GRAIL

AN INDUSTRY BRIEFING FROM GRAIL

What's Changing in the Methodology You Sell

Your platform budget is two instruments pretending to be one. The methodology your firm is famous for sits in forty partner heads and in a shared drive nobody has indexed. Your scorecard measures utilisation and billable hours, and neither is the 2026 question. A briefing for technology and product leaders at mid-market European strategy and management-consulting boutiques.

PREPARED BY

Johan Grönstedt · GRAIL

FOR

European mid-market strategy and management-consulting technology leaders

DATE

April 2026

EXECUTIVE SUMMARY

Monday 9:15, partnership capital review

Every technology or knowledge leader at a mid-sized European consulting boutique has had the same Monday. Utilisation sits at seventy-one percent, below the seventy-five percent healthy threshold. Industry EBITDA just landed at 9.8 percent, a five-year low. Your Managing Partner opens the review with the dashboard. He is mid-sentence when your phone buzzes. A senior partner emails one line: "Client asked in yesterday's steering meeting what percent of last week's deliverable was AI-generated, and what that means for our rate card." Your Head of Practice Groups mentioned over coffee that two senior analysts are interviewing at Unity Advisory. Your CFO forwarded a note from the partnership committee. "Worth discussing the compounding-platform line in next week's meeting." You have a capital ask on the agenda for the first time in four quarters.

You have run the technology function at this firm for years. The pattern used to be simple. Build internal tools. Keep the knowledge base tidy. Support the analysts with research templates. Meet the partnership on IT cost and uptime. The business was understandable.

Something changed. What used to take an analyst week now runs on an agent afternoon. McKinsey's Lilli is used by seventy-two percent of McKinsey's workforce. Bain's Sage is an IP-trained copilot. Unity Advisory raised three hundred million from Warburg Pincus with the explicit pitch that they are AI born, not AI-retrofitted, and they eliminated the pyramid on day one. McKinsey Agents, BCG X, and Bain Vector are not experimenting. They are selling. DiligenceSquared runs AI voice agents replacing the half-million-dollar due-diligence engagements your senior partners used to price on a project sheet.

You are not running one firm. You are running two, and only one is on your partnership distribution. One firm sells partner hours and pyramid production. The other firm is the platform layer that decides whether McKinsey Agents absorbs half your methodology in 2028 or you ship senior-partner-plus-agent composites that outrun them. On one hurdle rate the first wins every quarter. On the partnership distribution the second does not exist, because it does not generate billable hours this year.

This briefing covers what is happening in your sector, what it means for the platform function you run, and the three questions your partnership should answer before the next capital review.

LENS ONE · YOUR INDUSTRY

Strategy consulting just lost the equation that priced the pyramid

The AI-natives in your category are not experimenting. They are building. Unity Advisory raised three hundred million from Warburg Pincus to build a consulting firm with no pyramid. DiligenceSquared replaces McKinsey and Bain due-diligence engagements at a tenth of the cost using AI voice agents. McKinsey Agents, BCG X, and Bain Vector productise the methodology layer the boutiques thought was their moat. Your clients' procurement teams run comparison quotes where your blended daily rate sits next to an AI-native's outcome-based number. They are not asking which is the better firm. They are asking which pricing mechanic their CFO can defend.

The moat your firm ran on for a decade has moved. Sixty to seventy percent of junior analyst work, benchmarking, financial modeling, market sizing, slide production, literature review, is AI-executable today. McKinsey's Lilli already runs across seventy-two percent of their workforce and compresses research synthesis by thirty percent. Bain's Sage is an IP-trained copilot. BCG's Deckster builds decks in minutes. Industry EBITDA dropped to 9.8 percent, a five-year low across four hundred and three firms surveyed. Firms that kept time-based pricing grew 2.1 percent annually. Firms that moved to value-based grew 8.7. Seventy-three percent of clients now prefer outcome-based pricing. The billing arbitrage window is twelve to eighteen months before clients start systematically asking what percent of each deliverable was AI-generated.

