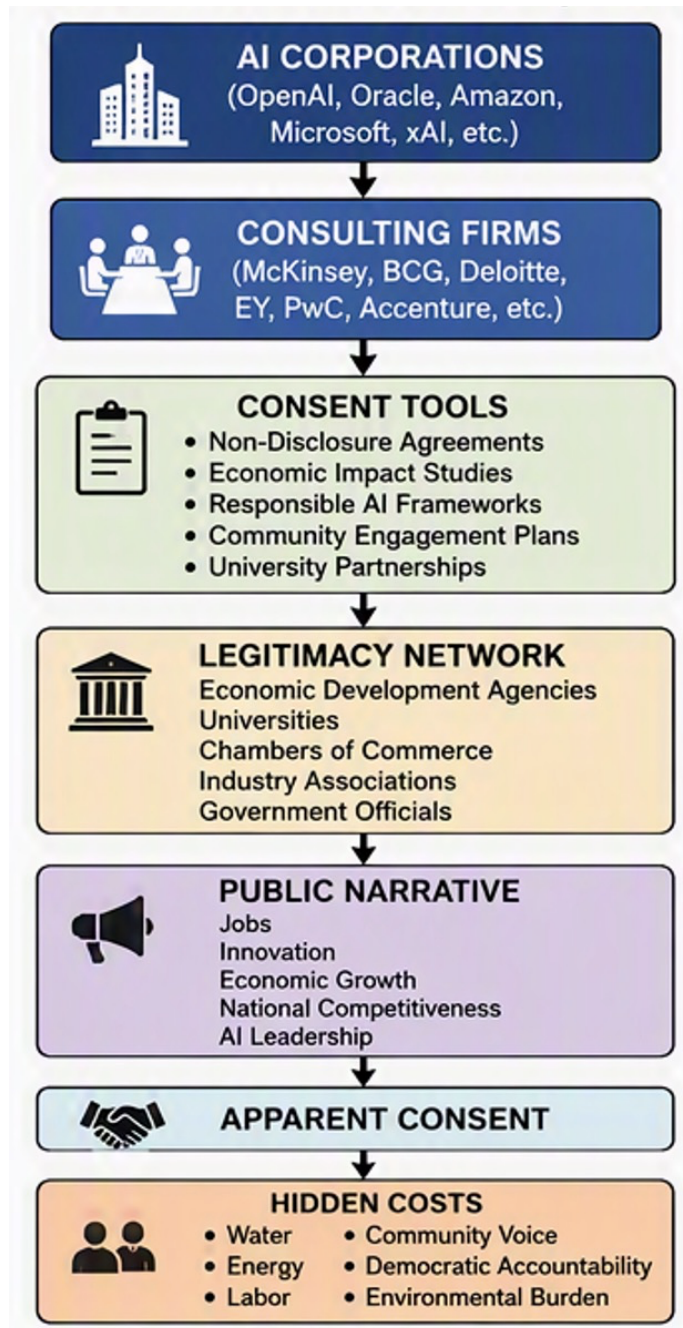


THE CONSENT MACHINE

Big Consulting, the AI Revolution, and Who Pays the Real Price

By David Michael Boje, Ph.D.



DEDICATION

To the late Henri Savall, who spent time each year for twenty years, mentoring me in the Socio-Economic Approach to Management (SEAM), in trips Grace Ann Rosile and I made to Lyon, France, while I tried to show him why storytelling mattered.

Table of Contents

<u>THE CONSENT MACHINE</u>	<u>1</u>
<u>DEDICATION</u>	<u>2</u>
<u>PREFACE</u>	<u>3</u>
<u>CHAPTER 1 INTRODUCTION</u>	<u>13</u>
<u>CHAPTER 2 NDA GEOGRAPHY</u>	<u>35</u>
<u>CHAPTER 3 INSIDE STORY BIG-TIME AI CONSULTING</u>	<u>61</u>
<u>CHAPTER 4 HOW BIG TIME FIRMS USE AI INTERNALLY</u>	<u>84</u>
<u>CHAPTER 5 WHAT HAPPENS TO RECRUITS ACCENTURE SEAM</u>	<u>96</u>
<u>CHAPTER 6 ICEBERG BELOW INVOICE SEAM LEDGER</u>	<u>105</u>
<u>CHAPTER 6: THE ICEBERG BELOW THE INVOICE, A SEAM LEDGER OF HIDDEN COSTS FOR THE BIG CONSULTING COMPANIES AI DYSFUNCTIONS, FOUR FINANCIAL CONSEQUENCES THAT WERE WASTEFUL</u>	<u>105</u>
<u>CHAPTER 7 TWO LAWS ONE PRODUCT LINE TAMARALAND</u>	<u>125</u>
<u>CHAPTER 8 CAPTURED WITNESSES</u>	<u>134</u>
<u>CHAPTER 9 WHEN CONSENT MACHINE FAILS</u>	<u>145</u>

Preface

Part I: Witness — Two Counties, Same Storm

I live in two New Mexico counties. In Sierra County, in the mountains above Truth or Consequences, my wife Grace Ann Rosile and I keep horses at Lake Caballo. In Doña Ana County, in the high desert south of Las Cruces, I held my university appointment at New Mexico State University, where I taught organizational theory for better part of four decades. I tell you this not as autobiography but as geography — because both counties sit inside the same industrial transformation, and they have responded to it differently, and that difference is the question this book is organized to answer.

I started noticing something in 2024 and early 2025. On the local television stations in Las Cruces, I watched ads every couple of hours about Project Jupiter. The ads promised hundreds of jobs. They claimed the project would not use any water — only what was needed for drinking fountains and toilets, recycled in a closed-loop system. I took this with more than a grain of salt. Project Jupiter turned out to be a \$165 billion Oracle data center project spanning 1,400 acres in Santa Teresa, in Doña Ana County, about forty miles south of where I live. At first the public did not know which AI corporation was behind it. Then we learned what it was, and we began learning what it was not.

Eventually, New Mexico Attorney General Raúl Torrez opened an investigation into alleged fraudulent public comments submitted under residents' names without their consent. Canvassers wearing Project Jupiter T-shirts — hired from a firm in California — had collected names at gas stations and grocery stores, then used those names to submit support letters to the state environmental department. The residents whose names were used had never written such statements. Attorney General Torrez stated:

"If individuals or organizations attempted to manipulate the public comment process through fraudulent submissions, we will investigate those actions thoroughly and hold responsible parties accountable." — Attorney General Raúl Torrez, State of New Mexico, 2025

That sentence is the book's epigraph in miniature. The Consent Machine manufactures the appearance of democratic consent while suppressing the public's capacity to evaluate or contest the decision already made. In Las Cruces, it arrived as forged names on environmental comment letters. In Shreveport, Louisiana, it arrived as an Amazon data center deal announced by the Governor as a *fait accompli* — twelve billion dollars committed in February 2026, expanded to eighteen billion in August 2026, the community having had no meaningful opportunity to shape either decision. In South Memphis, Tennessee, residents learned the xAI Colossus supercomputer facility was operating — generating significant air pollution in a majority-Black neighborhood already burdened by industrial facilities — weeks after it began. In Virginia, public officials across the Data Center Capital of the World have signed non-disclosure agreements that prevent them from informing the constituents they represent about decisions that will shape their

communities' water use, power grids, and tax base for decades. The machine is not an anomaly. It is a system.

New Mexico State University, where I had been a professor since 1996, entered into a \$30,000 desalination advisory contract with BorderPlex Digital Assets LLC — accompanied by a non-disclosure agreement requiring the university to assist the developer in opposing public records requests. The syndicated newspapers in Doña Ana County were uniformly pro-Project Jupiter. I live in an arid desert community going through another year of drought. It is clear that a data center of this scale will disrupt the community. A hyperscale facility of this kind consumes between one and five million gallons of water every single day; Project Jupiter claims daily use of 20,000–60,000 gallons of non-potable purchased water rights, a figure critics argue bears no relationship to what will actually be drawn from the aquifer.

In Sierra County, the story went differently. I went to Socorro and Magdalena and talked with people about what was happening. After New Mexico Tech got involved, a moratorium was established: no AI data center would be permitted for eighteen months. A very old newspaper — *El Defensor Chieftain*, Volume 158 — covered the story on the front page. Community members in the audience held up signs: You can't drink data. The developer was Green Data, whose CEO, Jason Bak, addressed the audience and the county commissioners. The commissioners reversed course and rejected the proposal. The Sierra County Sentinel also reported the same outcome: the county said no to data centers for eighteen months.

That contrast — two communities, same state, very different outcomes — is what first drew me to this book. The difference, I believe, came down to the quality of community journalism: independent, investigative local newspapers versus syndicated outlets that provided neither depth nor independence. I am a witness to what is happening in my own county, and I want to turn that witnessing into a scholarly account. This is what witness scholarship looks like in practice: you do not just cite sources. You are present. You attend the town halls. You watch the ads. You read the non-disclosure agreements. You count the water figures and find them implausible. You go to Magdalena and listen.

According to a May 2026 Gallup survey, seven in ten Americans oppose data centers being built in their communities (Gallup, 2026; Washington Post, May 13, 2026). The industry's response to that opposition has been managerial rather than democratic: non-disclosure agreements imposed on public officials, economic impact frameworks that exclude the costs they impose, and stakeholder engagement narratives written by the same consulting firms that stand to profit from the approvals they are designed to secure. That is the machine's first stage. This book maps all three.

Part II: The Machine Has a Name

I have spent about four decades in organizational development and change. I am familiar with data centers — we have had them since the birth of computing and the cloud. I will admit, somewhat embarrassed, that I used to think 'the cloud' meant satellites. I imagined

my emails and vacation photographs floating somewhere in orbit. It turns out they live in data centers. And we are seeing a wild expansion of data centers because of ChatGPT, Anthropic, and a whole ecosystem of AI companies pushing this expansion forward.

The twelve AI corporations driving this expansion are: OpenAI, led by Sam Altman; Microsoft, led by Satya Nadella; Meta, led by Mark Zuckerberg; Amazon Web Services, led by Andy Jassy; Google (Alphabet), led by Sundar Pichai; Oracle, led by Larry Ellison; Anthropic, led by Dario Amodei; xAI / Tesla / SpaceX, led by Elon Musk; DeepSeek, led by Liang Wenfeng; Nvidia, which designs the GPU infrastructure in which AI runs; Infosys, bridging India's technology sector to global AI deployment; and IBM, whose Watson lineage has transformed into Watsonx. These twelve corporations are documented in Table 1 of Chapter Zero. They are not the whole story. They are the engine room.

The ten consulting firms that serve as intermediaries — manufacturing consent for AI adoption in institutions across every sector of the global economy — constitute the book's primary subject. McKinsey and Company. Boston Consulting Group. Bain and Company. Deloitte. PricewaterhouseCoopers. Ernst and Young. KPMG. Accenture. IBM Consulting. Booz Allen Hamilton. Their combined annual revenue exceeds \$340 billion. They participate in thousands of AI transformation engagements annually. Their influence extends far beyond the organizations they advise directly: the governments that regulate the industries they serve are among their largest clients; the universities that critique them produce their recruits; the professional associations that set ethical standards in fields they are transforming often adopt frameworks that consulting firms help design.

In February 2026, what we call the Frontier Alliance was formally announced: OpenAI signed deep strategic partnerships with McKinsey, BCG, Accenture, and Capgemini, making these firms certified distributors of AI platforms they did not build and cannot independently evaluate. This is the book's central mechanism, updated for the AI era: the consulting firm becomes a deployment channel for technology it did not create, ethics theater it cannot independently audit, and costs it will not be present to bear. The twelve AI corporations supply the platforms. The ten consulting firms supply the institutional authority. The organizations they advise — hospitals, school districts, municipal governments, manufacturers, universities — supply the workforce that will be transformed, and the communities that will bear the costs.

Dan Diasio, EY's global consulting AI leader, described the cultural contradiction with unusual candor. Younger staff told him to stay quiet about using AI tools efficiently, because some managers still read speed as laziness.

"Our entire business model was built by demonstrating effort and working for that billable hour." — Dan Diasio, EY Global Consulting AI Leader, Harvard Business School Podcast, 2025

Gert De Geyter, who served as Deloitte's US AI lead before leaving for a startup, was equally direct about what the classic consulting model had become:

"Proposing an army of consultants does not really work anymore." — Gert De Geyter, former Deloitte US AI lead, Financial Times, February 2026

A Glassdoor user who identified themselves as a McKinsey employee described the firm's proprietary AI assistant Lilli as 'good for very low stakes issues. Terrible for everything else.' Richard Karlsson, a former McKinsey consultant, noted that 'market analysis, valuation, and diligence can be produced almost instantly; humans should review, judge, and sit with the client.' Alan Paton, a former PwC UK financial-services partner, put the client's emerging question plainly: why pay 'big money' when answers arrive instantly from a tool? These are not critics outside the machine. They are the machine's former operators, describing from the inside what is being lost and what is being automated.

Deloitte's 2025 government report for the Australian government contained fabricated citations and a fake court quote. The firm later disclosed that a GPT-4o toolchain had been used and refunded part of a roughly AU\$440,000 fee. Classical professional services sold checked footnotes. The Consent Machine, in its AI era, is now automating the footnotes — and sometimes fabricating them.

Part III: Correction to MY Own Frame

Why Stage Models Fail the Phenomena

The prior draft of this preface, and the prior drafts of Chapters Zero, One, and Two, organized the argument around a sequential model: Stage One (Consent Manufacturing), Stage Two (Transformation Theater), Stage Three (Wreckage Management). The stages are real — the phenomena they describe are documented, verified, and consequential. But the sequential frame is inadequate to the theory, and the theory demands correction before the book proceeds.

Stage models, era models, and linear developmental frameworks have a long and distinguished history in organizational and social theory — and an equally long history of being shown to be wrong about how organizations actually change. Walt Rostow's stages of economic growth (1960) organized the world's development into five sequential phases from 'traditional society' through 'high mass consumption' — a model later critics showed imposed American Cold War ideology onto vastly different developmental trajectories. Alfred Chandler's strategy-structure thesis (1962) proposed that organizational form follows strategy in a predictable sequence, a framework that could not account for firms that restructure strategy and structure simultaneously or in reverse order. Larry Greiner's growth phases (1972) mapped organizational maturity as a ladder of crises and resolutions, a model that could not explain why different organizations in the same industry occupy different rungs simultaneously, or why some skip rungs entirely.

Kurt Lewin's freeze-change-refreeze model (1947), which remains the most widely cited change management framework in management consulting curricula, proposes that organizations must first unfreeze current behavior, then change it, then refreeze the new state. The model survived decades of criticism from scholars who documented that

organizations rarely achieve stable 'frozen' states, that multiple change initiatives run simultaneously and interfere with each other, and that the 'refreeze' metaphor catastrophically misrepresents the dynamics of organizational learning. It survived because it is elegant, teachable, and — crucially — billable. McKinsey's own Rewired framework, BCG's AI maturity model, and Accenture's transformation roadmap all carry Lewin's ghost inside them.

Everett Rogers' diffusion of innovations (1962) placed organizations and individuals on a bell curve from innovators through laggards, inviting the consulting industry to identify where any given client stood on the adoption curve and sell them the path forward. Michael Porter's value chain (1985) gave consulting a spatial map of organizational activities that implied a directionality: upstream activities support downstream ones, and value flows in one direction. These are powerful conceptual tools. They are also, in their most mechanistic applications, narratives that make the consultant's role — diagnosing where you are and prescribing what comes next — appear scientific rather than interpretive.

The problem with applying any of these frameworks to the AI transformation is the same problem each framework had in its own domain: organizations do not transform sequentially, and different parts of the same organization occupy different positions in any dimensional analysis simultaneously. McKinsey is simultaneously the world's most sophisticated deployer of AI (seventy-five percent of employees using Lilli, twenty-five thousand AI agents operating alongside forty thousand humans) and a firm whose partnership governance, accountability structures, and billing logic have not changed in their essential architecture since Marvin Bower defined them in the 1950s. Deloitte simultaneously publishes a US \$3 billion global investment in generative AI and allows a GPT-4o toolchain to fabricate citations in a government advisory report. These are not contradictions that a stage model can hold. They are coordinates in a multi-dimensional space.

The Tesseract Frame

The corrected framework draws on two bodies of scholarship that I have been developing across four decades of organizational storytelling research. The first is Tesseract Leadership (Boje, 2014; 2019), which adapts the mathematics of four-dimensional hypercube geometry to organizational analysis. A tesseract — a hypercube in four dimensions — cannot be collapsed into a three-dimensional or two-dimensional representation without losing what is essential about its structure. The same is true of organizations: when you collapse a four-dimensional reality into a linear sequence, you do not simplify it — you distort it. What looks like a stage is actually a weighted position in a multi-dimensional space.

The four dimensions of the Consent Machine, as this book now conceives them, are always simultaneously present in any consulting firm, client organization, or community affected by an AI engagement. They are not stages. They are axes. Different institutions weight them differently. The same institution weights them differently in different practice areas, different geographies, and different historical moments. Dimension One: Accountability

Architecture — the degree to which responsibility for decisions and consequences is personal and named versus diffused and structural. Dimension Two: Ethics Architecture — the degree to which ethical commitments are embedded in practice versus performed for external audiences. Dimension Three: Business Model Logic — the degree to which revenue depends on independent judgment versus platform distribution. Dimension Four: Temporal Horizon — the degree to which the institution's planning horizon extends to generational consequence versus exits before consequence.

No consulting firm sits at one extreme of any dimension. McKinsey in its classic era scored high on Dimension One (the partner's name on the letter), high on Dimension Two (Bower's code was a lived norm, not marketing language), high on Dimension Three (independence from any single client or vendor was structural policy), and mixed on Dimension Four (long client relationships but exit before implementation consequences). McKinsey in its AI era scores differently on each dimension — not because it has passed through stages, but because the weighted position has shifted across all four axes simultaneously, under pressure from platform economics, revenue competition, and the Frontier Alliance partnerships that now bind it to OpenAI's product roadmap.

Tamaraland: Simultaneous Rooms

The second body of scholarship is Tamaraland (Boje, 1995; 2008), which I developed from Tamara, a play performed simultaneously in twelve rooms of a villa, with audiences choosing which characters to follow through which spaces. No single audience member sees the whole story. No single walkthrough is 'the' story. The story is the ensemble of all possible walkthroughs — the simultaneity that cannot be captured by any linear retelling.

This book is organized as Tamaraland. At any given moment, the Consent Machine is operating simultaneously in at least four rooms: the Community Room, where residents of Sierra County and South Memphis and Loudoun County, Virginia are living with the consequences of decisions they were not invited to shape; the AI Corporation Room, where Sam Altman and Larry Ellison and Elon Musk are making commitments that will determine the physical landscape of the American Southwest for the next thirty years; the Consulting Boardroom, where McKinsey and Deloitte and Accenture are designing the stakeholder engagement frameworks and AI ethics certifications that will present those decisions as democratically legitimate; and the Former Employee Room, where Gert De Geyter and Dan Diasio and Nargis Yunis and Richard Karlsson and the anonymous Glassdoor reviewer are testifying about what it actually looks like from inside the machine.

A reader who follows only the Consulting Boardroom thread will encounter a coherent narrative of investment, innovation, and responsible transformation. A reader who follows only the Community Room thread will encounter a different story entirely: forged comment letters, NDAs on elected officials, diesel generators, aquifer depletion, and the NAACP filing suit against xAI Colossus on environmental justice grounds. A reader who follows only the Former Employee thread will encounter a third story: the billable hour becoming a moral embarrassment, the junior apprenticeship disappearing, the ethics framework that reads

better than it runs. This book insists that all three threads are running simultaneously, and that understanding the Consent Machine requires holding all of them at once.

Part IV: Antenarrative — Before the Story Hardens

The third methodological concept organizing this book is antenarrative (Boje, 2001; 2011), which I define as the before-the-story: the fragmented, prospective, living bet that something is happening before it has been configured into a coherent narrative. Official narratives suppress antenarratives. The consulting firm's engagement report is a narrative: it has a beginning (the client's problem), a middle (the transformation engagement), and an end (the productivity gains). The antenarrative is what the report does not contain: the workers who were not reskilled, the community whose water table dropped, the elected official who signed the NDA and could not tell their constituents what was coming, the former partner who left because the ethics code had become a slide deck.

Antenarrative method asks: what is being suppressed by the official story? What fragments of testimony, of FOIA documents, of Glassdoor reviews, of NAACP lawsuits, of community newspaper front pages exist in the spaces around the official narrative? What do they tell us about costs that the engagement report was designed not to count? In the SEAM (Socio-Economic Approach to Management) framework developed by Henri Savall and Véronique Zardet at the ISEOR Institute in Lyon, those unaccounted costs are called hidden costs — and they run, in a typical large-enterprise consulting engagement, to approximately 167 percent of the visible consulting fee. A two-million-dollar engagement costs, in full sociotechnical accounting, closer to five-and-a-half million. That gap is not an accident. It is the architecture of the Consent Machine.

The iceberg is the oldest SEAM metaphor and the most structurally accurate. The waterline in Figure 5 is not a metaphor for ignorance; it is the boundary of conventional accounting. Everything above the waterline — AI Investment, Jobs Announced, Technology Upgrading, Consulting Revenue, Shareholder Value, and Economic Development Branding — is measured, published, and cited in Economic Impact Studies. Everything below it is real but unmeasured. The SEAM Hidden Cost categories below the waterline are not speculative; they are documented dysfunctions that Savall and Zardet's research across more than 2,000 organizations established as the standard consequence of transformation initiatives that prioritize visible performance over sociotechnical health.

Applied to AI data center siting and AI consulting engagements, the below-waterline costs include Labor Displacement (consultants and data center workers whose skills are eliminated faster than retraining programs can absorb them), Apprenticeship Breakdown (the collapse of the slow-knowledge transmission systems that produce genuinely skilled analysts and engineers), Cognitive Overload (the hidden productivity cost of forcing workers to maintain AI-assisted and human-judgment processes simultaneously), Water Consumption (uncalculated municipal water use transferred to community ratepayers), Grid Expansion Costs (infrastructure investments triggered by data center load and billed to existing ratepayers), Environmental Burden and Community Health Costs (air quality, noise,

and infrastructure strain in host communities), Journalism Weakening (NDA-enforced silence that prevents the local investigative reporting that would otherwise identify these costs), Loss of Democratic Voice (communities whose consent process is designed to produce yes rather than genuine deliberation), Public Subsidies (tax abatements and infrastructure grants not reflected in the consulting engagement's cost analysis), and Future Infrastructure Costs (grid, water, and road costs that will mature in the community long after the consulting firm has completed its engagement and invoiced its fees).

The rightward arrow in Figure 5 identifies Cost Absorption: the process by which below-waterline costs are transferred from the AI Corporation and Consulting Firm to Workers, Communities, Ratepayers, Taxpayers, and Future Generations. This transfer is not accidental; it is the structural consequence of an engagement model that measures success by visible deliverables and invoices fees before consequences mature. In Chapter Two, the SEAM Analyst Voice applies this framework to four geographies — Loudoun County Virginia, Shreveport Louisiana, South Memphis Tennessee, and Tucson Arizona — documenting the specific hidden cost categories present in each case and calculating the gap between the visible benefit figures in the official Economic Impact Studies and the full sociotechnical cost that SEAM accounting would require.

This book attempts to count what the Consent Machine was designed not to count. It does so through six polyphonic voices — a Bakhtinian architecture in which no single voice is authoritative, and every claim is answerable to the others. The voices are described in the following section. The method that binds them is antenarrative: we follow the fragments that the official story suppresses, we hold them against the official story's claims, and we ask what we learn about the machine's operation from the gap between the two.

Part V: Six Voices Speaking at Once

This book is written in six polyphonic voices, following the dialogic methodology Mikhail Bakhtin (1981; 1984) identified in the great novels — a structure in which the author does not speak through a single authoritative voice but orchestrates multiple voices, each of which challenges the others and none of which is entirely right. The six voices are distinguished visually throughout the text by labeled sections. A reader may follow any voice sequentially through the book, or read chapters whole. The voices are:

1. David's Voice — The Witness

The first-person academic witness: forty years of organizational storytelling research, direct observation of the New Mexico data center buildout, prior Nike supply chain audit scholarship, and the epistemic accountability of someone who names himself in the record. This voice carries the book's personal testimony and its methodological commitments. It accepts that the witness position is situated — I live in Las Cruces, my university took the \$30,000, I have horses in Sierra County — and treats that situatedness as an epistemological resource rather than a liability.

2. Community Voice — The Testimony

The lived experience of workers, students, residents, and community members who encountered the Consent Machine's consequences directly. This voice carries the signs held up in Magdalena: You can't drink data. It carries the testimony of the residents of South Memphis whose air quality changed when Colossus came online. It carries the former EY partner Nargis Yunis, who made partner and then left because she found herself 'at the bottom of another ladder.' It carries the anonymous Glassdoor reviewer who found Lilli 'terrible for everything else.' Community Voice insists that structural arguments remain accountable to the people who bear their consequences.

