

The Role of GenAI Startups in Reshaping Business Consulting

Massimiliano Dell’Orto, Giuseppe Guercio, Mattia Magnaghi, Antonio Ghezzi and Andrea Rangone

Politecnico di Milano, Department of Management, Economics and Industrial Engineering, Milan, Italy

massimiliano.luigi.dellorto@polimi.it

giuseppe.guercio@mail.polimi.it

mattia.magnaghi@polimi.it

antonio.ghezzi1@polimi.it

andrea.rangone@polimi.it

Abstract: This study examines how GenAI-native startups enter and position themselves in knowledge-intensive professional services and how their emergence reconfigures disruption dynamics. It adopts a qualitative-dominant multi-method design combining a dataset of 177 consulting-oriented GenAI startups with three semi-structured key-informant interviews. Three patterns emerge. First, startup formation accelerates after foundation models become widely accessible through APIs and cloud infrastructure, suggesting that GenAI operates as enabling infrastructure for a category of entrants rather than as a proprietary innovation owned by individual firms. Second, the evidence does not indicate displacement of incumbents. Startups occupy an intermediate position between consultancies and generalist AI providers, differentiating through domain expertise and deployment-ready applications while depending on upstream model suppliers that are also latent competitors. Third, the three cases combine recurring software revenues with enterprise agreements and usage-dependent infrastructure costs, producing heterogeneous commercial responses, including performance-based pricing in one case. The study qualifies Big Bang Disruption for knowledge-intensive professional services: its supply-side conditions hold, whereas near-zero marginal costs and immediate adoption do not. It also specifies the properties and boundary conditions of a triadic competitive configuration. The findings inform startup positioning above the model layer, cooperative incumbent responses, and investment in a capital-concentrated market.

Keywords: Generative AI, GenAI Startups, Business Consulting, Big Bang Disruption

1. Introduction

Generative artificial intelligence (GenAI) is reshaping the boundaries of knowledge-intensive work, extending automation beyond structured and codifiable tasks to synthesis, interpretation, and content generation. Business consulting represents a paradigmatic case: an industry historically reliant on human capital, institutional knowledge, and client relationships, and therefore resistant to prior forms of automation (Kubr, 2002). Recent estimates place the potential annual contribution of generative AI to the global economy between \$2.6 and \$4.4 trillion, with professional services among the most exposed domains (McKinsey & Company, 2023). Consulting firms are already integrating AI-enabled workflows to improve efficiency and redesign service delivery, and the conditions for competitive disruption in the industry are now in place.

A central feature of this transformation is the entry of a new class of market participants. GenAI-native startups are translating advances in large language models into targeted applications for specific consulting activities, entering a market long dominated by incumbent firms such as the Big Four and global strategy consultancies. At the same time they operate under competitive pressure from large AI infrastructure providers, including OpenAI, Google, and Anthropic, whose general-purpose models increasingly overlap with specialised solutions. The result is a triadic competitive structure in which new entrants sit between incumbent service firms and generalist GenAI providers.

Existing literature addresses these dynamics only partially, and from three directions. Research on GenAI and the future of work documents the exposure of knowledge-intensive professions to automation (Eloundou et al., 2023; Brynjolfsson et al., 2023), but at the macro level of aggregate employment rather than firm-level competition. Disruption research explains how technological discontinuities enable entrants to challenge incumbents (Christensen, 1997; Downes & Nunes, 2013), yet has been applied largely to product markets. Work on consulting characterises the pyramid delivery model and its knowledge economics (Hansen et al., 1999; Kubr, 2002) without addressing entrants whose core capability is externally supplied. What is not known is therefore specific: how GenAI-native startups position themselves within professional service markets, which consulting activities they target, and how their value propositions differ from both incumbent firms and generalist AI providers. This matters theoretically, because disruption frameworks derived from product markets may not

hold where expertise and trust govern exchange, and practically, because incumbents, entrants and investors are committing resources under competing assumptions about how this market will consolidate.

