Narrative Review Article

Impact of the Institutional Logics of Funding Partners on Shaping Outcomes in Young Firms

Kanishk Gupta

Heritage International Xperiential School, Mayom hospital South City 1 D block, Gurugram, Haryana, 122001, India

ABSTRACT

Through institutional theory, this review examines how the institutional logics of funding partners shape innovation, strategy and general outcomes in young firms, utilizing case studies in the AI industry. It investigates how young firms should consider the choice of their resource providers, specifically through factors that impact those choices besides the provision of capital. Clear trade-offs emerge through comparative analysis of the impacts of the institutional logics of 3 types of funding partners, namely venture capital, corporate and hybrids. Corporate logic can create substantial value in providing critical specialized complements (among other resources that would not have been accessible through alternative resource providers), specifically when uncertainty is high due to the vast amount of resources corporate partners possess, however they may restrict access to alternative partners and markets and tend to steer the young firm to align with their own strategic goals. Professional (venture capital) logic emphasizes discipline and pacing in addition to commercial innovation and is thus likely to aid these. However, young firms with ties to venture capital, carry the risk of information leakage to competitors and are unlikely to experience significant improvements in technical innovation. Hybrid logics may combine the commercialization benefits and take advantage of the resources of corporate partners while benefiting from the commercial discipline instilled by the professional logic of their venture capital partners. Professional logic is likely to aid mitigate the drawbacks of corporate logics, however, may lead to increased coordination costs and still carries the exposure that accompanies professional logic.

Keywords: Institutional Logics; Venture Capital; Funding Partners; Venture Funding; AI Startups; AI Ventures; Young Firms

INTRODUCTION

Entrepreneurship and strategy research has long treated external partners as shaping venture outcomes not only through capital, but also through the access they enable to resources and networks and the constraints they impose. In this view, different partner types can systematically alter what resources are available, what strategies are prioritized, and how quickly firms can commercialize. (1)

An active debate concerns corporate involvement: while corporate partners can unlock complementary assets and market pathways that young firms cannot build internally, they may also introduce strategic misalignment and dependence risks that redirect or constrain innovation. As a result, corporate partnering is often treated as leading to adverse outcomes, or contingent in its effects (2-4).

Institutional theory sharpens these debates by

explaining why partner relationships shape ventures structurally. Partners carry institutional logics and priorities that influence the perception of value, acceptable risk, and preferred governance. Importantly, young firms can face multiple logics simultaneously; outcomes may therefore depend on whether those logics are experienced as compatible or conflictual, an issue that is relevant to hybrid funding structures (5-6).

Startups select their resource providers through a combination of necessity and opportunity (2). Young firms rarely have the ability to develop and obtain all they require, and thus must rely on external providers for resources (2-3). This may often be due to meeting essential operational requirements, for example at the right scale, speed or cost. These choices are made using a combination of needs, strategic fitness and network access (2). Funding partners are one such category of resource providers that exert a particularly strong influence, however it is crucial to recognize they do not simply provide capital- they introduce expectations, priorities and norms that can shape the firm's trajectory (2). This means that the selection of a funding partner can shape not only the resources obtained but also the strategic tradeoffs a firm face, the additional providers and resources it can access, and the degree of autonomy it retains. For instance, venture capital (professional) logic typically emphasizes rapid scaling, commercialization, and exit driven growth, whereas corporate partner logic tends to prioritize strategic fitness, ecosystem integration, and long-term complementarity with the parent's business. These differences affect outcomes such as the speed of innovation, degree of openness, and strategic autonomy. Young firms must therefore choose potential resource providers based on not only the tangible resources (e.g. financial capital) they can provide, but also on the implications on the alignment of their strategic goals, incentives and objectives (2).

Most existing empirical work offers some expectations about differences across partner types, but it provides less clarity on the mechanisms through which funder logics shape resource access and strategic decision-making, and on the variation and nuance in these effects across industry contexts. Factors such as market maturity, context and the speed of innovation may lead to significant differences in outcomes. In emerging and relatively young industries like artificial intelligence, where resource requirements are unusually high (and expensive) (7), high uncertainty and innovation cycles are extremely fast (8), established literature may not fully capture the advantages and disadvantages for each funding type. In particular, the scale and cost of computing infrastructure may make corporate partnerships more attractive than theory would predict for early-stage AI firms, while the flexibility of VC funding could be more constrained by technical dependencies than in other industries (7, 9).

Furthermore, an underexplored area is how mixed funding structures (where a firm has both VC and corporate partners) influence outcomes. Theory suggests these hybrids face tension between the speed and exit oriented priorities of VCs and the integration driven goals of corporates. However, in practice, (specifically in the AI industry) the logics may coexist or reinforce one another (2). This is particularly relevant in AI, where critical complements such as computing capacity and distribution channels can become decisive constraints and where the governance design of partner relationships may materially shape innovation and commercialization pathways (7, 10).

This narrative review examines how the institutional logics of venture capital, corporate partners, and hybrid funding structures shape innovation, commercialization, strategic autonomy, and risk in young AI firms, and specifies the mechanisms through which those influences operate beyond the provision of capital. It uses a case-based synthesis of DeepL, OpenAI, and Cohere using publicly available secondary evidence to develop a structured, mechanism-oriented account of cross-case differences under distinct funding partner logics. This review therefore extends existing institutional logic expectations by arguing that in young AI firms, corporate logic can be substantially beneficial under conditions where specialized complements and commercialization channels are unusually decisive, and it surfaces a clearer theoretical account of hybrid governance rather than treating hybrid outcomes as an empirical regularity (2-3).

DEEPL (PROFESSIONAL LOGIC)

Pahnke, Katila and Eisenhardt explores How Partners' Institutional Logics Influence Innovation in Young Firms. The paper is an empirical study that focuses on how the institutional logics of distinct types of funding partners impact innovation in young firms, specifically focusing on how these institutional logics impact access to resources. The paper emphasizes the impact on the innovation of young firms beyond the financial resources gained through the relationship (2).

The paper concluded that ties to VCs (professional logic) have no significant effects on technical innovation

but significantly improved commercial innovation (2). This case uses that expectation as a baseline to examine how a professional logic configuration can shape outcomes. At the same time, a key drawback discussed in the literature is more specific to VC ties, because VCs often operate as intermediaries across portfolios and networks. Pahnke, McDonald, Wang & Hallen show that intermediaries can route opportunities and resources while simultaneously increasing exposure through network overlap and indirect ties. (11) This matters for professional logic because VC networks can expand access to information, expertise, status, customers, and complementary resources, while also creating competitor exposure risk through intermediation pathways that are structurally embedded in VC investing (11-12).