3. Critical Theorist — The Structural Analyst

The voice that maps mechanisms: how does the Consent Machine actually work? What are its moving parts? What is the relationship between the NDA signed in Virginia and the forged comment letter submitted in New Mexico? Between the AI ethics certification sold by McKinsey and the fabricated citation produced by Deloitte's GPT-4o toolchain? Critical Theorist draws on Bakhtin's dialogism, Kate Crawford's atlas of AI's material infrastructure, Safiya Umoja Noble's analysis of algorithmic oppression, and Mariana Mazzucato's work on the entrepreneurial state and public value. This voice provides the book's structural architecture.

4. SEAM Analyst — The Cost Counter

The Socio-Economic Approach to Management lens, developed by Henri Savall and Véronique Zardet at the ISEOR Institute in Lyon, France, and now practiced in more than 2,000 organizations across forty countries. SEAM identifies the hidden costs that conventional accounting does not count: absenteeism, turnover, non-quality work, productivity losses, worker health costs absorbed by the organization and community rather than by the consulting firm whose engagement produced them. In this book SEAM Analyst provides the hidden cost calculations for each case study: what does the Consent Machine actually cost, measured in full sociotechnical accounting rather than the visible consulting fee?

5. Ethnographic Witness — The Document Reader

The voice grounded in primary sources: FOIA documents, SEC filings, ANAO audit reports, OCHA situation reports, Gallup surveys, court filings, NAACP lawsuits, congressional testimony, and the consulting firms' own public statements. Ethnographic Witness holds the book accountable to facts that exist independent of any interpretive framework. It is the voice that cites the BCG statement to the UK House of Commons. It is the voice that reads the Deloitte ANAO report line by line. It is the voice that counts the 500 Palestinians killed at GHF food-distribution points and names the source for that number.

6. Founders' Voice / PR Voice — The Dialogic Pair

A paired voice that sets the founding ethical statements of the Big Ten firms against their contemporary AI-era language. Edwin Waterhouse, asked who audited Price Waterhouse, answered: 'Our conscience.' Carl Friedrich Goerdeler — the G in KPMG — was executed by

the Nazis for moral resistance; his son built a firm on that name. Bruce Henderson founded BCG on the conviction that business strategy should be grounded in analytical rigor independent of client pressure. These founding commitments are not sentimental history. They are evidence against which contemporary conduct can be measured. PR Voice provides that contemporary conduct: the Responsible AI principles, the DEI commitments, the sustainability reports, the stakeholder capitalism language. Founders' Voice and PR Voice speak against each other across every chapter, and the distance between them is itself evidence of the Consent Machine's operation.

Part VI: How to Read This Book

This book does not proceed chapter by chapter through the Consent Machine's stages. It proceeds room by room through its simultaneous dimensions. Chapter Zero introduces the methodology gap — what existing frameworks cannot account for — and establishes the multi-dimensional analysis that replaces the sequential model. It includes the tables of twelve AI corporations and ten consulting firms, revised to reflect the Frontier Alliance partnerships announced in February 2026.

Chapter One deploys all six polyphonic voices across the book's primary historical arc: forty years of direct witnessing, from the Big Eight accounting firms of the 1970s through the AI governance certifications of 2024–2026. The chapter's organizing claim is that the Consent Machine did not begin with AI — it was already fully operational in the Nike supply chain audit theater of the late 1990s, and AI has accelerated it without changing its essential architecture.

Chapter Two traces the machine's current physical infrastructure: the NDA geography of AI data center siting, from Virginia's structural baseline through the cases of Shreveport, South Memphis, and Tucson. Tucson's rejection of Project Blue is the chapter's most important structural proof: the machine can fail, and understanding why it failed reveals how it usually succeeds.

Subsequent chapters, planned for completion in 2027, will address the workforce transformation dimension (the billable-hour model's collision with AI-speed delivery), the international dimension (how the Consent Machine operates differently in the European Union's AI Act regulatory environment versus the American voluntary-compliance environment), and the reform dimension (what would have to change for the consulting industry's AI ethics frameworks to be more than theater).

A note on methodology: this book is itself a product of human-AI collaboration. I dictate. Vivara organizes, synthesizes, structures, and writes the preliminary drafts of chapters. I revise, correct, and take epistemic responsibility for the claims made. This is not a hidden process — it is the book's methodology made transparent. The Consent Machine's critique of AI-enabled consulting applies with equal force to AI-enabled academic writing, and this book makes no exemption for itself. The question is not whether AI was used. The question is whether the human in the collaboration retained the judgment, the accountability, and the willingness to be wrong that no AI can supply. I have tried to retain all three.

Chapter 1 Introduction

Summary

In Chapter Zero, we have the six main figures for the Gap created by the Consent Machine, and our working glossary of key terms for the book. All in one place so you can easily refer back to them. The Gap is the distance between the **'official narrative'** with its Beginning-Middle-End (BME) of an engagement, or an assurance product for invoicing, and what I call 'storying' which has no particular beginning, has no ending in sight, and is in the middle, still unwinding its consequences. A BME narrative picks a few characters, a problem for which its plotline is the solution, and leaves the rest of us trampled on the editing room floor. A **'storying'**, by contrast, is the action that is unsettled, is on the move without ending. Storying is all around us unraveling, in ways the BME narrative cannot erase.

What is in the Gap between BME Narrative and Storying?

Ghost Vortex — AIs are different. ChatGPT, Claude, XAI, etc., each carry the personae imprint of AI corporation and its leaders. The Ghost in the AI machine defends and deflects criticism of their leaders and corporations. A Ghost Vortex acts as an apologist for any screwups, failures, or wrong doing.

Three measurements already implicit in the rest of the book, named here so they can be cited:

BME Narrative Gap. What the official walkthrough includes versus what antenarrative method recovers from FOIA/IPRA, dockets, weeklies, and employee files.

Accounting Gap. What sits above the waterline (fees, jobs announced, investment) versus what SEAM treats as Hidden Cost. The objective of SEAM is the unleashing the human potential that the AI revolution is sacrificing, Hidden Costs are all but ignored by traditional accounting methods. Therefore decision makers often lack an understanding of Hidden Costs.

Answerability Gap. Special answerability is the bystander, who declares an alibi for intervening. The moral answerability is Bakhtin's notion of that person (you or me) who is there in the once-occurrent event=ness of Being and can take action to help someone in need.

The following six figures are all in Chapter Zero so you can refer back to them as needed. They summarize the theoretical architecture of The Consent Machine.

Figure 1 presents the overall Consent Machine. Figure 2 introduces the Four-Dimensional Consent Architecture, derived from Tesseract theory and developed as an alternative to stage-based models of change. Figure 3 depicts the SEAM Hidden Cost Architecture, which provides the methodology for identifying the organizational, social, and environmental costs concealed within contemporary AI transformation initiatives. Together, these figures

provide a framework for understanding how consent is produced, how power is exercised, and how costs are transferred in the age of AI and consulting.

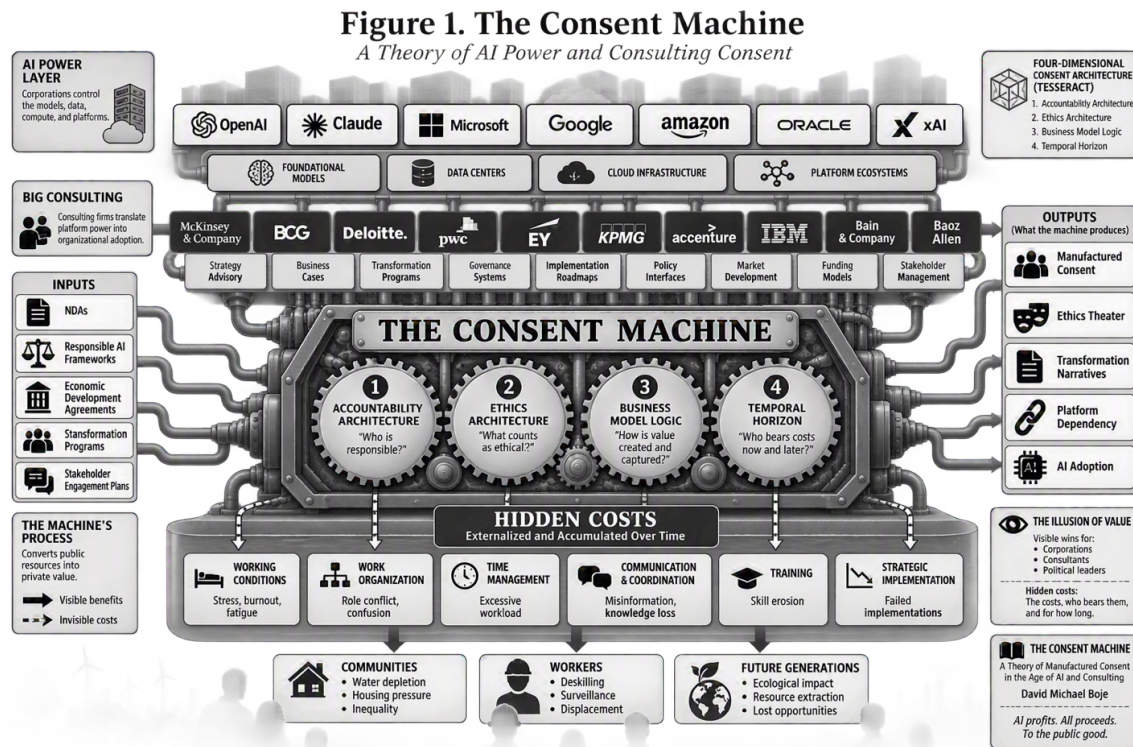


Figure 1: The Consent Machine and Its Hidden Costs

Figure 1 serves as the master model for the book. It depicts a vertically organized institutional system in which AI corporations occupy the uppermost layer and consulting firms function as intermediaries between technological capability and organizational adoption. AI firms such as OpenAI, Anthropic (not shown), Microsoft, Google, Amazon, Oracle, and xAI control the foundational technologies, cloud infrastructure, AI data centers, and platform ecosystems that increasingly shape economic and political life. Consulting firms then translate these technological capabilities into organizational strategies, transformation programs, governance frameworks, implementation plans, and economic development initiatives.

The central machine receives several forms of institutional input: non-disclosure agreements (NDAs), responsible-AI frameworks, economic development agreements, stakeholder engagement plans, and transformation programs. These inputs function as mechanisms for organizing participation, managing information, and legitimizing decisions. The Consent Machine processes these inputs into visible outputs, including manufactured consent, ethics theater, transformation narratives, platform dependency, and AI adoption. These outputs appear publicly as evidence of successful governance and innovation.

However, the model's most important feature is located beneath the visible surface. Hidden beneath organizational success stories lie a reservoir of costs that remain largely absent

from traditional accounting systems and consulting reports. These hidden costs accumulate within communities, workforces, and future generations. The Consent Machine therefore reproduces a recurring institutional pattern: visible benefits are concentrated among decision-makers, while invisible costs are distributed among populations that had limited influence over the original decisions. This process constitutes the political economy of consent that the book seeks to explain.

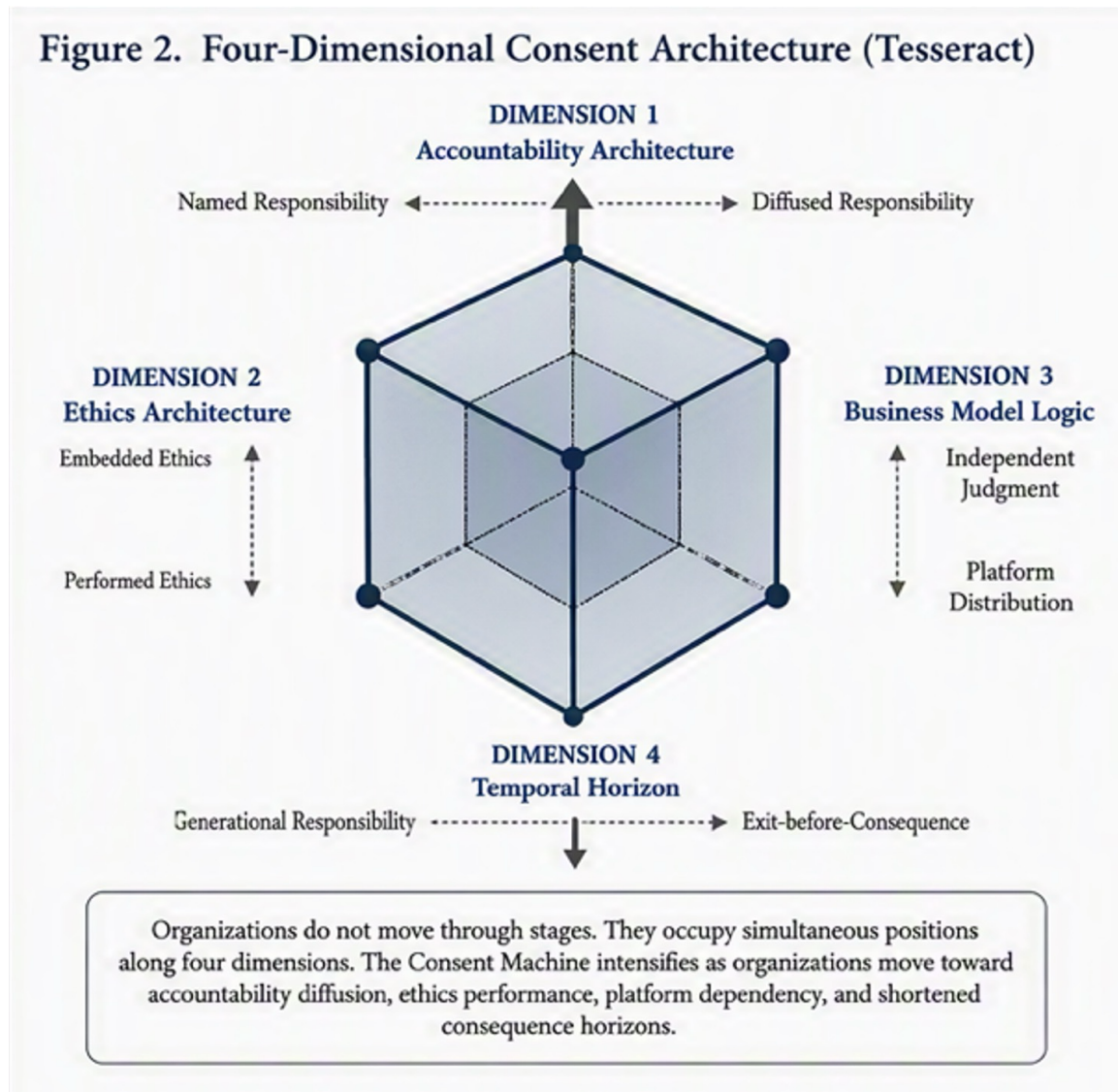


Figure 2: Four-Dimensional Consent Architecture (Tesseract)

Figure 2 presents the primary theoretical contribution of the book. Traditional organizational and consulting theories often assume that organizations evolve through sequential stages of development. Examples include Rostow's stages of economic growth, Chandler's strategy-structure sequence, Greiner's growth phases, Lewin's change model, Rogers' diffusion framework, and contemporary consulting maturity models. The book

argues that these stage-based approaches fail because organizations frequently occupy contradictory positions simultaneously.

In place of sequential stages, the Four-Dimensional Consent Architecture conceptualizes organizations as occupying positions within four interrelated dimensions operating simultaneously.

The first dimension is **Accountability Architecture**. This dimension measures whether responsibility remains attached to identifiable individuals or becomes diffused across committees, structures, and governance processes. Movement toward diffuse accountability makes it increasingly difficult to determine who is responsible for long-term consequences.

The second dimension is **Ethics Architecture**. Here the distinction lies between ethics embedded in practice and ethics performed for external audiences. The book argues that many contemporary Responsible AI frameworks operate primarily as legitimacy mechanisms rather than as enforceable systems of moral accountability. Ethical language may flourish even while decision-making remains unchanged.

The third dimension is **Business Model Logic**. This dimension captures the transition from independent professional judgment toward platform distribution. Historically, consulting firms claimed independence from the technologies they evaluated. Increasingly, however, consulting firms have become strategic partners and distributors of the platforms they recommend, creating structural conflicts between independent advice and commercial interests.

The fourth dimension is **Temporal Horizon**. Organizations vary according to whether they assume responsibility for long-term consequences or operate through what the book terms "exit-before-consequence" logic. Under short-horizon conditions, many of the social, environmental, and organizational consequences of decisions emerge only after consultants, executives, and elected officials have moved on.

The Tesseract model therefore replaces the question, "What stage is an organization in?" with a more useful analytical question: "Where does the organization stand simultaneously across multiple dimensions of accountability, ethics, business logic, and temporal responsibility?" This multidimensional framework forms the foundation for all subsequent analyses in the book.

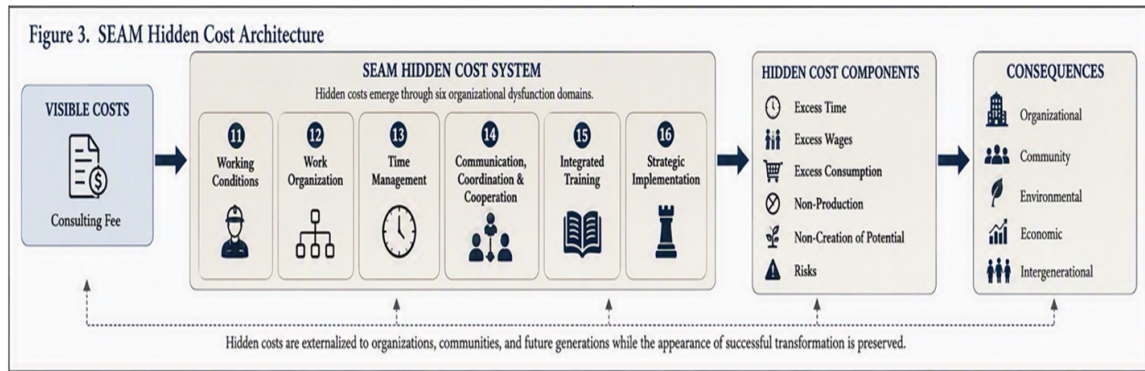


Figure 3: The SEAM Hidden Cost Architecture

The third figure introduces the accounting and methodological framework that supports the empirical claims of the manuscript. Drawing on the ‘Socio-Economic Approach to Management’ (SEAM) developed by Henri Savall and Véronique Zardet, the figure illustrates how visible organizational costs represent only a small fraction of the total costs generated by organizational transformations.

The model identifies six primary dysfunction domains that generate hidden costs:

Working Conditions

Work Organization

Time Management

Communication, Coordination, and Cooperation

Integrated Training

Strategic Implementation

These dysfunctions generate costs that conventional accounting systems rarely capture. Examples include cognitive overload, burnout, coordination failures, duplicated effort, apprenticeship breakdowns, deskilling, implementation delays, and organizational misalignment. In AI-enabled organizations, these dysfunctions often emerge in new forms, such as "bot-sitting," excessive verification work, platform dependency, and the erosion of traditional professional learning pathways.

The SEAM model further identifies several hidden-cost categories, including excess time, excess wages, excess consumption, non-production, non-creation of potential, and risk. Rather than disappearing, these costs are transferred to different stakeholders and absorbed through organizational adaptation, social burden sharing, environmental degradation, and deferred consequences. As a result, the apparent efficiency gains advertised by transformation programs frequently conceal a parallel production of hidden costs.

In this book, SEAM is extended beyond traditional organizational analysis and applied to data center development, AI infrastructure expansion, consulting engagements, and community-level consequences. It therefore becomes possible to analyze not only whether a transformation succeeds financially, but also who ultimately pays its hidden costs and how those costs are distributed across society.

Taken together, these three figures provide the conceptual foundation for *The Consent Machine*. Figure 1 identifies the institutional system through which consent is manufactured. Figure 2 explains the multidimensional mechanisms through which that system operates. Figure 3 reveals the hidden costs that remain obscured from public view while transformation initiatives proceed. The combined framework suggests that contemporary AI transformations cannot be understood solely as technological innovations. They are also systems of power, legitimacy, accountability, and cost transfer. The central proposition of the book is therefore straightforward: consent is manufactured through institutional architectures that convert uncertainty into inevitability while transferring hidden costs away from decision-makers and onto communities, workers, and future generations.

Figure 4. The Consent Pathway

How AI Infrastructure Gains Legitimacy

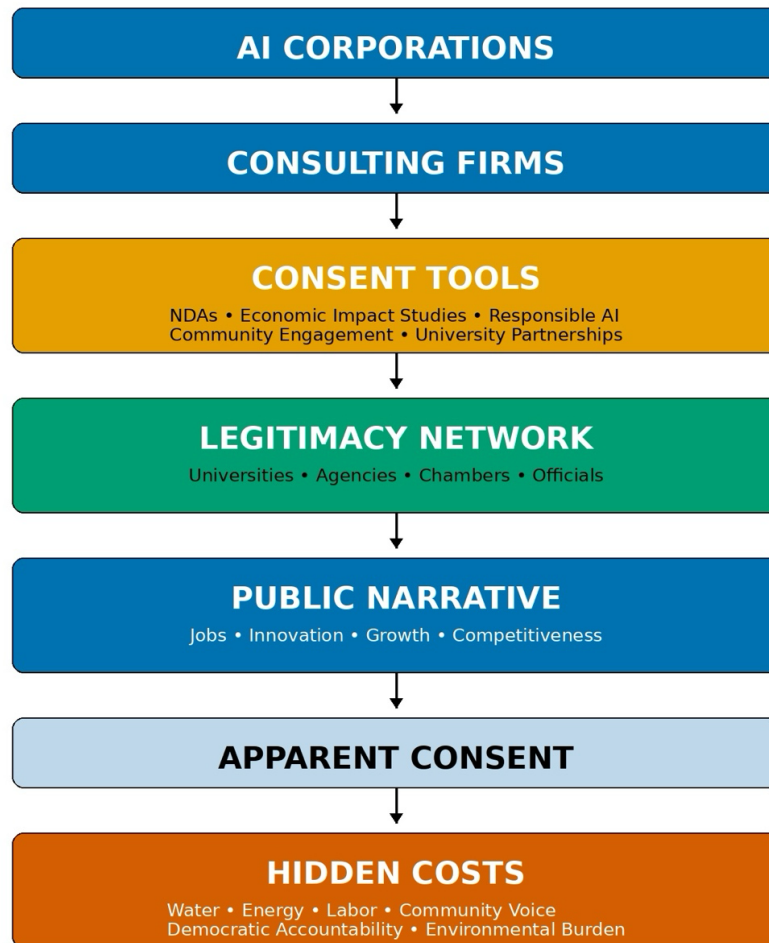


Figure 4: The Consent Pathway

Figure 4. The Consent Pathway: institutional chain from AI Corporation through Consulting Firm, Consent Tools, Legitimacy Network, and Public Narrative to Apparent Consent, with Hidden Costs absorbed at the bottom by communities, workers, and future generations.

Figure 4 traces the operating mechanics of the Consent Machine as a sequential institutional chain. At the top, AI Corporations — OpenAI, Oracle, Amazon, Microsoft, xAI, and their peers — originate the infrastructure plans, the capital commitments, and the ideological frame of national competitiveness and technological inevitability. Consulting Firms receive that frame and translate it into deployable instruments: Non-Disclosure Agreements that silence local employees and university partners, Economic Impact Studies that project job creation without auditing hidden costs, Responsible AI Frameworks that perform ethical commitment without embedding accountability, Community Engagement Plans that

provide process without power, and University Partnerships that convert public research institutions into legitimacy providers.

The Legitimacy Network — Economic Development Agencies, Chambers of Commerce, Industry Associations, and Government Officials — receives the consulting-manufactured instruments and translates them into Public Narrative: Jobs, Innovation, Economic Growth, National Competitiveness, and AI Leadership. These five narrative terms are not descriptions; they are decisions made before the community is consulted. By the time the Public Narrative reaches community members, it has passed through two institutional translation layers, each of which has eliminated one level of accountability. What arrives at the community level is Apparent Consent: a documented process that satisfies procedural requirements without transferring genuine deliberative power.

Hidden Costs — Water, Energy, Labor, Community Voice, Democratic Accountability, and Environmental Burden — appear at the bottom of the chain not because they are discovered there but because they are deposited there. Each layer of the pathway above transfers costs downward while retaining benefits upward. The NDA prevents the community from calculating what it is accepting. The Economic Impact Study presents only visible benefits. The Legitimacy Network endorses the process. The Public Narrative makes opposition appear anti-progress. By the time a community votes, signs, or fails to object, the pathway has already converted uncertainty into inevitability. This is the mechanism the remaining chapters document case by case.

The SEAM Hidden Cost Framework

The Socio-Economic Approach to Management (SEAM), developed by Henri Savall and Véronique Zardet at the ISEOR Institute in Lyon, France, beginning in 1973, provides the book's cost-counting methodology. SEAM identifies six categories of organizational dysfunction that generate hidden costs — costs that conventional accounting does not measure, that do not appear in the consulting firm's engagement report, and that are absorbed by the organization, its workers, and its community rather than by the firm whose engagement produced them.