The study is guided by the following research question: "How do GenAI-native startups enter and position themselves within knowledge-intensive professional service markets, and how does their emergence reconfigure traditional disruption dynamics?" It addresses this question through an exploratory qualitative investigation combining descriptive analysis of an original dataset of 177 GenAI startups with semi-structured founder interviews. The findings indicate that these firms are not displacing consulting incumbents but are instead occupying an intermediate position between them and generalist AI infrastructure providers, entering as the enabling infrastructure becomes available and adapting their commercial models to the variable costs it imposes. Theoretically, the study modifies Big Bang Disruption for knowledge-intensive professional services, where its supply-side conditions hold but its cost and adoption assumptions do not, and specifies the properties and boundary conditions of a triadic competitive structure. Practically, the results inform startup positioning beyond the model layer, incumbent responses that are more cooperative than competitive, and investment in a market of high capital concentration.

2. Literature Review

2.1 Generative Artificial Intelligence

Artificial intelligence initially relied on rule-based systems and statistical classifiers requiring explicit human logic. Deep learning shifted this by learning hierarchical representations from data (LeCun et al., 2015), while Transformer architecture (Vaswani et al., 2017) replaced sequential processing with self-attention, enabling large-scale training of models such as BERT (Devlin et al., 2019), the GPT family and multimodal systems with cross-domain capabilities. Generative artificial intelligence (GenAI) refers to models that generate meaningful content, including text, images, code and audio, from patterns learned during training (Feuerriegel et al., 2024). Unlike earlier systems focused on classification or retrieval, GenAI synthesises outputs from open-ended prompts and operates across tasks without task-specific redesign. Three properties are central here. Generativity enables knowledge-intensive tasks such as drafting analytical reports and strategic scenarios. Accessibility through APIs and cloud infrastructure lowers technical and financial barriers to AI-enabled applications. Scalability allows one model to support heterogeneous applications at marginal deployment costs approaching zero (Feuerriegel et al., 2024). Together, these properties extend automation into unstructured knowledge work involving synthesis and interpretation (Eloundou et al., 2023; Brynjolfsson et al., 2023), where business consulting creates value.

2.2 Big Bang Disruption and the Competitive Implications of GenAI

Disruption has been central to innovation management since Christensen (1997) showed how entrants can displace incumbents by targeting underserved segments with cheaper, initially inferior solutions that gradually improve and redefine the market. Downes and Nunes (2014), however, identify a distinct class of innovation that escapes this gradual logic. Big Bang Disruption (BBD) describes innovations entering markets with better performance, lower cost and immediate mass-market scalability, collapsing sequential diffusion stages and rendering incumbent advantages obsolete before defensive repositioning is possible.

At first sight, GenAI approximates the three BBD conditions, particularly on the supply side. First, unencumbered development: standardised AI infrastructure and API-based delivery reduce the cost and complexity of developing AI-enabled applications (Feuerriegel et al., 2024). Second, undisciplined strategy: from their initial public releases, generative models demonstrated cross-domain capabilities and performed competitively with specialised professionals in tasks from legal analysis to financial reporting without domain-specific redesign (Eloundou et al., 2023). Third, unconstrained growth: ChatGPT reached 100 million users within two months, the fastest adoption rate recorded for a consumer technology, followed by rapid enterprise adoption (Reuters, 2023).

Unlike Christensen's model, where incumbents retain time to respond because entrants initially underperform, BBD compresses or eliminates this window by competing immediately on performance and cost. This matters in knowledge-intensive industries, where incumbency rests on accumulated human capital rather than physical assets and language models can partly replicate work previously requiring substantial organisational resources.

2.3 Business Consulting: Incumbent Models and the Impact of GenAI

Business consulting is a paradigmatic form of knowledge-intensive work (Kubr, 2002). Incumbents have historically relied on three foundations: human capital organised through hierarchical delivery models, long-term client relationships and brand reputation. Their pyramid structure combines junior analysts handling data collection and analysis with senior professionals providing strategic judgment and client management. Growth therefore requires proportional expansion of human resources, while marginal costs remain tied to labour. Profitability depends on high billing rates and stable utilisation, a model long protected by the difficulty of codifying and automating consulting knowledge (Hansen et al., 1999). GenAI challenges this structure by making information-intensive tasks such as document analysis, regulatory review and market research partially automatable, while also supporting content and insight generation with narrowing quality gaps relative to human work (Eloundou et al., 2023).

