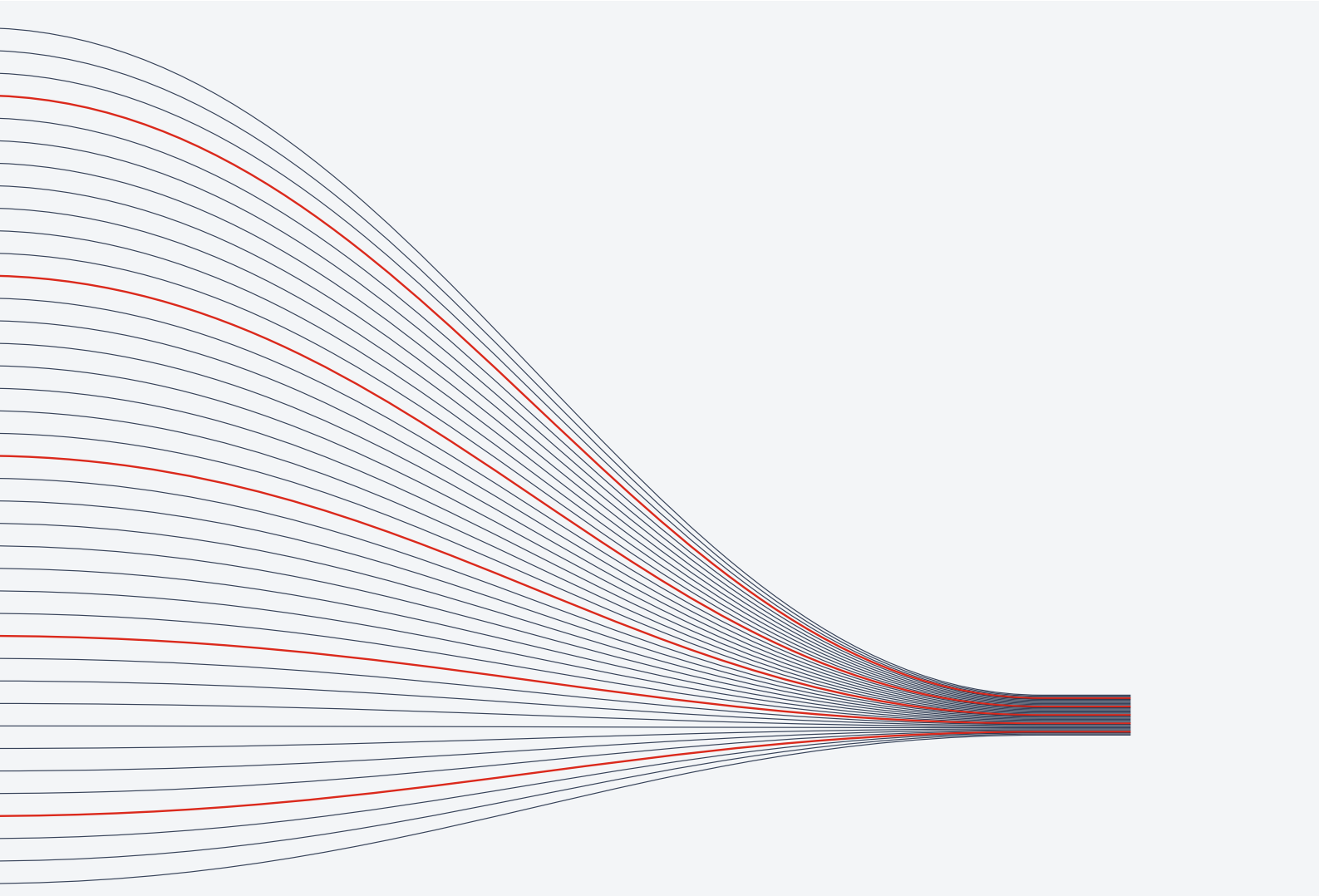
PAPER

Who puts AI into *production*.

Labs, capital and consulting: how the industry is changing, from boutiques to system integrators.



Ivan Ortenzi
Chief Innovation Evangelist | Bip

CONTENTS

Contents

01	Deployment has become a market	03
02	Three ways into the enterprise	05
03	What we know about the Hg deal	08
04	The compounding machine	10
05	An industry already under compression	12
06	Seven ways of doing consulting	14
07	The top and the networks	15
08	Volume under pressure	16
09	The squeezed middle	17
10	Prey, allies or competitors	18
11	Who gains, who has to change	19
12	The new unit economics	21
13	What to decide now	23
14	Three scenarios for 2027	24
	Sources	25

SUMMARY

Deployment has become a market 01

In 2026 the frontier labs moved down the value chain, all the way to the work of putting AI into production inside companies. In May Anthropic founded, with Blackstone, Hellman & Friedman and Goldman Sachs, the company that took the name Ode in July and serves mid-sized businesses. Also in May, OpenAI launched the OpenAI Deployment Company, which it controls and which has more than $4 billion of announced initial investment. In September Hg, one of Europe's largest software investors, extended its collaboration with Anthropic across its portfolio and, according to the Financial Times of 23 October, is preparing with Anthropic a new consulting firm for European companies [1–5].

The three initiatives differ in legal form, shareholders and market. They combine, however, the same four ingredients: direct access to the model, dedicated capital, engineers embedded with the client, and assets that carry over from one deployment to the next. For the first time, the companies that build the models are also building the capacity to put them into production, with financial and technical resources that no independent consulting firm can match on its own.

Consulting receives this news while already under pressure. The large strategy firms are holding entry-level pay flat for the third year running; clients are asking to share the savings that AI produces, as KPMG did with its own auditor, obtaining a fee 14% lower; the largest Indian IT services groups speak openly of AI deflation in their revenues [18–20]. The arrival of the labs adds a second pressure to the first. The industrialisable part of the value, meaning the design, build and operation of agentic systems, can shift towards players that own demand, models, deployment capacity and reusable intellectual property all at once.

A fifty-person boutique and a global integrator with hundreds of thousands of employees are equally exposed if all they sell is the time of qualified people. The industry will split along a line that cuts across the traditional categories and runs through what each firm owns: access to the decisions of top management, an asset that is reused from one client to the next, responsibility for the system in operation.