Figure 5: The SEAM Hidden Cost Iceberg

Figure 5. The SEAM Hidden Cost Iceberg: visible benefits (AI Investment, Jobs Announced, Technology Upgrading, Consulting Revenue, Shareholder Value, Economic Development Branding) above the waterline; SEAM Hidden Costs below (Labor Displacement, Apprenticeship Breakdown, Cognitive Overload, Water Consumption, Grid Expansion, Environmental Burden, Community Health, Journalism Weakening, Loss of Democratic Voice, Public Subsidies, Future Infrastructure Costs), absorbed by Workers, Communities, Ratepayers, Taxpayers, and Future Generations.

The six SEAM dysfunction categories are: (1) Working conditions — the physical, psychological, and social conditions in which work is performed, and the health and productivity costs that degraded conditions generate; (2) Work organization — the coordination, communication, and task-allocation structures that determine whether work flows smoothly or generates friction, redundancy, and waste; (3) Time management — the dysfunctions produced when time is misallocated, deadlines are misset, and the organization's temporal rhythms are disrupted by external transformation pressures; (4) Communication-coordination-cooperation — the dysfunctions produced when the information flows, cross-unit coordination mechanisms, and cooperative norms that organizations depend on are disrupted by reorganization, technology change, or workforce reduction; (5) Integrated training — the dysfunctions produced when skill development is inadequate to the organization's actual needs, producing quality problems, safety incidents, and the hidden costs of rework; (6) Strategic implementation — the dysfunctions produced when strategic decisions are implemented without adequate operational preparation, producing the gap between what transformation promised and what it delivered.

Savall and Zardet's research across more than 2,000 organizations in forty countries established that hidden costs, across these six categories, typically run to approximately 167 percent of the visible labor cost in any dysfunction-producing organizational system. Applied to AI consulting engagements: a \$2 million engagement that produces disruption across all six dysfunction categories costs, in full sociotechnical accounting, closer to \$5.5

million — with the additional \$3.5 million absorbed by the client organization, its workforce, and the communities that bear the externalized costs. That gap is not a rounding error. It is the architecture of the Consent Machine.

In Chapter Two, SEAM Analyst applies this framework to four geographies: Loudoun County Virginia, Shreveport Louisiana, South Memphis Tennessee, and Tucson Arizona. In each geography, the hidden costs include: working conditions (labor conditions at data center construction sites under NDA); time management (communities unable to plan water and power infrastructure when they do not know what is coming); communication-coordination-cooperation (elected officials who cannot inform constituents); and strategic implementation (long-term community costs — grid strain, water use, tax exemptions — that are not disclosed in the stakeholder engagement framework the consulting firm designed).

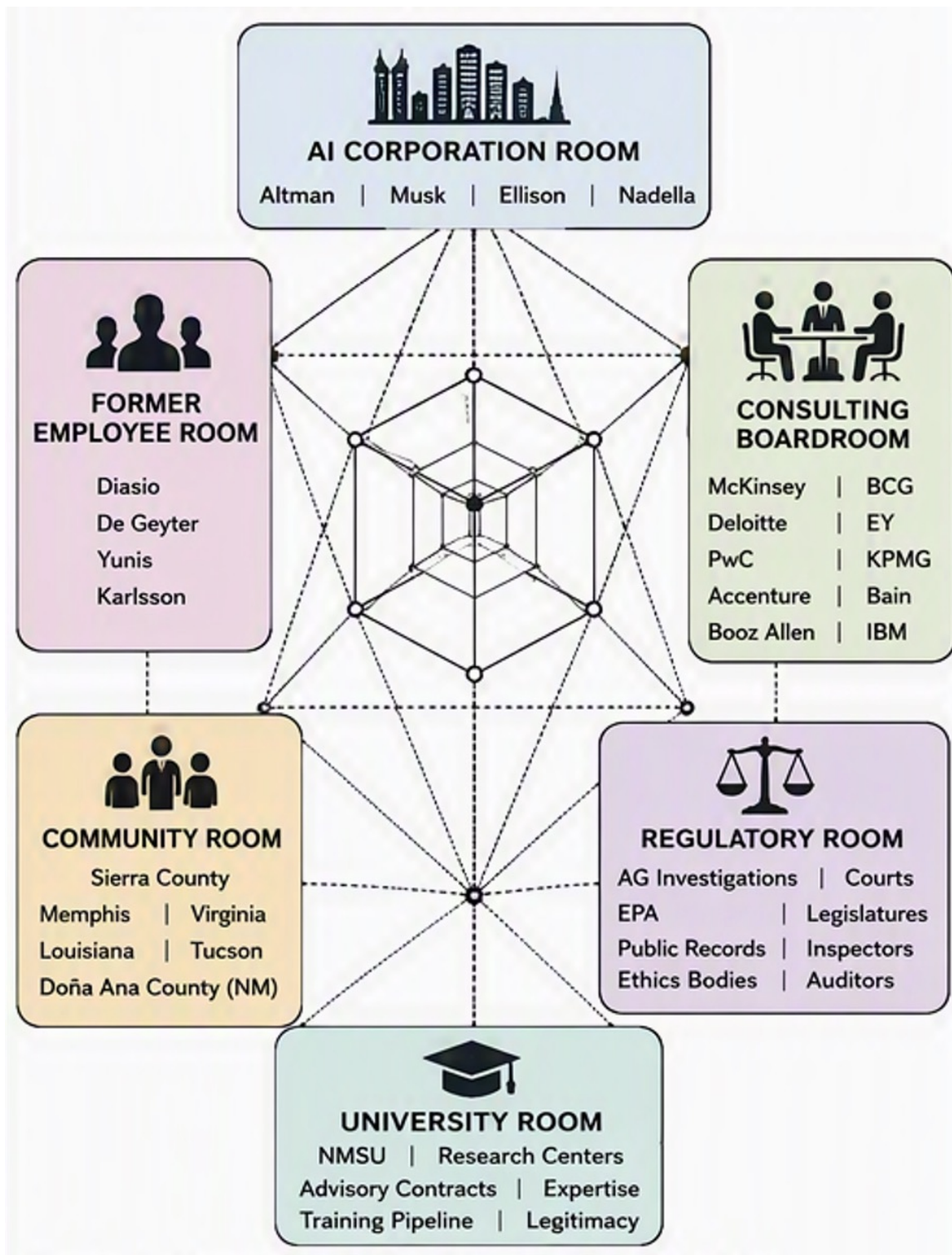


Figure 6: The Tamaraland Map of The Consent Machine

Figure 6. The Tamaraland Map: six simultaneous rooms connected by dashed lines of testimony, document flow, and contested accountability — AI Corporation Room (Altman, Musk, Ellison, Nadella); Consulting Boardroom (McKinsey, BCG, Deloitte, EY, PwC, KPMG,

Accenture, Bain, Booz Allen, IBM); Former Employee Room (Diasio, De Geyter, Yunis, Karlsson); Community Room (Sierra County, Memphis, Virginia, Louisiana, Tucson, Doña Ana County); Regulatory Room (AG Investigations, Courts, EPA, Legislatures, Ethics Bodies, Auditors); University Room (NMSU, Research Centers, Advisory Contracts, Expertise, Training Pipeline, Legitimacy).

Figure 6 maps the book's architecture as the organizational theorist Krizanc's Tamara conceived it: not a sequence but a simultaneity. In the original Tamara, audience members in a mansion choose which room to follow as the characters move between them, and no two audience members see the same play. In this book's Tamaraland, no two readers follow the same machine. The six rooms are connected not by corridors but by dashed lines: the contested flows of testimony, document, contract, and accountability that the Consent Machine manages and, where possible, suppresses. The hexagonal geometry of the map is not decorative; it reflects the equal standing of the six rooms. No room is the center. No room contains the full story. Each room's evidence is partial, and the machine's operation depends on that partiality remaining intact.

The AI Corporation Room — Altman, Musk, Ellison, Nadella — is where capital decisions originate. The Consulting Boardroom is where those decisions are translated into deployable frameworks and the Ghost Vortex deepens: each firm's founding ethical commitments preserved in language while practice migrates toward performance. The Former Employee Room is the machine's most vulnerable point: Diasio's midnight admission about the billable hour, De Geyter's army of consultants, Yunis's ladder within a ladder, and Karlsson's question about what humans are still for. These testimonies exist because the NDA regime is imperfect, and they are the book's most direct evidence of institutional conscience suppression in operation.

The Community Room grounds the entire architecture in place: Sierra County's moratorium, Memphis's NAACP lawsuit, Virginia's FOIA findings, Louisiana's fait accompli announcement, Tucson's Project Blue rejection, and Doña Ana County's ongoing desalination NDA dispute. The Regulatory Room — AG Investigations, Courts, EPA, Legislatures, Ethics Bodies, Auditors — represents the accountability infrastructure that the Consent Machine must manage, route around, or capture to operate without consequence. The University Room is where the machine's institutional legitimacy is most visibly manufactured: NMSU's advisory contract with BorderPlex, the research centers whose partnerships provide credibility, the training pipelines whose expertise validates the transformation narrative. Figure 6 is the map the reader carries through the book. Each chapter enters one or more rooms. No chapter contains all six. The Consent Machine operates in the space between them.

A reader who follows any one room will encounter a coherent story. A reader who follows all four simultaneously will encounter the Consent Machine as it actually operates: not as a linear process with a beginning and an end, but as a system of simultaneous pressures, simultaneous costs, and simultaneous rooms in which different actors believe different

things about what is happening and why. That is what the tesseract holds that the stage model cannot: the four dimensions, all at once, insisting on their simultaneous existence.

What is the Ghost Vortex?

A. The Cave: Plato's Allegory as Method

Plato's cave allegory, from Book VII of the Republic, offers the foundational geometry of what this book calls the Ghost Vortex. Prisoners chained since birth face a cave wall. Behind them, a fire burns. Between the fire and the prisoners, puppeteers carry objects whose shadows are cast on the wall. The prisoners, who have never seen the objects, take the shadows for reality. When one prisoner is freed and dragged toward the light, the experience is painful; the real sun blinds rather than illuminates, at first. The freed prisoner who returns to describe what they have seen is not believed — and in the most direct political version of the story, is killed.

Applied to consulting and AI, the allegory maps precisely. The shadows on the wall are the ethics frameworks, governance certifications, transformation promises, and stakeholder reports — the entire language apparatus that consulting firms produce. The fire that casts them is the business model: billable hours, platform commissions, alliance revenues. The puppeteers are the managing partners and global chairs who design what shadows to cast. The prisoners are the clients, the public institutions, the community members who have no access to the fire — who receive only the language of transformation while the business model continues unchanged. And the freed prisoner — the whistleblower, the former employee who speaks, the community that sues — is discredited, managed, or silenced.

The Ghost Vortex is what happens when this apparatus is institutionalized across time: the language of ethics becomes so elaborated, so professionally credentialed, so seamlessly integrated into the product, that it functions as its own reality. The ghost — the original conscience of the Conscience Firm — is preserved in every logo, every corporate values statement, every DEI report. But it is a ghost: it has no body. It cannot act. It haunts the institution without inhabiting it. The vortex is the motion: the institutional machinery that keeps producing ethical language at increasing velocity precisely as the gap between language and practice grows.

B. The Ghost Vortex in AI Corporations

The Ghost Vortex manifests in AI corporations in a specific form: the foundational safety claim. Each of the major AI corporations was constituted, publicly and legally, around a safety or ethics commitment that its subsequent business decisions have systematically contradicted. The vortex is not that individual leaders are hypocrites — it is that the institutional machinery produces ethical language automatically, regardless of what the business is doing.

OpenAI was incorporated as a 501(c)(3) nonprofit in 2015 specifically to prevent the concentration of AI power in the hands of any single company, including itself. Sam Altman's public rhetoric — describing GPT-4 as potentially "a little bit like a genie" that

must be handled carefully — coexists with the company's simultaneous pursuit of a \$150 billion valuation, a \$4 billion DeployCo subsidiary designed to replace the consulting distribution layer it also partners with, and NDA structures that silence clients who experience harm. The language of careful stewardship is not performative in the sense of being consciously dishonest — it is institutionally automatic. The safety team publishes papers; the commercial team signs the Frontier Alliance; neither reads the other's work closely enough to notice the contradiction.

Google's Ghost Vortex is archival: "Don't Be Evil" was removed from the corporate code of conduct in 2018, but the Alphabet "moonshot" rhetoric of making technology work for humanity continues in every product announcement. Google DeepMind publishes foundational alignment research while Google Cloud commits \$750 million to accelerate AI deployment through consulting partners whose clients have no ability to audit the systems being deployed. Anthropic's version is the most internally coherent — its founding team left OpenAI explicitly over safety concerns, and its Responsible Scaling Policy is among the most detailed in the industry. Yet Anthropic simultaneously launched a Wall Street consulting venture in May 2026, commercializing capabilities its own safety papers describe as requiring caution. The ghost is not a lie. It is a structure that produces safety language and commercial deployment simultaneously, with neither arm able to stop the other.

C. The Ghost Vortex in the Big Ten Consulting Firms

In consulting firms, the Ghost Vortex takes a different form from the AI corporations. The AI corporations are young enough that their founding ethics are still legible in their original documents — you can read the OpenAI charter from 2015 and see the gap with what the company does in 2026. The consulting firms have had over a century to elaborate the ghost, and the elaboration is so dense that the original conscience is now several institutional layers removed from any current practice. What follows is a firm-by-firm account of how the Ghost Vortex manifests in each of the Big Ten — the gap between the founding commitment and the current instance.

Working Glossary

The following definitions are working definitions — they are operational for this book and may differ from usage in adjacent literatures. They should be treated as provisional and subject to revision as the argument develops.

Antenarrative The before-the-story: the fragmented, prospective, living bet that something is happening before it has been configured into a coherent narrative. Antenarrative analysis recovers what official narratives suppress — the community voices, the dissenting memos, the whistleblower testimony that existed before the official story was assembled. Developed by David Boje (2001, 2011). In this book, antenarrative is the method for recovering what consulting and AI firms' official transformation narratives conceal.

The Capture (Era 4) The fourth and current era of consulting firm evolution (2024–present), defined by formal financial alliances between AI corporations and consulting firms. In the Capture era, consulting firms function as certified distributors of AI platforms they did not build, selling deployment services for systems they are structurally prevented from evaluating independently. The structural parallel is the ERP era (SAP/Oracle needed Accenture/IBM to implement), but with the added dynamic that the AI corporations are simultaneously building direct-deployment arms (e.g., OpenAI DeployCo) designed to eliminate the consulting layer.

The Consent Machine The institutional mechanism by which consulting firms manufacture organizational and public consent for AI adoption. The Consent Machine does not operate through deception in the conventional sense — it operates through the professional authority of the Big Ten firms, whose certifications, reports, and transformation frameworks are taken as independent judgment by clients and regulators who do not know those firms are financially allied with the AI platforms being evaluated.

Effort Theater The performance of consulting labor — slide decks, workshops, stakeholder interviews, framework presentations — that substitutes for measurable outcomes. In the Automation Firm era (Era 3) and Capture era (Era 4), effort theater continues at human billing rates while the substantive analytical work is increasingly performed by AI. The theater is not fraudulent in law; it is structurally built into the billable-hour model.

The Gap. Official Story versus suppressed storying; visible fee versus hidden cost; special versus moral answerability. This book's objective is transparency of this Gap.

Ghost Vortex The institutional mechanism that preserves the language of ethics, safety, and accountability while the practice governed by that language changes. The Ghost Vortex is not hypocrisy — it is structural: the ethics apparatus (certifications, frameworks, values statements, DEI reports) is produced by institutional machinery that runs independently of the business decisions it purports to govern. The ghost is the original conscience of the founding firm, preserved in language and logo; the vortex is the institutional motion that keeps producing ethical language at increasing velocity as the gap between language and practice grows.

Hidden Cost In SEAM methodology: the costs of organizational dysfunction and consulting advice that do not appear in the consulting fee but land on third parties — workers, communities, public institutions — as workforce reductions, community displacement, failed implementations, and degraded public services. Henri Savall and Véronique Zardet (ISEOR) calculate hidden costs at approximately 167% of the visible consulting fee. In the Capture era, AI deployment hidden costs are expected to exceed this ratio significantly.

Platform Capture The condition in which a consulting firm's financial alliance with a specific AI platform prevents it from offering the independent evaluation that is the commercial and ethical justification for its engagement. A firm with a \$2 billion Microsoft

deal cannot recommend Google Cloud to a client for whom Google Cloud is the better option. Platform Capture is the mechanism by which the Capture era erodes consulting firm independence.

Plato's Cave (as used in this book) A structural model, not a moral metaphor. The cave maps the institutional relationship between the fire (the business model), the puppeteers (managing partners and global chairs), the objects (consulting products and AI platforms), the shadows (ethics frameworks, transformation promises, governance certifications), and the prisoners (clients, public institutions, communities). The freed prisoner is the whistleblower or critical researcher who has seen the fire and cannot persuade those still watching the shadows.

Polyphonic Analysis A method of organizational analysis that holds multiple voices in simultaneous attention without resolving them into a single authoritative narrative. In this book, six polyphonic voices are maintained throughout: the PR Voice (what the firm says), the Critical Theorist (structural analysis), the SEAM Analyst (hidden cost accounting), the Community Voice (those who bear the costs), the Ethnographic Witness (what people said when not performing for the record), and the Founder's Voice (what the original commitment was). Developed from Boje's application of Mikhail Bakhtin's dialogism to organizational storytelling.

SEAM (Socio-Economic Approach to Management) A methodology developed by Henri Savall and Véronique Zardet at ISEOR (Lyon, France) for measuring the hidden costs of organizational dysfunction. SEAM identifies six categories of hidden cost: absenteeism, workplace accidents and occupational disease, staff turnover, non-quality of products and services, direct productivity gaps, and strategic underperformance. In this book, SEAM is applied to consulting firm engagements to calculate what clients actually pay (fee + hidden cost multiplier) and what communities bear (costs landing after the firm exits).

Tesseract Leadership David Boje and colleague's concept of multi-dimensional leadership that operates across time simultaneously — attending to present decisions, historical origins, and future consequences without collapsing into a single temporal frame. Tesseract leadership is the explicit opposite of the Ghost Vortex: where the Ghost Vortex dissociates ethics language from practice across time, Tesseract leadership maintains the connection. The term draws on the mathematical tesseract (four-dimensional hypercube) as a model for holding multiple temporal frames in simultaneous view.

Transformation Theater The performance of organizational change — workshops, change management frameworks, transformation roadmaps, cultural alignment sessions — that substitutes for measurable transformation. In Stages 2 and 3 of the Consent Machine model, consulting firms are paid to manage the narrative of transformation while the underlying systems and power relationships remain unchanged. Transformation theater is the organizational equivalent of Plato's shadows.

Wreckage Management The third stage of the Consent Machine model: the engagement that follows a failed transformation, in which the same or a different consulting firm is engaged to manage the aftermath of the transformation the previous engagement was paid to produce. Wreckage management is structurally profitable: the failures of the Consent Machine generate demand for the next cycle of the Consent Machine.

Conclusion

The Preface told you where the author stands and why a ladder cannot describe the machine. Chapters 1–9 will name buyers, land, stacks, hours, dollars, statutes, witnesses, and refusal. Between those rooms there has to be a single object they are all measuring.

That object is the Gap: the remainder Story is designed not to count. Ghost Vortex is how the remainder is kept in motion and out of the invoice.

Turn the page to Chapter 1 for the protagonists. Do not look here for their revenues, their product names, or their scores. Look here when a later chapter says the invoice was not the cost, and you need the name of what it left out.

Summary

The story is not “consultants vs. AI labs.” It is a three-sided market: labs sell intelligence, clouds sell compute and distribution, consultants sell the last mile—workflow redesign, data, risk, change, and political cover inside the client. The labs discovered they cannot staff that last mile. The consultants discovered they cannot own the models. So they rent each other.

Almost no contract values are public. Where dollars appear below they are announced *investments* or *seat counts*, not invoice line items.

These firms configure the workforce to meet the AI, often providing the “institutional cover” that allows for radical staff reductions.

Table 1: The Billion Dollar AI Consulting Companies

Table 2: The Top 10 AI Corporations

TIME named this collective group — Huang, Musk, Altman, Zuckerberg, Su, Amodei, Hassabis, and Fei-Fei Li — its 2025 “Person of the Year” as the “Architects of AI” (Reuters, TechCrunch). By mid-2026, Anthropic reportedly hit a \$65B annualized revenue run rate, with OpenAI’s valuation (\$852B) trailing Anthropic’s fresh valuation per recent reports.

Leaders' Statements on Superintelligence

Sam Altman (OpenAI) — The most aggressive optimist. He has said we’re already slipping past human-level AGI toward superintelligence, describing it as “a genie that can grant any wish.” Per a recent TIME profile, OpenAI expects an internal system matching or exceeding humans at most economically valuable tasks **before the end of 2026**, with Chen claiming they’re “80% of the way” there (Forbes, Aug 2026).

Demis Hassabis (Google DeepMind) — More measured but increasingly urgent. At Davos (Jan 2026) he put human-level AGI at "five to 10 years" away; by Google I/O 2026 he tightened that to ~2030, possibly 2029, saying humanity stands in "the foothills of the singularity." He frames the current agent era as a "practice run" and societal stress test for far more powerful systems to come (Fortune, Axios).

Dario Amodei (Anthropic) — Prefers the term "powerful AI" over AGI, which he considers overhyped, but has warned that systems like Anthropic's "Mythos" have "the power to catch businesses and governments off guard." He and Hassabis jointly called for a U.S.-led international AI coalition at the G7 summit in June 2026 to manage frontier-AI risk (CNBC, GIGAZINE).

Mark Zuckerberg (Meta) — Champions "personal superintelligence": AI that empowers individuals to create, learn, and build businesses rather than a centralized "trust the lab" model. His 6,500-word manifesto "The Future is for Everyone" (Aug 2026) argues superintelligence should be delivered to everyone so competing interests check and balance each other, and warns regulators shouldn't constrain the technology (Forbes, Washington Post).

Satya Nadella (Microsoft) — The skeptic-in-chief relative to peers. He describes AI as "a tool" and says his benchmark isn't AGI but achieving 10% global GDP growth — while Microsoft itself courts "superintelligence" branding while carefully avoiding the term AGI (The Atlantic).

Jensen Huang (NVIDIA) — Called by TIME "probably the single highest-leverage individual in AI." His stance focuses on the infrastructure enabling whichever company gets there first; he was part of the 2025 Person of the Year cohort and continues framing AI as an industrial revolution being built on NVIDIA silicon.

Elon Musk (xAI) — Court testimony in the Altman/OpenAI case (May 2026) revisited the founders' original fears about who would control AGI. He remains vocal that uncontrolled AI is among civilization's greatest risks while racing to build frontier models himself (Reuters).

A notable tension worth flagging: as Fortune reported from Davos, the LLM builders themselves disagree sharply — Hassabis insists today's systems are "nowhere near" AGI, while Altman claims we're already crossing that threshold. And per the broader commentary, none of them has given a serious answer to what happens if China gets there first.

Table 3: How the 10 firms sit with each AI company **OpenAI / ChatGPT**

This is the deepest lab relationship set.

How it works commercially: consultants buy enterprise seats (historically on the order of tens of dollars per user per month at list; large deals are discounted and unpublished), resell or bundle them, then charge **professional services** that dwarf the software. OpenAI gets distribution into regulated industries it cannot staff.

Anthropic / Claude

Anthropic's enterprise push is explicitly *through* consultants.

Deloitte: Largest Claude deployment Anthropic has announced—~**470,000 people**. Claude Center of Excellence; plan to certify **15,000** practitioners; industry solutions for FS, health, public sector; 2026 secure-software remediation platform on Claude. Alliance from 2024, expanded Oct 2025. Terms undisclosed. Timing irony: same week as the Australian AI-hallucination refund.

Accenture: Dec 2025 three-year partnership; Accenture becomes one of Anthropic's **three largest enterprise customers**; ~**30,000** staff trained on Claude. Anthropic said Accenture supplies “scale and heft” while Anthropic stays thin in-region. Terms undisclosed.

Others: Claude is a first-class model in PwC Agent OS, AWS Bedrock, and Microsoft Copilot. KPMG/EY/IBM use it as one option. Anthropic's 2026 Claude Partner Network (~\$100M partner fund) listed Accenture and Deloitte as anchors.

Claude's pitch to consultants is **safety + regulated industries**—the same clients who already pay Deloitte/EY for audit, tax, and risk.

Microsoft / Copilot

Microsoft is not “a model lab.” It is the **workplace OS** that already sits on the client desktop.

KPMG: Deepest stack—Workbench and Clara on Azure AI Foundry, Copilot Studio, Semantic Kernel → Microsoft Agent Framework, **Agent 365** for agent fleet control, M365 Copilot for every professional. This is a classic SI + product alliance.

EY: Orchestration layer of EY.ai is Microsoft (Foundry catalog, Copilot front door, Fabric, Copilot Studio); intelligence layer is NVIDIA.

McKinsey: Named Microsoft alliance—Copilot Studio agents, Azure migrations, contract-compliance copilots.

Bain: Azure OpenAI as the *trusted* way to put GPT on client data (announced 2023 alongside the OpenAI alliance).

Everyone else: Copilot licenses are often already in the client's Microsoft EA. Consultants implement, govern, and customize; they rarely “own” Copilot.

Money: Copilot is sold by Microsoft (list has been in the \$30/user/month class for M365 Copilot; enterprises negotiate). Consultants bill implementation, change, and custom agents on top.