US labour-market evidence indicates slower junior hiring alongside rising demand for senior professionals able to orchestrate AI-augmented workflows (Hosseini Maasoum and Lichtinger, 2025). GenAI therefore compresses the labour required to produce inputs for strategic judgment without replacing human expertise. Its impact concentrates on human-capital delivery, while client relationships and reputation remain less exposed. Incumbents consequently face both efficiency gains in junior work and erosion of the resource base on which traditional consulting models depend.

2.4 Literature Gap

Each of the three literatures reviewed above leaves this problem unaddressed, and their separation is itself part of the difficulty. Research on AI and the future of knowledge work establishes that generative models perform tasks previously considered resistant to automation, with measurable labour-market implications (Eloundou et al., 2023; Brynjolfsson et al., 2023; Hosseini Maasoum & Lichtinger, 2025), but treats exposure at the level of occupations rather than of the firms supplying the technology. The disruption literature provides frameworks for how discontinuities enable entrants to challenge incumbents, yet empirical applications to knowledge-intensive professional services remain limited, and GenAI-native startups have not been examined as a distinct category of entrant. Work on consulting models incumbent delivery economics but predates the arrival of competitors whose core capability is externally supplied.

This gap is consequential because these startups do not confront a single incumbent category. They must differentiate simultaneously from consulting firms offering human-intensive expertise and from generalist AI providers offering powerful but domain-agnostic models, a triadic configuration that remains underexplored in disruption research. The present study addresses that gap empirically, grounding the application of Big Bang Disruption in a knowledge-intensive professional service market.

More specifically, existing disruption models do not explain how entrants compete when the enabling technology is controlled by upstream providers that simultaneously act as suppliers and potential downstream competitors.

3. Methodology

This study adopts an exploratory, qualitative-dominant multi-method design (Yin, 1984; Lin, 1998) comprising two sequential phases: descriptive mapping of GenAI-based consulting startups and semi-structured interviews with three key informants. This design reflects the research question's dual requirements: breadth and comparability to map the market, and depth to interpret the strategic logics underlying observed patterns. The study is exploratory because this nascent segment lacks established constructs, performance baselines and comparable conditions. A large-scale survey would therefore presuppose categories still being identified, while quantitative performance analysis would be premature for predominantly early-stage, privately held firms without standardised public financial data.

Crunchbase was selected as the primary source because of its coverage of technology startups and use in entrepreneurship research. Its self-reported descriptive and sector classifications were treated as provisional and cross-checked against company websites, while financial fields were treated as verified. The advanced-search query combined five criteria: founding year from 2015 onwards; active status; classification under Artificial Intelligence or Generative AI; assignment to Management Consulting, Consulting, Compliance or Data Governance; and keyword-based exclusion of adjacent domains such as AR/VR/XR, advertising technology and

Web3. It returned over 3,000 firms, illustrating both the scale of AI entrepreneurship and the limitations of self-reported categories.

Refinement occurred in two stages. First, GPT-4.1 and Gemini 2.5 Pro, with web-search capability, screened candidates in batches of about 200 by comparing company descriptions with live websites and producing provisional classifications. These outputs informed but did not determine selection: every candidate was manually verified through company websites and other public sources. Manual review also resolved boundary cases, retaining firms whose core offering applied generative AI to a consulting domain and excluding those where AI was merely nominal or the entity was not a genuine startup; no duplicates were found. Inclusion required generative AI as a core offering component, primarily SaaS or cloud-based delivery, and consulting-related functionality. Deep-tech venture-capital portfolios, including Y Combinator, Sequoia Capital, Founders Fund and Italian and international early-stage investors, were also screened for overlooked firms. The final dataset contains 177 startups across Operations Consulting (88), Strategy Consulting (37), Compliance (35) and Data Protection (17). Section 4 reports the analysed dimensions; cumulative concentration ratios assessed funding concentration among the 96 funded firms.

The second phase comprised three semi-structured interviews with startup representatives selected as key informants because of their knowledge of strategic direction, business models and technological architecture. Purposive selection maximised variation in geography (two Italy-based, one US-based), specialisation (legal-tech compliance, strategy and deep-tech operations) and company stage rather than representativeness. The interviews therefore provide interpretive depth rather than generalisable estimates. The protocol covered five domains derived from the research question and census findings: founding rationale, business model and go-to-market strategy, technological architecture and model dependencies, perceived effects of GenAI on consulting, and growth and competitive outlook. All interviews were audio-recorded and transcribed.