DeepL is a company that emerged from the Linguee team and launched its neural machine translation (a form of artificial intelligence) service publicly in 2017, it is an AI language company that focuses on translations (13). Benchmark, an American venture capital firm, led DeepL's series A in 2018, procuring a 13.6% stake alongside existing investors such as b2venture (14). In January of 2023, it was estimated that DeepL raised over 100 million dollars at a 1-billion-dollar valuation, from venture capital investors such as IVP (leading the round), Bessemer Venture Partners, Atomico and WiL, among others (15). Public framing emphasized enterprise expansion and continued R&D. On May 22nd, 2024, DeepL hosted a funding round, raising 300 million dollars at a 2-billion-dollar valuation, led by Index Ventures, with participation from IVP, Atomico, ICONIQ Growth, among others. In this round, DeepL specifically emphasized research and product innovation, global expansion, hiring and commercial build out as the use case for the funds (16).

From 2018 onward, DeepL investor coalition is evidently VC dominant, which exhibits the impacts of professional logic (2). DeepL launched DeepL Pro in March 2018 with API access and CAT tool plugins aimed at "professional translators, companies and developers" (17). This is an explicit enterprise monetization path which includes paid plans and sets up developer integration, and stronger data-handling promises, rather than a long mass market phase with low profitability or usage.

Furthermore, in May 2024, CEO Jarosław Kutylowski noted that about 60% of DeepL's staff were technologists and that the company would be hiring more non-technical personnel going forward because they had still only captured very little of the addressable market. This indicates a deliberate build out of sales, marketing,

customer success, and operations alongside research, which is consistent with a professional (VC) logic emphasis on commercialization and rapid scaling (i.e., adding go to market capacity) (18). Adding Japanese and Chinese in March 2020 broadened DeepL's addressable enterprise demand into two high value markets. DeepL framed the milestone as opening access to "over a billion potential new users" (19). Then, In March 2021, DeepL added 13 European languages in one release (e.g., Bulgarian, Czech, Danish, Greek, Swedish) (20). This is a batch expansion of market coverage without changing the core product (translation), consistent with milestone paced scaling that VCs seek.

Pahnke et al. expects that professional logic pays attention to commercialization while keeping in mind milestones along with disciplined scaling (2). The scope of the product offering of DeepL, as well as adjacent product lines align with the theoretical prediction. The firm has positioned itself as a Language AI (including translation and writing) rather than a general purpose LLM, and its product evolution adds new offerings built around their translation core. In April 2024, DeepL launched DeepL Write pro, which was stated as being targeted to business communications with a per seat pricing, a standard B2B SaaS pattern (21). The aim aligns with professional logic, as this move was intended to increase revenue per account (commercialization).

However, their VC partners still retain significant amounts of control and authority in the form of board seats despite not having majority stakes (1). They carry the same indirect competitor exposure that Pahnke, Mcdonald, Wang & Hallen document (11), and especially given the extreme resource dependence of young firms that facilitates the disclosure of important information, it indicates that information leakage from their VC partners may harm their innovation.

Furthermore, in February 2024 (right after series B), DeepL completed a SOC 2 Type II report (full report available under NDA), a requirement for mid-market and enterprise IT buyers. However, this is considered a significant undertaking due to having high complexity and requiring significant time commitment and cost (22). The assurance investment does not add a lot of value for the consumer in terms of the product itself, but directly advances B2B commercial milestones by increasing market access (through contractable enterprise demand), akin to the sort of disciplined process Pahnke et al. tie to VC attention. Moreover, geographic and coverage expansion was accelerated in early to mid-2024, with their first US office being opened in February. Following

that, within the next few weeks DeepL pro expanded to 165 additional markets (now totaling 228) via a global rollout (23). The sequence establishes a pattern of rapid scaling that VCs push for, where they doubled down on their US offerings, widened the availability to capture demand and de-risked enterprise adoption.

Professional logic expects staged financing aligned along with discernable progress in order to achieve higher valuations and improve prospects of exit liquidity. For example, the funding round in May 2024 was communicated with over 100,000 organizational customers and named enterprise logos as evidence of growth/a milestone for investors and enterprise buyers (16) in order to justify the increase in valuation. In the same round, it states that the proceeds would go to product R&D, global expansion, hiring, and commercial build out, which align with professional logic (raising funding when momentum is strong, establish concrete growth goals specifically with regards to commercialization) The round also came 16 months after the January 2023 round but approximately doubled the valuation using the execution of visible milestones such as the Write pro launch, global expansion, and next generation models. Moreover, the timing of the May 2024 round also aligns with a broader AI funding surge, and exemplifies a standard VC tendency to raise rounds into hot markets in order to optimize pricing and dilution (24), i.e. VCs exploiting favorable funding windows and pacing ventures to arrive at rounds with important commercial milestones, as per Pahnke.

OPENAI (CORPORATE LOGIC)

Park & Steensma focuses on when firms partnering with CVCs may lead to improved outcomes as compared with VCs, and their research question centers on the contingencies under which corporate venture capital adds value and imposes constraints on young firms: "When does corporate venture capital add value for new ventures?" (3).

The paper concludes that CVC funding is particularly beneficial for new ventures when they require specialized complementary assets or operate in uncertain environments and conversely, tight links to a single corporate may constrain access to alternative partners and markets, dampening performance outside those conditions. (3) The paper adds nuance to the idea that corporate venture capital or venture capital partners are strictly superior for young firms' outcomes by outlining the specific conditions under which corporate

involvement leads to better outcomes for young firms (3). However, much of the institutional theory framing still treats corporate involvement as, at best, contingent, and often associated with dependence, strategic steering, and slow and complex organizational processes (2-4). This case therefore examines this view in the context of the AI industry.

The institutional logic (taken for granted norms, structures, and practices) of a resource provider (funder) is not dependent on the exact legal wrapper of the resource provider/capital provider, but the identity, norms and attention pattern of the funder (2). This paper extends the framework used in Pahnke, Katila & Eisenhardt to include both investments through a formal CVC arm (as discussed in Pahnke), or a direct equity investment. This is because corporate partners also bring corporate logic as the investing entity's identity and resource bundle are corporate (complementary assets, channels, governance routines), and those features orient the venture toward commercial integration with the corporation's assets and timelines.