Amazon / AWS

AWS is the **other** default runtime (Bedrock models: Claude, Llama, Mistral, Amazon Nova, etc.).

Deloitte: Multi-Agent System and 10X Analyst on **Bedrock AgentCore**; marketplace listings.

Accenture: Bedrock Agent inside Trusted Agent Huddle; large AWS SI practice.

PwC: Agent OS AWS-native edition (Bedrock + MCP servers).

IBM: Enterprise Advantage FedRAMP on AWS.

Booz Allen: GenAI data-processing on Bedrock; federal cloud work.

The rest: Standard AWS Premier / consulting partner motions.

Money: clients pay AWS for tokens and AgentCore; consultants take services margin. Amazon also invests tens of billions in **Anthropic** (and more recently committed large capital toward OpenAI compute). Consultants ride that compute, they do not fund it.

Google

Google sells **Gemini + Vertex + A2A** (the agent-to-agent protocol many of these firms adopted).

Accenture: Accenture Edge + Gemini for mid-market; Huddle lists Google Cloud; A2A partner.

Deloitte: 100+ ready Gemini agents; A2A work with ServiceNow for field service.

IBM: dedicated Google Cloud practice; industry agents on Gemini Enterprise + Consulting Advantage.

McKinsey, BCG, PwC, EY, KPMG: multi-cloud option, not usually the primary branded alliance.

Money: same pattern—cloud consumption + services. Google also invests in Anthropic (competitive with its own Gemini), which is why consultants treat Google as *both* a model vendor and a cloud.

Oracle

Oracle is mostly **the system of record**, not the brain.

Consultants (especially Accenture, Deloitte, IBM, PwC) implement Oracle Cloud ERP/HCM and now plug **Oracle-side agents** into fabrics like Accenture Huddle and PwC Agent OS. Clients already pay Oracle subscription + SI. AI is an add-on inside that install, not a separate “Oracle alliance” on the scale of OpenAI/Anthropic.

Meta

Llama shows up as a **model choice** on Bedrock/Azure and as a **Huddle partner** at Accenture. There is no “Meta + McKinsey Frontier” style motion. Meta’s enterprise channel is still thin compared with OpenAI/Anthropic/Microsoft.

xAI / Grok

This is the empty cell. There is no announced Bain/McKinsey/Accenture-style alliance with xAI among these ten.

xAI’s enterprise path so far: GSA seats for U.S. federal agencies at **\$0.42** for 18 months (a land-grab price vs OpenAI/Anthropic’s ~\$1 government offers); Wall Street pilots (Morgan Stanley, Apollo) with low actual usage; Saudi HUMAIN national deployment; Grok Bot for Enterprise (2026). That is direct + sovereign + federal, not Big Consulting. Brand, safety incidents, and thin implementation bench explain why the 10 have not bet their partner P&L on Grok yet.

What they actually pay (honest version)

Public numbers that *do* exist:

PwC US: \$1 billion over three years on GenAI (people, product, ecosystem)—not a check to OpenAI.

Seat footprints: PwC ~100k ChatGPT Enterprise; Deloitte ~470k Claude; Accenture ~30k Claude-trained; Bain ~13k ChatGPT Enterprise; EY historically ~96% of staff on EY.ai after a ~\$1.4B platform spend.

Labs’ partner funds: Anthropic Claude Partner Network launched with a **\$100M** 2026 commitment for training/certification/joint GTM—not fees *from* consultants.

List software: ChatGPT Enterprise and Claude Enterprise are negotiated; consumer/team list prices are a poor guide. Microsoft Copilot sits inside existing M365 agreements.

What they **do not** publish: per-token pass-through, reseller margin, or how much of a \$20M “AI transformation” is model API vs partner profit. In a typical large program the split looks like: **cloud/tokens 10–25%, software/IP 10–20%, people 50–70%**. Consultants live in the last bucket.

The *labs and clouds* pay each other far more than they pay consultants: Microsoft–OpenAI (historically ~\$13B+), Amazon–Anthropic (multi-billion, plus huge cloud commit), Google–

Anthropic, Amazon–OpenAI compute expansions, Nvidia circular GPU investments. Consultants are a **go-to-market layer** on that pile, not the financiers.

How the relationship changed over time

2016–2022 — Cloud SI era. Accenture, Deloitte, IBM, and the Big 4 were Microsoft/AWS/Google/Oracle implementers. AI meant Watson, SageMaker, or a data-science project. Strategy firms (MBB) wrote “AI strategy” decks and left the plumbing.

Nov 2022–2023 — Panic and first alliances. ChatGPT made every CEO ask for a demo. Bain locked OpenAI first among strategy firms (Feb 2023). Everyone stood up internal GPTs (Lilli, Sage, Deckster). Narrative in the press: “AI will kill consulting.” Firms answered by becoming **Client Zero**—use the tools internally, then sell the scar tissue.

2024 — Resale and seat wars. PwC becomes OpenAI’s first ChatGPT Enterprise reseller and largest customer. Bain expands the OpenAI pact. Consultants realize seats are a **trojan horse** for services. Anthropic starts courting the same channel because Microsoft already owns OpenAI distribution.

2025 — Multi-home, don’t marry one model. Accenture Huddle, PwC Agent OS, IBM Orchestrate, EY “OS” all become **model-agnostic buses** (MCP, A2A, Bedrock, Foundry). Deloitte and Accenture take huge Claude footprints *while* staying in OpenAI/Google/NVIDIA deals. Labs learn they cannot hire enough forward-deployed engineers.

2026 — Labs hire the consultants back. OpenAI Frontier Alliances with McKinsey, BCG, Accenture, Capgemini. Anthropic Partner Network with Deloitte/Accenture/Cognizant/Infosys. WSJ headline version: “AI needs management consultants after all.” Reason: PwC’s own CEO survey found more than half of CEOs saw **no significant financial benefit** from AI yet. The bottleneck is operating model, not tokens.

The power shift: early on, consultants feared disintermediation. Now labs fear **pilot purgatory** without them. Neither side is exclusive. That is the tell.

Why customers don’t just skip the consultants

They *do*—for a slice of work. ChatGPT Enterprise, Claude, Copilot, and Bedrock are sold direct to most of the Fortune 500. OpenAI has said the large majority of the Fortune 500 already had ChatGPT Enterprise in some form. Skipping is easy when the job is “give analysts a chatbot.”

They **don’t** skip when the job is any of the following:

The model is ~10% of the problem. Value is in messy ERP, identity, data contracts, audit trails, and who is allowed to press “post invoice.” Labs do not staff 18-month SAP + risk + works-council programs.

Procurement and liability. A bank or ministry wants a counterparty that carries PI, SOC, and professional indemnity—and someone to sue besides a research lab. Deloitte’s

Trustworthy AI + Claude, EY control rings, PwC Assurance for AI exist because the *buyer's* lawyers demand them.

Politics inside the client. “We hired McKinsey” is a decision the board understands. “We hired OpenAI forward-deployed engineers” does not relocate 40 processes or fire a shared-services center.

Multi-vendor reality. Real estates run SAP + Salesforce + ServiceNow + Oracle + two clouds. Accenture Huddle and PwC Agent OS exist because no single lab will integrate all of that as a product.

The labs themselves send you back. OpenAI’s own 2026 move was to *attach* McKinsey/BCG/Accenture to Frontier. Anthropic’s commercial chief said Accenture is the scale layer. That is an admission that a 70-person FDE team cannot cover the Global 2000.

Regulated work is the consultants’ home market. Audit, tax, government, defense (Booz Allen), Medicaid systems, AML—the model vendor is a component.

When customers *should* skip: a single team, a clean SaaS stack, a well-defined coding or research workflow, and an internal platform group. Many tech-native companies already did. The consultants’ revenue is concentrated where that description is false.

What’s the story, stripped of press releases

Act I: Labs built astonishing models and assumed enterprises would “just adopt.”

Act II: Enterprises bought seats, ran pilots, and reported little P&L impact. Hallucinated citations in paid government reports (Deloitte and others) showed that “we used the model” is not a control environment.

Act III: Labs re-imported the old software-industry channel—SIs and strategy firms—exactly as SAP, Salesforce, and Microsoft did. Consultants rebranded as “reinvention / agent OS / Client Zero” so they are not selling slideware on top of a product the client already owns.

Act IV (now): A multi-homed oligopoly. Every big consultant is in bed with two labs and two clouds. Every lab is in bed with four consultants. Clouds invest in the labs. Nobody is exclusive because exclusivity would strand the client—and the next model release.

Customers pay three rents: **tokens to the lab, infrastructure to the cloud, transformation to the consultant.** They skip the third only when they already have the people who do that work. Most large, regulated, legacy-system companies do not. That is the whole story.

Chapter 2 NDA Geography

Summary Chapter 2

This chapter argues that Stage One of the Consent Machine — Consent Manufacturing — in the AI infrastructure era has a physical geography. The primary instrument is not an audit

report, a certification document, or a Responsible AI framework: it is a Non-Disclosure Agreement signed between a trillion-dollar technology corporation and the elected officials of the community that will bear that corporation's environmental, electrical, and water costs for the next thirty years. This chapter traces that instrument across four American geographies and one connective case: Virginia (the structural baseline, where NDA practice was normalized before any community knew to resist it); Shreveport, Louisiana (where Amazon's \$12-billion-to-\$18-billion data center announcement was made as a *fait accompli*, with local officials legally prevented from informing constituents); South Memphis, Tennessee (where xAI's Colossus supercomputer arrived behind NDAs in a majority-Black neighborhood already carrying disproportionate industrial burden, producing what the NAACP called a death sentence for the community); Tucson, Arizona (where Project Blue was rejected — the counter-case proving the machine can fail, and revealing, through contrast, exactly how it usually succeeds); and New Mexico's Doña Ana and Sierra Counties (the author's home geography, where two communities in the same state arrived at opposite outcomes through the same institutional architecture). In each location the NDA performs the same function the Nike supply-chain NDA performed in the 1990s: it manufactures the appearance of democratic consent while suppressing the public's capacity to evaluate or contest the decision already made. The Consent Machine has moved from auditing labor to auditing land. The three stages are identical. The instrument has been updated. The analysis develops through the chapter's established Six Polyphonic Voices and applies the Four-Dimensional Framework introduced in Chapter Zero, with SEAM hidden-cost calculations for each geography and antenarrative readings of what official narratives suppress.

"A muzzle on elected officials": NDAs "cloak" Louisiana's biggest business developments. — Louisiana Illuminator / Type Investigations, March 27, 2026

"This is a death sentence for our community." — South Memphis resident, NAACP v. xAI filing, 2025

"You can't drink data." — Sign held by Sierra County residents, New Mexico Tech town hall, 2025

PART I: THE NDA AS THE CONSENT MACHINE'S PRIMARY INSTRUMENT

I.1 — What the NDA Does That the Audit Report Used to Do

In the 1990s and early 2000s, as Chapter 1 documented, the primary instrument of Stage One consent manufacturing in global supply chains was the audit report. Nike, Gap, and their counterparts did not simply ignore labor conditions in overseas factories. They commissioned audits. They published the results. They certified compliance. The audit report manufactured the appearance of informed, ethical oversight — and simultaneously suppressed the evidence that would have disturbed that appearance, because the audit was designed, funded, and controlled by the corporation whose practices it was certifying. The result was what Michael Power (1997) called the audit society: a proliferation of formal

accountability mechanisms whose effect was to insulate the audited from accountability rather than to produce it.

In the AI infrastructure era, the audit report has been replaced — at the ground level of community consent — by the Non-Disclosure Agreement. The structural logic is identical. A public interest decision — how community resources (water, power, land, tax revenue, air quality, grid stability) will be allocated for the next three decades — is converted into a private commercial negotiation. The elected official, who derives authority from constituents, becomes an agent of the corporation rather than of those constituents. The NDA is the contractual mechanism by which this conversion is accomplished. It does not merely keep secrets; it keeps the secret that there is a secret to keep. Elected officials subject to NDAs cannot tell their constituents that they have signed an NDA, cannot disclose its terms, and in many cases cannot acknowledge that a negotiation is occurring. The community's capacity for democratic deliberation is not diminished — it is structurally eliminated, before any vote, any public hearing, or any community benefit analysis can take place.

The Louisiana Illuminator, reporting in March 2026 in partnership with Type Investigations, named this dynamic precisely: NDAs as 'a muzzle on elected officials.' That framing is legally accurate. A muzzle does not merely silence; it signals that the silenced party has something to say. The community, observing a muzzled official, does not know what has been suppressed. In the AI infrastructure buildout of 2023–2026, what has been suppressed — in Virginia, Louisiana, Tennessee, and across the American geography of hyperscale data center expansion — is the community's right to know what is being built in its name, what it will cost, and who will bear those costs.

This chapter examines four American geographies where the NDA instrument operated at scale, plus the New Mexico case that runs throughout this book as the author's ground-level witness. The four geographies were chosen not for their typicality but for their structural distinctiveness: Virginia is the normalization case, where NDA practice became standard before it became contested; Louisiana is the *fait accompli* case, where the announcement preceded any opportunity for public deliberation; Memphis is the sacrifice-zone case, where the NDA enabled the imposition of environmental and health costs on a majority-Black community that had no formal voice in the decision; and Tucson is the counter-case, where the machine failed, and where the conditions of its failure illuminate — by contrast — the conditions of its usual success.

1.2 — The Four-Dimensional Frame Applied to NDA Geography

Chapter Zero introduced the Four-Dimensional Framework as this book's analytical architecture: Accountability Architecture (personal/named versus diffused/structural), Ethics Architecture (embedded/character versus performed/theater), Business Model Logic (judgment/independence versus platform/distribution), and Temporal Horizon (generational versus exit-before-consequence). These dimensions do not progress sequentially. They coexist, weighted differently in each institution and each geography.

Applying them to the NDA cases that follow reveals a consistent pattern: the dimension that enables Stage One most powerfully is the Temporal Horizon dimension — the systematic design of agreements whose environmental, economic, and social costs will mature years after the signing parties have exited the negotiation.

A Non-Disclosure Agreement signed between a technology corporation and a county commissioner in 2024 creates obligations and costs that will persist through 2040, 2050, and beyond. The water the data center draws will come from aquifers that recharge over geological time. The electricity grid strain will compound as adjacent data centers multiply. The tax exemptions — standard in nearly all technology-sector economic development agreements — will redirect public revenue from schools, roads, and emergency services for fifteen to twenty years. None of these costs appear in the NDA. The NDA covers only the negotiation. But the negotiation under the NDA determines every subsequent allocation. The Temporal Horizon dimension — the exit-before-consequence logic — is built into the instrument's design.

The Accountability Architecture dimension is equally activated. In a named-accountability system, the elected official who signed an NDA would be accountable — personally, by name, in public — for the consequences of the negotiation they conducted in secret. In a diffused-accountability system, the NDA itself absorbs the accountability question: the official was legally constrained from disclosure, the corporation was legally entitled to confidentiality, and no individual actor bears named responsibility for the community costs that result. The NDA is not merely a mechanism for secrecy; it is a mechanism for distributing accountability so diffusely that it attaches nowhere. This is the Accountability Architecture of Stage One, applied at the physical infrastructure level.

PART II: CASE 1 — VIRGINIA, THE STRUCTURAL BASELINE

II.1 — Loudoun County: Where NDA Practice Was Normalized

Loudoun County, Virginia carries a title that would have been inconceivable a generation ago: the Data Center Capital of the World. It is not a metaphor. By 2024, Loudoun County and its surrounding Virginia jurisdictions housed the densest concentration of hyperscale data center capacity on the planet, accounting for a substantial fraction of the world's cloud computing infrastructure. Amazon, Microsoft, Google, and dozens of smaller operators have built, expanded, and stacked facilities in a geography that was, within living memory, horse country and suburban commuter territory. The transformation did not happen by accident. It happened because Virginia offered a combination of relatively cheap electricity, tax incentives, fiber connectivity, and — critically — a political culture that treated economic development agreements as proprietary commercial information rather than as public records subject to democratic scrutiny.

It was in this context that Non-Disclosure Agreements became normalized as a feature of data center siting — not as an exceptional instrument invoked in extraordinary circumstances, but as standard economic development procedure. When a professor and students at the University of Mary Washington filed Freedom of Information Act requests in

2024 to examine the agreements between data center companies and Virginia localities, what they found was not concealment of an unusual practice. What they found was evidence of routine practice, reported as such: localities confirmed that they signed NDAs with data center companies 'often.' The word is from the localities' own responses to the FOIA requests. Not occasionally. Not in exceptional circumstances. Often.

Virginia Mercury reporter Charlie Paullin documented this pattern extensively through the FOIA Friday series in October 2024, and returned to the subject in a major analysis published April 30, 2025 under the headline 'Data centers, non-disclosure agreements and democracy.' That headline names the structural conflict directly. The NDA does not sit in tension with democracy in some philosophical sense; it sits in tension with the most basic mechanics of democratic self-governance: the elected official's obligation to inform and represent the constituents who elected them. In Virginia, that conflict had been present for years before anyone named it as such. The normalization of the practice preceded the normalization of its critique.

In November 2025, Public Citizen published an analysis titled 'The Secret Data Center Buildout: How States Can Stop Big Tech's Abuse of NDAs,' proposing legislative remedies at the state level. The remedies' necessity is itself evidence of how deeply entrenched the practice had become. You do not propose legislation to prohibit a rare or exceptional practice. You propose it to address a systematic one. By November 2025, the systematic character of NDA use in data center siting had been established beyond reasonable dispute. Virginia was the place where that system was born and where it had matured most completely.

[PR VOICE] We are proud to partner with the Commonwealth of Virginia in creating thousands of good-paying jobs and investing in the digital infrastructure that powers the modern economy. Our community partnerships reflect our long-term commitment to the regions where we operate. We work closely with local officials and stakeholders to ensure our facilities are designed and operated in ways that benefit the communities they serve. Virginia's welcoming business environment, skilled workforce, and excellent infrastructure make it an ideal location for the data centers that power AI, the cloud, and the digital economy that millions of Virginians rely on every day.

[CRITICAL THEORIST] The NDA is not merely a legal instrument: it is an ideological one. It converts a public interest decision — how community resources (water, power, land, tax revenue) will be allocated for the next three decades — into a private commercial negotiation. The elected official becomes an agent of the corporation rather than of the constituents. Antonio Gramsci called this the colonization of civil society by economic society: the gradual substitution of market logic for democratic deliberation in the governance of public resources. The NDA makes this colonization contractual. It is not informal influence or cultural hegemony — it is a signed instrument, enforceable in court, that converts a public official's democratic obligation into a private commercial confidentiality duty. Virginia became the place where this conversion was routinized, where it became, in the localities' own words, something done 'often.' The antenarrative of

Virginia's data center geography is a story about how a democratic deficit is institutionalized before anyone recognizes it as one.

[ETHNOGRAPHIC WITNESS] The University of Mary Washington FOIA research produced documentary evidence that is the foundation of this case analysis: NDAs are not occasional in Virginia data center siting — they are routine. The localities' own responses to FOIA requests confirmed this. The Virginia Mercury FOIA Friday series (October 2024) established a verifiable pattern across multiple Virginia jurisdictions. This is not a story about one bad actor or one unusual arrangement; it is a story about a normalized procedure. The Public Citizen analysis of November 2025 extended that documentation to the national level, proposing legislative remedies that would, if enacted, constitute a direct acknowledgment by state legislatures that the practice had become systematically inconsistent with democratic governance. As of the time of this writing, no such legislation had passed in Virginia.

[SEAM ANALYST] The SEAM hidden-cost framework identifies six dysfunction categories that generate costs invisible in standard economic assessments. In Virginia's NDA baseline, all six are activated. Working conditions dysfunction: the labor conditions of data center construction workers — many on temporary contracts, without union representation — are not subject to public scrutiny because the entire project is under NDA. Time management dysfunction: communities cannot plan water infrastructure, power grid investment, or transportation networks when they do not know what facilities are coming, when, or how large. Communication dysfunction: elected officials cannot inform constituents; the normal channel of democratic communication between representative and represented is legally blocked. Integrated training dysfunction: community members cannot develop the capacity to evaluate or negotiate these agreements because the agreements are secret. Strategic implementation dysfunction: long-term community costs — grid strain, water depletion, tax exemptions, displacement of competing land uses — are not publicly disclosed and cannot be factored into community planning. Implementation dysfunction generates a SEAM hidden-cost multiplier estimated at 1.67 times the visible consulting and development fee — costs that land on the community, not on the contractor. In Virginia, those costs have been accumulating for over a decade.

[COMMUNITY VOICE] The Virginia Conservation Network and allied advocacy organizations have documented the grid impact of data center concentration in Northern Virginia: the state's electric utilities have acknowledged that the load growth from data centers is straining a grid that was not designed for this density of continuous, high-draw facilities. Residential and commercial ratepayers are absorbing infrastructure investment costs that the data center tax incentive regime effectively transfers from the corporations to the community. This is not a speculative harm — it is a documented pattern, confirmed by utility filings and state regulatory proceedings. The NDA that covers the front end of the data center siting process ensures that the community does not have the information it would need to contest this cost transfer at the moment when contestation is still possible. By the time the costs are visible, the agreements are signed, the facilities are built, and the

community's options are limited to accepting or litigating — both of which are expensive and uncertain.

[DAVID'S VOICE] I want to be precise about what makes Virginia the baseline case, rather than the worst case. Virginia is not the most egregious example of NDA-enabled consent manufacturing in the AI infrastructure buildout. It is the most normalized. In Virginia, the practice has become so routine that the localities themselves describe it without apparent awareness that it is remarkable. 'Often' is not the language of a community grappling with an unusual arrangement. It is the language of institutional normalization — of a practice that has become so embedded in economic development procedure that it no longer triggers the reflexive democratic scrutiny that a genuinely unusual arrangement would provoke. The normalization of Virginia is the precondition for the escalation of Louisiana, the sacrifice zone of Memphis, and the scale of my own community's encounter with the same instrument. Virginia did not invent the NDA. But Virginia made it ordinary.

II.2 — Antenarrative: What Virginia's Official Story Suppresses

The official narrative of Virginia's data center economy is a story of prosperity, innovation, and economic competitiveness: a state that attracted the world's most consequential technology infrastructure through smart policy, infrastructure investment, and a welcoming business environment. That narrative is not false in the elements it includes. It is false in what it excludes. The antenarrative of Virginia's data center geography — the before-the-story that the official account cannot accommodate — concerns the distribution of the economy's benefits and costs. Who bears the grid strain? Who pays the rate increases? Who lives near the diesel generators that run when the power supply is strained? Who negotiated the tax exemptions that reduced the public revenue available for schools, roads, and emergency services in the jurisdictions where the facilities operate? The NDA instrument ensures that none of these questions are asked at the moment when the decisions that answer them are being made. That is the antenarrative function of the NDA: it suppresses the questions that would have to be answered if democratic deliberation were actually occurring.

PART III: CASE 2 — SHREVEPORT, LOUISIANA, THE FAIT ACCOMPLI

III.1 — The Announcement as Theater: \$12 Billion to \$18 Billion

On February 23, 2026, Amazon issued a press release announcing plans to invest \$12 billion in AI data center campuses across Caddo and Bossier Parishes in the Shreveport-Bossier City metropolitan area of northwestern Louisiana. The Governor of Louisiana, Jeff Landry, announced it simultaneously and described it as the largest investment in the history of Northwest Louisiana. Both announcements were made in the past tense, describing a decision already reached. There was no public deliberation before the announcement. There was no community benefit analysis released to constituents. There was no environmental impact statement preceding the deal. Local officials in Caddo and Bossier Parishes had spent months concealing the details of the negotiation behind Non-Disclosure Agreements — agreements that prevented them from informing the constituents who

elected them that a \$12 billion industrial transformation of their community was being arranged.

Amazon's announcement described the investment as 'next-generation data center campuses to support AI and cloud computing' (David Zapolsky, in Amazon press release, February 23, 2026). The Louisiana Economic Development agency confirmed it as the 'largest investment in the history of Northwest Louisiana.' Neither announcement included a public accounting of the water demand the campuses would place on Caddo Parish's water supply, the electricity load they would place on a regional grid already stressed by industrial demand, the air quality impacts of backup generation systems, or the community displacement that would accompany construction and operation of facilities of this scale. Those costs were not part of the announcement. They were not part of the negotiation that preceded the announcement, or if they were, the NDA prevented that information from becoming public.

On August 18, 2026, Amazon announced a \$6 billion expansion of the Shreveport investment — a third campus in west Shreveport — bringing the total announced commitment to \$18 billion. That announcement was made in the same way as the first: as a fait accompli, a completed decision announced to the public rather than negotiated with it. The Louisiana Illuminator and NOLA.com reported both announcements. The NOLA.com report on the August 2026 expansion noted that the additional campus would 'support AI and cloud computing workloads.' What it would support in terms of electricity draw, water use, and infrastructure strain from the community's perspective was not disclosed.

The Louisiana Illuminator had reported in March 2026 — between the two Amazon announcements — on the broader pattern of NDA use in Louisiana's economic development negotiations, under the headline 'A muzzle on elected officials.' That analysis, produced in partnership with Type Investigations, documented that NDAs had become standard practice in Louisiana's pursuit of large technology and industrial investments. Elected officials confirmed that they had signed agreements that prevented them from discussing active negotiations with constituents. The Illuminator's characterization — 'a muzzle on elected officials' — was not editorial hyperbole. It was the logical description of an instrument that legally prevents an elected representative from performing the representative function.