> *A boutique and a global integrator that sell only hours run the same risk.*

The impact by type of firm

TYPE	EXPECTED IMPACT	DEFENSIVE LEVER
Strategy firms	The base of the pyramid shrinks; day-rate selling gives way to fees linked to results.	Access to top management, alliances with the labs, proprietary assets.
Big Four	Fee pressure from clients; reputational risk on deliverables produced with AI.	Assurance, controls and trust as a service.
Global system integrators	Development, maintenance and managed processes are the most compressible volume.	The labs' preferred channel to large enterprises; running systems in operation.
Offshore IT services	Labour-cost arbitrage meets an even cheaper form of labour: agents.	Fixed-price contracts, platforms, physical AI.
Mid-sized integrated consultancies	Squeezed between the capital of the large groups and the specialisation of boutiques.	Vertical focus, neutrality across models, recurring services.
Digital and design agencies	Producing content and interfaces is among the most automatable activities.	Designing work shared between people and agents.
Boutiques and AI-native firms	Acquisition targets of the labs' companies, and their competitors.	Deployment engineers and reusable patterns.

What to watch over the next twelve months

The shape of the market will show in a few verifiable facts: the formal incorporation of the Anthropic–Hg company, with shareholders, capital and first clients; the appearance of lab-linked companies in tenders run by large European enterprises, today the integrators' territory; the trend in day rates for application development; the share of contracts priced on results or as a service; the requirements of the labs' partner programmes, which tie access to certifications and clients in production and will end up deciding who stays inside the channel. Chapter 14 returns to them as a set of indicators.

THE MARKET

Three ways into the enterprise

02

The frontier labs have more technological capability than companies manage to put into production. The bottleneck has moved from the model to integration: data, processes, controls, accountability. The three initiatives of 2026 answer the same problem in different ways, and they are best read together.

Ode was announced on 4 May by Anthropic together with Blackstone, Hellman & Friedman and Goldman Sachs, with backing from General Atlantic, Leonard Green, Apollo Global Management, GIC and Sequoia Capital [1]. On 15 July it started operating as a standalone company under its own brand, built on Fractional AI, an applied AI services firm acquired in May, with Anthropic engineers working alongside the team [2]. Its stated market is mid-sized businesses: regional banks, mid-sized manufacturers, local health systems, that is, organisations without the in-house resources to build and run frontier deployments [1].

The OpenAI Deployment Company was launched on 11 May with more than $4 billion of initial investment. OpenAI holds the majority and control; the round is led by TPG, with Advent, Bain Capital and Brookfield as co-leads. The investors also include three consulting and integration firms: Bain & Company, Capgemini and McKinsey & Company. OpenAI has agreed to acquire Tomoro, which brings around 150 engineers and deployment specialists [3].

The Anthropic–Hg initiative has two levels that need to be kept apart. The first is confirmed: on 21 September Hg extended its collaboration with Anthropic to bring Claude's agentic capabilities into its portfolio, through the Hg Catalyst incubator [4]. The second is reported by the Financial Times: a new independent consulting firm for European companies, with hubs in London and Berlin and a launch by the end of 2026 [5].

	ODE	OPENAI DEPLOYMENT COMPANY	ANTHROPIC–HG (EUROPE)
Status	Announced in May; brand launched in July [1, 2]	Announced in May [3]	Reported by the FT in October [5]
Capital	Amount not stated in company releases	More than $4 billion of announced initial investment [3]	Amount not public
Ownership	Standalone company founded by Anthropic with Blackstone, Hellman & Friedman and Goldman Sachs, with a consortium of investors	OpenAI majority and control [3]	Stakes not published; independent according to the FT
Initial capacity	Acquisition of Fractional AI; Anthropic engineers [2]	Agreed acquisition of Tomoro, around 150 specialists [3]	Team not announced; Hg's AI specialists are a pre-existing structure [6]
Stated buyer	Mid-sized businesses [1]	Organisations working on complex problems in demanding environments [3]	European companies, with a focus on B2B software according to the FT [5]
Type of source	Corporate	Corporate	Press

Four shared ingredients

All three have direct access to the technological frontier and to the lab's technical support, with a head start on new releases. Capital comes from financial investors, from the lab itself or from a private equity fund, and allows them to invest in assets and people before revenue pays for them, a luxury the day-rate model of consulting does not allow. The engineers to be embedded with clients were acquired, through Fractional AI and Tomoro, or were already in the ecosystem, as in Hg Catalyst; their worth lies in combining expertise in the model with work inside the client's processes.

The fourth ingredient is the least visible and it decides the economics: patterns, evaluations, components and procedures that carry over from one client to the next. Without reuse, a deployment company remains a day-rate services firm with a different shareholder.

Where the models differ

OpenAI controls its own company and targets organisations with complex processes, where the value of integration is highest. Ode addresses mid-sized businesses, a segment poorly served by the large consulting firms. Hg brings an ecosystem of software companies in which demand is already accessible and in which the fund can steer priorities and investment [8]. What sets the three models apart, then, is control of demand.

The relationship with existing consulting changes too. Anthropic states that large enterprises continue to be served by the system integrators in its network and that the new mid-market company will join that network [1]. OpenAI has brought three consulting and integration firms into the capital of its own. The Hg model, as the press describes it, relies instead on a portfolio of its own. The result is co-opetition: the same consulting firms are partners, investors and competitors of the labs, depending on the segment.

The fourth channel: partner networks

Alongside the new companies, the labs are investing in partner networks, through programmes that tie privileged access to scale and to proof of production. The table summarises the main announcements of the past twelve months.

PROGRAMME	WHAT IT OFFERS	WHAT IT REQUIRES OR IMPLIES	SOURCE
Accenture Anthropic Business Group (December 2025)	A group dedicated to Claude within Accenture.	Around 30,000 professionals to be trained; initial focus on regulated industries.	[13]
OpenAI Frontier Alliances (February 2026)	The Frontier platform and OpenAI engineers working alongside partners.	Strategy firms on operating model and change, integrators on systems. Partners: Accenture, BCG, Capgemini, McKinsey.	[12]
Claude Partner Network (March 2026)	$100 million in 2026 for training, technical support and joint market development.	Certification of professionals.	[9]
Services Track (June 2026)	Rising tiers of access and visibility.	Certified professionals, joint clients in production, public case studies.	[10]
Claude Frontier Academy (October 2026)	Training 10,000 Frontier Deployed Engineers by the end of 2027.	Announced partners include Accenture, Bain, Capgemini, Deloitte and McKinsey.	[11]

> *The market is shifting from those who own the model to those who run production at scale.*

This channel is open to every consulting firm, but its entry criteria favour those who already have scale: many certifications, clients in production, publishable cases. For mid-sized and small firms, taking part will become a condition of credibility in large-enterprise tenders, while their competitive advantage will have to come from what they build on top of the model.