Analysis followed a framework-based thematic approach. The five common domains formed the initial coding frame, allowing transcript segments to be assigned consistently despite non-linear conversations, while coding remained open to inductive themes, notably pricing, model dependency and relationships with incumbents and general-purpose AI providers. The two sources were integrated through triangulation, comparing interview themes with corresponding census dimensions so that qualitative evidence could interpret observed structural patterns, including the relationship between software-based distribution and interviewees' accounts of pricing and cost dynamics.

4. Findings

4.1 Market Structure and Geographic Concentration

Startup formation tracks the availability of enabling infrastructure rather than preceding it. Entry is sparse before 2015, shows isolated peaks in 2016 and 2020, then accelerates continuously from 2022 to a maximum of 37 firms in 2024, the window following the mass-market release of GPT-3.5/4 and Gemini; Geographically the sample concentrates in a few hubs, with a pronounced lead for the United States.

Table 1: Geographic distribution of the identified startups

Country	Firms (n=177)	Share
United States	62	35.0%
European Union	38	21.5%
United Kingdom	20	11.3%
Canada	9	5.1%
India	7	4.0%
Australia	4	2.3%
Other	37	20.9%
Total	177	100%

Italy is the largest single EU country (10 firms), included in the EU aggregate.

Capital, however, behaves differently. Among the 96 funded firms it is sharply concentrated: the eleven largest raised 80.9% of all capital (\$4.55B of \$5.62B), and the top twenty account for 92.3% (\$5.19B). The skew is also

national: within that top tier US firms hold roughly three quarters of the capital and UK firms a further 13.4%, while continental EU firms, numerically comparable across the full sample, hold under 6%. This asymmetry maps onto the growth trajectories reported in the interviews. The US strategy startup operated in a deeper, more risk-tolerant capital market, whereas both Italian firms described slower capitalisation; the deep-tech founder singled out the difficulty of raising for deep tech in Italy relative to France or the UK.

4.2 Competitive Positioning Between Incumbents and AI Providers

The funded subset (96 of 177 firms) underpins every metric that requires financial or go-to-market data, and the tables below state their base explicitly. By volume the sample tilts toward operations consulting, but capital intensity inverts the ranking: operations attracts the largest absolute capital through sheer firm count, while strategy consulting commands the highest average funding per firm.

Table 2: Cluster composition and reported funding

Cluster	Firms (n=177)	Share	Funded (n=96)	Total capital	Avg / firm
Operations Consulting	88	49.7%	41	\$3.01B	\$34.3M
Strategy Consulting	37	20.9%	18	\$1.55B	\$41.9M
Compliance	35	19.8%	23	\$0.98B	\$27.9M
Data Protection	17	9.6%	14	\$0.08B	\$4.7M
Total	177	100%	96	\$5.62B	

The 177-to-96 transition is itself informative: the regulated clusters convert a far larger share of their firms into funded status (compliance 23 of 35, data protection 14 of 17) than operations (41 of 88), indicating that investor selectivity concentrates on regulated niches even where absolute capital is smaller.

Beyond cluster, the funded firms share a recognisable solution and go-to-market profile.

Table 3: Industry orientation and solution architecture of the funded subset

Dimension	Category	Firms	Share
Industry orientation	Cross-industry	69	71.9%
	Single-industry	27	28.1%
Solution architecture	Vertical (task-specialised)	49	51%
	Horizontal (cross-functional)	47	49%

Table 4: Acquisition-channel occurrences in the funded subset

Dimension	Category	Count	Share
Acquisition channel*	Direct sales	84	41.6%
	Digital / SaaS	49	24.3%
	Consulting-mediated	35	17.3%
	System integrators	34	16.8%

* Non-exclusive: a firm may appear under several categories.

Table 5: Main industries served by the funded subset

Industry served (n=96)*	Firms
Finance	47
Healthcare	23
Manufacturing	20
Retail	19
IT	17
Technology	16

* Non-exclusive; cross-industry firms serve several sectors, so counts exceed 96.

The balance shifts by cluster: operations and data protection skew horizontal and cross-industry, whereas compliance skews vertical and concentrates on the legal function, consistent with task-specific regulatory tools. Direct enterprise sales leads in every cluster, and finance in all but strategy, where retail leads.