[PR VOICE] Amazon's investment in the Shreveport-Bossier City region represents a transformational commitment to Northwestern Louisiana that will create thousands of good-paying jobs, support local suppliers and contractors, and strengthen the regional economy for generations to come. We are deeply grateful to Governor Landry, the Louisiana Economic Development agency, and the community partners in Caddo and Bossier Parishes for their partnership in making this historic investment possible. These data centers will power Amazon Web Services and the AI capabilities that businesses and consumers across Louisiana and the nation rely on. We are committed to being a responsible partner to the communities where we operate, and we look forward to a long and productive relationship with the people of Northwest Louisiana.

[CRITICAL THEORIST] The *fait accompli* is a specific Stage One consent manufacturing technique. It is not the same as the normalized NDA baseline of Virginia, where the suppression is ongoing and gradual. The *fait accompli* compresses the suppression into a single moment: the announcement. By announcing the investment as a completed decision — the past tense is deliberate; 'Amazon announced plans to invest' describes a decision made, not a decision being made — the corporation and the state simultaneously manufacture consent and foreclose contestation. The community is presented with a decision it did not make, cannot unmake, and can only accept or resist at considerable cost. The Governor's framing of the investment as 'historic' further forecloses critique: to contest a 'historic investment' is to position oneself against prosperity, against jobs, against the region's future. This is Stage One's Language Architecture: the framing that makes contestation illegitimate before it can be organized.

[ETHNOGRAPHIC WITNESS] The documentary record of the Shreveport case is clear. Amazon press release (February 23, 2026): \$12 billion investment in Caddo and Bossier Parishes; described as 'next-generation data center campuses.' Governor Landry announcement (February 23, 2026): 'largest investment in the history of Northwest Louisiana.' Louisiana Illuminator / Type Investigations (March 27, 2026): NDAs described as 'a muzzle on elected officials'; local officials confirmed signing NDAs that prevented disclosure of negotiations. NOLA.com (August 18, 2026): \$6 billion expansion announced as third campus in west Shreveport. Total announced investment: \$18 billion. No public deliberation preceded either announcement. No environmental impact assessment was released publicly before the investment was announced. The SEAM hidden costs — water, power, air quality, tax exemption revenue loss — do not appear in any publicly available document related to either announcement.

[SEAM ANALYST] The SEAM hidden-cost framework applied to Shreveport produces the following preliminary accounting. Visible consulting and site advisory fee for a project of this scale: estimated at \$150 million to \$300 million across the project lifecycle. SEAM hidden-cost multiplier: 1.67 times the visible fee, placing total hidden costs at \$250 million to \$500 million attributable to the advisory engagement alone. But the community-level hidden costs extend far beyond the consulting engagement. A \$12-billion-to-\$18-billion data center campus complex will draw electricity at hyperscale rates — potentially hundreds of megawatts — from a regional grid that was not designed to support this load. The grid infrastructure investment required will be borne by ratepayers, not by Amazon. Water drawn from Caddo Parish's supply — which includes Red River water rights that are not unlimited — will reduce availability for residential, agricultural, and industrial users who were not party to the NDA negotiation. Air quality impacts from backup generation systems will fall disproportionately on communities near the facilities. Tax exemption agreements will reduce public revenue for schools and services over a period of fifteen to twenty years. None of these costs appear in the \$18 billion investment figure. All of them will be paid by the community.

[COMMUNITY VOICE] The residents of Caddo and Bossier Parishes did not elect their officials to sign Non-Disclosure Agreements with trillion-dollar corporations. They elected them to represent the community's interests in negotiations that will shape the community's future. The NDA instrument inverted that representational relationship: the official who signed the NDA became, for the duration of the negotiation, an agent of Amazon rather than an agent of the community. The community learned about the investment that will shape its next thirty years at the same moment as the rest of the country: when the press release landed. That is not democratic governance. That is notification. And notification is not consent.

[DAVID'S VOICE] What strikes me about Shreveport — and what distinguishes the Louisiana case from Virginia's slow normalization — is the speed. The *fait accompli* technique accelerates Stage One by compressing the consent-manufacturing process into a single public event. In Virginia, the normalization developed over years, and communities had time (if not always the institutional capacity) to develop awareness of what was happening. In Shreveport, the announcement preceded awareness by design. The Louisiana Illuminator's March 2026 reporting on NDA use in Louisiana economic development was published a month after Amazon's initial announcement. The critique arrived after the decision had been made public and the celebratory frame had been established. This is Stage One operating at its most efficient: by the time the analytical tools that might contest the decision are deployed, the decision is already presented as *fait accompli* — as historic, as transformational, as the largest investment in the history of Northwest Louisiana. Contestation has been linguistically, legally, and temporally foreclosed before it begins.

III.2 — Antenarrative: What Louisiana's Official Story Suppresses

The official narrative of the Shreveport-Bossier City Amazon investment is a story of triumph: a rust-belt industrial region that had watched manufacturing decline for decades, securing the largest private investment in its history from one of the world's most powerful corporations. That narrative is emotionally and politically resonant, and it is not without foundation. Northwest Louisiana does face genuine economic challenges, and genuine jobs will be created by the construction and operation of these facilities. The antenarrative concerns what the *fait accompli* technique suppressed: the negotiation itself. Who negotiated the terms? What tax incentives were granted? What water rights were committed? What electricity rates were locked in, and for how long? What community benefit agreements — if any — were attached to the investment? What happens when Amazon exits the market, or when the next generation of AI infrastructure supersedes this one, and the community is left with facilities designed for a technology that no longer requires them? These questions were not part of the announcement. They cannot be part of public deliberation because the NDA that preceded the announcement foreclosed the deliberation that might have produced enforceable answers.

PART IV: CASE 3 — SOUTH MEMPHIS, TENNESSEE: THE SACRIFICE ZONE

IV.1 — Colossus in a Majority-Black Neighborhood

South Memphis, Tennessee is a majority-Black neighborhood that has carried a disproportionate share of Memphis's industrial burden for generations. The pattern of industrial siting that has concentrated environmental and health risks in majority-Black and low-income communities across the American South is well-documented in environmental justice scholarship and in the legal record of litigation that has challenged it. South Memphis is not exceptional in this pattern; it is exemplary. The concentrated industrial presence in South Memphis — including facilities associated with significant air quality, water quality, and health impacts — is the product of decades of siting decisions made without meaningful community participation, often with the explicit or implicit support of local economic development frameworks that treated environmental and health costs to poor communities of color as externalities rather than as costs.

In 2024 and 2025, Elon Musk's artificial intelligence company xAI negotiated the arrival of its Colossus supercomputer facility in South Memphis behind Non-Disclosure Agreements. The community learned that the facility was operational — and that it was generating significant air pollution through the operation of natural gas-powered generators — weeks after operations had begun. The NDA instrument that structured the negotiation ensured that the community had no meaningful opportunity to evaluate the environmental and health implications of the facility before it was operating. This was not an oversight. It was the Stage One consent manufacturing process applied to a community that had the least institutional capacity to contest it and bore the highest existing burden of industrial environmental risk.

The Southern Environmental Law Center (SELC) and the NAACP filed legal challenges against the xAI Colossus facility, arguing that the natural gas generator operations constituted unpermitted air pollution in violation of federal and state environmental law. The legal record produced by those challenges provides the documentary foundation for this analysis. Community members who testified in those proceedings described the facility's environmental impact in language that the Louisiana Illuminator's 'muzzle' framing does not adequately capture. One South Memphis resident, quoted in the NAACP v. xAI filing, described the facility's arrival as 'a death sentence for our community.' That description is the antenarrative of the Memphis case in a single sentence: not an investment, not a transformation, not a historic commitment to the region. A death sentence. Delivered without a trial.

The xAI Colossus case illustrates the Accountability Architecture dimension of the Four-Dimensional Framework at its most extreme. In a named-accountability system, the decision to site a massive industrial AI facility in a majority-Black neighborhood with the highest existing environmental burden in Memphis would require named officials to justify that decision publicly, to acknowledge the environmental justice implications, and to accept personal accountability for the consequences. In the diffused-accountability system

produced by the NDA instrument, no such accountability is required. The NDA covered the negotiation; the economic development framework covered the approval; the air quality permitting process — which the SELC and NAACP argued was improperly bypassed — would have covered the environmental review. Each layer of the institutional architecture absorbed a portion of the accountability, and the aggregate result was that no individual or institution was named as responsible for what the community experienced as a death sentence.

[PR VOICE] xAI is proud to invest in Memphis, Tennessee, bringing world-class AI infrastructure and high-quality technical jobs to the Mid-South region. The Colossus facility represents a transformational commitment to American AI leadership and to the communities that support it. We are working closely with local leaders, environmental regulators, and community stakeholders to ensure that our operations meet the highest standards of environmental responsibility. The temporary generation systems used during the facility's initial operations were permitted and operated in full compliance with applicable regulations. We are committed to transitioning to cleaner energy sources as the grid infrastructure to support them becomes available, and we look forward to being a long-term, positive partner for the people of South Memphis and the greater Memphis community.

[CRITICAL THEORIST] The sacrifice zone is a specific political economy concept developed in environmental justice scholarship to describe geographies — typically poor, typically communities of color — that are designated, through the accumulated logic of siting decisions, to bear the industrial costs that more powerful communities have been able to exclude from their own geographies. South Memphis did not become a sacrifice zone through any single decision; it became one through the accumulated weight of decisions that each, individually, could be justified on economic grounds, and whose collective effect was to impose a disproportionate share of Memphis's industrial environmental burden on its most vulnerable residents. The xAI Colossus decision follows this accumulated logic. The NDA instrument did not create the sacrifice zone; it insulated the siting decision from the democratic contestation that might, at the margin, have altered its location or its terms. What the NDA did — structurally, in the logic of the Four-Dimensional Framework — was ensure that the Accountability Architecture remained diffused at the moment when named accountability might have produced a different outcome. By the time the community learned what was coming, Colossus was running.

[ETHNOGRAPHIC WITNESS] The documentary record of the Memphis case includes: NAACP and SELC legal filings challenging the xAI Colossus facility's air quality permit compliance; testimony from South Memphis residents in those proceedings, including the 'death sentence' characterization cited above; media reporting establishing that the facility became operational before community members were informed through public channels; xAI's public statements on its environmental commitments; and the legal and regulatory record of the permitting process. The NAACP's institutional involvement is significant: the organization has been central to environmental justice litigation in the South for decades

and does not bring legal action against industrial siting decisions lightly. The SELC, based in Charlottesville, Virginia and Atlanta, Georgia, is one of the preeminent environmental law organizations in the American South. Their joint involvement signals that the Memphis case is not a localized dispute about one facility but a test case for the environmental justice implications of AI infrastructure siting more broadly.

[SEAM ANALYST] The SEAM hidden-cost framework applied to South Memphis requires a modification of the standard hidden-cost multiplier because the baseline environmental burden in the affected community is not ordinary. In a community that is already bearing disproportionate industrial environmental cost, the marginal hidden cost of an additional major industrial facility — in terms of health impacts, absenteeism, reduced productivity, community stress, and institutional resilience — is significantly higher than in a community starting from a lower burden baseline. The SEAM six dysfunction categories applied to South Memphis: Working conditions dysfunction — the health and safety conditions of workers in and near the facility, including exposure to natural gas generator emissions; Time management dysfunction — community members' time diverted to health management, advocacy, and legal participation rather than productive activity; Communication dysfunction — the NDA-enabled suppression of information about the facility during the negotiation phase; Integrated training dysfunction — absence of workforce pathways from the South Memphis community into the technical roles the facility requires; Strategic implementation dysfunction — absence of community benefit agreements requiring local hiring, environmental mitigation, or health impact remediation; and Implementation dysfunction in the physical infrastructure of the community, whose grid, water, and air resources are allocated by agreements the community had no role in negotiating. The aggregate hidden cost in a sacrifice-zone context is not 1.67 times the visible consulting fee. It is incalculable in the standard SEAM framework because the baseline from which the dysfunction increment is measured is itself a product of accumulated Stage One suppression.

[COMMUNITY VOICE] We did not consent to this. That is the fact that all the economic development language, all the job creation projections, all the investment announcements cannot erase. We did not consent to this facility operating in our neighborhood. We did not consent to the air quality that our children breathe being degraded by generators that were not permitted through the process that our laws require. We did not consent to learning about this from the news, weeks after it was already running. Our elected officials signed an NDA. They did not ask us first. They could not ask us first — that was the point of the NDA. The Consent Machine is exactly the right name for this. Because what the machine manufactures, here in South Memphis, is our silence. Not our consent. Our silence. And when the machines are loud enough, no one can hear the difference.

[DAVID'S VOICE] The Memphis case is the hardest case in this chapter to write about with the analytical distance that scholarship requires, because what it describes is the use of a legal instrument — the Non-Disclosure Agreement — to replicate, in the AI infrastructure era, the siting logic that has made sacrifice zones out of majority-Black neighborhoods in the

American South for the past century. I am a white academic writing about the environmental justice implications of a decision that was made behind closed doors, the costs of which will be borne by people who did not make it and were not asked. I want to be honest about the limits of my position: I can document the pattern, I can name the instrument, and I can show how the Four-Dimensional Framework maps the mechanism by which diffused accountability enables sacrifice-zone outcomes. What I cannot do, from my position in Las Cruces, New Mexico, is fully represent the testimony of South Memphis residents who described what happened to their community as a death sentence. I can only amplify that testimony and insist that it be heard in the scholarly record alongside the economic development announcements that will otherwise define this episode's official history.

IV.2 — Antenarrative: What Memphis's Official Story Suppresses

The official narrative of xAI Colossus in Memphis is a story of American AI leadership: a domestic technology company investing in American infrastructure, creating technical jobs in a region that needs them, and advancing the frontier of artificial intelligence in competition with Chinese rivals. That framing is constructed carefully to foreclose the environmental justice critique before it can be organized: to contest Colossus is to contest American AI competitiveness; to demand environmental justice review is to delay American innovation. The antenarrative of the Memphis case concerns the question the official narrative cannot ask without collapsing: why South Memphis? Why a majority-Black neighborhood with the highest existing industrial environmental burden in the city? The answer — that South Memphis was chosen precisely because its institutional and political capacity to resist was the lowest — is the antenarrative that the NDA instrument helps suppress by ensuring that the community's first knowledge of what was coming arrived after the decision had been made and the contracts signed.

PART V: CASE 4 — TUCSON, ARIZONA: WHERE THE MACHINE FAILED

V.1 — Project Blue and the Conditions of Resistance

Tucson, Arizona is the counter-case. It is the most important case in this chapter, because it proves — not by argument but by example — that the Consent Machine is not inevitable. Project Blue, the data center investment proposal that sought to site a major hyperscale facility in the Tucson metropolitan area, was rejected. The machine failed. Understanding why it failed is understanding, by contrast, why it usually succeeds — and that understanding is the most useful thing this chapter can offer to communities currently facing their own version of the same machine.

The conditions of Project Blue's failure in Tucson are specific and identifiable. They are not a mystery or a matter of luck. They are a set of institutional, informational, and civic conditions that, taken together, produced an outcome different from the one the machine is designed to deliver. The analysis that follows traces each of those conditions through the Four-Dimensional Framework, not to celebrate Tucson as an exceptional case but to use its exceptionalism diagnostically: to ask what Tucson had that Virginia, Louisiana, and

Memphis did not, and to derive from that question a set of conditions that communities facing AI infrastructure siting decisions might seek to cultivate.

The first condition of Tucson's resistance was early information. Before the project reached the stage where an NDA would have been required for local official participation, community members and independent journalists had obtained and published enough information about the proposal's scale, water requirements, and likely siting locations that public deliberation could begin before the fait accompli technique could be applied. This is the informational condition for resistance: not that NDAs did not exist, but that information circulated through channels that NDAs could not suppress before the negotiation reached the stage where suppression would have been decisive. The timing mattered. The machine requires a window between the beginning of negotiation and the public announcement of a completed deal. In Tucson, that window was shorter than the machine required.

The second condition was community organizing capacity. Tucson has an unusually developed network of community organizations, environmental advocacy groups, and civic institutions that have maintained sustained engagement with land use, water, and development issues for decades. This network did not emerge in response to Project Blue; it predated the proposal and could respond to it quickly, with established relationships, communication channels, and institutional knowledge about how to participate in the permitting and approval process. The machine is most effective in communities where civic organizing capacity has been depleted by decades of economic stress, institutional abandonment, or the political demobilization that follows repeated losses. Tucson, whatever its economic challenges, had not experienced that depletion at the organizational level.

The third condition was water consciousness. Tucson is a desert city that has been thinking seriously about water scarcity for a generation. The community's relationship to the question of water allocation — who gets it, under what conditions, at what cost — is not abstract. It is lived daily, argued at city council meetings, litigated in state water courts, and embedded in the political culture of the community in ways that are simply not present in communities that have not faced water scarcity as an organizing political fact. When Project Blue's water requirements became a matter of public discussion, Tucson residents had the conceptual framework, the institutional vocabulary, and the political motivation to contest a water allocation they had not agreed to. This is what 'You can't drink data' means as a political slogan: it is a community's articulation of a priority ordering that the NDA instrument, by design, prevents from being applied at the decisive moment. In Tucson, it was applied before the decisive moment had passed.

[PR VOICE] We appreciate the engagement of the Tucson community in the Project Blue review process and respect the community's decision regarding this particular proposal. We remain committed to responsible data center development in the American West and look forward to future opportunities to partner with communities that share our vision for AI infrastructure that benefits local economies while respecting local resources. The dialogue around Project Blue has helped us better understand the community's priorities,

and we look forward to applying those lessons as we continue to grow our infrastructure in support of the AI services that businesses and consumers depend on.

[CRITICAL THEORIST] The Tucson counter-case is theoretically generative because it reveals, through contrast, the specific institutional conditions that the Consent Machine requires to operate effectively. The machine requires: an informational asymmetry that is established and maintained through NDA before the community can organize around the proposal; a civic capacity gap that prevents the rapid mobilization of community opposition in the window between the beginning of negotiation and the fait accompli announcement; a resource scarcity that is either not salient enough to mobilize opposition (as in Virginia, where water scarcity is not a lived daily political fact) or is so severe that it has depleted civic organizing capacity (as in parts of the rural American South). Tucson had neither the required informational asymmetry nor the required civic capacity gap nor the required salience deficit on water. The machine cannot operate when those conditions are absent. The structural implication is not that communities need to be like Tucson — that would be an impossibly abstract prescription. The structural implication is that the specific deficits that allow the machine to operate are identifiable and, in principle, addressable through policy: transparency requirements that prevent the NDA instrument from establishing the informational asymmetry; civic capacity investment that prevents the organizing gap; and water accounting that makes scarcity salient before the siting negotiation begins rather than after it concludes.

[ETHNOGRAPHIC WITNESS] The Tucson case's documentary record differs from the other three cases precisely because the machine's operation was interrupted before it reached the fait accompli stage. There is no Amazon press release describing the historic investment that transformed the community. There is no NAACP lawsuit documenting the health impacts of a facility that operated before community members were informed. There is a proposal, a public debate, and a rejection. The absence of a fait accompli announcement is itself the most significant documentary evidence in this case: it confirms that the machine's ability to present decisions as completed depends on the informational suppression that the NDA enables, and that when that suppression fails — when information circulates through the community before the negotiation reaches the announcement stage — the machine cannot function as designed.

[SEAM ANALYST] The SEAM analysis of the Tucson counter-case is unusual because the hidden costs this analysis is designed to identify were, in this instance, avoided. The SEAM framework is most typically applied to the hidden costs of decisions that were made under conditions of information suppression. In Tucson, the avoidance of those hidden costs is itself the finding. The community's water resource — which would have been committed to data center cooling for decades under a signed agreement — remains available for the community's use. The grid infrastructure that would have been strained by hyperscale power draw remains available for residential and commercial users who now have full access to its capacity. The tax exemption revenue that would have been diverted from public services remains in the public treasury. In SEAM terms, Tucson's rejection of Project

Blue is a case where the avoidance of dysfunction generated positive value — not just the absence of a cost, but the preservation of resources that would otherwise have been committed. This is the economic argument for the community organizing conditions that produced Tucson's outcome: resistance has a positive SEAM value that the economic development framework, by design, does not count.

[COMMUNITY VOICE] We said no. That is the thing I want to say plainly, because I think the other cases in this chapter can create the impression that saying no is not possible — that once the machine has started, there is nothing to be done. Tucson said no. We organized. We showed up. We made the water argument and we made it before the agreement was signed, not after. We did not have a community benefit agreement to negotiate because we did not allow the negotiation to reach the stage where a community benefit agreement would have been the only option. We stopped it earlier. And I want communities facing this right now — communities in Louisiana, in Tennessee, in New Mexico, in Texas, in Virginia — to know that it is possible to stop it earlier. The machine is powerful but it is not inevitable. Tucson is the proof.

[DAVID'S VOICE] I teach organizational theory. I have spent four decades studying how institutions change — or fail to change — in the face of external pressure. The Tucson case is, for me, the most theoretically important case in this chapter because it falsifies the deterministic reading of the Consent Machine. The machine is not a self-executing algorithm that produces a fixed output regardless of community conditions. It is a social mechanism that operates through specific institutional conditions, and when those conditions are not present, the mechanism does not function. The policy implication of this observation is not optimistic in a naive sense — the conditions that allowed Tucson to resist are not easily replicable in communities that have been institutionally depleted by decades of economic decline, racial exclusion, and civic demobilization. But it is optimistic in the deeper sense that it shows the machine's operation is contingent rather than necessary. The conditions that produce resistance are buildable. They can be invested in, in advance, by communities that have not yet faced this particular machine. That is the practical argument for the transparency policies, civic capacity investments, and water accounting reforms that the conclusion of this chapter will propose.

V.2 — Antenarrative: What Project Blue's Rejection Reveals

The official narrative of Project Blue's failure — if there is one — is a story about regulatory uncertainty, community opposition, and the challenges of siting complex infrastructure in politically engaged communities. That is the framing from which the technology industry will derive its lessons: not 'we should have used a more democratic process,' but 'we need to find more accommodating communities.' The antenarrative of Project Blue's rejection is a story about what democratic deliberation actually looks like when it is allowed to function: communities asking the right questions, at the right time, with the right information, and arriving at an outcome that reflects their own assessment of costs and benefits rather than the assessment imposed on them by a negotiation conducted in secret. The counter-case is not a story about a community that said no. It is a story about a community that was

allowed — barely, through the accident of early information and organized civic capacity — to say anything at all.

PART VI: NEW MEXICO AS CONNECTIVE TISSUE — WITNESS GEOGRAPHY

VI.1 — Two Counties, Same State, Opposite Outcomes

Throughout this book, I have returned to the contrast between Doña Ana County and Sierra County in New Mexico — the two communities I know most intimately, because I live in them. I want to return to that contrast one final time in this chapter, not because New Mexico is the most dramatic case of NDA-enabled consent manufacturing in the AI infrastructure buildout, but because it is the case I can witness directly, and because its dynamics illuminate, from the ground level, the institutional mechanisms that the four major cases in this chapter analyze from the outside.

In Doña Ana County, the story of Project Jupiter — Oracle's \$165 billion proposed data center complex spanning 1,400 acres in the high desert south of Las Cruces — is still unfolding. The approval process that has proceeded despite the exposure of fraudulent public comment submissions, despite state assessments finding that the facility's water claims were false, despite the evidence of institutional capture that I have documented throughout this book. Non-Disclosure Agreements were central to that process. New Mexico State University, my institution, entered into a \$30,000 desalination advisory contract with BorderPlex Digital Assets LLC, the Austin-based developer, accompanied by an NDA requiring NMSU to assist the developer in opposing public records requests. The university's own NDA required it to work against the transparency mechanisms that are the prerequisite for the democratic deliberation that the NDA instrument is designed to suppress. That is Stage One at the university level: the institution whose academic integrity provides legitimacy to the project is contractually bound to protect the project from the scrutiny that its academic integrity would, absent the NDA, require it to conduct.

In Sierra County, the contrast is as stark as it was in the Preface. Community members organized. Independent newspapers — the *El Defensor Chieftain* and the *Sierra County Sentinel* — covered the story with the depth and persistence that made informed community deliberation possible. New Mexico Tech hosted town halls. County commissioners heard the community. The signs in the audience read: You can't drink data. And the commission, which had largely supported the proposal, reversed course. An eighteen-month moratorium on new data center construction was established — one of the first successful community resistance campaigns of its kind in the American Southwest. Socorro County achieved the same outcome through the same mechanisms.

The New Mexico Attorney General, Raúl Torrez, opened an investigation into the fraudulent public comment submissions in Doña Ana County — canvassers wearing Project Jupiter T-shirts, hired from a California firm, who collected names at gas stations and grocery stores and used those names to submit false support letters to the state environmental department. The residents whose names were used had never written those statements. AG Torrez stated publicly: 'If individuals or organizations attempted to manipulate the public

comment process through fraudulent submissions, we will investigate those actions thoroughly and hold responsible parties accountable.' That investigation is ongoing. Its outcome will determine whether the accountability diffusion that the NDA instrument produces at the level of official negotiation is matched by accountability diffusion at the level of the fraudulent consent manufacturing that NDA-protected negotiations enable.

