihur & Zott, 2020) but neglects the possibility of bifurcated uncertainty and accumulation of strategically valuable resources. Prior research also frames (if at all) the transition from PP to RBV as a brief interval between the high uncertainty of new firms in nascent or disrupted markets and the moderate uncertainty of maturing firms in established markets (e.g., Furr & Eisenhardt, 2021; Tidhar et al., 2025). Yet, bifurcated uncertainty may lengthen this interval.

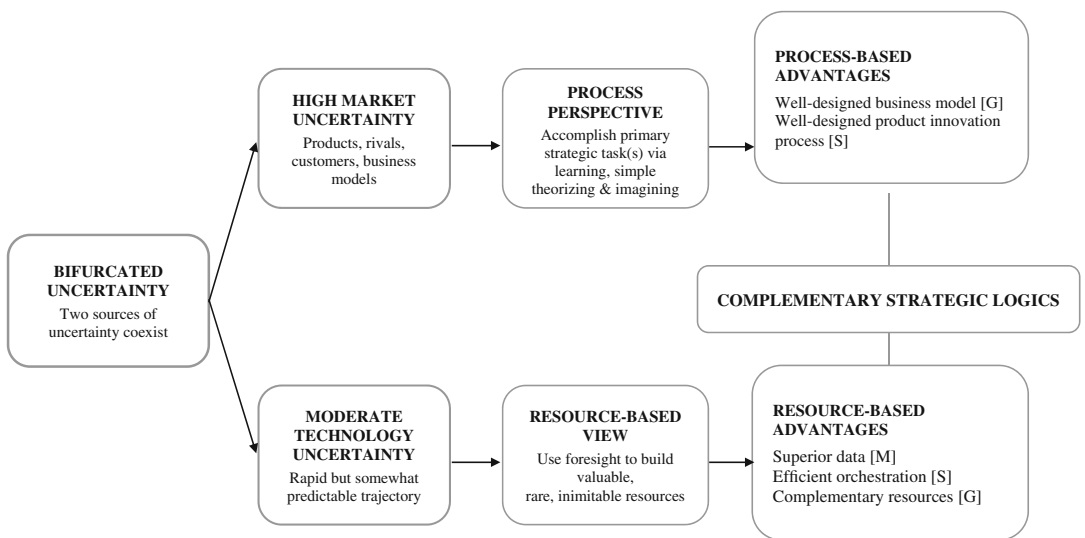


FIGURE 1 A theoretical framework of effective strategies for new GenAI firms. [G] denotes advantages broadly generalizable across many new firms. [S] denotes advantages closely tied to distinctive GenAI features. [M] denotes advantages that are moderately generalizable across many new digitally-based firms.

Second, we contribute insights into the link between strategy and technology by asking what is unique about entrepreneurship in the age of GenAI. In response, we introduce a *novel typology* of key GenAI features and their strategic implications that shape which strategies are effective. We also add a more precise theory of *technology strategy* when there is upstream technology dominance and downstream access for many firms as occurs in GenAI (also Kapoor & Raj, 2026). Finally, we contribute a *cross-technology comparison* of the strategic implications of GenAI for new firms versus other technology waves like software and predictive AI.

Overall, we contribute to the *larger GenAI conversation* that is clouded by head-spinning oscillation between hype and doom. We clarify that, although GenAI is clearly game-changing, well-known strategic logics (e.g., RBV, PP) and theories (e.g., agency, complexity, and contingency theories) are unexpectedly useful. We also cut through some ambiguity surrounding GenAI by grounding our framework in five illustrative new GenAI firms. While GenAI has significant societal implications that are not well-understood and that we do not at all wish to minimize, we highlight applications that may begin to de-mystify GenAI in a variety of settings.

We begin by outlining key features of GenAI and their strategic implications that help shape why some strategies are effective while others are not. We then turn to process-based advantages and resource-based ones, including illustrations from new GenAI firms that have achieved initial success. We conclude by discussing our major insights and contributions.

2 | STRATEGIC IMPLICATIONS OF GENERATIVE ARTIFICIAL INTELLIGENCE

Generative AI (GenAI) is a type of artificial intelligence that creates new content such as text, images, audio, and computer code (see Appendix A for a description of the GenAI industry). Although broader in scope, GenAI can be (and often is) equated with large language models (LLMs). Several features of GenAI have significant strategic implications for new GenAI firms and distinguish effective strategies of new GenAI firms. These strategic implications are summarized in Table 1.



TABLE 1 Strategic implications of GenAI for new GenAI firms.

GenAI feature	Description	Strategic implication
Low LLM switching costs	Switching among LLMs (and potentially other models) is relatively quick, easy, and inexpensive	Effectiveness of switching among LLMs to capture different functionalities, better tradeoffs of cost vs. speed vs. quality, less dependence on any model and higher performance
High (and volatile) variable costs	Significant (and volatile) variable costs (e.g., inference) are likely. Complex queries and AI agents raise variable costs. LLM subsidies mask true variable costs. Lean organizing w/ GenAI can lower variable costs	Pay careful attention to unit economics
Elusive learning economies	Self-reinforcing learning cycle is not automatic. Traditional tactics (e.g., A/B testing), prompt context and user feedback can help to learning	Actively cultivate learning economies
Rapid but partially predictable innovation	Technology trajectory of improved LLM capabilities, more complex and agent-based tasks, open-weight and open-source LLMs, lower per unit inference costs, disruptions (e.g., world models), etc. is somewhat foreseeable	Value of foresight to anticipate innovations and potentially exploit them
Hallucination	LLMs can create unreliable outputs. These outputs can be costly and limit productivity gains. Triangulating LLM outputs, RAG, collecting feedback on output quality and other tactics can help	Actively mitigate hallucination, especially in high-stakes settings like health care, banking and energy infrastructure

First, *switching costs* between LLMs (and potentially other models) are low, making moving among LLMs relatively quick, easy, and inexpensive. Users can switch to (1) capture different functionalities such as Claude for coding and ChatGPT for text analysis and (2) make better tradeoffs among cost versus speed versus quality. While LLM creators (and inference platform owners) will likely attempt to become “sticky” in the long run, switching costs remain low. Moreover, routing firms like OpenRouter are making switching even easier. Finally, low switching costs also limit the pricing power of LLM creators. Overall, the strategic implication for new GenAI firms is the *effectiveness of switching among LLMs* to (1) capture different functionalities, (2) make better tradeoffs at the frontier of cost versus speed versus quality, (3) reduce dependence on any one provider, and (4) improve performance.

Second, *variable costs* can be high (and volatile). This distinguishes GenAI from traditional software. Traditional software such as SaaS typically has high fixed (e.g., coding) and low (e.g., customer service) or even zero (e.g., manufacturing) variable costs. This cost structure favors growth and makes the unit economics of software attractive because the marginal cost of adding a customer approaches zero. In contrast, the marginal cost of adding a GenAI customer is not zero. Rather, new GenAI firms typically have both fixed (e.g., creation and training) and variable (e.g., inference) costs. Thus, while an individual GenAI task may be inexpensive (e.g., customer service at an airline), enterprise-scale volume can bring unexpectedly high costs. Similarly, more useful GenAI tasks and agent-based actions often add complexity, thus raising costs even when per-unit inference costs decline. Token usage by agents can be especially difficult to predict (Bai et al., 2026). In addition, since LLM creators currently subsidize users (e.g., charging subscription fees that do not cover costs), users may be misled about the true cost of their GenAI usage. Another challenge is that costs like inference can be volatile, even changing dramatically (Koning et al., 2026). Conversely, lean organizing using GenAI in internal workflows can reduce costs (Blank, 2026; Kim & Koning, 2026). Overall, the strategic implication for new GenAI firms is to pay *careful attention to unit economics*.

Third, GenAI often has *elusive learning economies*. By learning economies, we mean the improvement of knowledge or performance from usage. This contrasts with predictive AI (machine learning) with its self-reinforcing

algorithmic cycle that can become an engine of competitive advantage (Zott & Amit, 2026). The more a machine learning algorithm is used, the more data it generates, and the better performing it becomes, thereby attracting more users and so on. This path-dependent improvement has, for example, made the ever-increasing performance of Google's search engine virtually impossible for rivals like Bing to replicate. In contrast, the learning economies of GenAI are not automatic, especially given the complexity and high cost of re-training LLMs with new data and the often weak signals of whether GenAI results are useful. That said, traditional approaches to learning like A/B testing of product features can help. Creating profiles for users such as preferred prompts can leverage learning by adding context that can accelerate inference. Retrieval augmented generation (RAG) can also leverage learning by adding context from relevant databases in real-time to improve prompts rather than relying solely on static LLMs. User feedback about the quality of GenAI results can be accumulated and used to retrain LLMs. Yet, these approaches have limits. For example, AI agents can become confused by too much context (i.e., so-called "context rot"). Thus, a continually self-reinforcing cycle of learning that builds competitive advantage is not straightforward. Rather, LLMs are like new employees who lack accumulated knowledge and are not up-to-date on current information yet confidently answer every question (Amazon Web Services, 2026). The strategic implication for new GenAI firms is to *actively cultivate learning economies* rather than assume they will just appear.

Fourth, *rapid but partially predictable innovation* is likely to continue (Dey, 2026; Stanford HAI, 2026; Stokel-Walker, 2026). LLMs are increasingly intelligent while the unit price of intelligence and unit cost of inference are dropping. Simultaneously, users are demanding and using more intelligence. Thus, there is a rapidly advancing but also somewhat predictable technology trajectory of (1) improved functionality with new generations of LLMs and (2) lower per unit costs of both inference and older LLMs. In addition, open-weight LLMs are now highly capable and LLMs are becoming more diverse. Even disruptions in GenAI's trajectory are largely foreseeable although their exact timing is less so. For example, autonomous agents operating on top of LLMs have quickly moved to the forefront of GenAI (Dey, 2026). World Models (WMs) with their distinctive three-dimensional architecture, unique training data, deterministic outcomes, and new applications (e.g., autonomous vehicles, building design, video games) are next. The strategic implication for new GenAI firms is the *value of foresight* to anticipate these innovations and potentially exploit them.