Two features recur across all three interviews. The first is a reputation barrier: as young firms in a market built on long-term relationships, all three initially struggled to gain traction against incumbents. This surfaced as long sales cycles and caution toward startup vendors for the Italian legal-tech firm, the need to prove solidity and ROI as young graduates for the US strategy startup, and a perception of instability relative to established consultancies and hyperscalers for the deep-tech platform. The second is that each framed competition as two-sided rather than against a single rival, locating itself at an intermediate application layer between incumbent consultancies and generalist model providers. The Italian deep-tech CEO put it sharply: "Our competitor is not Deloitte, but Microsoft: they sell the bricks, we sell the house already built." The strategy startup's CTO held that rivalry with large AI providers is only partially relevant, since those firms optimise horizontally for breadth rather than domain depth, and contrasted a value-based, win-win model against incumbents' time-based billing. The legal-tech firm described a market polarised between generic LLMs carrying hallucination risk in regulated domains and static legal databases lacking interactivity, occupying the middle ground through a curated regulatory corpus.

This intermediate position can be summarised as a triadic structure that bilateral disruption models do not fully capture.

Table 6: Triadic competitive configuration

	Incumbent Consultancies	Generalist AI Providers	Genai-Native Startups
Core offering	Human expertise, institutional knowledge, client relationships	General-purpose models optimised for breadth	Domain-specialised applications on third-party models
Distinctive asset	Brand, delivery capacity, client trust	Model capability and infrastructure	Curated proprietary corpus, vertical workflows
Exposed weakness	High time-based cost, slower AI adoption	Breadth over depth, hallucination in regulated tasks, not deployment-ready	Dependence on external model providers
Relation to startups	Coopetitive: rival and distribution channel	Supplier and latent competitor	n/a

4.3 Commercial Configurations and Technological Dependencies

SaaS distribution predominates, but the token-based cost of inference conflicts with fixed subscription logic, a tension all three founders identified as a constraint on sustainability. Each resolves it differently, and none locates its defensibility in the model layer.

Table 7: Commercial configurations and technological dependencies of the three cases

Startup	Pricing Structure	Models and Infrastructure
Italian legal-tech	Tiered fixed-fee for professionals plus enterprise agreements	Multi-provider; RAG over an internally built regulatory corpus
US strategy consulting	Flat SaaS for SMEs; performance-based for enterprise, at 1-2% of incremental value, calibrated over an initial 8-week period	Multi-provider; multi-agent architecture with human validation by former consultants
Italian deep-tech	Annual contracts from 25 users, approx. €200,000/year, with modular skill components	Multi-provider and retrained open-source models; serving-agnostic

None of the three trains a proprietary foundation model, and all integrate several providers explicitly to avoid single-vendor dependency. Differentiation rests instead on a curated regulatory corpus, an orchestration layer with human validation, and ready-to-use vertical applications respectively, all located above the model layer.

4.4 Startups and Incumbent Consulting Firms

The evidence suggests a coopetitive rather than purely adversarial relationship with incumbents. Consulting firms may compete with GenAI-native startups by integrating similar technologies into their own delivery models, but they also operate as users, partners, and routes to market. Consulting-mediated relationships account for 17.3% of the channel occurrences coded in the funded subset. This pattern also appears in two interviews: the legal-tech platform reaches professionals and law firms through a combination of product-led and relationship-based channels, while the deep-tech platform positions consultants as users capable of configuring modular AI components for client-specific tasks.

5. Discussion

The evidence points to three interlocking patterns. Entry into GenAI consulting tracks the availability of transformer-based infrastructure rather than preceding it, forming a market at speed rather than displacing incumbents within it. The resulting firms position themselves between established consultancies and generalist AI providers, differentiating along both fronts at once. And their commercial architecture, predominantly SaaS, sits in tension with a token-based cost structure, producing the hybrid and performance-based adaptations observed in the interviews. Read together, these patterns qualify the disruption frameworks through which such entry is usually interpreted.