THE HG DEAL

What we know about the Hg deal

03

The industrial cooperation between Hg and Anthropic has been announced; the new European company remains a project reported by the press. Anyone taking decisions today should base them on the first and treat the second as a scenario, with indicators that track how it develops.

CLAIM	STATUS AT 30.10.2026	SOURCE AND CAVEAT
Hg–Anthropic collaboration extended on 21 September	Official Hg announcement: Catalyst and the AI teams work directly with Anthropic.	[4]
Independent European consulting firm	According to the FT of 23 October; incorporation, stakes, leadership and investment not verified.	[5]: a single primary source, picked up by several outlets.
Planned hubs in London and Berlin	According to the FT; target end of 2026, with no confirmed timetable.	[5]
More than 1,600 AI projects and around $260 million	Figures stated by Hg on its own portfolio; the $260 million is budgeted EBITDA impact.	[4, 6]: no independent attestation.
More than 150 specialists, around 100 Catalyst builders	Existing Hg capability, separate from the headcount of the new company.	[6, 7]
Ode operating as an independent company	Official confirmation: brand launched in July; acquisition of Fractional AI.	[1, 2]
OpenAI Deployment Company	Official announcement: OpenAI majority; more than $4 billion announced.	[3]

What Hg has already built

Hg has an ecosystem of software companies with mission-critical systems, vertical data and an incubator dedicated to AI [8]. At 31 December 2025 it reported more than 1,600 active AI projects across the portfolio, for around $260 million of budgeted EBITDA impact; Hg itself specifies that the figure aggregates catalogued initiatives and is not a forecast of absolute EBITDA improvement [6].

Operating capacity is concentrated in Hg Catalyst, the incubator launched in November 2025 to build AI products inside portfolio companies, with around 100 builders within a wider structure of more than 150 specialists dedicated to creating value with AI [6, 7]. It works like a software house inside the fund, with pricing, packaging and go-to-market developed together with the product.

Why capital changes the game

The owner of a portfolio can **originate demand, fund productisation, accelerate distribution and reuse patterns** from one company to another. The structural advantage is plausible, but there is no evidence yet of a lower customer acquisition cost: no comparable metrics on acquisition cost or margin have been published [4, 6, 8].

The model departs from the classic partnership between hyperscaler and system integrator. The lab brings intelligence and access to the frontier; the portfolio and the delivery system bring context, domain data and iteration on real work. The contractual role of the new company relative to Hg Catalyst is not known [4, 5].

What remains unknown

The new company's shareholdings and voting rights, the capital committed by Anthropic and the Hg funds, its leadership and initial headcount, its price list and revenue model, its launch clients and entry markets beyond the Hg portfolio are all undisclosed. Each of these facts will shift the assessment: a company focused on Hg's portfolio companies would be a value workshop inside the fund, while a company open to the European market would become a direct competitor of technology consulting.

MISTAKE TO AVOID

Attributing to the European initiative the more than $4 billion announced by OpenAI or figures circulated for Ode, or treating valuations reported on secondary websites as agreed values. These are different initiatives, with different levels of evidence.

INDUSTRIAL ECONOMICS

The compounding machine

04

The potential change lies in each deployment feeding the next and lowering its marginal cost. Engineers working at client sites have existed for a long time; what is new is a system that turns their work into reusable assets, data and patterns, under a single economic owner.

The economic cycle

Demand → business workflow → engineering and domain team → reusable asset → deployment → observability → product improvement → recurring contract → new demand. The cycle works if every step has an owner accountable for the result, if deployment telemetry flows back to improve the asset, and if investment in assets is decided on an explicit economic criterion, separate from the margin of the individual engagement.

The move from project to product is proven by real savings in cost and time on subsequent deployments. The proof is the second client served with less work at the same quality, and the third with less still. In traditional consulting, intellectual property is born inside engagements, at apparently zero cost, and nobody is tasked with maintaining it, versioning it and making it circulate. It is a long-standing limit of the trade, and the new market makes it visible.

Who captures the value

In a recomposed value chain, each player captures a different share. The table summarises what each one owns and where its margin sits.

PLAYER	WHAT IT OWNS	WHAT IT CAPTURES
Labs	Models, roadmap, technical support, now also deployment companies.	Model consumption; first standardisable installations.
Hyperscalers	Compute infrastructure and data platforms.	Cloud consumption, rising with every system in operation.
Private equity	Portfolios of companies and the right to set their priorities.	A channel into portfolio companies; value creation in the portfolio.
Lab-owned companies	Model, capital, deployment engineers, assets.	Repeatable deployments in the stated segments.
System integrators	Relationships with large enterprises, delivery capacity at scale.	Integration with existing systems; running systems in operation.
Strategy firms	Access to top management, method, reputation.	Investment decisions and transformation programmes.
Specialist boutiques	Expert engineers, accelerators, niche expertise.	Speed of execution; acquisition value.

> *A consulting firm without assets of its own risks remaining an integration supplier, holding the least scalable part of the margin.*

It is a possible scenario, with clear counterweights. Large enterprises retain a complexity of systems, regulation and organisation that no deployment company can standardise quickly. Anthropic itself acknowledges the role of the integrator network in large enterprises [1]. The strategic question every consulting firm has to ask, then, concerns the part of the chain in which it wants to own something: an asset, a recurring service, responsibility for a system in operation.

WHAT CAN BE REPLICATED

The capital stays with the funds and the model with the labs. The industrial discipline, on the other hand, can be replicated by anyone: structured access to demand, product and channel developed together, systematic evaluations, versioned assets, governance of systems in operation, measurable unit economics.

THE CONTEXT

An industry already under compression

05

The labs arrive while consulting is in the middle of a transformation that AI had already set in motion. Four signals, visible in financial results and in the news of the past twelve months, show where the traditional model is giving way: the pyramid, the price of the hour, the composition of demand and trust.

The pyramid narrows

The economic model of consulting rests on leverage: many young professionals, at contained cost, work under a few seniors who sell and vouch for the work. AI strikes precisely at the base. According to data gathered by Management Consulted and reported by the Financial Times, McKinsey and Boston Consulting Group have held entry-level pay flat for the third year running and the Big Four have not raised it since 2022; two Big Four executives estimated that graduate recruitment in the United Kingdom would halve [18].