OpenAI's capital structure is quite nonstandard. OpenAI, a nonprofit, in order to access resources, created a capped profit for profit subsidiary that is under the control of the nonprofit board (25). The analysis for this case focuses on Microsoft and the corporate logic it carries, because it is the single funding relationship with demonstrable and sustained influence on OpenAI. Since 2019, the deal has been structured as an ongoing operational partnership, through a tight relationship in terms of resource access and collaboration (26-27), and exhibits corporate logic. By contrast, participation from other investors has occurred largely through secondary tender offers that were primarily meant to provide liquidity for employees, arrangements that did not entail continued interaction or influence in OpenAI's operations (28). Moreover, OpenAI's governance structure gives sole control to the nonprofit board, further constraining the strategic relevance of investors who lack an operational role (25, 29). This can be corroborated through a regulatory review by the Competition & Markets authority of the UK, which concluded that Microsoft exerts a high level of material influence over OpenAI's commercial policy but not de facto control (30). Given these conditions, the OpenAI case is analytically best treated through the Microsoft tie for the purposes of this study as other funding partners are deemed to not have significant influence/impact through their identity, norms, strategy and attention patterns to OpenAI.

Microsoft has an estimated investment into OpenAI of over 13 billion dollars to date. The 2019 deal combined 1 billion dollars with exclusive Azure cloud provision and joint Azure supercomputing work. In 2020 Microsoft also obtained an exclusive GPT 3 license for their own products and services, explicitly positioning Microsoft's product stack as the first outlet for OpenAI's advances from then on (30). In January 2023 the firms announced a "multiyear, multibillion dollar" expansion and reaffirmed Azure as OpenAI's exclusive cloud for research, products and API services (27), further tightening infrastructural dependence during the commercialization surge after ChatGPT's breakout. From 2019 to 2025, OpenAI's training/inference were developed using the Azure infrastructure, the structuring of which made scaling easier but narrowed their strategic autonomy. The exclusivity meant that OpenAI could not operate core services over rival clouds even if better terms existed, because the corporate partner's strategic goals dominated the channel choice. The narrowing of subsequent exploration paths (resource providers) for OpenAI is in line with the intermediation discussed in de Vaan & Stuart. However, the configuration fits the contingency that Park & Steensma discusses when referring to when corporate partners can add value, in that the agreement offered specialized complementary assets in the form of bespoke supercomputing and preferential cloud access/credits in a period of high uncertainty in the AI industry among both technical and regulatory uncertainty. This indicates that corporate ties are beneficial despite the constrained access to alternative partners/markets.

In Jan 2025, they modified the terms: the OpenAI API remains exclusive to Azure, but new capacity moved to the right of first refusal model (allowing some non-Azure procurement), a contractual tweak that allows OpenAI to explore other cloud providers for new capacity, but Microsoft has the opportunity to match the terms. Microsoft is no longer OpenAI's exclusive cloud provider, even as API exclusivity persists (31). By mid-2025, Reuters reported that OpenAI added Google Cloud capacity (alongside Microsoft, Oracle, and CoreWeave), demonstrating bounded multi-provider capacity under ROFR, as a move to reduce their dependency/diversify (32). However, the ROFR was plausibly agreed to by Microsoft because it reduces Microsoft's obligation to shoulder all incremental capacity (to reduce their costs) while still preserving nearly the same amount of leverage (the choke points being the distribution and their claim on first capacity) through API exclusivity and first refusal on new capacity. For OpenAI, the benefit is a contractual hedge that partially, albeit in a limited manner, restores multi cloud optionality under scarcity, with the constraint of their API monetization continuing to route through Azure, and limits the impact of the constraints discussed in Park & Steensma.

In February 2023, Microsoft announced the launch of "new bing", which was confirmed to run on GPT 4 (33). Microsoft 365 was announced a month later embedding OpenAI models inside Word, Excel, Outlook, PowerPoint and Teams, effectively turning the Microsoft 365 installed base into the commercialization engine (25). GitHub Copilot illustrates an even earlier distribution, with the technical preview in June 2021 and general availability for business on Feb 14, 2023 (34). Routing capability first through Bing/M365/GitHub provided instant reach to a large number of users and enterprise credibility, accelerating adoption of their technology via Microsoft's channels (strategic alignment for Microsoft), aligning with the discussion by de Vaan & Stuart on higher trust and lower search costs (leading to faster adoption) but come with the risk of lower match quality. Moreover, Microsoft first sequencing (Microsoft's confirmation that GPT 4 ran in Bing before public acknowledgment, evidencing early internal access and rollout control) (33) can defer/slow down OpenAI direct routes in the release of products (consumer or API) even when the capability is ready. Although the consequences may not have necessarily been observed, in the fast cycle LLM market, even short lags matter for market share (primarily enterprise but mass market clients as well), reputation/brand image and revenue. These observations align with the outcomes that corporate logic would be expected to bring.

In May 2020 Microsoft unveiled an Azure supercomputer for OpenAI, which is estimated to involve approximately 10,000 NVIDIA V100 GPUs and 285,000 CPU cores, purpose built to train extremely large models (35). Azure then released Nvidia A100 clusters and, later, H100 fleets in general availability (36-37). The ability for them to integrate the GPUs at scale, given the cost of Nvidia GPUs from 2023-2025 (9-10), along with the Azure cloud credits provided in dedicated capacity (an estimated 10 billion dollars' worth of contribution till 2023), a computing subsidy that anchored training and inference on Azure and making sustained frontier scale R&D financially feasible (38), is a distinctive resource that other types of funding partners (venture capital) would simply have been unable to provide. The reasons for this are the access to the Nvidia GPUs at scale (both due to immense capital cost and the inability to procure

the GPUs), the ability to integrate it appropriately and Microsoft's control of the infrastructure at scale, thus making this a resource type (the computing infrastructure), specifically the mechanism of access to complementary assets, that enabled them sustain cutting edge technical innovation which would simply not have been achievable/accessible without one of a limited selection of corporate partners (in this case Microsoft).

A related consequence to the commercialization is OpenAI withholding architecture, computing, data and training method (weights) details in its public documentation, diverging from the 2018-2020 "open lab" practices and consistent with tighter IP control when commercialization is routed through a single dominant channel (39). The combination of Microsoft first deployment and Azure anchored monetization makes the option of being an open research lab, explicitly stated as one of their goals, unavailable if they wanted to secure Microsoft as partners as otherwise Microsoft would not have had much of a strategic reason to partner with OpenAI (External commentary at the time also underscored this shift toward opacity) (40).