> *Your deck templates ship on GPT-5 and commoditise in eighteen months. The case history, pattern recognition, and client-industry intelligence your firm has accumulated across fifteen years of engagements is the asset McKinsey Agents cannot replicate without your case base.*

The industry's speed mismatch is the third force. AI-natives build from scratch with no legacy, no partnership governance, no sixty partners with individual books and individual vetoes on capital allocation. Your firm has all three. You cannot pivot at their speed on their terms. What you can do is identify what is genuinely defensible in your methodology, stop defending what is not, and encode the moat they cannot build. The firms already moving are quiet about it. The ones that are rolling out ChatGPT Enterprise and calling it an AI strategy are the ones the partnership committee will be voting out of relevance in 2028.

Your platform function is feeling it every day

Your build engine. Your analysts run ChatGPT and Claude on every deliverable now. First-draft decks land on partner desks faster. Review time has not moved. Partners sign off on documents that used to be analyst-authored, some with the same rigour and some not. The review discipline that worked when a manager walked an analyst through three drafts is training a different practice on agent-generated volume. Twelve months in, the firm ships faster and catches less.

Your platform and data. McKinsey spent four years on Lilli and it now runs across seventy-two percent of their workforce. You cannot replicate four years in eighteen months. You can buy the research and deck-generation discipline from the tool layer and redirect the compounding platform capital into the layer the vendors cannot sell. Methodology encoded as executable infrastructure. A case-memory graph indexing fifteen years of engagements and tagging the pattern recognition underneath. A client-industry intelligence platform published for your top ten sectors. The tool layer commoditises. The methodology-and-case layer compounds.

Your product thesis. Your Managing Partner asked for the firm's three-sentence thesis last quarter. Your answer was the one from three years ago. Your rate card assumes AI-era deliverables price like analyst-authored ones. They do not. Unity Advisory and the Big Four's own internal AI arms run outcome-based delivery on the engagements your firm still prices by the day. That is a capital-allocation question about where the moat gets built, not a pricing one. Your Managing Partner and CFO need a joint document with you before clients force the conversation on their terms.

Your R&D bench. The routine work that trained analysts into managers into partners is now running on agents. No routine work, no training ground, no senior bench in 2030, no internal partner candidates the year after. The OpenAI and Anthropic pattern from late 2025 is to resume junior hiring specifically to pair juniors with seniors on judgment-intensive work. The consulting firms that cut the analyst layer for this year's utilisation have a cheaper function in Q4 and an empty bench in five years.

Your partnership-level authority. Every partner with a book of business has an implicit veto on platform investment that squeezes their practice this quarter. That is how the platform budget gets voted down four quarters in a row while McKinsey Agents keeps shipping. Your Q1 governance move is the three-page paper to the partnership committee that names the two budgets, the two hurdle rates, and the two kill criteria. Without it the platform capital compresses back into the IT line and the moat gets built by someone else.

BUILD. DEFENSIBILITY. CAPITAL.

The moat that matters in 2028 is not the deck library

Your sector is repricing from below because AI-natives are shipping outcome-based engagements at a fraction of your unit economics. Your function is compressing because an analyst week now runs on an agent afternoon, and the review bar that worked for analyst-authored decks cannot hold the volume agents produce. Same force, two altitudes. But there is a third thing happening most mid-market consulting CTOs have not named yet.

The moat that matters in 2028 is not the deck library or the methodology PDF on the partner wiki. It is the layer underneath. The deck template you ship this quarter on GPT-5 is on an eighteen-month obsolescence curve. The next foundation model rolls in 2027 and the deck-generation advantage commoditises. What does not commoditise is the case history your firm owns, the pattern recognition encoded into diagnostic agents trained on your IP, and the client-industry intelligence platform nobody else has indexed. McKinsey Agents does not have your firm's case base. A two-year-old AI-native cannot replicate fifteen years of engagements. The window closes in roughly eighteen months, which is the time between now and when McKinsey Agents has trained on enough publicly available consulting output to approximate your methodology layer.