Finally, *hallucination* remains a bottleneck. By sometimes creating unreliable outcomes, GenAI contrasts with most technological waves except perhaps quantum computing. LLMs typically generate plausible and well-structured content (often with a highly confident tone) that may be out of date, based on unreliable sources, or factually wrong (e.g., McCracken, 2026; Pavlus, 2025; Yoffie et al., 2024). Catching these errors is often costly, time-consuming, and limits the productivity gains of GenAI. Hallucination can be mitigated such as by requiring LLMs to provide sources for their results, triangulating results from multiple LLMs, and using RAG technology (above) (Thomke et al., 2025). Formal logic proofs and automated reasoning for GenAI output are on the horizon (Pavlus, 2025). Yet while these and other approaches can reduce hallucination, it is inherent in the probabilistic nature of LLMs. For a few tasks like software coding, hallucination is modest. For most new GenAI firms, the strategic implication is to *actively mitigate hallucination*, particularly in high-stakes settings such as health care, banking, and energy infrastructure where errors are highly consequential.

In summary, these features of GenAI have strategic implications for new GenAI firms that influence why some strategies are effective while others are not. In addition, while not unique to GenAI, the upstream dominance of LLM providers and inference platforms has similarities with the Microsoft and Intel (i.e., Wintel) dominance in the personal computing (PC) era and suggests a similar strategic implication of *differentiation via complementary resources*.

3 | STRATEGIES FOR NEW GENAI FIRMS

The size and range of GenAI opportunities have fueled the formation of numerous new GenAI firms (Kim & Koning, 2026; Pitchbook & National Venture Capital Association, 2026). Yet, many of these new firms are thin wrappers around well-known LLMs. This strategy may work for initial revenue growth but offers little long-term value. Thus, effective strategies are unlikely to emerge from relying on the same LLMs trained on the same data available to all.



Rather as per above, combinations of process-based strategies and resource-based strategies are more likely to be winning strategies. That is, many new GenAI firms face high uncertainty in nascent or disrupted markets that suggests the relevance of the Process Perspective as a strategic logic (Eisenhardt et al., 2026). At the same time, GenAI is changing with some predictability that suggests the relevance of the Resource-Based View as a strategic logic (Barney, 1991). We describe these advantages and ground them in illustrative new GenAI firms. Table 2 summarizes the five illustrative firms. Table 3 summarizes the strategic advantages and their illustrations in these firms.

4 | PROCESS-BASED ADVANTAGES

The Process Perspective (PP) is a theoretical lens on entrepreneurial success that builds on multiple process studies of how entrepreneurs succeed (or not) in high uncertainty markets (e.g., Bingham et al., 2019; Cohen et al., 2019; Eisenhardt et al., 2026; Ozcan & Eisenhardt, 2009). It uses the motivating analogy of the “entrepreneur as

TABLE 2 Illustrative new GenAI firms.

Firm	Founding	Founders	Domain	Initial product	Initial revenue	Initial users
Sierra AI	2023 US West	Two senior corporate executives (CRM, software engineering)	Enterprise sales and customer support	Product platform for users to build their own custom AI agents	Negotiated outcome-based pricing	Large enterprises, including in regulated markets like insurance, banking and health care
Moon.ai ^a	2023 US Midwest	Two recent science graduates. One executive	Enterprise software	AI super agents for routine tasks in enterprise software (e.g., documentation, code review, debugging)	Negotiated pricing	Large enterprises, especially with complex legacy IT systems
OpenEvidence	2022 US East	Serial entrepreneur. Research scientist	Medicine	GenAI platform to access medical research	3rd party advertising	Individual medical professionals. Later, hospitals and health systems
Anysphere (Cursor)	2022 US West	Four young engineers	Software engineering	GenAI coding assistant for in-line coding and debugging	Freemium	Individual software engineers. Later, corporate engineering teams and organizations
Counsel AI (Harvey)	2022 US West	Law firm attorney Research Scientist	Law	GenAI platform for legal tasks (e.g., document writing, document review, case law retrieval)	Seat-based pricing + sometimes revenue splits	Law firms. Corporate legal teams

^aA pseudonym.

TABLE 3 Strategic advantages of new GenAI firms: Descriptions and illustrations.

Advantage	Strategic logic	Description	How advantage is created	Illustrative firm
Well-designed business model	Process-based	- Business model with promising unit profitability and interdependent elements that are difficult to imitate. - Fewer errors from business-model misfits. - Better strategic understanding	- Address entire business model, including its major elements. - Learn, simply theorize and imagine new possibilities within each element. - Pay attention to end goal of unit profitability (incl. product-market fit and agent-outcome fit)	Sierra AI: - Outcome-based pricing from first principles. - Simple rules for customers and AI-agent jobs. - Complete business model (e.g., customers, agents, platform, sales) that fits together. - Paused growth to complete business model
Well-designed product innovation process		- Successful products on time w/product-market fit and promising product profitability. - Repeatable, semi-structured process balancing efficiency w/flexibility that is difficult to imitate. - Coevolving product innovation process w/GenAI advances	- Address entire product innovation process, including its major elements. - Learn, simply theorize and imagine new possibilities within each element. - Pay attention to end goals of product-market fit, schedule, and product profitability (incl. architecting lower compute and memory costs)	Moon.ai: - Vicarious learning from experienced advisors. - Modular teams “own” specific super-agents. - Semi-structures like simple rules (e.g., “only agents can code”), clear responsibilities for team leaders and CTO and high communication. - Quarterly reconfiguration of teams. - Ongoing CTO-led experimentation
Superior data	Resource-based	- Superior data that are valuable, rare and difficult to imitate. - Bolstered by mitigating hallucination. - Bolstered by cultivating learning economies	- Relevant and accurate data (Valuable). - Unique, often proprietary data (Rare). - Superior data become more valuable, rare and difficult to imitate w/learning economies (Inimitable)	OpenEvidence: - Contractual, symbiotic partnerships w/top-tier and trusted medical research journals (JAMA, NEJM) to access articles. - Mitigates hallucination by grounding results in actual articles plus RAG and other tactics. - Cultivates learning economies w/proprietary user feedback data
Efficient orchestration		Orchestration layer of evolving ensemble of LLMs and switching algorithms to match LLM and task is valuable, rare and difficult to imitate.	- Faster, more accurate, higher performance in focal task (Valuable). - Composing ensembles and switching algorithms is difficult (Rare).	Cursor: - Closed, custom and open LLMs. - Orchestration layer of evolving LLM ensembles and switching algorithms



TABLE 3 (Continued)

Advantage	Strategic logic	Description	How advantage is created	Illustrative firm
		• Exploits low switching costs. • Bolstered by causal ambiguity	• Causal ambiguity of private switching algorithms and ensemble composition (Inimitable)	for optimal tradeoff of cost vs. speed vs. quality. • Less dependence on any one LLM creator. • Foresight to anticipate GenAI technology trajectory (e.g., agents, open-weight LLMs)
Complementary resources		• Integration of complementary resources w/GenAI is valuable, rare and difficult to imitate. • Resources can be physical, intangible (e.g., domain expertise, IP) or organizational (e.g., R&D skills). • Bolstered by path dependent integration and potential user lock-in.	• Integrating complementary resources w/GenAI improves results (Valuable). • Integrating complementary resources w/GenAI well is challenging (Rare). • Path dependent integration and user lock-in, especially if applications embedded in user workflows (Inimitable)	Harvey: • Founders are an attorney and a research scientist. • Attorneys add legal expertise (e.g., regulatory requirements, tacit knowledge, legal contacts). • Attorneys well-integrated w/technical Gen AI staff on sales and product teams. • Head start over rivals on resource integration. • Often embed Harvey in client workflows

preschooler” and argues that successful entrepreneurs in high-uncertainty markets focus on accomplishing their primary strategic task(s) like designing a profitable business model (McDonald & Eisenhardt, 2020), orchestrating a coherent ecosystem (Hannah & Eisenhardt, 2018), and forming an effective strategy (Ott & Eisenhardt, 2020) to create advantage and succeed.

Process-based (PP) strategies fit with the *high uncertainty* of nascent and disrupted markets like those faced by many (or even all) new GenAI firms (Furr & Eisenhardt, 2021). Here, entrepreneurs lack a clear industry structure of rivals, buyers, and suppliers (Gao et al., 2026; Rindova & Fombrun, 2001). Product definitions are often vague or contested (Anthony et al., 2016; Zuzul & Tripsas, 2020). Demand is frequently ambiguous (Gao & McDonald, 2022). These markets are changing unpredictably and often at high velocity (Eisenhardt, 1989a).

PP argues that successful entrepreneurs navigate these markets by focusing on accomplishing their primary strategic task(s) by (1) stepping through all major elements of the task to complete the entire task, (2) iterating among learning, simple theorizing, and imagining new possibilities within each element to address its critical uncertainties, and (3) paying attention to the end goal (Eisenhardt et al., 2026). Advantage is created when there is an accomplished strategic task like an effective strategy or business model.

Two primary strategic tasks are particularly relevant for new GenAI firms. One is designing an *advantageous business model*. GenAI may superficially resemble software, but its distinctive technical features like LLMs and challenging unit economics suggest that effective GenAI business models are often not replicas of software business models or other familiar ones. Thus, while designing a business model is central to the success of most (or all) new firms

(e.g., Bocken et al., 2025; McDonald & Eisenhardt, 2020; Snihur & Zott, 2020; Tidhar et al., 2025), it is particularly critical for new GenAI firms because of GenAI's distinctive technical features, often high (and volatile) variable costs, and overall challenging unit economics.

A second primary strategic task is designing an *advantageous product innovation process*. Stunning advances in GenAI have dramatically disrupted software coding in new GenAI firms (and elsewhere) that is central to their product innovation process (Orosz, 2026; Thompson, 2026). Thus, while creating innovative products matters for most new firms (e.g., Bremner & Eisenhardt, 2022; Kirtley & O'Mahony, 2023), it is a primary strategic task for new GenAI firms because this task has been massively disrupted by the very GenAI innovations that underlie the entrepreneurial opportunities of these firms.

4.1 | Designing an advantageous business model

Designing an advantageous business model is a primary strategic task for new GenAI firms. Following others (Bocken et al., 2025; McDonald & Eisenhardt, 2020), we define a business model as the system of interconnected activities performed by the firm (and sometimes users and partners) to create and capture value. Consistent with PP, designing an advantageous business model involves (1) addressing the entire business model including its major elements, (2) learning, simple theorizing, and imagining new possibilities about each element, and (3) paying attention to unit profitability (including product-market fit), not just revenue and growth (Eisenhardt et al., 2026; Ott & Eisenhardt, 2020; Tidhar et al., 2025). The result is a well-designed business model with (1) promising unit profitability, (2) fewer errors from business-model misfits, and (3) better strategic understanding by the entrepreneurs (Hallen et al., 2026).