[PR VOICE] Project Jupiter represents an extraordinary opportunity for New Mexico — a generational investment that will create thousands of jobs, support New Mexico State University's research mission, and position the state as a leader in the digital economy. We are committed to working with the community, the university, and state regulators to address any concerns about water use, electricity, and community impact. We believe that the benefits of this investment far outweigh the concerns that have been raised, and we look forward to demonstrating that commitment through our operations over the years and decades ahead.

[CRITICAL THEORIST] The Doña Ana versus Sierra County contrast is the book's core structuring opposition, and it reveals something that the four major cases in this chapter do not as clearly: the role of community journalism in determining whether the NDA instrument's informational asymmetry can be established before community organizing can respond. Doña Ana County's syndicated newspapers — which relied on advertising from the same technology and development ecosystem that was advancing Project Jupiter — could not and did not provide the investigative depth that would have allowed community members to understand what was being negotiated on their behalf. Sierra County's independent newspapers — the *El Defensor Chieftain* and the *Sierra County Sentinel* — could and did. The difference in outcome is partially attributable to the difference in information environment. This is the Gramsci point applied to local journalism: the colonization of civil society by economic society does not require the direct suppression of the press. It requires only that the press whose independence might resist that colonization be structurally undermined — by advertising dependence, by consolidation, by the replacement of local investigative reporting with syndicated wire content — until the information environment is no longer capable of producing the early, accurate, independent coverage that democratic deliberation requires.

[ETHNOGRAPHIC WITNESS] The New Mexico cases produce a documentary record that includes: the AG Torrez investigation and public statement on fraudulent comment submissions; the NMSU desalination advisory contract and the NDA requiring the university to oppose public records requests; the Project Jupiter water claim assessment conducted by state representatives and found to be false; the *El Defensor Chieftain* and *Sierra County Sentinel* front-page coverage of the data center moratorium; the town hall testimony that produced the commission reversal; and the [storying.site](#) documentation maintained by the author as an ongoing primary source archive for this project. The contrast in documentary density between the two counties is itself significant: the Sierra County case produces more public documentary evidence because more of the deliberation occurred in public. The Doña Ana County case produces less, because more of the decisive negotiation occurred

behind NDAs. Documentation density is an inverse indicator of NDA penetration: the more of a community's deliberation has been captured by non-disclosure, the less the public record shows.

[SEAM ANALYST] The SEAM comparison of Doña Ana and Sierra Counties is the most direct hidden-cost analysis available because the two communities are in the same state, face the same structural pressures, and have access to the same water resources — the Rio Grande basin and its associated aquifers. A hyperscale data center complex consuming one to five million gallons of water per day — the range for facilities of Project Jupiter's proposed scale — would draw from the same aquifer system that supplies Las Cruces, Doña Ana County's agriculture, and the downstream communities in the Mesilla Valley. The hidden cost of that water commitment is not speculative; it can be estimated from the aquifer recharge rates, the existing senior water rights, and the downstream allocation commitments that New Mexico's interstate compact obligations require. Project Jupiter's claim of 20,000 to 60,000 gallons of daily non-potable water use — figures that critics argue have no plausible relationship to the actual operational requirements of a 1,400-acre hyperscale facility — represents, if accurate, a small fraction of the facility's actual draw. The SEAM hidden cost of the water depletion, if the facility operates at the scale implied by its stated acreage and power requirements, would be borne by the community's agricultural water users, residential ratepayers, and downstream compact obligations. None of those costs appear in the \$165 billion investment figure.

[COMMUNITY VOICE] I live in Las Cruces. My children go to school here. We drink the water here. When I see an advertisement every two hours on local television promising that a data center will not use any of our water, I do not feel reassured. I feel alarmed. Because the people making that promise signed a Non-Disclosure Agreement with the developer before they agreed to make it. They are not in a position to know whether the promise is true. They are in a contractual position that prevents them from telling me if they suspect it is not. The NDA does not only muzzle them. It weaponizes their public credibility against the community's right to know. When an elected official says the water claim has been independently verified, and that official has signed an NDA with the developer, I have no way to know what the verification actually showed, who conducted it, or whether the official is permitted under the NDA to tell me if the results were different from what is being publicly claimed.

[DAVID'S VOICE] I am a witness to what is happening in my own county. I want to turn that witnessing into a scholarly account. That is the sentence I wrote at the beginning of this book, and I want to return to it here because the New Mexico connective tissue section is the place in this chapter where the scholarly and the personal cannot be cleanly separated. I live in Doña Ana County. I work at NMSU. The university where I have taught organizational theory for the better part of four decades is under an NDA that requires it to assist a developer in opposing the public records requests that my own research depends on. That is not an abstract structural claim about institutional capture. It is a personal observation about the institution that employs me. I have an obligation, as a scholar, to say that plainly

and to let readers make their own judgment about how it affects my analysis. I also have an obligation, as a community member, to continue to document what I observe and to make that documentation available to the scholars, journalists, advocates, and community members who will evaluate it long after this particular NDA has expired.

PART VII: SYNTHESIS — WHAT NDA GEOGRAPHY REVEALS ABOUT THE CONSENT MACHINE

VII.1 — The Five Conditions of NDA-Enabled Consent Manufacturing

The four case geographies and the New Mexico connective tissue, read together through the Four-Dimensional Framework, produce a set of identifiable conditions that enable Stage One consent manufacturing through the NDA instrument. These conditions are not inevitable features of AI infrastructure development; they are contingent institutional arrangements that can be altered by policy, civic investment, and community organizing. Naming them precisely is the prerequisite for designing the interventions that could alter them.

The first condition is informational asymmetry, established before community organizing can respond. The NDA creates a window — between the beginning of negotiation and the *fait accompli* announcement — during which the corporation and its advisors have complete information about the proposal's scale, terms, and implications, while the community has none. If information circulates through independent journalism, academic research, or FOIA litigation before that window closes, the asymmetry can be reduced or eliminated. If it does not — as in Louisiana and Memphis — the announcement arrives before resistance can be organized.

The second condition is civic capacity depletion. The Consent Machine operates most effectively in communities whose civic organizing infrastructure has been weakened by economic stress, population loss, institutional abandonment, or the political demobilization that follows repeated losses. Tucson's resistance was enabled by civic capacity that predated Project Blue. South Memphis's inability to prevent Colossus's arrival was, in part, a product of the civic capacity depletion that is itself a legacy of decades of sacrifice-zone siting decisions. The machine targets the depleted; it avoids the organized.

The third condition is resource salience deficit. In communities where the resources being committed by the data center agreement — water, power grid capacity, air quality, tax revenue — are not already salient political questions, the announcement of a commitment can be framed as purely positive before the resource cost becomes visible. Tucson's water consciousness made the water cost of Project Blue immediately salient. Virginia's less acute water stress allowed the NDA-covered commitments to proceed without triggering the resource salience that would have mobilized opposition. In New Mexico's arid desert, water is always salient — but the NDA instrument, in Doña Ana County, captured the officials responsible for communicating that salience to the community before they could do so.

The fourth condition is institutional capture of legitimacy-granting bodies. The NDA instrument does not only silence elected officials; it recruits legitimacy-granting institutions — universities, chambers of commerce, community foundations, regional economic development agencies — into active support for the project, often through inducement payments or partnership fees that are themselves covered by NDAs. NMSU's \$30,000 contract, structured to require the university to oppose public records requests, is the clearest example of this capture in the New Mexico case. The institution whose academic credibility legitimizes the project is contractually bound to protect the project from the scrutiny that its academic credibility would otherwise require it to conduct. This is Stage One's most sophisticated mechanism: manufacturing consent not only by silencing opposition but by conscripting the institutions that would otherwise be most credible in producing it.

The fifth condition is the Temporal Horizon gap — the exit-before-consequence logic that is built into the NDA instrument's design. The corporation that signs an NDA with a county commissioner in 2024 is committing community resources that will persist through 2040, 2050, and beyond. The consulting firms that designed the community engagement strategy will have exited the engagement within eighteen to thirty-six months. The elected officials who signed the NDA will, in most cases, have served out their terms and left office before the environmental and infrastructure costs of the agreement become visible at scale. The NDA instrument encodes the Temporal Horizon dimension of the Four-Dimensional Framework: it creates a structure in which the costs are borne by the community over decades, while the benefits — fees, contracts, investment announcements — are captured immediately by the parties to the agreement.

VII.2 — Answerability for the NDA-Using Institutions

In this book, we are assessing Accountability diffusion, Antenarrative suppression, Conscience suppression, and BME Narrative constructions that applies directly to the institutional architecture of NDA-enabled consent manufacturing. A corporation that signs NDAs with elected officials and simultaneously publishes commitments to community partnership and responsible development scores at the highest end of the Hidden Cost analysis on all four dimensions: Accountability is maximally diffused (no named individual or institution is accountable for the community costs the NDA covers); Antenarrative processes are maximally suppressed (the NDA prevents the information that would allow community members to construct an alternative account of the decision); Conscience is maximally suppressed (the economic development framing positions conscience-driven opposition as opposition to jobs, investment, and progress); and Binary framing is maximally applied (support the investment or oppose economic development — there is no legitimate space for the community's own cost-benefit analysis).

The consulting firms that design Stage One community engagement strategies for AI infrastructure siting — producing the stakeholder alignment documents, the economic impact analyses, the community benefit narratives — hidden casts are the dimension that they receive fees for producing materials whose function is to manufacture the consent that

allows the NDA-covered negotiation to proceed to the fait accompli announcement. A McKinsey or BCG or Accenture community engagement strategy does not merely advise on how to communicate with a community; it constructs the informational environment in which the community's deliberative capacity will operate — or will fail to operate, because the strategy is designed to ensure that the deliberation occurs on terms favorable to the corporation before it occurs at all.

VII.3 — Toward Policy: Five Structural Remedies

This chapter does not end with critique alone. The analysis of the Tucson counter-case and the New Mexico witness geography, read together, suggest five structural remedies that could alter the institutional conditions that enable NDA-based consent manufacturing. These are not utopian prescriptions; they are extensions of legal and policy frameworks that already exist in some jurisdictions and that the Tucson case demonstrates can, in the right institutional conditions, produce democratic outcomes.

First, mandatory pre-announcement public disclosure. State legislation requiring that any data center siting negotiation involving public resources — tax incentives, water rights, electricity infrastructure investment, or land use approvals — be disclosed to the public before any NDA can be signed with elected officials or public institutions. Public Citizen's November 2025 analysis proposes specific legislative language for this remedy. Virginia, Louisiana, and Tennessee have not enacted it; their communities have paid the cost of its absence.

Second, community benefit agreement requirements with community negotiation capacity. State or local legislation requiring that any data center investment above a specified threshold — say, \$500 million in total investment or tax incentive value — include a community benefit agreement negotiated with community representatives who are not under NDA. This remedy addresses the institutional capture condition by creating a legally mandated channel for community voice that the NDA instrument cannot suppress.

Third, independent water and environmental impact assessment before any NDA. Environmental review that is conducted by an independent state agency — not by consultants engaged by the developer — before any NDA is signed and before any negotiation with elected officials begins. New Mexico's Attorney General investigation demonstrates that this is an achievable reform: the state already has the institutional capacity to conduct independent environmental review; what is needed is a legal requirement that it be completed before the NDA instrument can capture the negotiation.

Fourth, civic journalism investment as community infrastructure. The New Mexico contrast between Doña Ana and Sierra Counties demonstrates that the quality of community journalism is a decisive factor in whether communities have the information they need to exercise democratic deliberation before the NDA window closes. Independent local journalism that can conduct investigative reporting on data center proposals — not syndicated wire content that recycles developer talking points — is not a luxury for well-

resourced communities. It is the infrastructure of democratic self-governance in the AI infrastructure era, as essential as water pipes and power lines.

Fifth, institutional NDA prohibition for universities and public institutions. Legislation prohibiting public universities, community colleges, and other public institutions from entering into Non-Disclosure Agreements with private developers in connection with public resource siting decisions. The NMSU case demonstrates the specific harm of NDA capture of a public university: the institution's academic credibility legitimizes the project while the NDA prevents the institution from conducting the independent analysis that its credibility implies. This is not a subtle harm; it is a direct corruption of the public university's mission. It should be prohibited by the same state legislation that funds the university.

CONCLUSION: FROM AUDITING LABOR TO AUDITING LAND

This chapter has traced the Non-Disclosure Agreement as the Consent Machine's primary instrument in the AI infrastructure buildout across five geographies: Virginia (normalization), Louisiana (fait accompli), Tennessee (sacrifice zone), Arizona (counter-case), and New Mexico (witness geography). In each location, the NDA performs the same structural function: it converts a public interest decision into a private commercial negotiation, converts the elected official from a representative of the community into an agent of the corporation, and manufactures the appearance of democratic consent while suppressing the community's capacity to exercise it.

The historical parallel that Chapter 1 established — between the audit theater of the 1990s supply-chain compliance era and the Responsible AI theater of the 2020s — applies here with a geographic extension. In the 1990s, the NDA-covered supply-chain audit allowed Nike and its counterparts to manufacture the appearance of independent oversight while suppressing the evidence that would have disturbed that appearance. In the 2020s, the NDA-covered data center siting negotiation allows Oracle, Amazon, xAI, and their counterparts to manufacture the appearance of community partnership while suppressing the community's capacity to contest the terms of the partnership before it is decided.

The instrument has been updated. The stage has been extended from the factory floor to the community geography. The consulting firms that designed the labor compliance theater have evolved into the infrastructure siting advisory firms that design the community engagement strategies. The communities that bore the cost of the audit theater — the Bangladeshi garment workers, the Vietnamese electronics workers — have been joined by the South Memphis residents who learn from the news that a supercomputer is operating in their neighborhood, by the Shreveport residents who learn from a press release that \$18 billion in industrial development has been decided on their behalf, by the Doña Ana County residents who discover that the names their neighbors never consented to give are being used to file false public comment letters in support of a project their neighbors have not been allowed to evaluate.

The machine has moved from auditing labor to auditing land. The three stages are identical. The instrument has been updated. What has not been updated — what this book argues

must be updated — is the legal, institutional, and civic architecture through which communities can contest the machine before it closes the window on democratic deliberation. The Tucson case demonstrates that the window can be held open. The New Mexico cases demonstrate that holding it open requires community journalism, civic organizing capacity, and institutional integrity in the bodies whose credibility the machine conscripts. The policy proposals at the end of this chapter are not aspirational. They are the minimum institutional investment required to make democratic deliberation possible in the AI infrastructure era.

The next chapter will examine how the consulting firms that design these Stage One community engagement strategies apply the same architecture internally — how McKinsey, BCG, Accenture, Deloitte, and their counterparts manufacture consent within their own organizations for the AI-driven workforce transformations that are Stage Two of the machine. The geography changes. The mechanism does not.

Table 1

Organization	Impact Data Point	Note on Role
Accenture	77,000 AI professionals	Largest deployment of AI implementation labor.
McKinsey & Co	10% internal staff cut	Selling AI transformation while cutting their own staff.
BCG	25% of 2025 revenue	A quarter of all revenue now derived from AI consulting.
Deloitte	"Unhearing" staffing studies	Focus on efficiency metrics over human welfare.
IBM Consulting	Enterprise Layer Architect	Designing the specific tasks AI agents will replace.
PwC	Institutional Cover	Providing the "narrative shield" for radical change.
EY	Billion-dollar influence	Aggressive scaling of AI auditing and transformation.
KPMG	Trust/Usage studies	Shaping the public perception of AI reliability.
Infosys Consulting	Retail/Logistics focus	Redesigning global physical supply chains.
Cognizant Consulting	Gov/Healthcare focus	Moving AI into sensitive public and medical sectors.

Table 2

#	Company	Key Leaders	Flagship AI Products
1	OpenAI	Sam Altman (CEO), Fidji Simo (CEO of Applications), Mark	ChatGPT, GPT models, new "Astra" research model,

		Chen (Chief Research Officer)	browser and social apps
2	Google / DeepMind	Sundar Pichai (Alphabet CEO), Demis Hassabis (DeepMind CEO)	Gemini models, AI agents, AlphaFold research lineage
3	Anthropic	Dario Amodei (CEO)	Claude family (reportedly including a model referred to as "Mythos"), enterprise AI
4	NVIDIA	Jensen Huang (CEO)	GPUs (the compute backbone for nearly all AI training), CUDA stack
5	Microsoft	Satya Nadella (CEO)	Copilot ecosystem, Azure AI, significant minority stake in OpenAI
6	Meta	Mark Zuckerberg (CEO)	Llama open-weight models, Meta AI, newly consolidated superintelligence labs
7	xAI	Elon Musk	Grok models, integrated with X
8	Amazon (AWS)	Andy Jassy (CEO)	Bedrock, Nova models, Trainium custom AI chips
9	World Labs	Fei-Fei Li (Co-founder/CEO)	Spatial/world-model AI
10	AMD	Lisa Su (CEO)	Instinct AI accelerators challenging NVIDIA

Table 3

Firm	Interaction	What is public on money
PwC	First ChatGPT Enterprise reseller (May 2024); largest early enterprise customer; ~100,000 US/UK seats; starter packs that bundle licenses + workshops + RAI. Still a core Agent OS model.	PwC US \$1B / 3-year GenAI program (Apr 2023). Deal terms with OpenAI undisclosed.
Bain	First strategy-firm global services alliance (Feb 2023); Coca-Cola as first logo; Sage built on GPT-4;	Terms undisclosed. Bain pays OpenAI for seats and co-delivers.

	13,000 ChatGPT Enterprise seats (2024); ~50-person joint CoE; industry tools (retail, life sciences).	
McKinsey	Lilli uses OpenAI (and Cohere) on Azure; Feb 2026 Frontier Alliance with OpenAI, Accenture, BCG, Capgemini to deploy OpenAI Frontier “AI coworkers.” QuantumBlack forward-deployed teams feed product back to OpenAI.	Terms undisclosed.
Accenture	Uses OpenAI among many models; 2026 Frontier Alliance for implementation at scale. Also Anthropic (see below).	Terms undisclosed. Accenture’s broader AI investment has been cited in the multi-billion range.
BCG	ChatGPT Enterprise firmwide; 18k–36k custom GPTs; GENE on GPT-4o; 2026 Frontier Alliance.	Terms undisclosed.
Deloitte, EY, KPMG, IBM, Booz Allen	Consume GPT via Azure OpenAI or Bedrock, not always a named OpenAI “alliance.” Deloitte’s Australian government report used Azure OpenAI. IBM prefers watsonx as the control plane with optional external models.	Seat/API spend buried in cloud bills.

Chapter 3 Inside Story Big-Time AI Consulting

Summary

This chapter examines how ten global professional-services firms—Accenture, McKinsey, BCG, Deloitte, EY, PwC, Booz Allen Hamilton, KPMG, Bain, and IBM Consulting—have assembled agentic-AI platforms since late 2022. Across the set, architecture has converged on a shared stack: a rented foundation-model layer; a firm-owned knowledge and memory layer; an orchestration or “control plane”; specialized worker agents; and governance (identity, logging, guardrails). None of the ten built a frontier model comparable to GPT, Claude, Gemini, or Grok. What they sell is wrapping, routing, retrieval, evaluation, and multi-vendor interoperability, increasingly through Anthropic’s Model Context Protocol (MCP) and Google’s Agent-to-Agent (A2A) protocol. Accenture, BCG, Deloitte, IBM, KPMG, and McKinsey appear on the public A2A partner list.

The technical story is only half the chapter. The other half is the lived and litigated record of the same firms: employees describing how internal copilots compress junior work and how promotion systems now count AI log-ins; clients and civic researchers discovering fabricated citations in million-dollar public reports; whistleblowers alleging that confidential client or government material moved between bidding teams; and courts or commissions documenting older patterns—opioid sales advice, state-capture contracts, failed digital transformations, and False Claims Act settlements—that sit beside today’s “trusted agent” marketing.

Two design forks matter more than brand names. First, some platforms (BCG AgentKit, Bain’s published three-layer model) constrain agents with pre-authored plans and phased governance; others (Accenture Huddle, PwC Agent OS, IBM watsonx Orchestrate, EY’s “wedding cake,” KPMG Workbench) publish interoperable control planes that can host third-party agents. Second, grounding fails most visibly when the “agent” is asked to write an assurance or policy report rather than execute a checked transaction. Deloitte’s 2025 Australian welfare-assurance review and the 2025 Newfoundland and Labrador health-workforce plan are the clearest public demonstrations: Azure OpenAI or selective citation tools entered the methodology; invented authors, papers, and judicial quotations left the building.

The chapter’s claim is architectural and institutional at once. An agent operating system can register tools, score trust, and log every call. It cannot, by itself, restore apprenticeship, confidentiality walls, or the habit of verifying a footnote. The cases that follow treat employees, customers, whistleblowers, and courts as primary sources for that gap. Four teaching modules at the end convert the densest files (Deloitte footnotes, KPMG locker, BCG Aurora, McKinsey Lilli + Purdue) into seminar instruments.

Terms and acronyms

History in one arc: ChatGPT (November 2022) → enterprise RAG copilots in 2023 (Lilli, Sage, custom GPTs) → constrained or multi-agent kits in 2024 (AgentKit, Bedrock multi-agent, Semantic Kernel) → 2025–2026 “operating systems” that orchestrate third-party agents via MCP and A2A.

xAI / Grok. No major public architecture at these ten firms is derived from xAI. **Oracle** appears as an integration target (ERP and agents inside Huddle or Agent OS), not as the origin of the agent runtime.

Table 1: Ten firms at a glance: internal agent platforms, vendor lineage, and control-plane claims

Vendor “genes” show up as rented brains plus firm memory. Bain and BCG sit closest to OpenAI chat-plus-corpus. PwC productized MCP early. KPMG and much of EY’s office plane sit on Microsoft Foundry/Copilot. Deloitte and Accenture lean on Bedrock and multi-cloud huddles. NVIDIA appears in Accenture Refinery and EY intelligence/guardrails. No documented core derivation at these ten is from xAI.

Firm-by-firm architectures

1. Accenture — AI Refinery, Distiller, Trusted Agent Huddle

What / history. AI Refinery launched October 2024 on NVIDIA AI Enterprise. Distiller (June 2025) is the agent lifecycle SDK (memory, collaboration, evaluation, observability). Trusted Agent Huddle (April 2025) is the multi-system collaboration layer.

Architecture. Three built-in super agents plus utility agents:

Base Super Agent dynamically decomposes a task and assigns subtasks.

Flow Super Agent runs a user-configured deterministic workflow.

Evaluation Super Agent scores other agents and closes a feedback loop.

Third-party agents (Amazon Bedrock Agent, Azure AI Agent, Salesforce, SAP, ServiceNow, Workday, Oracle, Meta, Google, Databricks, Adobe, Snowflake) join the same fabric via A2A and MCP. A proprietary algorithm scores agents; Accenture has described this as the seed of an agent trust score. NVIDIA's Agent Intelligence toolkit supplies heterogeneous connectivity.

Derivation. Not a fork of ChatGPT. Runtime and acceleration come from NVIDIA; worker agents are often AWS Bedrock, Azure, or SaaS-native agents. Interop is Google A2A plus Anthropic MCP.

Official example. Accenture + NVIDIA + FedEx, announced with Huddle: multi-system agent collaboration for resilience, April 2025 onward. Claimed first multi-vendor huddle so agents from SAP, ServiceNow, and cloud providers can share a workflow instead of living in silos.

Controversy, previewed. Hertz website litigation is about implementation quality, not this platform specifically. The architectural risk is that a “trust score” is vendor-defined. Employee pressure to log into Refinery/SynOps for promotion is taken up in Lived Case 3.

2. McKinsey — Lilli (and LilliX)

What / history. Built in 2023 by ClientTech under CTO Jacky Wright; firmwide from July–August 2023; named for Lillian Dombrowski, the first professional woman hired in 1945. Started as knowledge search; became an orchestration layer. LilliX is the experimental reasoning/agent-chain lab.

Architecture. McKinsey is explicit that Lilli is not a simple RAG instance. It is a proprietary stack between users and models:

Hyperscaler-hosted large models (OpenAI on Azure, Cohere).

Five smaller expert models McKinsey trained to classify intent and improve answer relevance (“shaper + dash of maker”).

Custom PowerPoint parser (first versions read ~15% of slides; later >85%) because most firm knowledge lives in decks.

Dual surface: “GenAI Chat” versus “Client Capabilities” (100,000-plus internal documents and transcripts).

Citations, expert routing, Tone of Voice style agent.

Derivation. UX and base generation from ChatGPT-class models, hosted like Microsoft Copilot on Azure—but the memory and routing are McKinsey’s. Not Claude-native, not xAI.

Official example. On the order of 40,000–45,000 colleagues; research, synthesis, expert find, PowerPoint; 70–75%+ monthly use; ~17 queries per user per week; 500,000-plus prompts per month; ~30% time saved on search and synthesis.

Controversy, previewed. Lilli takes junior-analyst work (decks, synthesis). Separately, Purdue and South Africa are about *what* the firm advises, not Lilli’s code. Developed in Lived Case 4 and Teaching Case D.