Independent VCs (professional logic) reliably impose milestone discipline and time pacing (1-2), whereas corporates (corporate logic) supply complements and mediate access through their own structures and contracts. In OpenAI's case, the resource bundle included capital, computing, and channels of distribution as resources, a bundle that accelerated both commercialization (Bing/M365/GitHub first) and technical innovation (frontier training runs on corporately controlled infrastructure) (26, 30, 35). However, as discussed, this led to concentrating dependence (exclusive workloads, then API exclusivity) on Microsoft and immense costs and leverage that they likely would not have incurred in the possibility of accessing the same or similar resources through a VC partner (26, 40). This indicates that when a critical resource (for innovation or otherwise) is scarce, expensive, or practically accessible only through a limited selection of resource providers, despite the disadvantages brought about by corporate logic, as discussed in Pahnke, the potential upside when those resources are accessible (in this case through contractual agreements) can be net accelerative and lead to significantly improved outcomes even as it narrows autonomy (3-4, 35, 37).

COHERE (HYBRID LOGIC(S))

Pahnke, Katila & Eisenhardt (2015) notes that hybrid logics, or ties to both VC and CVC partners, lead to improved technical innovation (2). However, the paper focuses on a single, mature industry and leaves the discussion of hybrid logics (VC and corporate ties) and the discussion of specific mechanisms/contextual factors underdeveloped. (2) Similarly, Park & Steensma does not explore in depth the internal venture processes through which these mechanisms are enacted, nor does it explore the implications of the presence of multiple funding partners (hybrids) (3).

Moreover, something that is underdeveloped in established literature may be the specific exploration of how partner logics interact with and shape venture outcomes, especially under hybrid logics, where theoretical discussion regarding the impacts remain bleak. This paper builds on existing literature by adding more depth to the specific mechanisms through which distinct partner logics operate and influence outcomes, particularly on the potential value of corporate partners on firm outcomes and discusses the relevance of institutional logics in the selection of resource providers in the AI industry (2-3, 5-6).

Cohere, founded 2019 by Aidan Gomez, Ivan Zhang and Nick Frosst, is an example of hybrid logics, a VC led governance coalition along with deeply embedded corporate partners for computing and distribution (41). In Cohere's case, the board and staged financing are primarily through VCs, where Index Ventures led Series A in 2021 and took a board seat (Mike Volpi) (42), with Radical Ventures also on board (43), embedding professional logic attention to pacing and commercialization inside the firm. Cohere has many large enterprises and cloud partners (Oracle, AWS, Salesforce, Nvidia etc), each bringing corporate complements (distribution, compliance layers, and hyperscale computing) at various stages, but structured without exclusivity, which is part of Cohere's risk mitigation strategy (44). Cohere structures these ties without exclusivity to preserve access to alternative partners and markets while still securing specialized complements, to reduce the constraint the tight single corporate ties create which Park & Steensma discusses (3).

Cohere deliberately avoided exclusive cloud deals and "worked to provide technology through multiple cloud platforms" and avoiding taking heavy financing "in the form of credits" tied to a single provider (45), a stance it has reiterated in 2024, stating it has "steered clear of exclusive deals with cloud providers" despite its close ties with Oracle (46). Rather contracting for computing and distribution using several corporate partners, namely Oracle Cloud Infrastructure (OCI) for training

and enterprise distribution through Oracle's channels and AWS retail and enterprise distribution via Amazon Bedrock and SageMaker JumpStart (Cohere has used Google cloud a significant amount to train its LLMs on their TPU supercomputers as well, however this was structured as a "technology partnership" as opposed to an investment, making it so Google is not a funding partner) (47) . Cohere uses contractual agreements to maintain flexibility and reduce dependency on providers of critical resources, which aligns with professional logic to allow them a greater degree of autonomy in terms of obtaining more favorable deals and pacing in terms of the ability to control the quantity of resources more freely, while still allowing them access to the distinctive resource type(s) that venture capital provider could not have (mitigating downside of corporate logic).

In mid-2023, Oracle and Cohere launched OCI generative AI and integrated Cohere's models in Oracle Fusion Cloud Applications and NetSuite involving use cases in finance, HR, supply chain, sales and services etc., stating over 50 embedded use cases (48). In June 2023, Salesforce ventures announced their investment into Cohere (49), embedding Cohere's LLMs (among others) in its Einstein GPT (50), and providing Cohere as an option in its selectable AI provider in Slack (51). AWS distribution channels happen through Bedrock and Jumpstart (52-53). The wide range of distribution channels allows Cohere to have broad enterprise reach without having to build their own SaaS platforms/ stacks. The pattern aligns with corporate logic, with corporate partners routing new technologies into their systems to capture value and also with the fast paced, commercialization pressures typical of professional logic. Analogous to the referral strength theory by de Vaan & Stuart, Cohere combines the assurance heavy (higher trust from partners' reputation) embedded pathways for commercialization through their corporate partners, leading to faster adoption, and the broader pathways (e.g. the API) that professional logic would push them toward, which allows them to benefit from the resource networks of both their corporate and venture capital partners. The 'hybrid' approach introduces significant risk of channel capture/dependence on partner roadmaps (due to the effect of corporate logic on commercialization, potentially losing the ability to monetize self-sufficiently effectively), however the use of distribution channels from several corporate partners mitigates this risk substantially. Moreover, likely attributable to the "checks and balances" brought about by professional logic (2), this approach significantly

improves adoption/reach (large base of enterprise clients through these distribution channels), allowing Cohere to benefit from the complementary assets which enhance commercialization of corporate partners, illustrating the effectiveness of hybrid logic.

In this instance, at least in the short run the corporate distribution is a significant benefit as it is hard to attract clients at scale (and therefore generate revenue) offering it (their models) as an independent (standalone) service; however, as discussed, in the long run the development of dependency is a substantial risk especially if the strategic goals of a corporate partner change (2, 4).

Moreover, because hybrids and Cohere specifically retain significant involvement from VCs (per their basis of strategy) (2) in the form of board seats despite not having majority stakes (1), they carry the same indirect competitor exposure that Pahnke, Mcdonald, Wang & Hallen document, which indicates that information leakage from their VC partners can harm their innovation (11).