Sierra AI illustrates. Co-founded by two highly experienced executives in 2023, Sierra provides AI agents for the enterprise. Specifically, Sierra offers a platform that enables enterprises to construct their own unique AI agents for many facets of customer experience. Sierra sells to large firms that typically have legacy IT systems, including some that operate in complex, regulated markets like health care and banking. Sierra's customers include more than 40% of the Fortune 50 firms. While many large firms struggle with effective GenAI implementation (e.g., Challapally et al., 2025; Kilminster et al., 2026), Sierra's customers seem to be succeeding despite often operating in demanding situations.

A key to Sierra's advantage is its *business model* design process. These entrepreneurs began by returning to “first principles” to theorize how the features of AI agents might affect their business model. While these agents may be more intelligent in the future, they are now especially effective for well-structured, repetitive tasks. These tasks typically have clear outcome metrics—i.e., the agent either resolves the problem or not, gets the sale or not and so on.

The critical insight was that clear outcome metrics fit with *outcome-based pricing*. In effect, these entrepreneurs used agency theory reasoning with the enterprise as the “principal” and Sierra's AI agent as the “agent.” Given clear outcome metrics, agency theory predicts that it is efficient for the enterprise (principal) to pay Sierra and its AI agent (agent) on the outcome (e.g., piece-rate pay) (Eisenhardt, 1989b). Outcome-based pricing is attractive to enterprises because they only pay for what they get. By contrast, the usual software business models like per-seat and usage pricing are not well-correlated with the preferred outcomes of enterprises. Outcome-based pricing works well for vendors like Sierra because they can charge a premium (e.g., value-based pricing) given that they carry the outcome risk, another clear prediction from agency theory (Eisenhardt, 1985). With outcome-based pricing, Sierra is paid by its customers only when its AI agent autonomously resolves an issue. When an AI agent has a successful outcome without human intervention, Sierra earns a pre-negotiated fee while its customer avoids substantial expense (often \$10–20 per task). But if the AI agent must escalate to a human, the customer pays nothing. Thus, Sierra focuses on *agent-outcome fit*, not just product-market fit.

Given outcome-based pricing, Sierra's entrepreneurs readily theorized their preferred customers and jobs for AI agents including simple rules for these elements: (1) *large customers* with high transaction volumes that can cover the



upfront negotiation costs and generate ongoing revenue and (2) *routine jobs* with stable structures in well-documented workflows and outcomes such that AI agents can reliably accomplish them and cost savings can be accurately determined.

In contrast, a difficult element to design was the *agent platform*. This product platform enables enterprises to construct their own AI agents. Since Sierra earns no revenue until an AI agent accomplishes its jobs, rapid implementation of the platform by customers is highly desirable. At the same time, the platform needs to be flexible for a variety of jobs (albeit well-structured ones) that may require agents to take actions in the physical world (e.g., arrange shipments), not just respond to queries like a chatbot. For example, Sierra's agent platform had to work with a global bank to ensure that its AI agents were multi-lingual and with a roadside assistance company to give its AI agents sophisticated voice recognition for use on noisy highways. Further, the platform needs to drive down expensive costs like inference to generate favorable unit economics. This combination of speed, complexity and cost at the cutting-edge AI agent technology was difficult to design.

Another major business model element was *sales and customer support*. Since Sierra earns nothing unless its AI agent accomplishes its job, these entrepreneurs recognized the importance of ensuring that its AI agents would quickly succeed. An easily implemented platform design (above) helped. These entrepreneurs also added “forward deployed engineers.” FDEs work with customers to ensure that the AI agents are effective and spot opportunities for new applications. Finally, given the high interdependence among business model elements, Sierra paused growth which added time to complete the entire business model and ensure that the elements fit together (Hallen et al., 2026; Tidhar et al., 2025).

Overall, Sierra followed a superior business model design process by (1) addressing the key elements of the entire business model, (2) learning, simple theorizing and imagining within each element, (3) paying attention to the end goal of favorable unit economics (not just revenue and growth), and (4) pausing growth until achieving fit among these elements. Consistent with PP, the result was an advantageous business model (i.e., valuable, rare, inimitable given interdependent elements) with promising unit profitability that was core to Sierra's early success.

4.2 | Designing an advantageous product innovation process

Employees use GenAI in myriad ways like writing emails, summarizing information, and brainstorming. Significant variation occurs within firms, across firms within the same industry and across industries (e.g., Jacobides & Ma, 2026) as well as across skill levels and occupations (e.g., Brynjolfsson et al., 2025). Amid this complex landscape, software coding stands out as an unusually favorable application (Orosz, 2026; Thompson, 2026). Its readily verified outputs, short feedback loops, widely available training data, and digital environment make coding an ideal GenAI application (McCracken, 2026). Indeed, Cursor, Claude Code, and other GenAI coding assistants have disrupted how individual software engineers code (Orosz, 2026). For example, AI agents can now write the entire code base of a new product. Since individual software engineers can manage multiple AI agents that work 24×7, this transition from human coders to AI agents has the potential for extraordinary time compression and cost efficiency.

The massive disruption in software engineering is particularly relevant for new GenAI firms. Coding is at the heart of their creating innovative products, making time compression and cost efficiency both attractive and essential. Moreover, new GenAI firms hold a coding advantage over established firms. These GenAI-natives can code on a clean slate (i.e., greenfield coding). This is much easier than updating legacy systems that are so prevalent in established firms (i.e., brownfield coding) (Thompson, 2026). New GenAI firms can also right-size and right-skill their employees from the start (Kim & Koning, 2026). In contrast, established firms have legacy systems that complicate exploiting GenAI.

While the disruption of software engineering creates opportunities for new GenAI firms, it remains highly uncertain how to design an effective product innovation process at the firm level. Many new GenAI firms allow their software engineers to do individual experiments. This “anything goes” approach is often fun for individual software

engineers (Brown & Eisenhardt, 1997). Yet, it also limits the production of useful code, creates problems like inconsistent and buggy code bases, can swamp software engineers with too much information, and more generally limits the flow of innovative products (Blank, 2026; Bremner & Eisenhardt, 2022; Brown & Eisenhardt, 1997).

Consistent with PP, designing an advantageous product innovation process involves (1) designing the entire process including its key elements, (2) learning, simple theorizing and imagining new possibilities within each element, and (3) paying attention to the end goals of product-market fit, schedule, and product profitability. The result is likely to be a flow of successful products that fit the market, are on time, and have promising profitability.

Moon.ai (a pseudonym) illustrates. Launched in 2023 by three founders, Moon builds AI “super-agents” (i.e., highly capable AI agents) for enterprise software. In large enterprises, software engineers work on complex, legacy systems that can be difficult to understand fully but that must operate with high reliability and at high transaction volumes. Moon's super-agents take on many of the most boring yet necessary jobs of enterprise software like code review, documentation, refactoring the code base and debugging. This leaves engineers in large enterprises with time for more valuable work that better fits their expertise. Moon's super-agents have excelled in benchmark tests, thus enabling the firm to attract high-status investors, top software talent, and enterprise customers sometimes over much larger firms like Anthropic.

A key to Moon's advantage is how these entrepreneurs designed their *product innovation process*. They began designing their product (i.e., super-agents) with a small team of software engineers who worked closely together. This process initially worked well. Yet over time, they realized that Moon needed a more repeatable and structured process for creating a flow of successful products. Since these entrepreneurs did not know much about firm-level product innovation processes, they chose to learn vicariously from several experienced advisors. While GenAI has disrupted software engineering, they theorized that not everything had changed.

Building on this learning, the entrepreneurs theorized about how GenAI changes (or not) product innovation. They realized that traditional ideas like exploiting the modularity of software and assigning software teams to modules still made sense (Baldwin & Clark, 2000). So, they divided their software engineers into teams with each team responsible for a specific super-agent. They now have about 30 software engineers in six teams. These entrepreneurs also realized that paying attention to critical outcomes by assigning responsibility for them still mattered in the age of GenAI (Brown & Eisenhardt, 1997, 1998; Clark & Fujimoto, 1991). So, the CTO took responsibility for setting product priorities and profitability while team leaders were given responsibility for meeting schedules and assigning tasks. Given high uncertainty, there was also flexibility. For example, while team leaders became responsible for assigning tasks to their teams, they could flexibly assign those tasks as they wished. Yet, while there was flexibility in how the teams worked, they had to adhere to *simple rules* like “only agents can code.” In effect, these founders used contingency and complexity theory reasoning to implement semi-structures that balance efficiency (e.g., clear roles, simple rules) with flexibility to fit high uncertainty markets (Bingham & Eisenhardt, 2011; Davis et al., 2009; Reineke et al., 2025).

To tackle the remaining elements and refine others, the CTO coordinated experimentation (both serial and parallel) among the software engineers to address uncertainties like ideal team size, optimal # of coding agents per engineer, and ideal new tasks for software engineers like architecting lower compute and memory costs that are now critical in the age of GenAI. The CTO also elaborated the team-assignment element of the product innovation process by keeping team composition fixed while adding a simple rule for quarterly team reconfiguration.

Finally, the entrepreneurs ensured that high communication was an essential element in their product innovation process, echoing their early success and fitting traditional product innovation (e.g., Clark & Fujimoto, 1991). For example, they specified required days in the office, use of Slack and Notion, and office seating that mingled engineering and marketing. They also imagined a novel interview approach that reinforced communication. This 2-day interview requires potential hires to work on real coding problems with Moon's software engineers. It provides valuable information about whether the candidate is highly skilled, communicates well, and can cooperatively get work done at an exceptional level. This approach contrasts with the initial emphasis on traditional hiring criteria (e.g., elite education) that proved ineffective.



Overall, Moon designed an advantageous product innovation process by (1) addressing the major elements of the entire process, (2) learning, simple theorizing and imagining within each element, and (3) paying attention to the end goals of product-market fit, schedule, and product profitability. The result was an advantageous product innovation process (i.e., valuable, rare, difficult to imitate given interdependence and causal ambiguity) that yielded a flow of successful new products that was core to Moon's early success. Further, Moon has continued experimentation to coevolve their product innovation process with GenAI advances.

5 | RESOURCE-BASED ADVANTAGES

Resource-based (RBV) strategies fit with moderately uncertain markets (Furr & Eisenhardt, 2021; Ross & Lieberman, 2026). Here, entrepreneurs have the time and foresight to build resources and update them as the market unfolds. In the case of GenAI, a somewhat predictable technological trajectory enables entrepreneurs to anticipate and build advantageous resources, especially as their business model and product innovation process begin to coalesce. Three sources of resource-based advantage are particularly relevant for new GenAI firms.