In early 2026 McKinsey's chief executive described an organisation of around 40,000 people working alongside around 25,000 AI agents, with 1.5 million hours of research and synthesis automated in a year; client-facing roles had grown by about a quarter, internal support roles had shrunk by a similar amount [17]. The emerging shape is narrower at the base and wider in the middle, with professionals directing the work of agents.

The price of the hour

Clients have understood that AI lowers the cost of producing professional services, and they are asking to share the benefit. In February the Financial Times reported that KPMG International obtained from its auditor, Grant Thornton, a fee 14% lower, arguing that AI would make the work faster [19]. The request came from one of the Big Four, a firm that sells audit services itself.

In April the chief executive of HCLTech attributed to what he called AI deflation a revenue decline of between 3% and 5% in the following year [20]; chapter 8 returns to the Indian IT services groups. At the opposite end, McKinsey states that around a quarter of its global fees are already linked to results [16]: the response of the strongest firms is to move the unit of price from the hour to the outcome.

Demand grows, in a different shape

Source Global Research forecasts 6% growth for the consulting market in 2026, with AI advice and implementation at 15% and strategy at 4% [22]. Gartner expects IT services spending to grow by 5.3% in 2026, against 15.5% for software and 62.5% for data centre systems [23]. Technology budgets are growing, and the share going to services is growing more slowly than the share going to infrastructure and software.

Inside the firms, the revenue mix is shifting accordingly. Boston Consulting Group closed 2025 with $14.4 billion in revenue and reports that services focused on technology

and AI exceed 40% of the total [15]. Accenture, between the last quarter of fiscal 2025 and the first of fiscal 2026, recorded more than $900 million in business optimisation costs, mainly for employee severance [14].

Trust becomes a cost

AI used without controls has already caused significant reputational incidents. In October 2025 Deloitte Australia agreed to partially refund the Australian government for a report worth 440,000 Australian dollars that contained a fabricated quote from a court ruling and references to non-existent studies [24]. In 2026 EY Canada withdrew a report on the security of loyalty programmes after an external review found 16 of its 27 sources to be fabricated, misattributed or unreachable [25].

> *When producing a document costs little, the client pays for the guarantee that it is true.*

If the cost of producing a report tends towards zero, what keeps its value is the checking of sources, signed accountability and the ability to defend an analysis before a board committee or a regulator. Firms with a culture of control, and with a reputation to protect, can turn these into services of their own.

MAPPING THE INDUSTRY

Seven ways of doing consulting

06

Talking about "consulting" in general hides the differences that matter. Firms that sell professional services to companies have different revenue models, assets and clients, and the new players hit them in different places. I distinguish seven types; many groups combine more than one.

TYPE	PREVAILING REVENUE MODEL	ASSETS IT CAN DEFEND	WHERE THE NEW MODELS HIT
Strategy firms	High-fee projects, with a growing share of results-based fees.	Access to top management, method, reputation, alliances with the labs.	Analysis and synthesis at the base of the pyramid.
Big Four	Audit, tax and advisory; much day-rate capacity in consulting.	Assurance, risk, regulation, widespread presence.	Day-rate advisory; fee pressure across all services.
Global system integrators	Multi-year contracts for development, application management and processes.	Relationships with large enterprises, delivery at scale, alliances.	Application development and maintenance; managed processes.
Offshore IT services	Capacity at competitive cost; managed services contracts.	Scale, process discipline, long-term relationships.	Everything that can be codified and repeated.
Mid-sized integrated consultancies	Projects and day-rate capacity for large national companies.	Proximity to the client, sector expertise, flexibility.	AI projects for the mid-market; application development.
Digital and design agencies	Creative and product projects; retainers.	Understanding of end customers, quality of experience.	Production of content and interfaces.
Boutiques and AI-native firms	Specialist, technically intensive projects.	Expert engineers, accelerators, domain niches.	The same trade as the lab-owned companies, with less capital.

Two axes for reading the map

The first variable is **how much revenue depends on hours** . The more turnover is tied to days of codifiable work, the more direct the compression: AI reduces the hours needed, and the client knows it.

The second variable is **how much the firm owns beyond its people** : relationships with top management, reusable assets, platforms, managed services contracts, standing in the labs' programmes. If an integrator has already codified its components and delivers in eight weeks work that used to take twenty, at the agreed price, the difference stays with it as margin. A firm that sells only hours loses it from its revenue.

STRATEGY FIRMS AND BIG FOUR

The top and the networks

07

Strategy firms and the Big Four face the change from opposite positions. The former sell mainly decisions and have already shifted their mix towards technology and implementation; the latter bring into consulting a huge day-rate capacity and a reputation built on assurance.

Strategy firms

The large strategy firms are the ones that have most closely accompanied the labs' entry. McKinsey and Boston Consulting Group are partners in OpenAI's Frontier Alliances; Bain & Company and McKinsey are among the investors in the OpenAI Deployment Company and among the partners of Anthropic's Claude Frontier Academy [3, 11, 12]. They are carving out the role of those who define the operating model and govern the change, leaving the connection with systems to the labs' engineers and to the integrators.

Strategy in the narrow sense accounts for less than 20% of McKinsey's work, according to its global leader for technology and AI, and around a quarter of its fees, as noted, are linked to results [16]; at Boston Consulting Group, services focused on technology and AI exceed 40% of revenue [15]. These are firms that anticipated compression and turned it into an offering.

The leverage of the pyramid, which funds margins and the training of future partners, weakens when agents do the work of analysis and synthesis, and it is from here, more than from the new players, that the risk to these firms comes. The freeze on entry-level pay and the cuts in support roles are the first signs of a model that has to find a new balance between people, agents and price [17, 18]. Direct exposure to the new players remains medium: no deployment company competes today for the board's mandate.

The Big Four

For the multidisciplinary networks, the KPMG–Grant Thornton case showed that regulated services such as audit are also entering a negotiation over AI savings [19], while graduate recruitment is falling [18]. In the advisory lines, where much capacity is sold by the day, compression hits more directly than in the strategy firms.

The Deloitte Australia and EY Canada incidents show the specific risk of these networks: an error in a deliverable produced with AI damages a brand that lives on trust and verification [24, 25]. The same risk points to their most natural opportunity. Assurance of AI systems, controls, compliance and risk management are services consistent with their historical trade, and they become more in demand as agentic systems enter regulated processes.

The Big Four also have a constraint the other types do not: independence rules towards audit clients limit some forms of alliance and joint selling. It slows down alliances and, at the same time, gives the client a guarantee that other players find hard to offer.