Cohere's product roadmap shows priority for features that suit the needs of enterprise clients and the platform partners (strategic alignment with its corporate partners). For example, the large context windows (128k tokens) (54), tenancy isolation (in the form of dedicated clusters on OCI and AWS) (55) (tenancy isolation to ensure privacy for enterprise clients and business use cases) and following AWS' extensive privacy compliances (56) are assurances which enable better commercialization by making enterprise sales cycles faster and entry into regulated/high compliance markets (expanding TAM) which also align with basis of attention of professional logic, the hybridization and balance between professional and corporate logic can be observed here. This also fits the framework used in de Vaan & Stuart, in that the young firm would likely have higher trust and act faster to align with their corporate partners' distribution channels' needs, thus shortening enterprise sales cycles.

However, the influence of their corporate partners results in the specific product roadmaps of the corporate partners (alignment) influencing Cohere's priorities, for example the trust layer integration on Salesforce (Even Salesforce refers to the data masking involved as a "double edged sword") (57), privacy measures such as tenancy isolation on OCI and AWS bedrock privacy requirements imposing significant, though not necessarily an extreme reallocation of focus to ensure compliance (58), an example of the coordination frictions Pahnke ties to corporate logic. Together, these illustrate the impact of hybrid, or rather dual logic pressures that require

compliance with multiple ecosystems, however this is an instance where the logics "complement each other with appropriate checks and balances" (2) and actually work to Cohere's benefit due to the aforementioned benefits in improving commercialization.

The functionality of corporate partners as a means of accessing resources that would have been otherwise inaccessible through other resource providers (VCs) is evident when discussing the computational resources and specialized knowledge (as per Pahnke et.al.,) that corporate partners carry (complementary assets such as computational resources at scale, GPU access, specialized knowledge) (2). Specifically, Oracle provides the highest end computing infrastructure for model training and inferencing (bare metal H100/H200/Blackwell fleets, high performance cluster networking etc.) and for enterprise distribution and AWS offering H100 instances for deployment to consumers (59-61), which aid technical innovation dramatically. Furthermore, Cohere placed models on the Nvidia API catalogue (62), Nvidia, stating that it has been working closely with Cohere (among other companies) for the TensorRT-LLM optimization program, which intends to optimize LLM inference (63). The benefits of this are observable through Cohere's 2025 documentation for the Command A, which emphasizes small GPU count serving (as few as two A100 or H100 GPUs for the base) and illustrates the drastic improvements in efficiency (64) (an R&D program geared to enterprise scale efficiency rather than brute force frontier scaling), in simpler words significant improvements in technical innovation at much lower costs.

Moreover, the approach of nonexclusive multi cloud contracting brought about by the influence of professional logic (although they do have significant dependence on Oracle) partially mitigates the risks associated with a prominent corporate partner by improving bargaining power and flexibility in the long run. However, it is notable that in order to pursue this there is a tradeoff in maintaining multi-platform relationships that creates overhead costs in separate integrations, certifications and security mandates as well as keeping feature parity across partner APIs and deployment targets. This is exacerbated in the case of private deployments, where the cohere must support each customer in managing infrastructure, drivers, and containerized stacks (65).

Corporate partners often hold valuable resources and complementary assets, however as emphasized in Pahnke, katila and Eisenhardt (2015) accessing those resources is fettered by corporate routines and internal coordination frictions (2). A method to reduce the risk

of being unable to or to reduce the friction to access the vast resources corporate partners bring can be through the use of contracting, which includes built in guarantees for access to those resources, as seen in both the OpenAI case and the Cohere case. A further benefit to this, is that the use of non-exclusive contracts can help firms to reduce dependency on its corporate partner(s) and allow them to explore the open market for resources, one of the most significant drawbacks of corporate logics noted in Park & Steensma. However, a downside to hybrid funding is that it leads to indirect ties to competitors that may create indirect information leakage similar to that of only having VC partners, as opposed to exclusively having corporate partners.

Hybrid logics can improve outcomes for young ventures because the different logics counterbalance each other, the "appropriate checks and balances" Pahnke discusses. Corporate partners have extremely valuable resources and complementary assets that can enhance commercialization of new ventures (3), and the disciplined focus that professional logic brings per their basis of attention, can help young firms benefit from and enjoy greatly improved outcomes from these.

CROSS CASE COMPARISON

A comparison of DeepL (professional logic), OpenAI (corporate logic), and Cohere (hybrid logic) shows that funding-partner logics primarily reshape outcomes by changing access to complementary assets, the constraints placed on autonomy, and the route through which commercialization is achieved (Table 1). In the OpenAI case, corporate partnering is decisive because it provides critical complements, particularly computing and distribution which supports frontier-scale technical progress and a fast partner-routed route to market, but it also concentrates strategic dependence through platform alignment and routing constraints. DeepL reflects a contrasting configuration in which professional logic supports disciplined scaling through enterprise productization and assurance-building, resulting in particularly strong commercial innovation. However, this involves lower autonomy in day-to-day governance (2), since high VC involvement and milestone pacing can shape managerial decision making directly. Cohere's hybrid structure combines VC discipline with access to a portfolio of corporate complements, using non-exclusive, multi-partner relationships to secure much of the same critical resources as OpenAI, without involving a significant loss in strategic autonomy due to the multi-

Table 1. Comparison of outcomes on notable factors across funding partner types.

	Resource types provided	Autonomy	Innovation impact	Commercialization pathways	Risks / limitations
Professional logic (DeepL)	Financial capital, governance expertise, scaling legitimacy, market signaling	Low (high direct governance influence; milestone pacing)	Technical: Insignificant Commercial: Very High	Governance/pacing only (no distribution channel)	Intermediation exposure
Corporate logic (OpenAI)	Bottleneck complements (computing capacity, distribution platforms), ecosystem access, infrastructure embedding	Medium (less day-to-day control; higher strategic dependence)	Technical: Very High Commercial: High	Partner routed (fast)	Lock-in; alignment pressure; constrained autonomy
Hybrid logic (Cohere)	Mixed complement portfolio (multi-channel infrastructure/distribution) with VC discipline and legitimacy	Medium (reduced single-partner lock-in; higher governance complexity)	Technical: High Commercial: High	Multi-channel partner route (fast and diversified)	Coordination burden; dual accountability; Intermediation exposure

Note regarding autonomy ratings: They refer to the firm's strategic discretion over core technical and commercialization decisions given its funding partner(s). Low indicates reduced autonomy due to high day-to-day governance involvement, including direct control over decision-making and milestone pacing. Whereas, the rating of 'medium' indicates less day to day oversight but high strategic constraints. Further explanation on the same can be found in the individual case studies.

partner strategy which supports both enterprise-oriented technical and commercial innovation. VC involvement does, however, reduce autonomy in operations to some extent.