First, since GenAI is a digital information technology, *superior data* are often a valuable, rare and inimitable source of resource-based advantage. Mitigating hallucination and cultivating learning economies further bolsters the advantage of superior data. Second, given low LLM switching costs and high (and volatile) variable costs like inference, *efficient orchestration* of LLMs and other foundation models (i.e., optimizing the tradeoff of cost versus speed versus quality of results) also provides a resource-based advantage. Third, when upstream technologies limit downstream differentiation as in the GenAI industry, *complementary resources* that fit with specific vertical markets and are integrated with GenAI can offer a resource-based advantage.

5.1 | Advantage from superior data

Superior data offer a source of resource-based advantage for new GenAI firms. By *superior data*, we mean data that are valuable (e.g., relevant, accurate), rare (e.g., unique), and difficult to imitate (Barney, 1991; Dinçkol et al., 2026). By fitting with the focal GenAI application, relevant data create value by improving the quality of GenAI results as do accurate data (e.g., edge cases). Training on valuable data also decreases hallucination, a major impediment to high-quality GenAI results. When these data are also rare and difficult-to-imitate (e.g., proprietary, bolstered by cultivating learning economies over time), they further enhance competitive advantage.

An example is **OpenEvidence**. Founded by a serial entrepreneur and research scientist, OpenEvidence provides a GenAI platform for medical advice. This platform is free and available only to medical professionals, not patients. It relies on the well-known advertising revenue model that WebMD used so successfully to earn revenue from medical device and pharmaceutical companies via precisely targeted ads (Tidhar & Eisenhardt, 2020). Since its founding in 2022, OpenEvidence has built an increasingly large user base among doctors, nurses, and other medical professionals, including currently roughly 750,000 verified clinicians, over 40% of all U.S. physicians logging in daily, and approximately 27 million clinical queries per month (as of April 2026) (Perlo, 2026).

These entrepreneurs initially self-funded, which allowed focus on the business model design process without pressure from investors to grow. They also joined the Mayo Clinic's accelerator that provided rare access to practicing physicians for exploring product-market fit and to Mayo's research arm for studying GenAI performance. OpenEvidence was initially seen as "ChatGPT for doctors." Yet, these entrepreneurs soon realized that LLMs were too inaccurate for medical protocols to treat patients. Instead, they recognized that physicians value high-quality research articles (especially for patients with multiple conditions) and that LLMs trained on superior data could provide this information. Their analogy became "Google search" and then "Bloomberg Terminal." This latter analogy was especially apt given that Bloomberg Terminal provides accurate, highly valued, and trusted information (not advice) to finance professionals.

A key advantage is *superior data*. These entrepreneurs began with an OpenAI chatbot as the base LLM. They then trained their LLM with research articles from top-tier, peer-reviewed medical journals like Journal of the American Medical Association (JAMA) and New England Journal of Medicine. They later created their own LLMs and added other sources of superior data such as guidelines from the American College of Cardiology and the National Comprehensive Cancer Network. Superior data help to ensure that OpenEvidence delivers high-quality results. These data also enhance the trust that medical professionals have in the accuracy of results because they are easily understood signals of quality that these professionals recognize.

In addition to high-quality and trusted results, superior data also lessen *hallucination* which is particularly consequential in medical applications. OpenEvidence also added other approaches to mitigate hallucination such as by programming its models to reject low-confidence answers, providing high transparency with direct citations to actual research articles (rather than medical advice), and relying on RAG (retrieval-augmented generation) that uses highly relevant, updated information to provide better prompt context before engaging with LLMs. Finally, the venture cultivates *learning economies* such as by enhancing prompt context and more importantly, accumulating proprietary feedback data from its many users and then periodically re-training its models, thus building a self-reinforcing feedback loop that bolsters advantage.

OpenEvidence also illustrates the significance of *partnerships* with established organizations that own the superior data that new firms typically lack (Dinçkol et al., 2026). In contrast, large incumbents in verticals like health care, banking and insurance often have proprietary data that are impossible for others to obtain for training LLMs without content partnerships (Ozcan et al., 2026). So, while the appropriability regime for data is weak compared with other IP regimes (e.g., software copyrights, patents), proprietary data that are contractually secured can create a valuable, rare and inimitable advantage (Teece, 1986). Importantly, these partnerships are also symbiotic (Dinçkol et al., 2026; Ozcan et al., 2026). They benefit both the incumbent firms and new GenAI firms by providing revenue for the former (e.g., license fees, website traffic) and advantageous business models (e.g., superior data) for the latter.

In the case of OpenEvidence, its contractual partnerships with top-tier medical research journals for their unique, proprietary data produced valuable, trusted GenAI results that others could not match. These results spurred viral adoption as individual doctors increasingly used OpenEvidence and spread the word to others. In turn, these journals benefitted from licensing fees and increased website traffic. Although others ultimately licensed these data, OpenEvidence furthered its advantage with early viral adoption and cumulative learning economies.

Overall, superior data are advantageous for new GenAI firms because they (1) lead to more valuable (i.e., relevant and accurate) results, (2) achieved with less hallucination, and (3) are typically rare (e.g., unique) and difficult to imitate (e.g., proprietary, bolstered by deliberately cultivating learning economies that accumulate over time).

5.2 | Advantage from efficient orchestration

Efficient orchestration of LLMs (and potentially other models) offers a second source of resource-based advantage. A general-purpose LLM is broadly useful for many applications, but not optimized for any specific application. In contrast, a custom LLM is designed to fit a specific application and so is often right-focused. By *right-focused* we mean a model that is adapted to and trained on relevant content for the application (e.g., coding vs. customer service). A right-focused LLM often starts with a general-purpose LLM, then is “distilled” into a more focused LLM, and finally trained on application-relevant data. A custom LLM is also often right-sized. By *right-sized* we mean a model that is designed to be no larger than necessary. In contrast, a general-purpose model with billions of parameters is expensive to train and run. While such models can produce amazing results, they are often bigger and more capable than many applications require. Thus, a custom LLM can often produce better and faster results with lower training and inference costs than a general-purpose one.



Efficient orchestration of LLMs (and potentially other models) has emerged to exploit *low switching costs* among increasingly diverse LLMs. Orchestration involves creating an “orchestration layer” of an evolving ensemble or set of LLMs. These LLMs can be custom, general-purpose or both. Orchestration occurs by algorithmically switching among LLMs in the ensemble to choose the best match for the task (e.g., matching task complexity with LLM complexity). Thus, efficient orchestration optimizes the tradeoff of cost versus speed versus quality of results. In doing so, it improves *unit economics* (e.g., giving higher-paying users higher quality outcomes). Overall, efficient orchestration creates value with an evolving ensemble of LLMs and switching algorithms that choose the most favorable one. Yet orchestration is often difficult to do well and so rare. These private ensembles and algorithms are also interdependent and causally ambiguous, making them difficult to imitate.

An example is Anysphere's **Cursor**. Four young engineers founded Anysphere in 2022 and launched Cursor in 2023. Cursor is a leading GenAI coding assistant for individual software engineers. It is an interactive, human-in-the-loop tool for serious software development, not casual vibe coding (Dey, 2026). Its aim is superior coding (e.g., better accuracy, speed and overall performance) for individual software engineers. Cursor excels at in-line coding and debugging. It has a free introductory version and several paid tiers. It has been widely adopted by millions of software engineers in many of the world's most sophisticated engineering firms. Cursor is also among the fastest products ever to reach \$100M annual recurring revenue.

A key advantage is **efficient orchestration**. These entrepreneurs began with GenAI for mechanical engineering applications. Within a few months, they realized that mechanical engineering is a poor application of LLMs for several reasons, like being three-dimensional and physics-based rather than text-based where LLMs excel. They then pivoted to software engineering applications (Cursor) which fit much better with both LLMs and their own skills. These entrepreneurs initially planned to improve Cursor by riding the wave of increasing functionality in general-purpose LLMs like Claude. When they realized that software has unique features that can be exploited by custom LLMs, they detoured into creating custom LLMs. For example, since the next lines of software code are much easier to predict than next words of text, they created a custom LLM that can autocomplete code that could take 10 to even 30 min to code, a predictive impossibility with text. To better create custom LLMs, they hired model specialists to complement their software engineers. Importantly, these entrepreneurs have attributed their most significant product breakthroughs to their custom LLMs.

Over time, Cursor's entrepreneurs artfully developed an orchestration layer to switch algorithmically among its ensemble of LLMs (their own and others' LLMs) to better match the LLM with the task (e.g., matching LLM cost with task pricing). By orchestrating its ensemble of LLMs and evolving its composition, Cursor achieves its aim of superior coding (e.g., better accuracy, speed and overall performance for users) as well as better unit economics than general-purpose LLMs by lowering training and inference costs. Cursor also uses cost-cutting tactics from computer science (e.g., caches, open-weight models) to manage latency, memory, compute and other costs that are critical for the unit economics of GenAI. For example, Cursor used an open-weight LLM (Kimi) as its base LLM. This enabled model development at far lower cost than training a model from scratch or building on a proprietary LLM.

Overall, efficient orchestration is advantageous for new GenAI firms because it (1) creates value by improving user outcomes, (2) improves unit economics by exploiting low LLM switching costs to choose LLMs that optimize cost versus speed versus quality, (3) often reduces training and inference costs by using custom or open-weight/open-source LLMs, (4) is rare because of private, unique ensemble composition and switching algorithms, (5) is difficult to imitate given causal ambiguity, and (6) limits dependence on any one LLM creator.

Cursor also illustrates the classic strategic rivalry of “*best of breed*” versus “*best of platform*” in its competition with Anthropic's Claude Code. This rivalry refers to the best standalone application versus best application within a product platform for a given product category (e.g., Zoom vs. Microsoft Teams). A platform rival to Cursor, Claude Code is also a high-performing code assistant that relies on efficient orchestration (McCracken, 2026). On the one hand, Anthropic optimizes Claude Code using ensembles of its own LLMs. On the other hand, it also locks into its models, making Claude Code potentially less effective. Thus, Anthropic optimizes Claude Code within its proprietary LLM models, exploits its unique insights into these models, and attempts to be the “best of platform” solution.