In 1981, a former Salomon Brothers bond trader left his firm with a severance check. Reuters owned distribution. Thomson owned the data licensing. He did not try to out-distribute them. He built a terminal that was both a product and his internal research platform, and he spent a decade encoding data and workflow into a layer Reuters could not replicate without a ten-year investment they were never going to make. By the mid-1990s the terminal was the moat. Distribution did not change. The unit of judgment had.

> *You can tell which camp a firm is in from a single question. Is the platform budget you defended at the last partnership meeting a plateau budget, a compounding budget, or one budget pretending to be one when it is two? Most mid-market incumbents cannot answer. The ones that can have already made the shift.*

You have twelve to eighteen months. The technology leaders who claim the authorship of this decision in Q1 author the decade. The ones who wait until the next McKinsey Agents enterprise release lands in a client procurement conversation inherit a seat that has narrowed to IT uptime and tool rollouts, and a scorecard that stopped mattering the quarter a senior partner took an offer from Unity Advisory.

TWO PATHS

What most firms are doing. What the winners are doing instead.

Every consulting-boutique CTO has had the AI conversation. Most are making the same four decisions. A smaller group is making different ones.

THE DEFAULT PATH	THE WINNERS' PATH
Roll out ChatGPT Enterprise firm-wide Analyst throughput up. Review time flat. Partners sign off on first-draft decks. Foundation rots under a green utilisation dashboard.	**Buy the tool layer. Redesign the review bar.** Partner sign-off as floor. Case review on agent diagnostics. Eval harness tracking partner acceptance rate on first draft.
Keep platform spend inside the IT line Every quarter a partner with a squeezed practice vetoes the capital. The compounding budget never accumulates. McKinsey Agents ships the moat first.	**Split platform capital into two budgets** Two hurdle rates, two kill criteria. Plateau delivers utilisation. Compounding builds the 2028 methodology moat.
Price AI-era deliverables on the blended day rate Client sees analyst-week fee for agent-afternoon work. First named rate-card challenge arrives from procurement. Margin compresses.	**Run one practice on outcome-based delivery** Parallel P&L. Joint Pricing Architecture with Managing Partner and CFO. Value-based for new work, day rate for installed book.
Cut the analyst layer for utilisation this quarter Cheaper function in Q4. Empty succession bench by 2031. No internal partner candidates the firm can absorb.	**Pair analysts with senior partners on pattern work** OpenAI and Anthropic resumed junior hiring in late 2025. Eighteen-month apprenticeship on judgment, not five years on benchmarking.

> *Same partnership. Same twelve months. Opposite outcomes. The difference is whether the CTO authors the capital allocation or signs the IT requisition.*

THREE QUESTIONS

Build Velocity. Product Defensibility. Platform Capital Allocation.

Three questions every consulting-boutique CTO is tracking. The third is the crux. The first two are how you earn the right to answer it.

01 · BUILD VELOCITY

Is our deliverable speed shipping partner-defensible work, or first drafts that fall over in the client CFO's hands?

Analysts run ChatGPT and Claude on every deck. Throughput is up. Review time has not moved. McKinsey spent four years on Lilli before it ran across seventy-two percent of their workforce. You cannot replicate four years in eighteen months. Buy what the tool layer can carry. Own the review discipline that makes it ship.

02 · PRODUCT DEFENSIBILITY

What does our methodology do that McKinsey Agents or Unity Advisory cannot copy?

Your deck library commoditises every eighteen months when the next foundation model rolls. Your case history and client-industry intelligence compound. Your methodology sits in forty partner heads and in a shared drive nobody has indexed. The moat sits underneath the tool layer, not above it. The window to build it is eighteen months.

03 · PLATFORM CAPITAL ALLOCATION

Is our platform budget one instrument or two?