The cases also imply distinct levels of risk. Corporate logic can accelerate deployment through embedded channels but creates the risk of dependence, and the pressure of alignment with the corporate partners reduces autonomy. Professional logic carries intermediation exposure and can constrain autonomy through direct governance influence. Whereas, Cohere's structure reduces single-partner lock-in by diversifying routes to market, but it introduces coordination burden and dual accountability as the venture must satisfy both VC governance expectations and corporate integration demands.

CONCLUSION

Across the cases, it is clear that the institutional logics of funding partners shape young firms and their outcomes through far more mechanisms than the provision of capital. The most notable mechanisms across the logics relate to the use of contracting, degree of control, access to complementary assets, partner mediated routing of opportunities and assurance/legitimization routines.

The DeepL case illustrates the functioning of professional logic in the AI industry. The emphasis on disciplined commercialization, milestone-paced development and adjacent productization (commercialization benefits) focused on the strengths of the venture are clearly observable. The benefits to this strategic discretion and speed are in shaping product scope and pricing without having to align to a corporate partner's strategic needs. However, enterprise credibility had to be built through the firm's own assurance investments (audits and certifications) and accumulated market reputation. Moreover, the structural exposure associated with professional logics in the form of indirect exposure to competitors and the potential for information leakage may be a problem. Overall, the case exemplifies a setting where professional logic delivers commercialization and discretion and the observable outcomes and analysis yield outcomes that align well with established literature with slight nuance brought about by the nature of the industry being analyzed.

The OpenAI case illustrates how corporate logic functions in the AI industry. Because of the strategic alignment that corporate partners seek and the constraints that corporate logics impose, the compromise

is observable on OpenAI's optionality and autonomy in making strategic choices, and the associated costs, such as in their product roadmap, control over the release of IP, release of new technologies, and the constraints on their distribution (exclusivity) and training channels (dependence). However, despite the drawbacks of corporate logic, the access to and scale of the critical resources and complements that are only available within the vast resources of a corporate partner (which otherwise would not have been accessible), as well as the commercialization benefits (2) brought about through their ties to a corporate partner it has aided them in achieving higher technical innovation and led to generally improved outcomes.

This contrasts with established literature regarding the value of corporate partners relative to their venture capital counterparts, and supports the view that corporate partnering can be substantially beneficial when specialized complements and commercialization channels are unusually decisive. The Cohere case illustrates the enactment of hybrid logics in the AI industry. Hybrid logics, as a combination of corporate and venture capital partner logics, can result in appropriate checks and balances that allow favorable exploitation of the benefits brought about by each partner, as well as the mitigation of the costs of both, but incur costs of added coordination. The drawbacks of corporate logic still hold, namely in the compromise of strategic autonomy through the frictions and costs associated with the alignment with the ecosystems of their corporate partners, shaping of their product roadmap, reduced optionality in the distribution channels and higher bureaucracy due to the assurance routines required for various corporate partner channels. However, the hybrid logics are able to significantly mitigate the impacts and severity of these drawbacks, while still allowing Cohere to access the critical resources and complements that could not have been without the corporate partners, which has allowed them to achieve both higher levels of commercial and technical innovation. This is achieved through the professional logic of VCs providing the appropriate "checks and balances" to the corporate logic of the corporate partners, namely through the use of nonexclusive, multi partner contracting to maintain optionality and strategic leverage, which allows them to secure favorable terms for computing, distribution and integration (through their ability to access the "open market" (2), the basis of attention and strategy of venture capital improving their discipline and commercialization while maintaining their strategic flexibility.

CONFLICTS OF INTEREST

The author declares that there are no conflicts of interest regarding the publication of this article.

REFERENCES

1. Kaplan SN, Strömberg P. Financial contracting theory meets the real world: an empirical analysis of venture capital contracts. *Rev Econ Stud.* 2003; 70 (2): 281-315. doi:10.1111/1467-937X.00245.

2. Pahnke EC, Katila R, Eisenhardt KM. Who takes you to the dance? How partners' institutional logics influence innovation in young firms. *Adm Sci Q.* 2015 Dec; 60 (4): 596-633. https://doi.org/10.1177/0001839215592913

3. Park HD, Steensma HK, Kim PH. When does corporate venture capital add value for new ventures? *Strateg Manag J.* 2012; 33 (1): 1-22. https://doi.org/10.1002/smj.937

4. Dushnitsky G, Lenox MJ. When does corporate venture capital investment create firm value? *J Bus Ventur.* 2006 Nov; 21 (6): 753-772. https://doi.org/10.1016/j.jbusvent.2005.04.012

5. Thornton PH, Ocasio W, Lounsbury M. The institutional logics perspective: a new approach to culture, structure, and process. Oxford: Oxford University Press; 2012. https://doi.org/10.1093/acprof:oso/9780199601936.001.0001

6. Besharov ML, Smith WK. Multiple institutional logics in organizations: explaining their varied nature and implications. *Acad Manage Rev.* 2014; 39 (3): 364-381. https://doi.org/10.5465/amr.2011.0431

7. Stanford Institute for Human-Centered Artificial Intelligence (HAI). Artificial intelligence index report 2024 [Internet]. Stanford: HAI; 2024. Available from: https://hai-production.s3.amazonaws.com/files/hai_ai-index-report-2024-smaller2.pdf (accessed on 2025-12-16).

8. Stanford Institute for Human-Centered Artificial Intelligence (HAI). Artificial intelligence index report 2025: chapter 1 research and development [Internet]. Stanford: HAI; 2025. Available from: https://hai.stanford.edu/assets/files/hai_ai-index-report-2025_chapter1_final.pdf (accessed on 2025-12-16).

9. Cottier B, et al. How much does it cost to train frontier AI models? [Internet]. Epoch AI; 2024. Available from: https://epoch.ai/blog/how-much-does-it-cost-to-train-frontier-ai-models (accessed on 2025-12-16).

10. Nellis S, et al. Nvidia's supply snags limit deliveries even as demand booms. Reuters [Internet]. 2024. Available from: https://www.reuters.com/technology/

nvidias-supply-snags-hurting-deliveries-mask-booming-demand-2024-11-21/ (accessed on 2025-12-16).