Conversely, Cursor relies on a flexible choice of LLMs from its ensemble and attempts to be the “best of breed” solution.¹

Finally, Cursor illustrates the *value of foresight* to anticipate technological innovations and potentially exploit them. Since their initial success with LLMs for individual software engineers, these entrepreneurs have anticipated GenAI's technology trajectory by early adoption of innovations like agents and open-weight LLMs. They have also extended their business model by adding enterprise sales with subscription pricing for large engineering organizations. Similarly, OpenEvidence has added agent-based capabilities for an expanded range of medical tasks and enterprise sales with subscription pricing to hospitals and health systems to complement their ad-based business model that targets individual professional users.

5.3 | Advantage from complementary resources

Complementary resources offer a third source of resource-based advantage. Complementary resources such as physical plant, intangibles (e.g., domain expertise in banking), and organizational capabilities (e.g., R&D, sales) can be useful and even necessary to gain the value of new technologies (Teece, 1986; Tripsas, 1997; Wu et al., 2014). Consistent with this argument, new GenAI firms that integrate complementary resources with GenAI can potentially gain a resource-based advantage. For example, new GenAI firms that serve verticals like law and insurance can combine domain expertise (e.g., workflow, regulatory requirements, tacit knowledge) with GenAI technology and expertise to create a resource-based advantage.

An example is **Harvey**. Harvey is a GenAI platform for the legal profession, including law firms and corporate legal teams. Law fits particularly well with LLMs because it is highly textual with massive troves of written documents and with many complicated (and even byzantine) yet repeatable workflows. Founded in 2022, Counsel AI developed Harvey and named it after the lead character (Harvey Specter) in the legal drama *Suits*. Harvey converts simple legal tasks (e.g., M&A documents, case law retrieval, estate planning drafts, private equity solicitations) into modules that are chained together to accomplish complex tasks. Harvey was among the first firms to use OpenAI and runs on secure inference infrastructure (e.g., Microsoft's Azure) that fits with high-privacy applications like law. It uses several revenue models including seat-based pricing and revenue splits for high-efficiency, low-margin legal tasks. Harvey is used by the majority of the 100 largest US law firms and major corporations like Comcast and Bridgewater. Harvey is an early leader in GenAI for the legal profession with about \$200M in annual recurring revenue (Harvey, 2026; Martin & Shrivastava, 2026).

A key advantage is **complementary resources**. Harvey combines complementary legal resources (e.g., knowledge of case law, tacit legal knowledge, networks of contacts) with its GenAI resources. This complementarity began with its founding team. Its two founders were roommates—one a law firm attorney, the other a research scientist from DeepMind. They replicated this complementarity with their employees. About half are software engineers and scientists while the other half are attorneys, often from Big Law (i.e., major law firms and large corporate legal teams). Since Harvey targets Big Law customers because of their scale and ability to pay, these attorneys are particularly useful.

A key factor in Harvey's success is the integration of these attorneys with GenAI professionals in sales and product teams. Given the distinctive cultures of “suits” and “techies,” this integration can be time-consuming. Yet, it is highly relevant because attorneys are very effective in sales given their tacit knowledge of the legal profession and contact networks. They speak the “language” of the legal profession, fit with its cultural norms, and understand its ethics and the importance of relationships among attorneys. These attorneys also add to product teams because they understand the byzantine workflows and nuanced hierarchies of lawyers. They can pinpoint the structured yet time-consuming tasks for which GenAI excels. The combination of legal and GenAI resources is especially effective in prioritizing product features. For example, employees who are also attorneys understand the nuanced chain of downward delegation of legal tasks and upward chain of review that is common in the legal profession. This tacit



knowledge enables them to work well with Harvey's technical team to figure out where best to apply GenAI and how to mitigate hallucinations given the cultural nuances of extensive and layered legal reviews. It also helps to embed Harvey in client workflows that can create lock-in.

Overall, complementary resources are advantageous for new GenAI firms because they (1) lead to more valuable (i.e., relevant and accurate) results, (2) are rare because of the challenge of integrating these resources with GenAI technology and professionals, and (3) difficult to imitate given the path dependence of resource integration and potential client lock-in.

Finally, Harvey illustrates *multiple sources of resource-based advantages*. In addition to its advantage from complementary legal resources, Harvey also has the advantage of *superior data*. In addition to training its LLMs on case law, Harvey trains its LLMs on its own proprietary data and high-quality legal data from sources like EDGAR and its LexisNexis partnership as well as its own proprietary synthetic data in legal areas for which there is little actual data. Finally, Harvey has built an advantage around *efficient orchestration* such as an evolving ensemble of LLMs from Anthropic, OpenAI, and others.

Similarly, the other firms—Sierra, Moon.ai, OpenEvidence, and Anysphere (Cursor)—also have multiple resource-based advantages like (1) complementary resources in their vertical domains of enterprise CRM, enterprise software, medicine, and individual coding, respectively, (2) efficient orchestration (especially Cursor), and (3) superior data (especially OpenEvidence).² They often also have process-based advantages like a process leading to well-designed business models. Overall, multiple advantages amplify the defensibility of competitive advantage.

6 | DISCUSSION

We began by observing a tsunami of new GenAI firms (e.g., Kim & Koning, 2026; Pitchbook & National Venture Capital Association, 2026) with many offering little strategic differentiation. These firms face a difficult situation (also von Krogh et al., 2026). On the one hand, they must accomplish primary strategic tasks like designing an advantageous business model while navigating highly uncertain markets. On the other hand, they need sufficient foresight to develop strategically advantageous resources that fit with a moderately predictable technology trajectory.

Our primary contribution is a theoretical framework for how GenAI entrepreneurs create effective strategies. Here, we identify major sources of advantage. An unexpected insight is that, although GenAI is clearly game-changing, well-known strategic logics (e.g., RBV, PP) and theories (e.g., agency, contingency, and complexity theories) are surprisingly still useful. We also deepen understanding of these strategic logics with concrete illustrations from new GenAI firms that have achieved early success. Thus, our framework is both theoretically reasoned and empirically grounded in actual firms.

More broadly, we contribute a better understanding of strategy and technology by asking what is unique about entrepreneurship in the age of GenAI. In response, we add a new typology of GenAI features and their strategic implications, insights into technology strategy when firms face upstream technical dominance and downstream rivalry, and a cross-technology comparison of the strategy of new GenAI firms versus other technology waves like software and predictive AI.

6.1 | Effective strategies for new GenAI firms

Our primary contribution is a theoretical framework for how GenAI entrepreneurs create effective strategies, including major sources of advantage. These entrepreneurs follow a temporal sequence of (1) initially addressing high market uncertainty and process-based advantages (i.e., well-designed business models and product innovation processes), (2) adding resource-based advantages (i.e., superior data, efficient orchestration, complementary resources) to keep pace with a moderately predictable technology trajectory, and (3) relying on both logics as the

technology advances. Thus, the Process Perspective and Resource-Based View are *complementary* (not competing) strategic logics that together form complete strategies.

A novel contribution is the *varied generalizability* of these advantages. Some generalize across many new firms. For example, while designing an advantageous business model is critical for new GenAI firms because of challenging unit economics, it is also important for most (or even all) new firms. Similarly, complementary resources like expertise in a vertical market are widely advantageous. In contrast, other advantages relate closely to GenAI. For example, designing an advantageous product innovation process is particularly relevant for new GenAI firms because software coding is massively upended by GenAI and yet is at the heart of this process. Similarly, efficient orchestration is particularly advantageous for new GenAI firms because it exploits low switching costs among models and enhances unit economics. Finally, superior data moderately generalizes to many new digitally-based firms.

We also contribute the novel concept of *bifurcated uncertainty*—i.e., high uncertainty of nascent or disrupted markets and moderate uncertainty of a somewhat predictable technology trajectory. In contrast, much entrepreneurship research focuses on navigating high uncertainty (e.g., Andries et al., 2013; Camuffo et al., 2024; Hannah & Eisenhardt, 2018; Ott et al., 2017; Sarasvathy, 2024) but neglects the possibility of bifurcated uncertainty and accumulation of strategically valuable resources. Prior research frames (if at all) the transition from PP to RBV as a brief interval between the high uncertainty of new firms in nascent or disrupted markets to the moderate uncertainty of maturing firms in established markets (e.g., Furr & Eisenhardt, 2021; Tidhar et al., 2025). In contrast, bifurcated uncertainty lengthens this interval for new firms like GenAI ventures that face this uncertainty.

Overall, we contribute a novel theoretical framework for forming effective strategies in which we identify major sources of advantage. Here, PP and RBV are complementary strategic logics. While these may not be the only effective strategies for new GenAI firms (and the illustrative firms may ultimately fail), these strategies are likely to be among the effective ones. We also introduce bifurcated uncertainty that generalizes the framework to other new firms that face similar uncertainty. Finally, we crystallize the fundamental role of uncertainty in determining relevant strategic logic(s) (Davis et al., 2009; Ross & Lieberman, 2026).

6.2 | What is unique about entrepreneurship in the age of GenAI?

More broadly, we also contribute to the strategy and technology literatures by clarifying what is unique about entrepreneurship in the age of GenAI. First, we contribute a *novel typology* of major GenAI features and their strategic implications that shape which strategies are likely to be effective. These features include low LLM switching costs, high (and volatile) variable costs, elusive learning economies, rapidly evolving but somewhat predictable innovation, and hallucination. These features, in turn, have strategic implications for new GenAI firms such as the critical importance of unit profitable business models, switching among LLMs, and deliberately cultivating learning economies to develop superior data.

Second, we contribute a more precise *theory of technology strategy* when firm-level advantage is influenced by structural position in the industry stack or more formally, layered technical system. Specifically, when upstream technology dominance exists and downstream access is open to many, downstream firms should differentiate by skewing their strategies toward advantages that the dominant technologies cannot easily provide (also Kapoor & Raj, 2026). This advice is amplified when upstream technologies are evolving and are neither controlled nor easily appropriated by downstream firms, as in GenAI. For example, the structural position of new GenAI firms in the Applications layer favors differentiation with advantages like a well-designed business model, superior data, and complementary resources in vertical markets that rivals and upstream firms are less likely to replicate.

Third, we contribute a *cross-technology comparison* of the strategic implications of GenAI for new firms versus other technological waves (Table 4). Given GenAI's reliance on software, a key comparison is with *software*. Yet, new software firms have very low variable costs leading to attractive unit economics that favor growth. In contrast, new GenAI firms have much higher (and more volatile) variable costs that create difficult unit economics that favor



TABLE 4 New GenAI firms versus new applications firms in other technology waves.