ASSESSMENT

Strategy firms: medium exposure, with the risk concentrated in the pyramid model more than in competition from the labs. Big Four: medium-high exposure in the day-rate advisory lines, with a way out in trust as a service.

INTEGRATORS AND IT SERVICES

Volume under pressure

08

Global integrators and IT services groups are the most exposed to compression, and the integrators are also the labs' preferred channel to large enterprises.

Global integrators

Application development, maintenance and managed processes are the most compressible activities in the whole sector, and they account for a significant part of the integrators' revenue. Accenture, with around 814,000 people and $74.18 billion in revenue in fiscal 2026, expects growth of between 3% and 6% in local currency for the following year and lists among its risk factors the possibility that AI reduces demand for its services [14].

The integrators are responding by entering AI from the side of running systems in operation. Accenture has created a dedicated group with Anthropic, with around 30,000 professionals to be trained, and is a partner in OpenAI's Frontier Alliances [12, 13]. Capgemini acquired WNS for $3.3 billion, with the stated aim of transforming managed business processes through agentic AI [21]. The logic of the move is to buy the base of processes to be run and transform it with agents before others do.

Anthropic states that large enterprises are served through the network of integrators [1], and the labs' partner programmes reward precisely the scale of certifications and of clients in production [10]. The integrators therefore start with an advantage; their problem is to move from revenue proportional to people to revenue proportional to results faster than deflation shrinks the volume.

Offshore IT services

The model of the large Indian groups rests on labour-cost arbitrage: the same activities, performed with industrial discipline, at a lower cost. Agents introduce an even cheaper form of labour. The fiscal 2026 results show it: TCS closed with revenue down 0.5%, HCLTech put the effect of AI deflation on the following year's revenue at between 3% and 5%, Wipro reported lower margins on some contracts, while Infosys expects deflation to become a factor in the future [20].

The ways out indicated by the companies themselves are fixed-price contracts, in which the savings produced by AI stay with the supplier, proprietary platforms and new markets such as physical AI: TCS cites AI-based inspection systems for an electronics manufacturer [20].

ASSESSMENT

Global integrators: medium-high exposure on volume, with the sector's widest opportunity in running systems in operation. Offshore IT services: high exposure, because their cost advantage is the first target of agents.

THE MIDDLE TIER

The squeezed middle

09

Mid-sized integrated consultancies risk finding themselves in the hardest position. They are too small to compete with the capital of the large groups and the labs, and too large to live on the specialisation of a boutique. Digital and design agencies share part of this pressure.

Mid-sized integrated consultancies

I mean national or European firms, from a few hundred to a few thousand people, that combine strategy, technology and implementation for large companies in their own country. Their historical advantage is proximity: knowledge of processes, relationships with the leaders of national champions, a flexibility the large groups do not offer. The new models are eroding it from several directions.

From below, the lab-owned companies are targeting precisely mid-sized businesses and, in the case of the Anthropic–Hg company, European companies [1, 5]. From above, the partner programmes reward scale, with tiers linked to the number of certified professionals and clients in production [10]: a mid-sized firm struggles to reach the tiers that guarantee privileged access. From the side, the global integrators use their alliances with the labs to enter projects they used to leave to smaller players.

There remain, however, spaces in which the middle tier starts with an advantage. Large enterprises that use several models and several clouds are looking for an independent party to hold responsibility, which no lab-controlled company can be. Regulated and asset-intensive sectors such as energy, infrastructure, utilities and manufacturing require domain expertise and presence in the field. The convergence of information and operational technology, where agents meet plants and networks, is ground the new players have barely occupied.

A firm in this tier that keeps selling day-rate capacity will see revenue and margins compressed. If instead it focuses on verticals, builds assets in a few domains, takes on the running of systems in operation and turns neutrality across models into a measurable service, it can occupy a space the large players cover poorly. The middle tier is likely to consolidate, through mergers or acquisitions by larger groups.

Digital and design agencies

Producing content, interfaces and prototypes is among the most rapidly automatable activities, and agencies that live on production feel compression directly. The design trade, however, has a natural extension in the new context. Designing how people and agents work together, which decisions remain human, how a customer or an employee experiences a partly automated service: these are design problems, and they come before the engineering ones.

ASSESSMENT

Mid-sized integrated consultancies: high exposure for those that sell hours, medium for those that own vertical assets and recurring services. Digital and design agencies: high exposure on production, with an opportunity in designing hybrid services.

BOUTIQUES AND AI-NATIVES

Prey, allies or competitors

10

AI engineering boutiques have become the most contested resource in the market. The two deals that shaped the new models are acquisitions of boutiques: Fractional AI for Ode, Tomoro for the OpenAI Deployment Company [2, 3]. Their value lies in engineers able to put agentic systems into production and in the patterns accumulated on projects.

An ambivalent position

For AI engineering boutiques, the arrival of the labs is an exit opportunity and a competitive threat at the same moment. Valuations are rising, because labs, funds and large groups are looking for deployment capacity that is already formed. Those not acquired, however, find themselves competing with companies that do the same work with privileged access to the model, capital to invest in assets and, in Hg's case, a portfolio of demand already under control.

Domain boutiques are in a different position: specialists in regulation, in a sector or in a function, whose value lies in knowledge the models do not have and that deployment companies must buy. Their exposure is lower, and AI can widen their market, because it multiplies the projects that need specific expertise to be put into operation safely.

The AI-native entrants

Firms born around AI, including those controlled by the labs, have no day-rate revenue to protect, organise themselves around agents and have direct access to the technology. Their limits are just as concrete. They know less about the complexity of large enterprises' systems and organisations. They have yet to demonstrate economics and reuse at scale. Above all, when they belong to a lab, they carry an obvious conflict of interest: they are unlikely to recommend a competitor's model, even when cost, performance or sovereignty would suggest it.

> *Neutrality across models is a service that no lab-owned company can sell.*

This conflict opens the space for neutrality as a service. Companies that use several models will want someone to compare them, govern them and ensure continuity when one changes its price, performance or terms. This role is closed by design to the lab-owned companies and open to every other type, provided it is backed by genuinely multi-model architectures and portability clauses.

ASSESSMENT

AI engineering boutiques: medium-high exposure, offset by strong acquisition value. Domain boutiques: low-medium exposure. AI-native firms: entrants with technical advantages and limits in scale, enterprise experience and independence.

COMPARATIVE ASSESSMENT

Who gains, who has to change

11

Compared side by side, the types fall into three groups. The matrix rates four pressure factors and derives a net exposure from them. It is a qualitative judgement, built on the evidence of the previous chapters, to be read as a map of directions more than as a measurement.