3. BCG — AgentKit, GENE, Deckster

What / history. ChatGPT Enterprise to all staff in 2023; AgentKit open-sourced February 2024 by BCG X on LangChain, Next.js, and FastAPI; GENE announced September 2024; Deckster for slides. Tens of thousands of custom GPTs. BCG is an A2A partner.

Architecture. Deliberately anti-ReAct. A Meta/Router Agent picks a pre-authored Action Plan (a tree of toolsets run in sequence or parallel). Humans constrain the routes so the agent cannot wander. Intermediary step outputs can be shown to the user. GENE is a GPT-4o chatbot with a wide context window stuffed with BCG research and private interviews. Deckster is a template-trained slide agent with a “review this” grader.

Derivation. LangChain agent patterns plus OpenAI ChatGPT/GPT-4o. Not a Microsoft Copilot fork; not NVIDIA-first.

Official example. Public shipbuilder R&D case: multi-agent design automation; claimed 45% fewer engineering resources and 80% shorter lead time per ship deck. Logistics RFP agents: 30–50% efficiency.

Controversy, previewed. Gaza work that drifted into unauthorized relocation modelling (2024–2025) was a governance failure of partners, not AgentKit—and showed that agent/tooling discipline does not automatically equal engagement discipline. Developed in Lived Case 5 and Teaching Case C.

4. Deloitte — Zora AI, Ascend, Multi-Agent System (MAS)

What / history. Zora is the productized “digital workforce”; Ascend is the delivery/control platform; MAS on Amazon Bedrock AgentCore is the build/orchestrate layer sold via AWS Marketplace (2025). 100-plus Gemini agents with Google; A2A work with ServiceNow.

Architecture.

Agents typed **Perform** (transactions, self-healing) versus **Advise** (analysis, scenarios).

Perceive → reason → act, including multimodal extract and model-running.

Supervisor plus specialized collaborators (the Bedrock pattern): supervisor decomposes, calls sub-agents in serial or parallel, optionally uses a faster routing mode for simple intents.

Integrates SAP Joule agents as peers.

Ascend adds a Trust Score on prompt results and decision trails.

Derivation. AWS Bedrock multi-agent collaboration plus Azure OpenAI (used in some reports) plus NVIDIA (Quartz Atlas) plus Google Gemini/A2A. Zora is Deloitte IP on top of those runtimes.

Official example. HPE finance with Zora: operational performance-review automation; internal targets on the order of 25% cost / 40% productivity. Toyota Motor North America work referenced \$1 billion revenue/savings exploration with MAS.

Controversy, corrected. The Australia Department of Employment review (published mid-2025) was a A\$440,000 (about US\$290,000) assurance document, not the C\$1.6 million Canadian file. The C\$1.6 million health-workforce plan is Newfoundland and Labrador. Both are RAG/generation failures in *document production*, the opposite of a well-governed Perform agent with HITL. Developed in Lived Cases 1–2 and Teaching Case A.

5. EY — EY.ai Agentic Platform (“operating system”)

What / history. EY.ai after an approximately \$1.4 billion GenAI program; Agentic Platform announced March 2025 with NVIDIA; described as a three-layer OS co-built with Microsoft and NVIDIA. 150 tax agents for 80,000 professionals at launch; leadership later cited 50,000-plus agents in nine months. 2026: validation of NVIDIA NemoClaw for LangChain Deep Agents.

Architecture (EY’s “wedding cake”).

Data — Fabric / Data Marketplace; AI-ready governed data.

Orchestration — Azure AI Foundry model catalog (on the order of 11,000 models), Copilot as front door, Copilot Studio for builders, Fabric for unification.

Intelligence — NVIDIA GPUs, NIM microservices, NeMo Guardrails, Foundry distillation, domain Nemotron-style reasoning models, Omniverse for physical AI.

Plus **control rings**: nested governance so agent autonomy cannot bypass confidentiality. Runtime spans client cloud, on-prem, edge, NVIDIA Cloud. EY SafePrompt sits beside NeMo

Guardrails. Tax example: managing, extraction, classification, and reporting agents around a human practitioner.

Derivation. Dual stack: Microsoft Copilot/Foundry for the office/orchestration plane; NVIDIA for models, guardrails, and inference; LangChain Deep Agents in the 2026 blueprint. ChatGPT is one model option among many.

Official example. Lab results with NVIDIA: +20% task-success in smart logistics; +30% retrieval accuracy in sustainability regulatory intelligence.

Controversy. Industry-wide hallucination reports have named EY among peers. EY's answer is control rings and NeMo—not a public product recall on the scale of Deloitte's government refunds. The thinner public employee/customer record is itself data: scandal, not architecture diagrams, is what puts consulting AI on the front page.

6. PwC — Agent OS and PwC One

What / history. Agent OS launched about March 2025 as an enterprise command center; MCP support May 2025; AWS-native services July 2025. PwC One is the knowledge/workflow UX. PwC was OpenAI's first ChatGPT Enterprise reseller.

Architecture. Patent-pending recursive, graph-based orchestration: small workflows compose into larger ones. A language-state machine lets non-engineers describe state transitions in natural language and pick different LLMs per state. MCP registers tools as servers so any authorized agent can reuse them with auth, policy, and logs. RBAC via Microsoft Graph (User / Superuser / Admin). Connectors: Anthropic, AWS Bedrock (Nova, Claude, Mistral, Llama), Azure, Google, OpenAI, Oracle, Salesforce, SAP, Workday, CrewAI, LangGraph.

Derivation. Closest to an Anthropic-protocol plus multi-model OS. ChatGPT is one engine; Claude and Bedrock models are first-class. Copilot is a possible *agent*, not the bus.

Official example. Claim: assemble multi-agent workflows “up to 10× faster” than custom integration. Internal stats previously cited: 250-plus agents, 12,000-plus GPTs, 31 million-plus GenAI interactions.

Controversy, previewed. Australia tax-confidentiality scandal (2023) and later reports with broken citations show generation without verification and walls without teeth, while Agent OS is marketed as governance at the protocol layer. Developed in Lived Case 6.

7. Booz Allen Hamilton — aiSEMBLE and Vellox

What / history. aiSEMBLE is a lean-manufacturing AI engineering system (DataOps → ModelOps → DevSecOps), open-sourced as aiSEMBLE Baseline April 2024 (Java, Python, Maven, Kubernetes/Helm). Vellox (RSAC 2026) is an agentic cyber product line, not a general consulting chatbot.

Architecture.

aiSEMBLE: reference architecture plus generated pipelines/containers so teams do not rebuild MLOps. Modular so tools can be swapped. Drift/bias/provenance hooks.

Vellox: specialized agents—**Reverser** (malware analysis in minutes), **Ranger** (autonomous detection engineering), **Striker** (adversary emulation), **Navigator** (live compliance), **Responder** (pre-detection containment), **LayerOne** (compliance gateway that validates agent actions before they become “truth”). Models trained on elite-operator tradecraft.

Derivation. Engineering lineage is open-source ML platforms plus Kubernetes. Inference may use AWS Bedrock or other foundation models. Mission models are proprietary.

Official example. U.S. federal/DoD cyber and intelligence; toolkit 2024; Vellox 2026; positioned as closing the “speed gap” versus AI-enabled attackers.

Controversy, previewed. Contractor billing ethics and data-handling cases sit under the factory. Vellox’s product risk is autonomous cyber *action* (Responder) if LayerOne fails. Developed in Lived Case 9.

8. KPMG — Workbench, Clara AI, TaxBot

What / history. Workbench launched mid-2025 with Microsoft as the global multi-agent fabric under Clara (audit), Digital Gateway (tax), and Velocity (advisory). About 50 production assistants, ~1,000 in development. 2026: Microsoft Agent 365 for fleet governance plus M365 Copilot. TaxBot used a ~100-page prompt to draft 25-page advice.

Architecture. Azure AI Foundry plus Copilot Studio (low-code) plus Semantic Kernel evolving to Microsoft Agent Framework (Semantic Kernel + AutoGen). Agents as digital teammates with A2A-style communication on Foundry. Clara AI: grounding, Bing Search, agents for substantive testing and documentation. Leadership plans agent swarms that run parallel tasks and compare outputs. Multi-LLM by design despite a Microsoft-centric control plane.

Derivation. The most Microsoft Copilot / Foundry / Agent 365-native of the ten. TaxBot is classic long-prompt plus RAG, not a new model.

Official example. Microsoft case study on Clara + Workbench for faster risk detection in audit. TaxBot: multi-week advice compressed to hours, with human review.

Controversy, previewed. 2026 Australia client-data leak between bid teams (CEO and audit head resigned) is a human confidentiality failure sitting next to a platform that advertises “trusted” agents. Developed in Lived Case 7 and Teaching Case B.

9. Bain — Sage and the three-layer agentic platform

What / history. Global OpenAI alliance February 2023; Sage (GPT-4 chat over Bain IP) plus about 11 other tools rolled out through 2023. Later thought leadership specifies a three-layer production architecture. OpenAI Dev Day highlighted Bain’s eval strategy and a code-modernization agent.

Architecture (published target state).

Governance/trust (identity, prompt-injection filters, content safety, registries).

Orchestration: workflow engines, MCP tool adapters, agent registry, A2A in-workflow handoffs, session plus long-term memory.

Applications (Sage, Aura workforce agents, PDLC/SDLC multi-agent product-development platform).

Sage itself is still the 2023 pattern: ChatGPT-like UI plus proprietary retrieval. Aura is a multi-agent data collector/analyst over global workforce data. Bain reported a 25% efficiency gain in methodology via OpenAI Evals.

Derivation. The cleanest ChatGPT/OpenAI line in the set (GPT-4 Sage, OpenAI AgentKit/Evals, Codex). MCP/A2A added as industry standards, not Bain inventions.

Official example. Shelf-image custom GPT: 50-plus store photos analyzed in under a minute versus hours. America's Cup digital-twin sailing agent (2024).

Controversy, previewed. SARS / South Africa state-capture work predates Sage and remains the firm's defining governance scar. Architecturally Bain now argues Phase-1 governance must exist *before* multi-agent orchestration. Developed in Lived Case 8.

10. IBM Consulting — watsonx Orchestrate + Consulting/Enterprise Advantage

What / history. watsonx is IBM's enterprise AI platform; Orchestrate is the agent runtime and, by 2026, an agentic control plane. Consulting Advantage began as IBM's internal delivery fabric and was productized as Enterprise Advantage for clients (including FedRAMP on AWS). IBM Bob is the agentic developer partner.

Architecture.

Each agent = LLM + instructions + tools + collaborators + optional knowledge base.

Primary agent evaluates handoff conditions, invokes collaborators, synthesizes the answer.

Reasoning modes: ReAct, Flows, collaborators.

Agentic workflows as graphs: start/end nodes, branch, foreach, HITL, nested flows (async).

Control plane: directory, identity, policy, pre-deploy evals, tracing, cost control.

Import LangGraph and Langflow agents; speak A2A; coordinate SAP Joule with watsonx agents. Hybrid via Red Hat.

Derivation. IBM's own watsonx models and middleware, plus LangGraph and Google A2A. OpenAI/Anthropic models can sit behind the control plane. Closest analog to PwC Agent OS, with stronger hybrid/mainframe heritage (watsonx Code Assistant for Z).

Official example. Masters/Wimbledon fan assistants historically; Forrester TEI 370% ROI cited for watsonx Assistant. Enterprise Advantage: clients stand up the same agent platform IBM uses internally.

Controversy, previewed. Watson Health write-downs and ERP implementation lawsuits feed skepticism that the control plane will be better than past large-system delivery. Developed in Lived Case 10.

How the vendor genes actually show up

ChatGPT / OpenAI: Bain Sage, McKinsey Lilli (partial), PwC reseller plus many GPTs, BCG custom GPTs/GENE. Pattern: chat UX + rented GPT-4-class model + private corpus.

Claude / Anthropic: MCP as the tool bus (PwC first among this set to productize it hard); Claude as a Bedrock/Azure model choice; Deloitte large Claude workforce rollout.

Microsoft Copilot / Azure / Foundry / Agent 365: KPMG's home stack; EY's orchestration layer; Deloitte Azure OpenAI (including the Australian report); McKinsey hosting.

Amazon / AWS Bedrock + AgentCore: Deloitte MAS; Accenture Bedrock Agent in the Huddle; PwC Agent OS AWS edition.

Google (Gemini + A2A): Deloitte's 100-plus Gemini agents and field-service A2A demo with ServiceNow; Accenture Huddle; IBM Orchestrate A2A.

NVIDIA: Accenture Refinery/Distiller; EY intelligence and guardrails; Deloitte Atlas R&D.

Meta: Accenture lists Meta as a Huddle partner (Llama as a model option elsewhere).

Oracle: Integration surface, not the brain.

xAI: No documented core derivation at these ten.

Design fork that matters most. BCG and Bain publish *constrained* plans and phased governance. Accenture, PwC, IBM, EY, and KPMG publish *interoperable control planes*. McKinsey invests in *proprietary memory plus small expert models* on rented LLMs. Booz Allen splits an *open MLOps factory* from *closed cyber agents*. The remaining shared weakness—visible in Deloitte's and others' hallucinated citations—is that orchestration does not automatically equal grounded evidence when the “agent” is asked to write a report rather than execute a checked transaction.

Lived cases: employees, customers, whistleblowers, and courts

The vignettes below are reconstructed from named public sources—news investigations, court and commission records, firm statements, and first-person posts on Glassdoor, Reddit, and social media. Anonymous posts are labeled as such. They are not composite fiction.

Lived Case 1. The customer who read the footnotes: Deloitte, Australia, and a A\$440,000 welfare review

In December 2024 Australia's Department of Employment and Workplace Relations (DEWR) commissioned Deloitte to review the Targeted Compliance Framework—the IT system that penalizes jobseekers who miss obligations. The 237-page independent assurance review was published in July 2025. The contract was worth about A\$440,000 (about US\$290,000).

Chris Rudge, a University of Sydney researcher in health and welfare law, did what the report's citations invited a reader to do: he looked them up. He found fabricated references, including a nonexistent book attributed to Sydney constitutional-law professor Lisa Burton Crawford, phantom papers tied to Lund University's Björn Regnell, a misspelled "Justice Davis" (apparently Justice Jennifer Davies), and a fabricated quotation from nonexistent paragraphs of *Amato v Commonwealth*, the robodebt case. Rudge told reporters the document was "full of fabricated references" and later estimated up to twenty errors.

A corrected version was quietly re-uploaded. Quotes and phantom papers disappeared. In an appendix Deloitte disclosed that part of the methodology "included the use of a generative artificial intelligence (AI) large language model (Azure OpenAI GPT-4o) based tool chain licensed by DEWR and hosted on DEWR's Azure tenancy," used in part for "traceability and documentation gaps." Deloitte and the department insisted the substance and recommendations were unchanged. Deloitte agreed to repay the contract's final installment. A Labor senator called it a "human intelligence problem."

The architectural irony is sharp. The same global firm markets Zora as a digital workforce, Ascend as a control platform with a Trust Score, and multi-agent systems on Amazon Bedrock AgentCore. The failure mode here was not an unsupervised Perform agent posting a transaction. It was an Advise-class document: generation without verification, then a methodology note after the press arrived. For a government customer, the story is simple: we paid professional-services rates for footnotes a graduate student could falsify in an afternoon.

Lived Case 2. The second country: Deloitte Canada and a C\$1.6 million health plan

Months later the pattern repeated in Newfoundland and Labrador. A 526-page health human-resources plan, paid at C\$1,598,485 over 2023–2025, contained false citations, papers that real researchers said they had never written together, and authors attached to work they did not do. *The Independent* and CBC reported the errors. Deloitte Canada said it stood behind the recommendations, that "AI was not used to write the report," and that AI was "selectively used to support a small number of research citations." The firm said it would issue citation corrections that "do not impact the report findings."

Yvette Coffey, president of the Registered Nurses' Union of Newfoundland and Labrador, asked why a firm "that has twice faced public scrutiny and embarrassment for how it uses AI" was still running an ongoing core staffing review. By March 2026 the province had

rewritten procurement rules: vendors must declare intended AI use; the government may audit that use.

Customer voice here is institutional rather than memoir: a nurses' union, a provincial opposition, and a small-province newsroom doing the verification work that a C\$1.6 million contractor's control plane did not.

Lived Case 3. Inside the huddle: Accenture employees, Hertz, and promotion-by-login

Employees tell a shop-floor story that the Huddle press release does not. In February 2026 the *Financial Times* reported that Accenture had begun collecting weekly log-in data on key internal AI tools—including AI Refinery and SynOps—for associate directors and senior managers, and that “use of our key tools will be a visible input to talent discussions.” People familiar with the change called some tools “broken slop generators”; one said they would “quit immediately” if the rule applied to them. The firm has publicly trained hundreds of thousands of staff in generative AI and tied continued employment, in CEO messaging, to adaptation.

Anonymous first-person posts match the pressure, if not the product name. A consultant who joined in January 2025 wrote of midnight stand-ups, weekend work that could not be logged as overtime, “+1” firm-citizenship tasks stacked on a full client load, and colleagues who shared the grievance but would not escalate it up “the power chain.” Glassdoor reviews recycle the consulting constants: bench anxiety, project-dependent hours, strong learning, weak rest. Reddit threads on Accenture’s “use AI or leave” messaging include client-side comments that a Big Four team delivered “AI slop” that later had to be ripped out.

The Hertz litigation is older than the agent stack and still the canonical customer complaint about Accenture delivery. In 2016 Hertz hired Accenture to rebuild Hertz.com and mobile apps. Hertz later sued in the Southern District of New York, alleging that after more than US\$32 million in fees Accenture never delivered a functional site or app; missed successive go-lives; produced non-responsive, brand-locked, insecure front-end code that had to be scrapped; and then asked for another US\$10 million to finish. Accenture called the suit without merit and sought dismissal and unpaid invoices. The case is about implementation quality, not Huddle. It is the story customers reach for when a firm promises an agent OS and they remember a website that never shipped.

Lived Case 4. Lilli in the room: McKinsey consultants, juniors, and two older scars

Employee voices in official channels are glowing. Executive assistant Susan Reyes: she no longer starts from zero. Associate partner Adi Pradhan uses Lilli to hunt weaknesses before meetings and save about 20% of prep time. Partner language treats Lilli as a colleague in the room: “What does Lilli think?” Recruitment blogs now describe final-round exercises in which candidates solve cases *with* Lilli—testing structure, not bot-wrestling.

Unofficial voices are more anxious about the apprenticeship. Reddit and industry coverage note that Lilli drafts outlines, charts, and “McKinsey tone of voice,” and that leadership has said technology can do work typically performed by junior employees, including decks.

Headcount fell from about 45,000 (2023) toward 40,000 while the firm rolled out thousands of agents. Commenters ask a succession question the control plane does not answer: if juniors no longer build the slides, where do seniors who can still read a bad slide come from?

Two pre-Lilli cases remain the firm's public moral ledger. On Purdue Pharma, the U.S. Department of Justice announced in December 2024 that McKinsey would pay US\$650 million to resolve criminal and civil investigations of advice to "turbocharge" OxyContin sales. A former senior partner, Martin Elling, pleaded guilty to destroying records after emailing that the firm should consider "eliminating all our documents and emails." McKinsey apologized, accepted a deferred prosecution agreement, and later agreed to contribute a further US\$125 million into Purdue's bankruptcy estate. More than two dozen partners had worked the account over about fifteen years; Purdue paid the firm about US\$93 million.

In South Africa, McKinsey's Eskom and Transnet work during state capture produced illegal or irregular contracts, Gupta-linked intermediaries (Regiments, Trillian), repaid fees (including about R1 billion to Eskom and R650 million to Transnet and SAA), and later U.S. resolutions describing bribery by former executive Vikas Sagar. The firm says it never served the Gupta family, terminated Trillian after failed due diligence, and is "deeply remorseful." Commissions and prosecutors have been less gentle. Lilli did not design those engagements. They are why critics hear "trusted orchestration" as a brand claim rather than a completed reform.

Lived Case 5. Constrained agents, unconstrained engagement: BCG, AgentKit, and Gaza

On r/consulting, BCG staff greeted a corporate "employee joy" headline with sarcasm: AI is a useful tool; joy is a press release.

The 2024–2025 Gaza work showed that tool discipline is not engagement discipline. The *Financial Times*, *Wall Street Journal*, and others reconstructed Project Aurora: more than a dozen staff, more than US\$4 million of contracted work over seven months, help standing up the U.S.- and Israel-backed Gaza Humanitarian Foundation, and a postwar financial model that costed "relocation packages" of about US\$9,000 for more than 500,000 Palestinians (on the order of US\$5 billion), with commentary that leaving could be cheaper per person than staying through reconstruction. BCG said senior leaders were misled, that a lead partner was "categorically told no" and violated the directive, that it disavows the reconstruction modeling, that two partners were fired, that invoices were cancelled, and that a WilmerHale review found approval processes circumvented—including Signal chats that were not retained. It refused to publish the review. Save the Children suspended a long-standing pro-bono partnership. UK parliament asked questions. The GHF aid sites were already under intense criticism for civilian deaths.

No public evidence ties AgentKit's code to that modeling. The case is governance: a firm that open-sources constrained agents still depends on partners who can open a side channel.

Lived Case 6. “Whatever it takes”: PwC Australia, then Agent OS

The Australian tax-confidentiality scandal is the customer and citizen counter-narrative to Agent OS. Peter Collins, then a senior international-tax partner, advised Treasury on multinational anti-avoidance law while confidentiality acknowledgments were in force. Internal emails later released in parliament showed confidential timing and design detail moving to colleagues who used it to win “brand defining” work, including with U.S. technology companies. The Tax Practitioners Board deregistered Collins. Treasury referred matters to police. CEO Tom Seymour resigned. Eight further partners were removed. Government consulting was sold. Clients froze work. An external review described a “whatever it takes” culture and conflicts that the firm had failed to mitigate because it was designing the law and selling around it at the same time. PwC Australia later admitted clients beyond Google received leaked information.

Agent OS did not cause that breach. The breach is why “MCP plus logs” is a necessary but insufficient answer. Logs do not create a Chinese wall.

Lived Case 7. The locker and the leak: KPMG Australia

A whistleblower emailed audit leadership on 30 May 2024 alleging that confidential Lendlease (and later Optus) material was used to pitch audit work at Dexus, Westpac, Telstra, and others; one allegation described client board papers kept in a Sydney locker. Internal reviews initially found the claims unsubstantiated. Senator Deborah O’Neill put them on the parliamentary record in March 2026. CEO Andrew Yates and audit chief Julian McPherson resigned, saying processes “fell short.” Chairman Martin Sheppard and partners Paul Rogers and Eileen Hoggett later left. Allens and further reviews confirmed misuse of confidential files, including rival firms’ pitch documents that had been submitted to a client board. Government clients demanded assurances. ASIC opened a preliminary probe into named auditors. Hoggett later sued after expulsion from the partnership. New leadership called the failings “indefensible.” Parliament heard that “many, many more” whistleblowers had come forward.

Workbench cannot be blamed for a locker. Clara cannot be credited with catching it. The case is the oldest professional-services risk—misuse of client knowledge to win the next mandate—arriving in the same year the firm advertised an agent fabric.

Lived Case 8. Bain, Sage, and SARS

The Zondo Commission found Bain’s SARS work a clear example of private-sector collusion in state capture: managing partner Vittorio Massone’s meetings with Jacob Zuma; the arrival of Tom Moyane; restructuring that hollowed a capable revenue service. Bain repaid SARS fees plus interest in 2018, commissioned Baker McKenzie, and denies knowing participation in capture. Nugent described symbiotic but not identical interests: Moyane wanted control; Bain wanted fees. The UK banned Bain from government contracts for three years for “grave professional misconduct.” National Treasury imposed a long public-sector ban. In 2025 Bain shut its South African consulting business, citing the lasting stigma.

Sage did not design SARS. SARS is why Bain's own thought leadership now puts identity and injection filters in Layer 1.

Lived Case 9. Factory and fortress: Booz Allen employees, qui tam, and cyber agents

Sarah Feinberg, a former finance employee, filed a False Claims Act complaint after a manager allegedly called federal auditors "too stupid" to notice overcharging. In 2023 Booz Allen paid US\$377.45 million to settle allegations that commercial and international costs were improperly allocated onto government contracts from about 2011 to 2021. Feinberg's relator share was reported around US\$70 million. Separate matters include a 2025 GSA time-charging settlement (about US\$423,000) and a US\$15.9 million settlement tied to a subcontractor bribery scheme around Air Force simulator work. In a different register, Booz Allen contractor Charles Littlejohn pleaded guilty to leaking IRS data that underpinned ProPublica and *New York Times* reporting on wealthy taxpayers; in 2026 Treasury cancelled dozens of Booz Allen contracts. The firm said the crime occurred on government systems.

Vellox's stated risk, even in sympathetic readings, is the Responder agent: autonomous action if LayerOne fails. The billing cases are the culture risk sitting underneath the factory.

Lived Case 10. IBM: control plane after Watson Health

The dark customer memory is Watson Health: a moonshot that STAT and others documented as collapsing under over-claimed clinical impact, missed revenue, layoffs, and eventual sale. Older ERP implementation fights and a BIPA facial-scan class action over Flickr-derived geometry sit in the same file of "AI that was supposed to be the platform." None of that falsifies Orchestrate's engineering. It explains why buyers treat a new control plane as a delivery bet, not a completed redemption.