Technology wave	Basis of the comparison	Differences from GenAI	Distinct strategic implications for new GenAI firms
Software	• Reliance of GenAI on software	• Low variable costs create attractive unit economics that favor growth. • Deterministic outcome where incorrect result can be traced to a fixable defect. • No hallucination to mitigate	• Greater importance of unit economics for GenAI
Predictive AI (machine learning)	• GenAI's AI cousin	• Attractive unit economics of software. • Learning economies readily achieved—positive loop of more usage yields more data, yields better algorithms (e.g., Google vs. Bing). • No hallucination to mitigate	• Active cultivation of learning economies for GenAI
Personal computer (PC)	• Similar position in the industry stack. • Dominant upstream providers in Microsoft and Intel (i.e., Wintel)	• Attractive unit economics of software. • Wintel enables fewer opportunities to differentiate than GenAI, so potentially higher rivalry in PCs	• High relevance of differentiation for both. • Differentiation easier for GenAI (e.g., business models, orchestration)
Mobile	• New technology leads to new dominant business models	• Significance of marketplaces and network effects. • New firms often apps w/free, paid or freemium revenue models. • Algorithmic control by marketplace owners (e.g., AppStore). • Attractive unit economics of software	• Mostly unrelated to strategies for GenAI
Social	• New technology leads to new dominant business models	• Significance of network dynamics and network effects. • New firms often network owners (e.g., Facebook, Instagram, Snapchat). • Attractive unit economics of software	• Mostly unrelated to strategies for GenAI
Quantum computing	• Similar position in the industry stack. • Dominant upstream providers (e.g., Google, Microsoft). • Rapid yet partially predictable technology trajectory. • Problematic probabilistic results. • Potentially high variable costs	• Multiple similarities. • More nascent than GenAI	• Potential strategic similarities with quantum firms in future

attention to unit profitability. Another comparison is with GenAI's cousin, *predictive AI* (machine learning). Yet, new predictive AI firms have the attractive unit economics of software and readily benefit from learning economies (Zott & Amit, 2026) while new GenAI firms have difficult unit economics, elusive learning economies, and hallucination, all of which complicate strategy. An intriguing comparison is with *quantum computing*. Here, new applications firms also face dominant technology suppliers, a rapidly yet somewhat predictably evolving technology, problematic

TABLE 5 Implications for entrepreneurs.

	Do	Don't
Big picture	Build thick and defensible wrappers around orchestrated LLMs	Build thin wrappers around well-known LLMs
Business model design	Return to first principles to think carefully about how productive but imperfect agents can lead to re-imagined business models	Mindlessly adopt familiar business models like freemium or subscription from software
Product innovation process design	Return to first principles to think carefully about how GenAI (including agents) changes software engineering (e.g., managing coding agents, efficient token usage, switching among LLMs, architecting lower memory and compute costs)	Automatically choose traditional ways of writing software and organizing software engineering
Superior data	Add superior data using contractual relationships w/proprietary data owners, mitigate hallucination, and cultivate learning economies like prompt context and user feedback loops	Treat access to proprietary data as sufficient
Efficient orchestration	Develop an orchestration layer (w/an ensemble of models and switching algorithms) which is thicker and more advantageous than single models	Stop at one or a few LLMs
Complementary resources	Add complementary resources and integrate them w/GenAI technology and professionals	Settle for a few resources that are weakly integrated w/GenAI
Over time	Constantly consider updating and/or redesigning advantages	Assume designs and resources are settled

probabilistic outcomes, and potentially high variable costs, all of which fit GenAI. Thus, while quantum is nascent, its applications firms may have strategic similarities with new GenAI firms.

In summary, GenAI has distinctive features and strategic implications for new firms that differ from other technologies. Thus, the specifics of effective strategies are likely to differ. Yet at a more abstract level, our framework of PP and RBV as complementary strategic logics likely generalizes across technologies, especially when bifurcated uncertainty is relevant.

6.3 | Implications for entrepreneurs

We also contribute implications for entrepreneurs (Table 5). The clearest one is that new GenAI firms will not likely succeed with thin wrappers around well-known LLMs. Rather, entrepreneurs who return to the first principles of GenAI to design their processes will improve their odds of success. More subtly, entrepreneurs who reinforce their advantages also improve their odds. For example, an efficient orchestration layer offers a thicker, more defensible advantage than simply a custom LLM. Finally, entrepreneurs who coevolve their advantages with advancing GenAI technology are likely to outlast those that treat these designs as settled.

7 | CONCLUSION

Many entrepreneurs have launched new GenAI firms with little strategic differentiation. Yet, we argue that entrepreneurs who approach these opportunities with clarity regarding relevant strategic logics will increase their likelihood of success. In contrast, entrepreneurs who chase the “hype” without attending to (1) process-based advantages of a well-designed business model and product innovation process and (2) resource-based advantages of superior data, efficient orchestration, and complementary resources will find the road to success much harder.



ACKNOWLEDGMENTS

We appreciate the excellent comments of Shaker Zahra (Ed.), our reviewers, Shaz Ansari, Jiayi Bao, Susan Cohen, Chris Coleridge, Chuck Eesley, Matthew Grimes, Benjamin Hallen, Mana Heshmati, Arvind Karunakaran, Karla Sayegh, Daniel Wilde, and Guankai Zhai as well as seminar participants at the UNLV Troesh Conference, Cambridge Judge Business School, and ION Lab Bocconi University. We also thank the Stanford Technology Ventures Program.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Kathleen M. Eisenhardt  <https://orcid.org/0000-0002-4879-1679>

Xilan Zhang  <https://orcid.org/0000-0002-0941-9081>

ENDNOTES

- ¹ This is a common strategic rivalry for new GenAI firms, especially successful ones (e.g., Harvey (next section) competes with OpenAI). These new firms may win, coexist, lose or be acquired (e.g., SpaceX recently acquired Cursor).
- ² A question is whether new GenAI firms have a resource advantage based on extreme human capital (e.g., individuals with very rare and highly valued skills, commanding seven figure salaries or more). While all firms benefit from talent and often hire carefully like Moon.ai, new GenAI firms typically do not need this very rare level of talent. Rather, these individuals usually work for the creators of frontier foundation models.

REFERENCES

- Allen, R. T., Heshmati, M., Lenox, M., McDonald, R. M., & Perez, M. (2026). *How human representations shape AI-augmented strategies* [Working paper]. Brigham Young University.
- Allen, R. T., & McDonald, R. M. (2026). How well can AI do strategy? Empirical benchmarking using strategy simulations. *Strategy Science*, 11(1), 93–117.
- Amazon Web Services. (2026). *What is RAG (retrieval-augmented generation)?* <https://aws.amazon.com/what-is/retrieval-augmented-generation/>
- Andries, P., Debackere, K., & van Looy, B. (2013). Simultaneous experimentation as a learning strategy: Business model development under uncertainty. *Strategic Entrepreneurship Journal*, 7(4), 288–310.
- Anthony, C., Nelson, A. J., & Tripsas, M. (2016). “Who are you?...I really wanna know”: Product meaning and competitive positioning in the nascent synthesizer industry. *Strategy Science*, 1(3), 163–183.
- Bai, L., Huang, Z., Wang, X., Sun, J., Mihalcea, R., Brynjolfsson, E., Pentland, A., & Pei, J. (2026). *How do AI agents spend your money? Analyzing and predicting token consumption in agentic coding tasks* [Working paper]. Digital Economy Lab, Stanford University.
- Baldwin, C. Y., & Clark, K. B. (2000). *Design rules: The power of modularity* (Vol. 1). MIT Press. <https://doi.org/10.7551/mitpress/2366.001.0001>
- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120.
- Bingham, C. B., & Eisenhardt, K. M. (2011). Rational heuristics: The “simple rules” that strategists learn from process experience. *Strategic Management Journal*, 32(13), 1437–1464.
- Bingham, C. B., Howell, T., & Ott, T. E. (2019). Capability creation: Heuristics as microfoundations. *Strategic Entrepreneurship Journal*, 13(2), 121–153.
- Blank, S. (2026, March 17). Your startup is probably dead on arrival. *Steve Blank*. <https://steveblank.com/2026/03/17/your-startup-is-probably-dead-on-arrival/>
- Bocken, N. M. P., Heidenreich, S., Spieth, P., Tucci, C., & Zott, C. (2025). Innovating for impact: Harnessing business model innovation to tackle grand challenges. *Strategic Entrepreneurship Journal*, 19(4), 665–681.
- Bremner, R. P., & Eisenhardt, K. M. (2022). Organizing form, experimentation, and performance: Innovation in the nascent civilian drone industry. *Organization Science*, 33(4), 1645–1674.