5.1 Theoretical Implications

The first contribution specifies how the GenAI context extends the Big Bang Disruption. The original BBD formulation largely assumes that the disruptive offering is controlled by the market entrant (Downes & Nunes, 2014). Here the innovation originates upstream: none of the interviewed firms trains a proprietary foundation model, and all integrate several external providers. Disruption is therefore a market-formation event rather than a firm-level one, lowering barriers for an entire category of entrants simultaneously rather than advantaging any single firm, as the founding-year distribution indicates. This has a consequence BBD does not anticipate. Because the same infrastructure reaches incumbents on equal commercial terms, incumbents are not technologically excluded; what constrains their response is a pyramid delivery model whose profitability depends on leveraged junior labour and stable utilisation (Hansen et al., 1999; Kubr, 2002). The compressed adjustment window BBD attributes to technological lag is here produced by revenue architecture.

The evidence also qualifies BBD's condition of unconstrained growth, which precludes marginal costs approaching zero. Token-based inference imposes a variable cost floor that rises with usage, and the pricing adaptations documented above are responses to it. A second attenuation operates on the demand side: adoption in professional services is gated by institutional credibility rather than product quality, as the reputation barrier reported by all three founders and the dominance of direct enterprise sales both indicate. The pattern is therefore one of asymmetric diffusion, in which BBD's supply-side conditions hold while its demand-side prediction does not. Entry is immediate and categorical; adoption remains incremental and trust-mediated. This reconciles a market formed at Big Bang speed with incumbents that have not been displaced.

The second contribution concerns a triadic configuration absent from traditional entrant-incumbent models, which do not explicitly theorise an upstream provider that both enables entry and can expand into the entrant's application layer (Christensen, 1997). Three properties distinguish it. Dual-role dependency: the AI provider is at once critical supplier and latent competitor, so rivalry runs vertically as well as horizontally. Asymmetric mobility: providers can extend downstream into domain applications at low incremental cost, whereas entrants cannot move upstream, and the multi-provider integration adopted by all three firms is best read as supplier diversification substituting for unavailable vertical integration. Two-front differentiation: value must be established against incumbents on cost and delivery terms and against providers on deployment-readiness and curated domain data, the distinction captured in the founder's contrast between selling bricks and selling the house already built. Together these produce a value-capture squeeze between an upstream cost floor and a price ceiling anchored by incumbent alternatives, which makes business model innovation a condition of survival rather than an accompaniment to positioning.

The configuration is bounded. It should be expected where upstream capability is concentrated, the entrant does not own it, the domain requires knowledge that general models do not encode, and regulatory or trust requirements sustain a defensible application layer. Where these weaken, the structure should revert toward bilateral competition: if upstream capability commoditises into a utility with abundant substitutes, if providers

absorb the application layer, or if domain corpora become publicly available. On this reasoning the regulated clusters, where proprietary and continuously updated corpora are hardest to substitute, should prove the most durable. These are propositions for testing rather than established results.

5.2 Practical Implications

For founders, the asymmetric mobility described above implies that defensible advantage cannot reside in the model layer, which is accessible to all entrants on equal terms. It depends instead on assets built around it: proprietary datasets, domain expertise, and integration into client workflows, which is where the interviewed firms locate their differentiation.

For incumbents, the relationship with GenAI entrants is more accurately coopetitive than competitive. The constraint they face is not access to the technology but the delivery economics of their current model, and the consulting-mediated sales channel observed in the dataset indicates that entrants also depend on them for distribution. Integrating AI-native solutions can therefore capture efficiency gains while preserving the trusted-advisor role that generative AI does not replace (Kubr, 2002).

For investors, the concentration of capital documented here, and its geographic skew toward the United States, indicates that ecosystem conditions continue to shape trajectories in AI-enabled sectors independently of firm-level positioning.

5.3 Limitations and Future Research

Three limitations qualify these findings. First, the dataset relies primarily on Crunchbase, which combines verified financial records with self-reported descriptive fields, so classification inconsistencies are possible and firms without public funding records may be underrepresented.

Second, the evidence base is supply-side by design: founders offer the most direct view of entry and positioning, and of the coopetitive dynamics at the interface with clients and incumbents, but the perspectives of clients, incumbent firms, and practising consultants fall outside its scope. Characterising the broader market impact of GenAI on consulting, as distinct from the entry of the startups driving it, would require a design incorporating those viewpoints.

Third, three interviews provide interpretive depth on the census patterns rather than generalisable estimates, and the boundary conditions proposed above are derived analytically rather than tested. Future research could examine them through larger interview samples, longitudinal tracking of startup trajectories, or comparison across national innovation ecosystems as both the technology and the market mature.