Plateau capital compresses analyst production at lower unit cost. Compounding capital builds the 2028 methodology moat. On one hurdle rate the first wins every quarter. On one scorecard the second does not exist. The CTO who walks into the partnership meeting with one budget runs the same programme every peer is running. The one who walks in with two, each defended separately, authors the decade.

FOUR MOVES

What the firms winning this play are actually doing

Four concrete moves any consulting-boutique CTO can start this quarter. Each addresses one sub-function. Pulled together, they become one architecture the partnership builds on.

1

Build engine: absorb agent velocity without the review bar collapsing

Instrument partner-acceptance rate on first-draft deliverables, not just turnaround time. Case review mandatory on every agent-generated diagnostic above complexity threshold. Evaluation harness tracking partner sign-off rate as a first-class metric. Senior-to-junior ratio rebalanced around the composite rather than preserving the 2022 team shape. The dashboard is no longer utilisation. It is review depth against deliverable volume.

2

Platform and data: build the moat underneath the tool layer

Three concrete programmes that stack over eighteen months. Methodology encoded as executable infrastructure: the firm's diagnostic patterns wired into retrieval agents trained on your IP. A case-memory graph indexing fifteen years of engagements and tagging the pattern recognition underneath. A client-industry intelligence platform published for your top ten sectors, maintained as structured metadata the partnership can walk into a renewal conversation with. The tool layer commoditises. This layer compounds.

3

Product thesis: run one practice on outcome-based delivery

Do not migrate the whole book. Keep day-rate for the installed book that does not want to switch. Ship outcome-based on a separate P&L for new engagements and for clients explicitly asking. Write the cannibalisation trigger and kill criteria in Q1: if outcome revenue on an account exceeds day-rate revenue, trigger the migration conversation. The delivery architecture supports both pricing models natively. The joint Pricing Architecture Document with your Managing Partner and CFO is signed before the first outcome-based engagement closes.

4

R&D bench: rebuild the analyst pathway around senior-plus-agent pairing

Your 2024 decision to freeze analyst hiring saves Q1. It empties the 2031 partner bench. Reverse it. Analysts in 2026 do not start on benchmarking, deck drafting, or market sizing. The agent layer absorbed that work. They start on pattern recognition, exception-handling, and paired composite work with a named senior partner for their first eighteen months. The career curve compresses from five years to months because the senior's judgment is legible in the agent-and-analyst review loop. By 2029 you have a mid-tier bench the competition cannot hire away because it did not exist anywhere else to hire from.

The only firm that joins methodology authorship to working infrastructure

The consulting-boutique CTO has four places to go. None can do what we do.

The Big Four digital arms (McKinsey Digital, BCG X, Deloitte, Bain X) ship a standardised AI-governance framework scaled across a sector book. Their cost structure requires reusability. Your methodology is idiosyncratic. Worse, three of the four are shipping McKinsey Agents, BCG X, and Bain Vector at your clients directly.

Enterprise AI-platform vendors productising the methodology layer (Harvey-for-consulting entrants, Glean, Hebbia, Rogo) sell a methodology-as-service across twenty firms at once. Your methodology becomes their training corpus. Their product works; it is just not yours.

In-house from-scratch platform build inside partnership governance gets voted down every quarter by a partner whose practice group gets squeezed this cycle. The compounding capital never accumulates. The moat gets built by someone else.

Individual partners using ChatGPT is not a platform strategy. No shared methodology layer, no case graph, no client-industry intelligence that compounds. The firm's IP stays in forty heads and McKinsey Agents ships the same diagnosis without your partners.

> *The four options you have already looked at. None can run the capital-allocation conversation the seat requires. GRAIL is the fifth.*

WHAT YOU GET

- Senior advisory that speaks partnership-capital, not transformation-deck
- Working infrastructure your team uses the same week
- Custom agents built on your methodology, your cases, your IP
- The start of a methodology-and-case-memory moat your firm owns
- No vendor lock-in at any phase of the engagement

WHY IT MATTERS

- GRAIL is itself the post-transformation consulting firm you are becoming
- Business first, technology second
- Three weeks from first meeting to working infrastructure
- Your team keeps the infrastructure, you own the IP
- Every next step earned by what just shipped, never assumed

Three sessions with your partnership. Working infrastructure from week one.