- Brown, S. L., & Eisenhardt, K. M. (1997). The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Administrative Science Quarterly*, 42(1), 1–34.
- Brown, S. L., & Eisenhardt, K. M. (1998). *Competing on the edge: Strategy as structured chaos*. Harvard Business School Press.
- Brynjolfsson, E., Chandar, B., & Chen, R. (2025). *Canaries in the coal mine? Six facts about the recent employment effects of artificial intelligence*. Stanford Digital Economy Lab. https://digitaleconomy.stanford.edu/wp-content/uploads/2025/08/Canaries_BrynjolfssonChandarChen.pdf
- Camuffo, A., Gambardella, A., Messinese, D., Novelli, E., Paolucci, E., & Spina, C. (2024). A scientific approach to entrepreneurial decision-making: Large-scale replication and extension. *Strategic Management Journal*, 45(6), 1209–1237.
- Challapally, A., Pease, C., Raskar, R., & Chari, P. (2025). *The GenAI divide: State of AI in business 2025*. MIT NANDA. https://asociace.ai/wp-content/uploads/2025/08/ai_report_2025.pdf
- Clark, K. B., & Fujimoto, T. (1991). *Product development performance: Strategy, organization, and management in the world auto industry*. Harvard Business School Press.
- Cohen, S. L., Bingham, C. B., & Hallen, B. L. (2019). The role of accelerator designs in mitigating bounded rationality in new ventures. *Administrative Science Quarterly*, 64(4), 810–854.
- Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2009). Optimal structure, market dynamism, and the strategy of simple rules. *Administrative Science Quarterly*, 54(3), 413–452.
- Dell'Acqua, F., McFowland, E., III, Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Kraymer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science*, 37(2), 403–423.
- DeSantola, A., Gulati, R., & Zhelyazkov, P. I. (2023). External interfaces or internal processes? Market positioning and divergent professionalization paths in young ventures. *Organization Science*, 34(1), 1–23.
- Dey, V. (2026, February 27). Andrew Ng says AGI is decades away—And the real AI bubble risk is in the training layer. *Fast Company*. <https://www.fastcompany.com/91499247/andrew-ng-agi-decades-away-interview>
- Dinçkol, D., Fraser, J., & Ozcan, P. (2026). *AI-driven entrepreneurship in regulated industries: A study of UK fintech ventures* [Working paper]. University of Oxford.
- Eisenhardt, K. M. (1985). Control: Organizational and economic approaches. *Management Science*, 31(2), 134–149.
- Eisenhardt, K. M. (1989a). Making fast strategic decisions in high-velocity environments. *Academy of Management Journal*, 32(3), 543–576.
- Eisenhardt, K. M. (1989b). Agency theory: An assessment and review. *Academy of Management Review*, 14(1), 57–74.
- Eisenhardt, K. M., Ott, T. E., Hannah, D. P., McDonald, R. M., Hallen, B. L., Ozcan, P., & Bingham, C. B. (2026). *Entrepreneur as preschooler: The process perspective on entrepreneurial performance* [Working paper]. Stanford University.
- Furr, N. R., & Eisenhardt, K. M. (2021). Strategy and uncertainty: Resource-based view, strategy-creation view, and the hybrid between them. *Journal of Management*, 47(7), 1915–1935.
- Gao, C., DeSantola, A., & Guilbeault, N. (2026). “Clean” meat? Regulatory entrepreneurship and jurisdictional contestation in a nascent industry. *Administrative Science Quarterly* (in press). <https://doi.org/10.1177/00018392261477694>
- Gao, C., & McDonald, R. (2022). Shaping nascent industries: Innovation strategy and regulatory uncertainty in personal genomics. *Administrative Science Quarterly*, 67(4), 915–967.
- Hallen, B. L., Tidhar, R., & Eisenhardt, K. M. (2026). The risks of scaling too fast. *Harvard Business Review*, November–December (in press).
- Hannah, D. P., & Eisenhardt, K. M. (2018). How firms navigate cooperation and competition in nascent ecosystems. *Strategic Management Journal*, 39(12), 3163–3192.
- Harvey. (2026, March 25). *Harvey raises at \$11 billion valuation to scale agents across law firms and enterprises* [Press release]. <https://www.harvey.ai/blog/harvey-raises-at-dollar11-billion-valuation-to-scale-agents-across-law-firms-and-enterprises>
- Hatmaker, T. (2026, February 5). Why one Anthropic update wiped billions off software stocks. *Fast Company*. <https://www.fastcompany.com/91487960/why-one-anthropic-update-wiped-billions-off-software-stocks>
- He, V. F., & Puranam, P. (2026). *Improving abductive reasoning*. SSRN preprint: <https://ssrn.com/abstract=6594299> or <https://doi.org/10.2139/ssrn.6594299>
- Huang, J. (2026, March 10). AI is a 5-layer cake. *NVIDIA Blog*. <https://blogs.nvidia.com/blog/ai-5-layer-cake/>
- Jacobides, M. G., & Ma, M. D. (2026). *What are generative AI's competitive implications? A longitudinal documentation of director expectations* [Working paper]. London Business School.
- Kapoor, R., & Raj, M. (2026). Value creation and capture in artificial intelligence (AI)-enabled ecosystems. *Strategic Entrepreneurship Journal* (in press).
- Kilminster, W., Ozcan, P., & Eisenhardt, K. M. (2026). *Theorizing AI adoption in established firms: An empirical-based, LLM-augmented approach* [Working paper]. University of Oxford.
- Kim, H., & Koning, R. (2026). *AI-native firms* (Working Paper 26-090). Harvard Business School. <https://doi.org/10.2139/ssrn.6905079>



- Kirtley, J., & O'Mahony, S. (2023). What is a pivot? Explaining when and how entrepreneurial firms decide to make strategic change and pivot. *Strategic Management Journal*, 44(1), 197–230.
- Koning, R., Keller, N. T., & Kim, H. (2026). *Gamma: Slides in the blink of AI* (Case No. 9-826-001, rev. February 2026). Harvard Business School. <https://store.hbr.org/product/gamma-slides-in-the-blink-of-ai/826001>
- Kost, D. (2025, November 12). AI companies don't have a profitable business model. Does that matter? *Harvard Business Review*. <https://hbr.org/2025/11/ai-companies-dont-have-a-profitable-business-model-does-that-matter>
- Martin, I., & Shrivastava, R. (2026, February 9). Legal AI startup Harvey in talks to raise \$200 million at \$11 billion valuation. *Forbes*. <https://www.forbes.com/sites/iamartin/2026/02/09/legal-ai-startup-harvey-in-talks-to-raise-200-million-at-11-billion-valuation/>
- McCracken, H. (2026, February 13). AI is still both more and less amazing than we think, and that's a problem. *Fast Company*. <https://www.fastcompany.com/91491419/matt-shumer-something-big-is-happening>
- McDonald, R. M., & Eisenhardt, K. M. (2020). Parallel play: Startups, nascent markets, and effective business-model design. *Administrative Science Quarterly*, 65(2), 483–523.
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192.
- Orosz, G. (2026, February 24). The future of software engineering with AI: Six predictions. *The Pragmatic Engineer*. <https://newsletter.pragmaticengineer.com/p/the-future-of-software-engineering-with-ai>
- Ott, T. E., & Eisenhardt, K. M. (2020). Decision weaving: Forming novel, complex strategy in entrepreneurial settings. *Strategic Management Journal*, 41(12), 2275–2314.
- Ott, T. E., Eisenhardt, K. M., & Bingham, C. B. (2017). Strategy formation in entrepreneurial settings: Past insights and future directions. *Strategic Entrepreneurship Journal*, 11(3), 306–325.
- Ozcan, P., Dinçol, D., & Zachariadis, M. (2026). The rise of platforms in regulated industries: Ownership, curation, and the asymmetric strategies of challenger versus incumbent banks. *Research Policy*, 55(7), 105499.
- Ozcan, P., & Eisenhardt, K. M. (2009). Origin of alliance portfolios: Entrepreneurs, network strategies, and firm performance. *Academy of Management Journal*, 52(2), 246–279.
- Pavlus, J. (2025, December 4). Amazon takes on AI's biggest nightmare: Hallucinations. *Fast Company*. <https://www.fastcompany.com/91446331/amazon-byron-cook-ai-artificial-intelligence-automated-reasoning-neurosymbolic-hallucination-logic>
- Perlo, J. (2026, May 13). Most U.S. doctors are quietly using this AI tool. Few patients know about it. *NBC News*. <https://www.nbcnewyork.com/news/national-international/most-us-doctors-use-this-ai-tool-few-patients-know/6501320/>
- PitchBook, & National Venture Capital Association. (2026). Q4 2025 PitchBook-NVCA venture monitor. <https://nvca.org/wp-content/uploads/2026/01/q4-2025-pitchbook-nvca-venture-monitor.pdf>
- Reineke, P., Katila, R., & Eisenhardt, K. M. (2025). Decentralization in organizations: A revolution or a mirage? *Academy of Management Annals*, 19(1), 298–342.
- Rindova, V. P., & Fombrun, C. J. (2001). Entrepreneurial action in the creation of the specialty coffee niche. In C. B. Schoonhoven & E. Romanelli (Eds.), *The entrepreneurship dynamic: Origins of entrepreneurship and the evolution of industries* (pp. 236–261). Stanford University Press. <https://doi.org/10.1515/9781503618527-011>
- Ross, J.-M., & Lieberman, M. B. (2026). Strategic management of uncertainty: An integrative thread for our field. *Strategic Management Review* (in press).
- Sarasvathy, S. D. (2024). Lean hypotheses and effectual commitments: An integrative framework delineating the methods of science and entrepreneurship. *Journal of Management*, 50(8), 3035–3063.
- Schneider, J., Meske, C., & Kuss, P. (2024). Foundation models. *Business & Information Systems Engineering*, 66(2), 221–231.
- Snihur, Y., & Zott, C. (2020). The genesis and metamorphosis of novelty imprints: How business model innovation emerges in young ventures. *Academy of Management Journal*, 63(2), 554–583.
- Stanford HAI. (2026). *Ecosystem graphs for foundation models* [Interactive database]. <https://crfm.stanford.edu/ecosystem-graphs/>
- Stokel-Walker, C. (2026, February 18). New AI models are losing their edge almost immediately. *Fast Company*. <https://www.fastcompany.com/91494576/new-ai-models-are-losing-their-edge-almost-immediately>
- Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, 15(6), 285–305.
- Thomke, S., Eisenhauer, P., & Sahni, P. (2025). Addressing GenAI's quality-control problem. *Harvard Business Review*, 103(5), 60–67.
- Thompson, C. (2026, March 12). Coding after coders: The end of computer programming as we know it. *The New York Times*. <https://www.nytimes.com/2026/03/12/magazine/ai-coding-programming-jobs-claude-chatgpt.html>
- Tidhar, R., & Eisenhardt, K. M. (2020). Get rich or die trying... finding revenue model fit using machine learning and multiple cases. *Strategic Management Journal*, 41(7), 1245–1273.

- Tidhar, R., Hallen, B. L., & Eisenhardt, K. M. (2025). Measure twice, cut once: Unit profitability, scalability, and the exceptional growth of new firms. *Organization Science*, 36(1), 88–120.
- Tripsas, M. (1997). Unraveling the process of creative destruction: Complementary assets and incumbent survival in the typesetter industry. *Strategic Management Journal*, 18(S1), 119–142.
- Volmar, E., & Eisenhardt, K. M. (2025). Mavericks and diplomats: Bridging commercial and institutional entrepreneurship for society's grand challenges. *Organization Science*, 36(2), 572–600.
- von Krogh, G., Weiser, A.-K., & Tinguely, P. (2026). "Borrowing intelligence to build a business": The promise of open-source AI for entrepreneurship. *Strategic Entrepreneurship Journal* (in press).
- Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, 5(2), 171–180.
- Wilde, D., & Packard, M. (2026). *When is foresight useful? A contingency theory of foresight alignment* [Working paper]. Indiana University.
- Wu, B., Wan, Z., & Levinthal, D. A. (2014). Complementary assets as pipes and prisms: Innovation incentives and trajectory choices. *Strategic Management Journal*, 35(9), 1257–1278.
- Yoffie, D. B., Dan, O., & Corsi, E. (2024). *AI21 Labs in 2023: Strategy for generative AI* (Case No. 9-724-383, rev. January 23, 2024). Harvard Business School. <https://store.hbr.org/product/ai21-labs-in-2023-strategy-for-generative-ai/724383>
- Yotzov, I., Barrero, J. M., Bloom, N., Bunn, P., Davis, S. J., Foster, K. M., Jalca, A., Meyer, B. H., Mizen, P., Navarrete, M. A., Smietanka, P., Thwaites, G., & Wang, B. Z. (2026). *Firm data on AI* (Working Paper 34836). National Bureau of Economic Research. <https://doi.org/10.3386/w34836>
- Zott, C., & Amit, R. (2026). Whither business model innovation strategy in the age of AI? *Strategic Entrepreneurship Journal* (in press).
- Zuzul, T., & Tripsas, M. (2020). Start-up inertia versus flexibility: The role of founder identity in a nascent industry. *Administrative Science Quarterly*, 65(2), 395–433.