TYPE	PRESSURE ON FEES AND VOLUMES	COMPETITION FROM NEW PLAYERS	FRAGILITY OF THE PYRAMID	ASSETS TO BE BUILT	NET EXPOSURE
Strategy firms	Medium	Low	High	Low	Medium
Big Four	High	Medium	High	Medium	Medium-high
Global system integrators	High	Medium	Medium	Low	Medium-high
Offshore IT services	High	Medium	High	Medium	High
Mid-sized integrated consultancies	High	High	Medium	High	High
Digital and design agencies	High	Medium	Medium	High	High
AI engineering boutiques	Medium	High	Low	High	Medium-high
Domain boutiques	Low	Low	Low	Medium	Low-medium

The first group: who can gain

Strategy firms and the most advanced global integrators have what the new market rewards: access to top-management decisions, capital to invest in assets and platforms, structured alliances with the labs. For them, the compression of hours can become margin, provided they change the unit of price. Their main risk is organisational: managing the transition of the pyramid without losing the ability to train future senior people.

The second group: who has to transform

The Big Four, integrators on the volume side and offshore IT services have scale and relationships, but a significant part of their revenue depends on codifiable hours. For them the question is one of speed: converting day-rate contracts into contracts priced on results or as a service before deflation erodes the base. The moves already visible, such as Capgemini's acquisition of WNS and the groups dedicated to the labs, go in

this direction [13, 21].

The third group: who risks being squeezed

Mid-sized integrated consultancies without assets, agencies that live on production and generalist technical boutiques share a fragility: they sell what the new players can do with more capital and access to the model. They can survive by specialising, in a domain, in a regulated sector, in multi-model management, or by joining a larger group. Domain boutiques are the exception in the group.

> *Domain knowledge stays scarce even when technique becomes abundant.*

What it means for buyers of consulting

For client companies, the recomposition of the market is first of all an opportunity to negotiate. Those who buy professional services can ask to see the savings produced by AI reflected in the price, and move contracts from the hour to the result where measurement is possible. The risk to manage is dependence: on a single model, when production is entrusted to a lab-controlled company, or on a supplier without assets, which charges by the day for work that others know how to industrialise. Every supplier should be asked two questions: what does it own, beyond the people it provides, and who answers for the system once it goes into operation.

THE ECONOMICS OF CONSULTING

The new unit economics

12

AI-native consulting requires economics different from those of day rates. Investment in assets, operating costs and recurring revenue coexist with project work and must be measured separately. The figures in this chapter are illustrative.

Four profit and loss accounts

ACCOUNT	WHAT IT MEASURES	MISTAKE TO AVOID
Client	Revenue, acquisition and retention cost, contribution of the relationship.	Assigning all the margin to the client, with no product cost.
Product	Recurring revenue and services, development, roadmap, achievement of outcomes.	Products without an owner and without support allocation.
Deployment and run	Engineering cost, capacity utilisation, consumption, service levels and incidents.	Day-rate work hidden under the “product” label.
Intellectual property	Development costs, asset maintenance, return on reuse.	Free reuse that penalises whoever invests in the asset.

The indicator everything else rests on

Net contribution per workflow in production = set-up fee + service fee + variable components – deployment engineering – cloud and model consumption – operation, quality and incidents – amortisation of intellectual property – partner costs. The indicator must be calculated on a defined unit of outcome, including supervision and exceptions, which in agentic systems are the most underestimated cost item. Cost per token says little about business value; cost per successfully completed task and contribution per workflow tell you whether a system produces margin.

Price follows production

The further work moves into operation, the closer the unit of price moves to the result. A diagnosis is sold at a fixed price, with accepted deliverables. A system under construction is paid by milestones and acceptance gates, then per task completed at agreed quality and cost. A managed service is paid per service level. Results-based fees work only with a baseline measured before go-live and robust attribution.

The work of Brynjolfsson and Petropoulos for the Stanford Digital Economy Lab explains why hybrid contracts remain the most common form: agent telemetry improves the measurement of results, but does not on its own solve the problem of attributing them to the intervention [26].

An example of reuse

A first deployment requires €200,000 of technical work, of which €80,000 is invested in reusable components. The second, at an independent client, costs €135,000 thanks to those assets, plus €20,000 of adaptation: €155,000 against €200,000, or 22.5% less. The example shows why reuse

must be measured at the second independent client: that is where you see whether an asset transfers.

The cost that escapes the business case

In business cases for agentic systems, the most often overlooked item is the human work that remains around the agent: checking uncertain cases, handling exceptions, correcting errors, maintaining evaluations when processes or the model change. A system that completes 90% of tasks on its own can cost more than expected if the remaining 10% requires experienced people and long lead times. For a consulting firm selling on results, this item decides the margin, and for the client it decides whether the promised saving reaches the income statement.

FOUR INDICATORS TO READ TOGETHER

Reuse at the second deployment, time from signature to production, contribution per workflow, share of recurring revenue. Rising contribution with flat reuse signals margins earned on bespoke projects; rising reuse with flat client results signals assets that do not affect value.

DECISIONS

What to decide now

13

Every consulting firm, whatever its type, has to take over the next twelve months a set of decisions that the old model allowed it to postpone. I list them in the order in which they are best tackled.

01 **What to own.** Choose the part of the value chain in which to build ownership: the client's decision, a reusable asset, a recurring service, responsibility for a system in operation. Without this choice every other decision remains tactical.

02 **Unit of price.** Define for each offering a signable scope, an acceptance criterion and a verifiable unit of price, and progressively reduce the weight of day rates.

03 **Ownership of assets.** Give every strategic asset an owner, a cost, a life cycle and rules of reuse, with a development budget separate from engagement margins.

04 **Deployment engineers.** Build a Frontier Deployed Engineers capability measured on cases in production, paired with domain expertise, with career paths and incentives consistent with reuse.

05 **Relationship with the labs.** Join the partner programmes while preserving multi-model architectures, transparency on consumption and portability of data and configurations.

06 **Shape of the pyramid.** Redesign the entry, training and progression of young professionals in an organisation where agents do the groundwork.

07 **Trust and control.** Introduce systematic checks on every deliverable produced with AI and assess whether to turn them into a service for clients.

08 **Capital structure.** Decide on acquisitions, spin-offs or new companies only after documenting the economics and traction of the first products.