6. Conclusion

This study asked how GenAI-native startups enter and position themselves within knowledge-intensive professional service markets, and how their emergence reconfigures disruption dynamics. Drawing on an exploratory design combining a census of 177 startups with three founder interviews, it offers supply-side evidence on the competitive dynamics of AI-enabled consulting markets.

Entry is enabled by external infrastructure rather than by proprietary innovation, lowering barriers for entire categories of entrants at once (Eloundou et al., 2023; Feuerriegel et al., 2024). Rather than displacing incumbents, these firms occupy an intermediate position between consultancies and generalist AI providers, differentiating against both. Their commercial models are constrained by the same infrastructure that enables them, as token-based costs sit uneasily with recurring SaaS revenue and drive hybrid and performance-based adaptations.

Two contributions follow. First, the study modifies Big Bang Disruption for knowledge-intensive professional services: entry follows the framework's supply-side logic, but token-based costs and trust-mediated adoption qualify its predictions of unconstrained growth and immediate diffusion. Second, it specifies a triadic competitive structure, and the conditions under which it holds, in which startups differentiate simultaneously from incumbent service providers and from technology infrastructure firms.

Whether that structure persists depends on conditions this study identifies but cannot test: whether upstream capability commoditises, whether providers absorb the application layer, and whether domain corpora remain proprietary. Tracking these boundaries as the market consolidates is the natural extension of this work.

Ethics Declaration

This study draws on anonymised qualitative data from voluntary semi-structured interviews with informed startup representatives. Formal ethical approval was not required under the institutional guidelines and national regulations applicable at the time of data collection. Nevertheless, the research followed established ethical standards for studies involving human participants.

AI Declaration

Artificial intelligence tools were used under human supervision for preliminary screening and language editing. GPT-4.1 and Gemini 2.5 Pro supported the provisional classification of publicly available company information, while all inclusion and exclusion decisions were manually verified against company websites and other sources. The authors retain full responsibility for all methodological, analytical, and interpretive decisions.

References

- Brynjolfsson, E., Li, D. and Raymond, L.R. (2023) 'Generative AI at work', NBER Working Paper 31161.
- Christensen, C.M. (1997) *The innovator's dilemma: When new technologies cause great firms to fail*. Boston: Harvard Business School Press.
- Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. (2019) 'BERT: Pre-training of deep bidirectional transformers for language understanding', *Proceedings of NAACL-HLT 2019*, pp. 4171-4186.
- Downes, L. and Nunes, P.F. (2013) 'Big Bang disruption', *Harvard Business Review*, 91(3), pp. 44-56.
- Downes, L. and Nunes, P.F. (2014) *Big Bang disruption: Strategy in the age of devastating innovation*. New York: Portfolio Penguin.
- Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2023) 'GPTs are GPTs: An early look at the labor market impact potential of large language models', *arXiv:2303.10130*.
- Feuerriegel, S., Hartmann, J., Janiesch, C. and Zschech, P. (2024) 'Generative AI', *Business & Information Systems Engineering*, 66, pp. 111-126.
- Hansen, M.T., Nohria, N. and Tierney, T. (1999) 'What's your strategy for managing knowledge?', *Harvard Business Review*, 77(2), pp. 106-116.
- Hosseini Maasoum, S.M. and Lichtinger, G. (2025) 'Generative AI as seniority-biased technological change: Evidence from U.S. résumé and job posting data', *SSRN Working Paper*.
- Kubr, M. (2002) *Management consulting: A guide to the profession*. 4th edn. Geneva: International Labour Organization.
- LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', *Nature*, 521(7553), pp. 436-444.
- Lin, A.C. (1998) 'Bridging positivist and interpretivist approaches to qualitative methods', *Policy Studies Journal*, 26(1), pp. 162-180. doi: 10.1111/j.1541-0072.1998.tb01931.
- McKinsey & Company (2023) *The economic potential of generative AI: The next productivity frontier*. Available at: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>
- Reuters (2023) 'ChatGPT sets record for fastest-growing user base', 1 February. Available at: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) 'Attention is all you need', *Advances in Neural Information Processing Systems*, 30, pp. 5998-6008.
- Yin, R.K. (1984) *Case study research: Design and methods*. Beverly Hills, CA: Sage.