How to cite this article: Eisenhardt, K. M., & Zhang, X. (2026). New wine, old bottles? Effective strategies of new GenAI firms. *Strategic Entrepreneurship Journal*, 1–26. <https://doi.org/10.1002/sej.70056>

APPENDIX: BRIEF SUMMARY OF THE GENAI INDUSTRY A

Generative AI (GenAI) is a type of artificial intelligence that creates new content such as text, images, audio, and computer code. This technology is widely expected to have transformative implications that are possibly on the scale of general-purpose technologies like electricity and the Internet. Although broader in scope, GenAI can be (and often is) equated with large language models (LLMs). Finally, the GenAI industry is often conceptualized as a five-layer stack or, formally, a layered technical system (Figure A1) (Huang, 2026). We sketch these layers next.

A1 | GenAI stack

The most visible (and perhaps most important) layer of the stack is *Models*, now dominated by LLMs. Some firms operating in this layer created the initial foundational LLMs like OpenAI (ChatGPT), Google (Gemini), and Anthropic (Claude) that currently are driving the emergence and evolution of GenAI (Schneider et al., 2024; Stanford HAI, 2026). LLMs are very large, deep-learning neural networks that can “understand,” summarize and generate human language, computer code, and other human-like content (e.g., Dell’Acqua et al., 2026; Yoffie et al., 2024). Trained on massive data sets to identify patterns and relationships, LLMs at a very basic level rely on transformer architecture to predict the next likely word or piece of content in a sequence. Thus, they are probabilistic. Further, the specific training data and scale of the neural network are central to LLM performance.

Building these LLMs requires massive computing resources, large teams of researchers, and enormous datasets. This capital intensity creates steep barriers to entry for most would-be entrepreneurs. Frontier LLM creation is thus

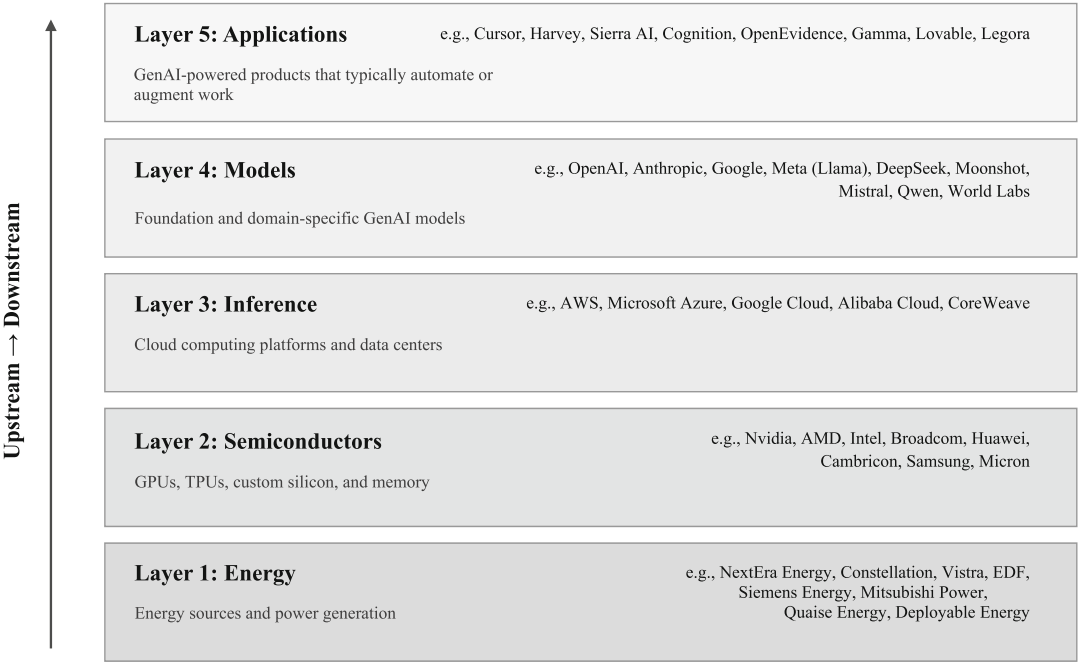


FIGURE A1 Generative AI (GenAI) industry stack.

confined to a handful of exceptionally well-funded organizations, whether Big Tech or heavily capitalized ventures. These models are typically proprietary and closed. Yet the layer is far from settled. Capable open-weight models such as DeepSeek, Kimi, Zhipu AI's GLM, and Alibaba's Qwen are now significant (Stokel-Walker, 2026). Moreover, leading GenAI actors like Nvidia are endorsing open-weight models. At the same time, the foundational LLMs are becoming cheaper (on a per unit of intelligence basis) and more capable (Kost, 2025; Stanford HAI, 2026). Further, some new LLM models are seeking differentiating niches. For example, Thinking Machines' Inkling is an open-weight model that corporations can customize for their own use. World Models that use an entirely different structure are next. Finally, increasing use of "distillation" is shortening the window of technical advantage for LLM creators (Stokel-Walker, 2026). The leading proprietary models are responding by extending their product lines across varying sizes, capabilities, and price points. While a few entrepreneurial firms have entered, they typically have marquee founders and substantial investments such as AI pioneer Fei-Fei Li's World Labs and former OpenAI CTO Mira Murati's Thinking Machines. Overall, rapid innovation, increasing differentiation, savvy and wealthy rivals, and the potential for rapid imitation make LLM creation a highly competitive space where entrepreneurial firms rarely enter.

Inference refers to the platforms that are also at the center of the GenAI stack. These firms provide the infrastructure to execute LLMs for both training and inference. Big Tech cloud computing providers like Amazon (AWS), Alphabet (Google Cloud), Microsoft (Azure), and Alibaba Cloud are among the prominent players (Yoffie et al., 2024). This layer carries substantial fixed capital (e.g., data centers, GPUs) and variable (e.g., electricity to process tasks) costs. Although a few entrepreneurial firms entered early on by repurposing assets built for other uses (e.g., CoreWeave, founded as a cryptocurrency miner in 2017), the substantial capital requirements have kept most out. Firms in this layer also face significant operational challenges (e.g., weather, supply shortages, zoning restrictions, public opposition). Overall, dominant and large incumbents, a mostly undifferentiated product, huge capital costs, complex operations, and dependence on concentrated semiconductor suppliers make this a highly rivalrous and difficult layer in which to compete that most (or all) entrepreneurial firms avoid.

Semiconductors supply the GPUs, TPUs, and custom silicon that enables computation and memory at the inference layer. Much of the innovation in the GenAI stack occurs at this layer and the Models layer. Nvidia has long dominated the market for AI chips although other firms like AMD are making in-roads. Some chip makers are also optimizing on non-traditional dimensions such as memory over speed and biology-based architectures given the technical opportunities for differentiation at this layer. In addition, BigTech firms in the U.S. are increasingly designing their own chips (e.g., Google's TPUs, Amazon's Trainium, Microsoft's Maia, Meta's MTIA), while major Chinese firms like Huawei also design chips. In turn, these designers depend on a small number of foundries, notably TSMC and more recently China's SMIC, to manufacture their chips. The scale of investment at this layer is typically prohibitive for entrepreneurial firms. The exceptions are rare, heavily capitalized, and usually led by high-profile founders (e.g., Unconventional AI, founded in 2025 by team led by Naveen Rao (former head of AI at Databricks) with a \$475 million seed round). Overall, high investment costs, long product development times, dominant and large incumbents, and limited access to manufacturing foundries make this a difficult layer for entrepreneurial firms to enter.

The *Applications* layer consists of firms that create GenAI-powered products that rely on LLMs. Unlike the middle layers, the Applications layer is attractive for entrepreneurial firms. Examples include Gamma, Lovable, Sierra AI, Anysphere (Cursor), Counsel AI (Harvey), and OpenEvidence. The reasons for this attractiveness include (1) numerous opportunities, especially in a variety of vertical markets (e.g., health, law, small business, coding, and military) and (2) low barriers to entry like relatively modest investment costs. For example, these firms can access LLMs through APIs rather than spending capital to train one. They can often also run very lean organizations (Kim & Koning, 2026), further lowering investment costs. These low barriers to entry and numerous opportunities can create rivalrous vertical markets. A key to success is moving past thin wrappers around well-known LLMs to thick, defensible wrappers by building process-based and resource-based advantages. That said, firms at this layer may face encroachment, especially by model creators, thus creating potential best-of-breed versus best-of-platform rivalry. Simultaneously, entrepreneurial firms may also be attractive acquisition targets.

Energy sits at the bottom of the GenAI industry stack. It supplies the energy sources and electrical capacity that power GenAI. Both incumbents and entrepreneurial firms compete here. Established utilities provide most of this power in the U.S. (e.g., NextEra, Constellation, GE Vernova). Deep-tech startups are advancing futuristic sources in nuclear (Deployable Energy), geothermal (Quaise Energy), battery (Form Energy), and other forms of energy. Several have direct support from BigTech. Amazon backed X-energy, which was listed publicly in 2026, Alphabet contracted with Kairos Power for small modular reactors, and Microsoft agreed to buy power from Helion. There is similar activity in countries like China and the European Union. While Energy is an attractive layer for entrepreneurial firms given its numerous technical opportunities, substantial available investment capital, and high demand if the technology is successful, the strategies of these firms are not dependent on the features of GenAI. Rather, they are typically deep-tech startups that have their own economics, high investment requirements, long timelines, substantial technical challenges and demand outside the GenAI stack that distinguish them from the new GenAI firms in the Applications layer that we study.