11. Pahnke EC, Katila R, Eisenhardt KM. Exposed: venture capital, competitor ties, and entrepreneurial innovation. *Acad Manage J.* 2015; 58 (5): 1334-1360. https://doi.org/10.5465/amj.2012.0777

12. de Vaan M, Stark D, Vedres B. Referral triads. *Adm Sci Q.* 2025; 70 (1): 157-193. https://doi.org/10.1177/00018392241296140

13. DeepL Translator. In: Wikipedia [Internet]. Available from: https://en.wikipedia.org/w/index.php?title=DeepL_Translator&oldid=1323198825 (accessed on 2025-11-16).

14. Taylor Wessing. Taylor Wessing advises Benchmark on investment in DeepL [Internet]. 2022. Available from: https://www.taylorwessing.com/fr/insights-and-events/news/media-centre/press-releases/2018/12/taylor-wessing-bert-benchmark-bei-investment-in-deepl (accessed on 2025-11-16).

15. Lunden I. DeepL, the AI-based language translator, raises over $100M at a $1B+ valuation. TechCrunch [Internet]. 2023. Available from: https://techcrunch.com/2023/01/11/deepl-the-ai-based-language-translator-raises-over-100m-at-a-1b-valuation/ (accessed on 2025-11-16).

16. DeepL. DeepL announces $300 million investment at $2 billion valuation fueled by global demand for AI language solutions [Internet]. 2024. Available from: https://www.businesswire.com/news/home/20240522689624/en/DeepL-Announces-%24300-Million-Investment-at-%242-Billion-Valuation-Fueled-by-Global-Demand-for-AI-Language-Solutions (accessed on 2025-11-16).

17. Heise Online. DeepL Pro: Neuer Aboservice für Profi-Übersetzer, Firmen und Entwickler [Internet]. Available from: https://www.heise.de/news/DeepL-Pro-Neuer-Aboservice-fuer-Profi-Uebersetzer-Firmen-und-Entwickler-3998601.html (accessed on 2025-11-16).

18. Lunden I. AI language translation startup DeepL nabs $300M on a $2B valuation to focus on B2B growth. TechCrunch [Internet]. 2024. Available from: https://techcrunch.com/2024/05/22/deepl-the-ai-language-translation-startup-nabs-300m-on-a-2b-valuation-to-focus-on-b2b-growth/ (accessed on 2025-11-16).

19. DeepL. DeepL translator learns Japanese and Chinese [Internet]. 2020. Available from: https://www.deepl.com (accessed on 2025-11-16).

20. Marking M. DeepL adds 13 European languages as traffic continues to surge. Slator [Internet]. 2021. Available from: https://slator.com/deepl-adds-13-european-languages-as-traffic-continues-to-surge/ (accessed on 2025-11-16).

21. DeepL. DeepL launches new AI-powered offering, DeepL Write Pro, to supercharge business communication [Internet]. 2024. Available from: https://www.businesswire.com/news/home/20240425816210/en/DeepL-Launches-New-AI-powered-Offering-DeepL-Write-Pro-to-Supercharge-Business-Communication (accessed on 2025-11-16).

22. Nelson E. A step-by-step guide to getting a SOC 2 report [Internet]. Hyperproof; 2024. Available from: https://hyperproof.io/resource/a-step-by-step-guide-to-soc-2-report/ (accessed on 2025-11-16).

23. DeepL. DeepL Pro expands global reach to 165 new markets [Internet]. 2024. Available from: https://www.deepl.com (accessed on 2025-11-16).

24. Metinko C. Generative AI translation startup DeepL locks up $300M. Crunchbase News [Internet]. 2024. Available from: https://news.crunchbase.com/ai/deepl-generative-ai-language-translation-music-startups-funding-suno/ (accessed on 2025-11-16).

25. OpenAI. Our structure [Internet]. Available from: https://openai.com/our-structure/ (accessed on 2025-11-16).

26. Microsoft Source. OpenAI forms exclusive computing partnership with Microsoft to build new Azure AI supercomputing technologies [Internet]. 2019. Available from: https://news.microsoft.com/source/2019/07/22/openai-forms-exclusive-computing-partnership-with-microsoft-to-build-new-azure-ai-supercomputing-technologies/ (accessed on 2025-11-16).

27. Microsoft Blogs. Microsoft and OpenAI extend partnership [Internet]. 2023. Available from: https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/ (accessed on 2025-11-16).

28. Rennolds N. OpenAI valued at $80 billion following deal that allows staff to cash out their shares, report says. Bus Insider [Internet]. 2024. Available from: https://www.businessinsider.com/openai-valued-at-80bn-following-deal-employees-cash-out-report-2024-2 (accessed on 2025-11-16).

29. Spataro J. Introducing Microsoft 365 Copilot - your copilot for work. Microsoft Blog [Internet]. 2023. Available from: https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/ (accessed on 2025-11-16).

30. Scott K. Microsoft teams up with OpenAI to exclusively license GPT-3 language model. Microsoft Blog [Internet]. 2020. Available from: https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/ (accessed on 2025-11-16).

31. Wiggers K. Microsoft is no longer OpenAI's exclusive cloud provider. TechCrunch [Internet]. 2025. Available from: https://techcrunch.com/2025/01/21/microsoft-is-no-longer-openais-exclusive-cloud-provider/ (accessed on 2025-11-16).

32. OpenAI lists Google as cloud partner amid growing demand for computing capacity. Reuters [Internet]. 2025. Available from: https://www.reuters.com/business/media-telecom/openai-lists-google-cloud-partner-amid-growing-demand-computing-capacity-2025-07-16/ (accessed on 2025-11-16).

33. Microsoft Bing. Confirmed: the new Bing runs on OpenAI's GPT-4 [Internet]. 2023. Available from: https://blogs.bing.com/search/march_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4 (accessed on 2025-11-16).

34. Dohmke T. GitHub Copilot for Business is now available. GitHub Blog [Internet]. 2023. Available from: https://github.blog/news-insights/product-news/github-copilot-for-business-is-now-available/ (accessed on 2025-11-16).

35. Moorhead P, Strategy MI. Microsoft builds massive supercomputer for OpenAI. Forbes [Internet]. 2020. Available from: https://www.forbes.com/sites/moorinsights/2020/05/20/microsoft-builds-massive-supercomputer-for-openai/ (accessed on 2025-11-16).

36. Raines K. Azure high-performance computing at Supercomputing 2020. Microsoft Azure Blog [Internet]. 2020. Available from: https://azure.microsoft.com/blog/azure-high-performance-computing-supercomputing-2020/ (accessed on 2025-11-16).