We start with your Managing Partner and your senior partnership because the change cannot outrun the partnership's personal understanding of what is possible. Three workshops. Each delivers standalone value. Each earns the right to the next.

WORKSHOP 01

The mirror

Walk through what is happening in your sector and inside your platform function. Build the first piece of infrastructure in the room: a capital-allocation agent that takes your real platform budget, splits it two ways, and runs alternatives analysis against McKinsey Agents, Unity Advisory, and Harvey-for-consulting. Your partnership leaves with a shared view and the first working artefact in hand.

WORKSHOP 02

The augmented partner

A half-day build. Each senior partner leaves with five to ten pieces of working infrastructure tuned to their practice. Client-preparation agent. Proposal generator trained on your firm's IP. Benchmarking agent. Case-memory retrieval prototype. Pattern-recognition scaffolding. Infrastructure your partners use the same week, not decks about infrastructure.

WORKSHOP 03

The new model

Refined infrastructure based on real use. The partnership document that splits the platform budget into two instruments with two hurdle rates and two kill criteria. The boundary agreement with the Managing Partner on who authors the AI thesis. The joint Pricing Architecture Document with your CFO. The eighteen-month methodology-moat build plan.

> *Three weeks. No IT project. No partnership vote required to start. The infrastructure your firm is already running on when we finish.*

FIVE QUESTIONS

For your next partnership capital review

The questions where your partnership cannot agree on an answer are the ones worth an hour on the agenda. Ask these honestly.

> *Is your platform budget today one instrument or two? If one, what does the partnership see that splits plateau from compounding, and what is the kill criterion on each?*

> *Name the AI-native in your category. What is their pricing mechanic, and at what percent of your unit economics are they quoting? If you cannot name the top three, that is the first thing this briefing changes.*

> *If two senior partners left for Unity Advisory tomorrow, how many months would it take to reconstruct the pattern recognition and client-industry intelligence behind your core engagements?*

> *ChatGPT adoption is firm-wide. Partner-acceptance rate on first-draft deliverables has not been measured. Where did the freed analyst hours go, and is your review bar still the one you want the next generation trained on?*

> *Industry EBITDA at 9.8 percent, a five-year low. Firms on time-based pricing growing 2.1 percent. Firms on value-based growing 8.7. Is your Q1 boundary agreement with the Managing Partner written, or is it waiting until after the next McKinsey Agents enterprise release?*

> *If any of these produces a disagreement in your partnership, that is the conversation worth having before the next capital review.*

JOHAN GRÖNSTEDT · GRAIL · JOHAN@GRAIL.WORKS · GRAIL.WORKS

IN THIS SERIES

There’s a version of this briefing for every seat at the table.

GRAIL has produced industry briefings calibrated to the specific reality of each C-suite role.

CEO

Revenue. Margin. Valuation.

Three numbers your board reads. AI is rewriting all three in your industry.

CFO

Capital. Forecast. Story.

What you allocate this year is the story your board reads next year.

COO

Cheaper. Better. Faster.

The triple threat that used to be a tradeoff is now table stakes.

CHRO

Capacity. Capability. Promise.

The biggest talent transition of a career. Compete on capacity, not headcount.

CPTO

YOU ARE READING

Moat. Speed. Allocation.

The platform moat that survives 2028 is being chosen this year.

CCO

Margin. Pace. Position.

The whole GTM playbook is being rewritten in real time, across every commercial team at once.

Visit grail.works for the full series.

GRAIL

An AI-native management consultancy helping European mid-market leadership teams figure out what fundamentally changes about their business as AI enters their industry.

www.grail.works