Four ways to build the capability

OPTION	WHY IT CAN WORK	BEST SUITED TO
Building in-house	Maximum control over standards, intellectual property and culture.	Firms with concentrated demand and investment discipline.
Alliance with the labs	Training, roadmap access, joint selling, reputation.	Integrators and strategy firms; for mid-sized firms, as specialist partners.
Acquiring a boutique	Immediate deployment capacity, as Fractional AI and Tomoro show [2, 3].	Groups with capital and the ability to integrate technical talent.
Separate company	Dedicated P&L and value proposition; external capital possible.	Groups with a recurring portfolio and clear rules on client ownership.

The most robust sequence combines in-house building and alliances in the first months, and keeps acquisitions and spin-offs open until data on reuse and margin justify a commitment of capital.

SCENARIOS

Three scenarios for 2027

14

How the market evolves remains open. Three plausible scenarios describe the possible directions; observable signals make it possible to tell in good time which one is unfolding.

SCENARIO	OBSERVABLE SIGNALS	WHO BENEFITS	WHO SUFFERS
1. Ecosystem-led	The labs keep working through partners; the Hg company operates mainly in private-equity-owned software.	Integrators and strategy firms with structured alliances.	Mid-sized firms outside the top tiers of partner programmes.
2. Lab expansion	The lab-owned companies win large enterprises, make acquisitions and launch managed platforms.	Acquired boutiques; players neutral across models.	Integrators on the development side; technical boutiques not acquired.
3. Platform commoditisation	Tools automate development and evaluation; rates fall; vertical players enter.	Those selling results and managed services; domain boutiques.	Those selling day-rate capacity, in every type.

Some choices hold in all three scenarios: owning reusable assets, moving the unit of price towards the result, staying neutral across models, measuring economics per workflow. What changes is the speed: in scenario 2 the priority shifts to running systems in operation and to the relationship with top management; in scenario 3, from selling projects to selling services.

The indicators to track

Structure of the new companies. Stakes, capital and clients of the Anthropic–Hg company; acquisitions by Ode and the OpenAI Deployment Company; entry into new European countries.

Prices. Day rates in application development, the share of results-based fees declared by the large firms, clients' requests to share savings.

People. Graduate recruitment, entry-level pay, the ratio of agents to people declared by firms, demand for and pay of deployment engineers.

Channel. Requirements of the labs' partner programmes, presence of lab-owned companies in public and private tenders of large European enterprises, consolidation deals in the middle tier.

Trust. Frequency of incidents on deliverables produced with AI, the emergence of assurance services for agentic systems, contract clauses on liability for agents' errors.

REFERENCES

Sources

1 Anthropic, “Building a new enterprise AI services company with Blackstone, Hellman & Friedman, and Goldman Sachs”, 04.05.2026.
https://www.anthropic.com/news/enterprise-ai-services-company

2 Business Wire, “Anthropic, Blackstone, and Hellman & Friedman Introduce Ode with Anthropic, an Enterprise AI Services Firm”, 15.07.2026.
https://www.businesswire.com/news/home/20260715205134/en/Anthropic-Blackstone-and-Hellman-Friedman-Introduce-Ode-with-Anthropic-an-Enterprise-AI-Services-Firm

3 OpenAI, “OpenAI launches the Deployment Company”, 11.05.2026.
https://openai.com/index/openai-launches-the-deployment-company/

4 Hg, “Hg extends collaboration with Anthropic to bring agentic AI capabilities across the portfolio”, PR Newswire, 21.09.2026.
https://www.prnewswire.co.uk/news-releases/hg-extends-collaboration-with-anthropic-to-bring-agentic-ai-capabilities-across-the-portfolio-302884668.html

5 Financial Times, article on Anthropic and Hg (Hammond, Heal), 23.10.2026.
https://www.ft.com/content/df70bf20-e809-48d0-8485-f8f514a7fbc4

6 HgCapital Trust, “Hg reaffirms AI leadership and announces intention to increase strategic investment in HgT”, 04.06.2026.
https://www.hgcapitaltrust.com/news-insights/news/2026/04-06-2026

7 Hg, “Hg launches Catalyst: a new AI incubator”, 17.11.2025.
https://hgcapital.com/insights/hg-launches-catalyst-a-new-ai-incubator/

8 Hg, corporate homepage, accessed 30.10.2026.
https://hgcapital.com/

9 Anthropic, “Anthropic invests $100 million into the Claude Partner Network”, 12.03.2026.
https://www.anthropic.com/news/claude-partner-network

10 Anthropic, “Introducing the Services Track and Partner Hub of the Claude Partner Network”, 03.06.2026.
https://www.anthropic.com/news/services-track-partner-hub

11 Anthropic, announcement of the Claude Frontier Academy, 02.10.2026.
https://www.anthropic.com/news/claude-frontier-academy

12 Computerworld, “OpenAI partners with consulting giants to deploy enterprise AI agents”, 23.02.2026.
https://www.computerworld.com/article/4136024/openai-partners-with-consulting-giants-to-deploy-enterprise-ai-agents.html

13 Anthropic, “Accenture and Anthropic launch multi-year partnership”, 09.12.2025.
https://www.anthropic.com/news/anthropic-accenture-partnership

14 Accenture, “Accenture Reports Fourth-Quarter and Full-Year Fiscal 2026 Results”, Form 8-K, 01.10.2026.
https://www.sec.gov/Archives/edgar/data/0001467373/000146737326000037/q4fy26earnings8-kexhibit.htm

15 Boston Consulting Group, press release on 2025 revenue, 23.04.2026.
https://www.bcg.com/press/23april2026-bcg-revenue-22nd-consecutive-year-growth

16 Business Insider, “AI is reshaping how McKinsey makes money”, 17.11.2025 (syndicated by Yahoo Finance).
https://finance.yahoo.com/news/ai-reshaping-mckinsey-makes-money-195132745.html

17 International Business Times UK, on statements by Bob Sternfels (HBR IdeaCast and CES 2026), 13.01.2026.
https://www.ibtimes.co.uk/50-50-future-mckinseys-ceo-reveals-ai-agents-are-set-match-human-staffing-1770406

18 Financial Times, “Top consultancies freeze starting salaries as AI threatens pyramid model”, republished by The Irish Times, 01.12.2025.
https://www.irishtimes.com/business/2025/12/01/top-consultancies-freeze-starting-salaries-as-ai-threatens-pyramid-model