37. Microsoft Tech Community. New Azure NC H100 v5 VMs optimized for generative AI and HPC workloads [Internet]. Available from: https://techcommunity.microsoft.com/blog/azurehighperformancecomputingblog/new-azure-nc-h100-v5-vms-optimized-for-generative-ai-and-hpc-workloads-is-now-ge/4087034 (accessed on 2025-11-16).

38. OpenAI has received just a fraction of Microsoft's $10 billion investment. Semafor [Internet]. 2023. Available from: https://www.semafor.com/article/11/18/2023/openai-has-received-just-a-fraction-of-microsofts-10-billion-investment (accessed on 2025-11-16).

39. OpenAI, et al. GPT-4 technical report [Internet]. arXiv; 2024. Available from: http://arxiv.org/abs/2303.08774 (accessed on 2025-11-16).

40. Xiang C. OpenAI's GPT-4 is closed source and shrouded in secrecy. VICE [Internet]. 2023. Available from: https://www.vice.com/en/article/openais-gpt-4-is-closed-source-and-shrouded-in-secrecy/ (accessed on 2025-11-16).

41. Cohere. About our company [Internet]. Available from: https://cohere.com/about (accessed on 2025-12-16).

42. Cohere. Cohere raises $40 million in Series A financing. GlobeNewswire [Internet]. 2021. Available from: https://www.globenewswire.com/news-release/2021/09/07/2292856/0/en/Cohere-Raises-40-Million-in-Series-A-Financing-to-Make-Natural-Language-Processing-Safe-and-Accessible-to-Any-Business.html (accessed on 2025-12-16).

43. Bresge A. Jordan Jacobs on AI's implications and why he backed Cohere. Univ Toronto News [Internet]. 2024. Available from: https://www.utoronto.ca/news/jordan-jacobs-ai-s-profound-implications-responsible-development-and-why-he-backed-cohere (accessed on 2025-12-16).

44. AI startup Cohere in talks to raise roughly $500M to $1B. Reuters [Internet]. 2024. Available from: https://www.reuters.com/technology/ai-startup-cohere-talks-raise-roughly-500mln-1bln-source-2024-01-19/ (accessed on 2025-11-16).

45. Hu K, et al. Nvidia and Salesforce double down on AI startup Cohere in $450 million round. Reuters [Internet]. 2024. Available from: https://www.reuters.com/technology/nvidia-salesforce-double-down-ai-startup-cohere-450-million-round-source-says-2024-06-04/ (accessed on 2025-11-16).

46. Google Cloud. Accelerating language model training with Cohere and Google Cloud TPUs [Internet]. Available from: https://cloud.google.com/blog/products/ai-machine-learning/accelerating-language-model-training-with-cohere-and-google-cloud-tpus (accessed on 2025-11-16).

47. Oracle. Oracle adds new generative AI capabilities to Oracle Fusion Cloud Applications Suite [Internet]. 2024. Available from: https://www.oracle.com/news/announcement/oracle-adds-generative-ai-capabilities-to-fusion-applications-2024-03-14/ (accessed on 2025-11-16).

48. Salesforce Ventures. Welcome, Cohere! [Internet]. Available from: https://salesforceventures.com/perspectives/welcome-cohere/ (accessed on 2025-11-16).

49. Salesforce. Salesforce announces AI Cloud [Internet]. 2023. Available from: https://www.salesforce.com/news/press-releases/2023/06/12/ai-cloud-news/ (accessed on 2025-11-16).

50. Slack. Cohere AI [Internet]. Available from: https://slack.com/marketplace/A056U9XEQ1W-cohere-ai (accessed on 2025-11-16).

51. Amazon Web Services. Cohere models - Amazon Bedrock [Internet]. Available from: https://docs.aws.amazon.com/bedrock/latest/userguide/model-parameters-cohere.html (accessed on 2025-11-16).

52. Amazon Web Services. Available foundation models - Amazon SageMaker AI [Internet]. Available from: https://docs.aws.amazon.com/sagemaker/latest/dg/jumpstart-foundation-models-latest.html (accessed on 2025-11-16).

53. Oracle. Cohere Command R+ (08-2024) [Internet]. Available from: https://docs.oracle.com/en-us/iaas/Content/generative-ai/cohere-command-r-plus-08-2024.

htm (accessed on 2025-11-16).

54. Oracle. Overview of generative AI service [Internet]. Available from: https://docs.oracle.com/en-us/iaas/Content/generative-ai/overview.htm (accessed on 2025-11-16).

55. Amazon Web Services. Data protection - Amazon Bedrock [Internet]. Available from: https://docs.aws.amazon.com/bedrock/latest/userguide/data-protection.html (accessed on 2025-11-16).

56. Sanghavi T. LLM data masking: silver bullet or double-edged sword? Salesforce Blog [Internet]. 2025. Available from: https://www.salesforce.com/blog/llm-data-masking/ (accessed on 2025-11-16).

57. HashiCorp. 2024 HashiCorp state of cloud strategy survey [Internet]. Available from: https://www.hashicorp.com/en/state-of-the-cloud (accessed on 2025-11-16).

58. Cohere. Deployment options: SaaS, cloud API, VPC, on premise [Internet]. Available from: https://cohere.com/deployment-options (accessed on 2025-11-16).

59. Oracle. Oracle bare metal GPU service [Internet]. Available from: https://www.oracle.com/cloud/compute/gpu/ (accessed on 2025-11-16).

60. Amazon Web Services. Cohere Command R and R+ available in Amazon SageMaker JumpStart [Internet]. 2024. Available from: https://aws.amazon.com/blogs/machine-learning/cohere-command-r-and-r-are-now-available-in-amazon-sagemaker-jumpstart/ (accessed on 2025-11-16).

61. Cohere. Cohere's Command R Enterprise model coming to ai.nvidia.com [Internet]. Available from: https://cohere.com/blog/cohere-launch-nvidia (accessed on 2025-11-16).

62. NVIDIA. NVIDIA TensorRT-LLM supercharges large language model inference on NVIDIA H100 GPUs [Internet]. 2023. Available from: https://developer.nvidia.com/blog/nvidia-tensorrt-llm-supercharges-large-language-model-inference-on-nvidia-h100-gpus/ (accessed on 2025-11-16).

63. Cohere. Release notes [Internet]. Available from: https://docs.cohere.com/changelog (accessed on 2025-11-16).

64. Cohere. Private deployment overview [Internet]. Available from: https://docs.cohere.com/docs/private-deployment-overview (accessed on 2025-11-16).