19 Financial Times, “KPMG pressed its auditor to pass on AI cost savings”, republished by The Irish Times, 06.02.2026.
https://www.irishtimes.com/business/2026/02/06/kpmg-pressed-its-auditor-to-pass-on-ai-cost-savings

20 The Register, fiscal 2026 results of TCS, Infosys, Wipro and HCLTech, 28.04.2026.
https://www.theregister.com/2026/04/28/tcs_infosys_wipro_hcl_fy26/

21 Capgemini, “Capgemini completes the acquisition of WNS”, 17.10.2025.
https://www.capgemini.com/news/press-releases/capgemini-completes-the-acquisition-of-wns-and-creates-a-global-leader-in-agentic-ai-powered-intelligent-operations/

22 Source Global Research, “Service line forecasts for 2026: where are clients set to spend?”, 16.03.2026.
https://sourceglobalresearch.com/market-update/service-line-forecasts-for-2026-where-are-clients-set-to-spend

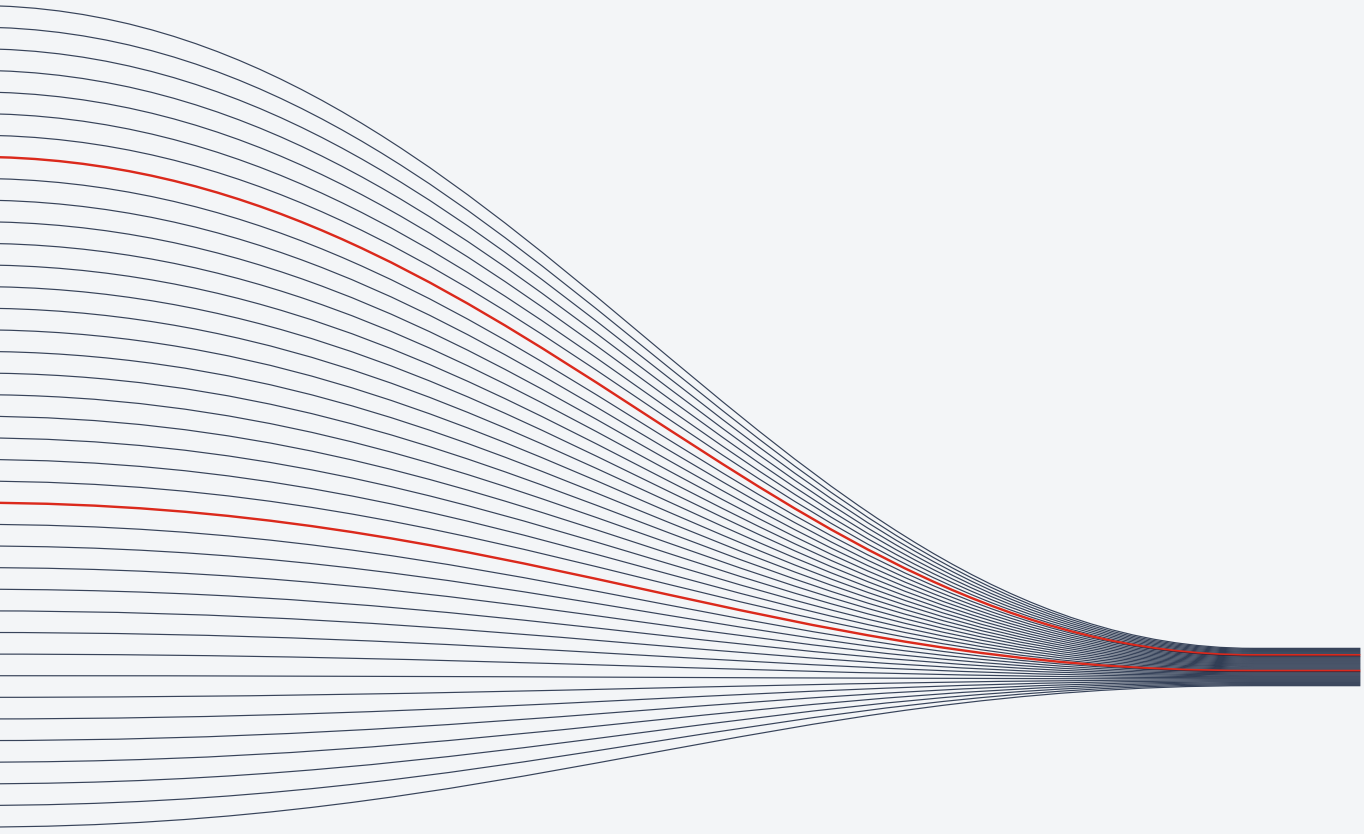
23 Gartner, “Gartner Forecasts Worldwide IT Spending to Grow 14.2% in 2026”, 27.07.2026.
https://www.gartner.com/en/newsroom/press-releases/2026-07-27-gartner-forecasts-worldwide-it-spending-to-grow-14-point-2-percent-in-2026-totaling-6-point-37-trillion

24 Associated Press, “Deloitte to partially refund Australian government for report with apparent AI-generated errors”, republished by CTV News, 07.10.2025.
https://www.ctvnews.ca/business/article/deloitte-to-partially-refund-australian-government-for-report-with-apparent-ai-generated-errors/

25 Computing, “EY cybersecurity report pulled after probe finds AI hallucinations”, 2026.
https://www.computing.co.uk/news/2026/ai/ey-cybersecurity-report-withdrawn-ai-hallucinations

26 Brynjolfsson, E. and Petropoulos, G., “Pricing Consulting Services in the Age of Agentic AI: A Strategy Map for Executives”, Stanford Digital Economy Lab, 23.08.2026.
https://digitaleconomy.stanford.edu/publication/pricing-consulting-services-in-the-age-of-agentic-ai-a-strategy-map-for-executives/

NOTA Exposure assessments and scenarios are the author’s qualitative judgements, based on the sources listed. Press sources are flagged as such; where an article was republished by other outlets, the version consulted is cited. All links were checked on 29 and 30 October 2026; the Financial Times article requires a subscription.



Who puts AI into production

Labs, capital and consulting.

Ivan Ortenzi is Chief Innovation Evangelist at Bip, where he leads Corporate Innovation and AI strategy projects. He is academic coordinator of the Master in Innovation Management at 24ORE Business School, teaches on Executive MBA programmes at several Italian business schools, is the author of six books published by FrancoAngeli and il Mulino, and hosts the Future Proof podcast.

The views expressed are the author's own and do not necessarily represent the position of Bip. Data updated to 30 October 2026.

© 2026 Ivan Ortenzi