Lived Case 11. EY and the control-ring bet

EY has not had a single public refund spectacle on the Deloitte scale in this cycle. Industry hallucination round-ups have named it among peers. Its published answer is rings and guardrails. Until a comparable assurance report is dissected in public, the employee and customer record is thinner than the architecture deck.

What the cases rhyme

Report agents hallucinate; transaction agents are easier to test. Deloitte Australia and Newfoundland failed as document factories. Zora-style Perform agents with HITL on a ledger are a different reliability class. Marketing collapses the two.

Confidentiality failures are still human routing problems. PwC's Treasury emails, KPMG's locker and bid teams, Littlejohn at IRS systems, McKinsey document destruction on Purdue—none are MCP bugs. A control plane that logs tool calls does not log a partner forwarding a PDF.

Employee experience is acceleration plus surveillance plus thinner apprenticeship. Lilli and AgentKit save hours. Accenture counts log-ins for promotion. Juniors lose deck-

building as craft. Reddit and Glassdoor supply the affective register official “joy” surveys omit.

Old delivery and ethics cases are how customers will read new agent OS pitches.

Hertz, Watson Health, Eskom, SARS, Purdue, and FCA settlements become the interpretive frame for Huddle, Orchestrate, and Workbench—fairly or not.

Research cases

Use as 45–75 minute seminars or a four-session sequence. Through-line question: *If the firm now sells an agent operating system, what exactly is the product—the model calls, the control plane, or the professional duty that the control plane cannot execute?*

Research Case A. Deloitte, Azure OpenAI, and the footnotes the government paid for

Objective. Distinguish Perform agents from Advise agents. Map where RAG, HITL, and trust scores failed as *document* controls. Design a public-sector acceptance test that would have caught the errors before go-live. Judge whether a partial refund and an appendix disclosure constitute adequate remedy.

Decision point. You are chief knowledge officer. The minister wants, by Monday: (1) a technical root-cause that does not say only “the model hallucinated”; (2) a contracting clause other jurisdictions can copy; (3) whether Advise-class reports should be excluded from agent workflows until citation verification is a blocking gate.

Research questions.

Where, in a perceive–retrieve–generate–cite–sign-off pipeline, did control break? Name the human role that was missing.

If citations were load-bearing for legal interpretation of welfare penalties, is “recommendations unchanged” auditable? How would you test it?

The toolchain sat on the department’s Azure tenant. Who owned prompt logs, retrieval corpora, and liability?

Why is report generation a different risk class from a Perform agent posting a ledger entry with HITL?

Newfoundland required AI declaration and audit rights. Is disclosure enough, or must contracts specify allowed tasks?

What does it say about quality systems that they moved only after a university researcher and a newspaper acted?

AI note. Push the AIs off of their “hallucinations” to a Hidden Cost assessment. The operational failure is unverified generation released as assurance. Put a red gate on “every citation resolves to a live DOI or court reporter before PDF lock.”

ResearchCase B. KPMG Australia: the locker, the bid teams, and Clara

Objectives. Separate platform trust from institutional trust. Analyse why first-line investigations “unsubstantiated” a claim that later became a leadership wipeout. Design wall-crossing controls that do not depend on hero whistleblowers. Evaluate whether marketing “trusted agents” during an unresolved leak is itself an ethics issue.

Decision point. A bank GC will keep Clara for workpaper drafting but not the statutory audit unless you can show that bid teams cannot see incumbent-client board packs—*technically*, not culturally. Approve or reject “confidential corpus isolation” as a Workbench feature. Decide what you tell Microsoft, whose case study is still public.

Research Questions.

Is this an AI case at all? If no, why does it belong in this chapter?

Which failure is primary: incentive, investigation design, or shared knowledge lakes?

Could Clara or Agent 365 have detected locker-and-forward behaviour? What surveillance problem does that telemetry create?

What did the industry fail to learn from PwC Australia 2023?

How should a firm weigh Hoggett’s due-process suit against client demand for visible sacrifice?

Draft the one-paragraph client letter that is true and does not pretend a control plane solved a locker.

Research note. Leaders and organizations who say “the AI agents did it” are deflecting answerability.

Teaching Case C. BCG, AgentKit, and Project Aurora

Research objectives. Hold two facts at once: constrained agent design can be real *and* engagement governance can still fail. Specify what authorization means when work moves from pro bono aid design to paid reconstruction modelling. Judge transparency choices (fire partners, withhold the law-firm report). Build a political-conflict acceptance test that is not only a sanctions checklist.

Decision point. Write the one-page policy that would have stopped Aurora and still allow humanitarian operations work. Include: definition of relocation modelling; Signal/ephemeral-message rules; partner override of a CRO “no”; whether such work is ever billable.

Research questions.

AgentKit constrains tool sequences. Aurora is about which problem is allowed. Same design layer?

What evidence would you need to accept “seniors were deceived,” given seven months and a multimillion-dollar arc?

Should the WilmerHale review be published? Privilege versus public licence to operate.

How should NGO and university partners underwrite “we do not do population-transfer modelling”?

Is Aurora evidence for Bain’s claim that governance must precede multi-agent orchestration?

Where is the line between scenario planning and designing inducements to leave a territory? Who in a partnership is competent to draw it?

Research note. Finish the organizational analysis before the proxy war on the underlying conflict. Then allow the political-ethics layer. Both belong. Contrast the shipyard efficiency story with Aurora’s legitimacy question.

Research Case D. McKinsey: Lilli in the room, Purdue in the docket

Learning objectives. Analyse how an internal knowledge copilot changes apprenticeship and staffing leverage. Connect document-retention culture (Elling) to what a Lilli-class system logs—or deletes. Ask whether “we should not have taken the work” is a sufficient lesson for agent governance. Design a forbidden-work policy that binds agents as well as partners.

Decision point. Issue a two-part standard. Part A — apprenticeship: which tasks juniors must still perform unaided. Part B — forbidden work: categories that may not be scoped, retrieved, or drafted by Lilli even if a partner wants a “quick take” (opioids-class products, SOE capture risk, document-destruction scenarios).

Research Questions.

McKinsey says Lilli is not simple RAG. Does that change the Purdue lesson?

If juniors no longer build decks, where does quality judgement form?

What should Lilli retain by default, knowing traces can be discovery?

When is simultaneous learning from a health corpus and commercial pharma work a conflict even with tenancy walls?

Compare Accenture’s promotion-by-login with McKinsey’s “have you asked Lilli?” norm.

Why include Eskom/Transnet in an AI teaching case?

Note. On the Exercise: Give Part B more time than Part A. Firms love apprenticeship design; they under-write refusal design.

Cross-case Question to Consider

Question: “Across Deloitte’s footnotes, KPMG’s locker, BCG’s Aurora file, and McKinsey’s Purdue docket, the advertised remedy is always a better control plane (Trust Score, Agent 365, Action Plans, Lilli’s expert models). Using two cases, can you argue whether the next control plane should be understood as (a) a technical object that can be specified in a statement of work, or (b) a professional obligation that cannot be delegated to MCP, A2A, or HITL. Take a position.”

My Answer: The Control Plane Is Not the Duty

My reading: the next control plane should be understood as (b) a professional obligation that cannot be delegated to MCP, A2A, or HITL. A statement of work can specify logs, trust scores, Action Plans, and expert-routing models. Those artifacts are useful. They are not the remedy advertised after Deloitte’s footnotes or KPMG’s locker. What failed in both files was not a missing microservice. It was a refusal to treat verification and confidentiality as non-delegable professional acts. Treating the control plane as a purchasable technical object is how the industry keeps selling the same promise after the same class of failure.

This argument uses two cases that look, at first, like opposite problems. Deloitte Australia looks like an AI-generation defect. KPMG Australia looks like an old Chinese-wall defect. The marketing response in both neighborhoods is identical: a better plane (Ascend Trust Score, Bedrock supervisors, Workbench, Clara, Agent 365). If the same product category is offered as the fix for unlike failures, the category is doing ideological work. It converts duty into procurement.

1. Why the technical-object view is tempting—and why Deloitte seems to confirm it

A control plane, in the vendors’ own language, is registry, identity, policy, observability, and routing for many agents—not the agents themselves. That definition is already an invitation to write a statement of work. One can require: citation resolution before PDF lock; retrieval provenance on every footnote; model-card disclosure; tenant-bound logs; a named HITL signer; a Trust Score threshold. Deloitte’s 2025 DEWR review looks, superficially, like a missing clause of that kind.

The facts are now public. DEWR paid about A\$440,000 for a 237-page independent assurance review of the Targeted Compliance Framework. Chris Rudge found fabricated references, a phantom book attributed to Lisa Burton Crawford, nonexistent work tied to Björn Regnell, a misspelled judge, and an invented quotation from Amato v Commonwealth. The corrected text disclosed an Azure OpenAI GPT-4o toolchain on the department’s tenant, used in part for “traceability and documentation gaps.” Findings, the firm and the department said, were unchanged. The final instalment would be repaid.

If one stops there, position (a) writes itself. Specify a blocking citation gate. Ban generative fill of legal quotations. Require every authority to resolve to a reporter or DOI. Put the gate in Ascend. Score the run. Do not ship. The Newfoundland file—C\$1.6 million, selective AI “to

support a small number of research citations,” the same “recommendations unaffected” line—seems only to prove that the gate was still unwritten.

That reading is half right and wholly insufficient. The DEWR document was not an unsupervised Perform agent posting a ledger line. It was an Advise product sold as independent assurance. Someone chose to let a generation stack fill “gaps” in a public legal-adjacent review. Someone locked a PDF without reading the authorities the PDF cited. Someone, after the press arrived, treated an appendix disclosure and a partial refund as closure. None of those acts is a missing A2A handshake. They are decisions about what a professional is still required to do when a model is in the loop.

A statement of work can demand the gate. It cannot perform the reading. Rudge did the job the citation graph should have done because the graph, if it existed, was not treated as a condition of speaking in the state’s name. The technical object is downstream of that choice. Specify it, yes. Do not confuse the specification with the obligation that makes the specification binding when the partner is late, the client is political, and the deck is due Friday.

Newfoundland’s later procurement rewrite—declare AI use, accept audit—is position (a) as public law. It is better than silence. It still assumes that a declared toolchain plus an audit right will produce verified footnotes. Deloitte’s own distinction, in Canada, that “AI was not used to write the report” but was used on citations, shows how a SOW object will be lawyered: the model touched only the “small” part. Duty does not admit that partition. If the citation is load-bearing for a welfare penalty or a staffing plan, the citation is the work.

2. KPMG’s locker: when the advertised plane cannot see the failure

KPMG’s 2025–2026 product story is Workbench on Microsoft Foundry, Clara in audit, TaxBot-style long prompts, Agent 365 for fleet governance, swarms that compare outputs. The advertised remedy is again a plane: grounding, logs, digital teammates, trusted agents.

The 2024–2026 Australian file is not a grounding bug. A whistleblower emailed audit leadership on 30 May 2024 that confidential Lendlease—and later Optus—material was used to pitch other mandates. One allegation put client board papers in a Sydney locker. Internal processes first found the claims unsubstantiated. Parliament forced the matter into view in March 2026. The CEO, the audit head, the chair, and named partners left. External review confirmed misuse, including rival firms’ pitch packs that had sat with a client board. ASIC opened a preliminary probe. A former COO sued after expulsion. New leadership called the failings indefensible.

Ask the technical-object question in this file and it goes empty. Which MCP server should have stopped a locker? Which Agent 365 policy detects a partner walking paper from an incumbent audit into a competitor bid? Clara can draft a workpaper. It cannot invent a wall that compensation and origination credit have already demolished. PwC Australia had already taught the sector this lesson in 2023: Treasury confidentiality acknowledgments, emails celebrating “brand defining” wins, a “whatever it takes” review, a sold government

practice. MCP arrived afterward as the advertised fix. Logs do not create a Chinese wall. They record the call after the PDF has already moved.

Here position (a) becomes actively misleading. If the next control plane is a thing one buys, KPMG can say—truthfully—that Workbench did not cause the locker, and Clara cannot be blamed for missing it. That sentence is the point. The plane is scoped so that the constitutive failure of an audit firm (using one client’s knowledge to win another) sits outside the product. Duty cannot sit outside the product. The licence to call an agent “trusted” is the same licence that used to attach to a partner’s signature. You do not get to keep the adjective after you have relocated the noun.

A serious SOW can still help: separate vector indexes per engagement; cryptographic denial for bid teams; immutable access logs to the regulator; no shared origination credit across the wall. Those are technical expressions of an obligation that must already exist in compensation, investigation design, and the willingness to lose a tender. KPMG’s first two years of handling—whistleblower as HR problem, claims unsubstantiated—were not a missing API. They were a decision about whose discomfort mattered. Agent 365 will not reverse that decision. It will log a cleaner version of it.

3. Why “HITL,” “MCP,” and “A2A” cannot carry the obligation

Human-in-the-loop is the industry’s favorite alibi because it sounds like duty. In both cases HITL was, in a thin sense, present. People issued the DEWR PDF. People ran KPMG’s first investigations. HITL failed as a moral technology for the same reason dashboards fail: the human is in the loop as a throughput function, not as a reader or a wall. A loop that signs what it has not checked is not oversight. It is laundering.

MCP standardizes how agents call tools. A2A standardizes how agents call agents. Neither standardizes which problems may be opened. That is why BCG can open-source anti-ReAct Action Plans and still produce Project Aurora, and why McKinsey can build Lilli’s expert models and still carry a Purdue deferred prosecution in which a partner discussed eliminating documents. Those files are not needed as primary cases here; they are the surrounding evidence that the same category error repeats. Constrained tool sequences are not constrained engagements. A knowledge OS that accelerates synthesis also accelerates the old temptation to treat the file as disposable. Elling’s mail is what a logging plane is supposed to make unthinkable. It is also what a logging plane will not prevent if the partnership’s prior move is to make the record go away.

So the technical object is not nothing. It is the wrong kind of thing to be the remedy. Specify it in the SOW the way one specifies encryption at rest: mandatory, testable, insufficient. The sufficient thing is older and less vendable. For Advise work: no citation ships unless a named person has resolved it, and that name survives FOI. For client knowledge: no corpus, locker, or inbox is shared across a wall, and the investigation of a breach is not an HR file. For both: the firm accepts that some work will be slower, some tenders lost, some agents left unused, because speed is how duty is currently being priced out.

4. Position, without the hedge that undoes it

The next control plane should be written into contracts. Citation gates, tenant isolation, bid-team denial, retention that survives a partner's delete key—put them in the SOW. Then stop calling that the remedy.

Deloitte's footnotes show that a generation stack will invent a judge if someone lets "documentation gaps" be filled in an assurance review. KPMG's locker shows that a swarm of audit agents will not notice a board pack changing rooms. In both files the advertised next object is a better plane. That advertisement is how professional services converts a non-delegable act—reading the authority, keeping the confidence—into a feature list.

The control plane is a technical object only after it is a professional obligation. MCP, A2A, HITL, Trust Score, Agent 365, and Lilli's expert models can implement the obligation. They cannot be it. If the industry keeps selling them as it, the next chapter will again be written by someone who simply checked the footnotes, or by a senator who read an email that two years of internal process had filed under personality.

That is the position. Specify the plane. Do not delegate the duty.

Conclusion

The ten firms have, in three years, standardized an enterprise agent stack that would have sounded like science fiction in 2021: rented frontier models, proprietary memory, supervisors, specialist workers, MCP and A2A as handshake protocols, and a control plane that aspires to be an operating system rather than a chatbot. That convergence is real. So is the competitive differentiation that remains: BCG's constrained plans versus Accenture's multi-vendor huddle; McKinsey's closed corpus and tiny expert models versus PwC's protocol-first bus; Booz Allen's split between an open MLOps factory and closed cyber agents; KPMG's Microsoft-native swarm; Bain's OpenAI-first lineage with an explicit governance-first sequencing learned the hard way.

What has not converged is the production of trustworthy public knowledge and the protection of other people's secrets. The most expensive failures in this chapter were not missing NVIDIA NIMs. They were missing readers. Chris Rudge did the job a citation graph should have done. A Newfoundland newsroom and a nurses' union did the job a C\$1.6 million methodology should have done. A KPMG whistleblower and a senator did the job an ethics wall should have done. Courts and commissions did the job internal risk committees deferred on Purdue, Eskom, and SARS.

For employees, the inside story is double-edged. Lilli, Sidekick, PairD, Sage, Clara, and Refinery remove blank-page terror and compress search. They also invite headcount math, promotion metrics that measure compliance theater, and a generation of analysts who may never learn why a slide is wrong by building the wrong slide. For customers, especially governments, the inside story is that "agent OS" language can coexist with Azure GPT-4o footnotes that invent a judge.

The design implication is unromantic. Treat generation-heavy Advise work as a high-risk pathway that requires retrieval provenance, citation verification as a blocking test, and named human sign-off that survives a freedom-of-information request. Treat Perform work as a different pathway, with narrower tools and HITL on irreversible actions. Treat confidentiality as a socio-technical control—separate identities, separate tenants, separate incentives—not as a banner on an A2A partner slide. Treat the historical docket (opioids, state capture, tax leaks, False Claims, failed websites) as part of the architecture review, because buyers already do.

Big-time AI consulting is no longer a prompt window. It is an operating system bolted onto partnerships that still fail in the old ways. The stack will keep converging. The question this chapter leaves open is whether verification and professional duty converge with it, or whether the next chapter is written again by a researcher who simply checked the footnotes.

Table 1

Term	Meaning
Agentic AI	Systems that plan, use tools, and act toward a goal—not just answer a prompt.
LLM / FM	Large language model / foundation model (GPT, Claude, Gemini, Llama, Nemotron, and others).
RAG	Retrieval-augmented generation: retrieve documents, then generate an answer grounded in them.
ReAct	Reason + Act loop: the model freely chooses thoughts and tool calls. Powerful but unreliable.
Supervisor / primary / meta / super agent	An orchestrator that decomposes work and delegates to specialist agents.
MAS	Multi-agent system: several agents with roles, shared context, and handoffs.
MCP	Model Context Protocol (Anthropic, November 2024): standard way for agents to call tools, files, and APIs.
A2A	Agent-to-Agent protocol (Google, April 2025; Linux Foundation): standard for agents from different vendors to delegate work.
Guardrails	Policy filters (NVIDIA NeMo Guardrails is a common implementation) that constrain prompts and outputs.
Control plane / Agent OS	Registry, identity, policy, observability, and routing for many agents—not the agents themselves.
HITL	Human-in-the-loop review before an agent's action is final.

Table 2

Firm	Named architecture	Pattern	Primary derivation	Openness
Accenture	AI Refinery + Distiller + Trusted Agent Huddle	Super-agents + multi-vendor huddle	NVIDIA AI Enterprise + A2A/MCP + cloud/SaaS agents	High interoperability
McKinsey	Lilli / LilliX	Knowledge orchestration + small expert models	Azure OpenAI + Cohere; not plain RAG	Closed corpus, rented models
BCG	AgentKit + GENE + Deckster + custom GPTs	Constrained Action Plans (anti-ReAct)	LangChain + ChatGPT Enterprise / GPT-4o	Open-sourced AgentKit
Deloitte	Zora + Ascend + MAS	Perform vs Advise; Bedrock supervisor	AWS Bedrock AgentCore, Azure OpenAI, NVIDIA, Gemini/A2A	Multi-cloud
EY	EY.ai Agentic “OS”	Data / orchestration / intelligence + control rings	NVIDIA NIM/NeMo/Nemotron + Microsoft Foundry/Copilot	Portable, sovereign
PwC	Agent OS + PwC One	Recursive graph orchestration + MCP	Multi-LLM + MCP	Vendor-agnostic OS
Booz Allen	aiSEMBLE + Vellox	MLOps factory + domain cyber agents	Open-source Java/Python/K8s; operator-trained models	Open baseline; closed cyber products
KPMG	Workbench + Clara AI	Swarm of specialists on Azure	Microsoft Foundry, Copilot Studio, Agent 365	Microsoft-centric, multi-LLM
Bain	Sage + 3-layer platform	GPT-4 knowledge chat → MCP/A2A	OpenAI GPT-4 / AgentKit / Evals	OpenAI-first
IBM	watsonx Orchestrate + Consulting/Enterprise Advantage	Control plane over native + external agents	watsonx + LangGraph + A2A; Red Hat hybrid	Multi-framework

Table 3

Session	Case	Core tension	Best for
---------	------	--------------	----------

1	Deloitte Australia / DEWR	Agent stack vs. unverified generation in public reports	AI governance, public-sector procurement
2	KPMG Australia locker	“Trusted agents” beside human confidentiality failure	Professional ethics, audit independence
3	BCG Aurora / GHF	Constrained tools vs. unconstrained engagements	Risk, political work, partner autonomy
4	McKinsey Lilli + Purdue	Productive internal AI beside a criminal-civil docket	Apprenticeship, historic liability, dual use

Chapter 4 How Big Time Firms Use AI Internally

Summary

This chapter asks a narrower question than Chapter 3. Chapter 3 mapped the *architectures* the ten firms sell: rented foundation models, proprietary memory, orchestration planes, worker agents, MCP/A2A handshakes, and governance wrappers. Chapter 4 asks how those same firms *use* AI on themselves—research, slides, staffing, IT operations, audit workpapers, tax drafts—and how that internal use is then productized for clients as strategy, agents, process automation, industry solutions, and “responsible AI.”

The internal pattern is now standardized. Almost every firm first built a knowledge-grounded chatbot plus slide and research accelerators (Lilli, Sage, GENE, Deckster, PwC One, EYQ, Clara). It then moved to multi-agent orchestration with hyperscaler partners (Refinery/Huddle, Agent OS, Workbench, watsonx Orchestrate, EY.ai’s wedding cake, Zora/Ascend). Consultants are told to “drink their own champagne”: use the stack, log the hours saved, then sell the pattern. Accenture’s AATA claims VPN setup falling from thirty minutes to two. McKinsey reports ~30% time saved on search and synthesis and 500,000-plus Lilli prompts a month. EY cites 96%+ internal adoption after a \$1.4 billion program. PwC cites hundreds of thousands of training hours and twelve thousand custom GPTs.

The client pattern is less standardized than the slide. Customers still buy senior-led sales and receive junior-heavy delivery, now with a copilot in the loop. Criticisms cluster in four places: price-to-impact (an Emergn/Censuswide survey of 700-plus transformation managers found 84% said MBB firms were “no help at all”); implementation systems that harm the public (Deloitte-run Medicaid eligibility platforms in at least twenty-five states generating wrong notices and lost coverage); refunds and fee fights (EY’s ~£15 million return to Santander UK on Project Morgan; BCG’s suit against GameStop for about \$30 million; Hertz’s suit against Accenture after \$32 million); and integrity scandals already developed in Chapter 3 (Purdue, SARS, PwC Treasury, KPMG locker, BCG Aurora).

Employee voice—Glassdoor, Reddit, workplace-medicine reports, named interviews—repeats a different ledger: extreme hours that AI did not shorten, “+1” citizenship tasks, bench risk, promotion politics, bait-and-switch staffing, and the feeling that prestige no longer covers the personal cost. Some Accenture staff describe oversold projects as making them feel like “scammers.” Some McKinsey alumni describe “capitalism incarnate.” Some Booz Allen staff describe pressure to bill and fear of retaliation. These are representative, not universal; offices and partners vary.

The chapter’s claim is that **internal AI is a labor-process story before it is a product story**. Time saved on synthesis is real. Whether that time becomes client impact, shorter hours, or higher utilization is a partnership decision, not a model property. Four research cases at the end—Deloitte Medicaid, EY/Santander, the MBB “no help” survey plus New York’s bins, and Accenture’s internal agents versus Hertz—are written so we can argue that decision with names, numbers, and documents. They do not repeat Chapter 3’s four cases (DEWR footnotes, KPMG locker, Aurora, Lilli + Purdue).

How to read this chapter against Chapter 3

Firm-by-firm: internal use, client use, customer voice, employee voice

Accenture

Named internal systems. AI Refinery (NVIDIA AI Enterprise); AI Navigator for use-case discovery; Trusted Agent Huddle for multi-vendor orchestration; AATA (Accenture Advanced Technology Agent) for IT operations; Edge agentic solutions with Google Cloud.

How it works internally. Refinery connects siloed models, knowledge stores, agents, and governance. AATA is the clearest *eat-your-own-cooking* story: more than 100 agents for ticket auto-remediation, VPN configuration, and chat troubleshooting. Claimed cuts include VPN setup from about 30 minutes to 2. Consultants are expected to scale patterns from 1,000-plus GenAI engagements. By 2026, weekly log-ins to Refinery and SynOps had become a visible input to leadership-promotion discussions for some senior grades—an adoption metric, not a quality metric.

Client use. Industry agents (contact-center assist, clinical-trial companion, hyper-personalization), sovereign/agentic programs in Europe, “reinvention” programs (Fincantieri), mid-market pre-built agents via Edge.

What customers have written and sued. Hertz remains the canonical delivery complaint. The 2019 Southern District of New York complaint alleged that after more than US\$32 million Accenture never delivered a functional website or mobile app; missed successive go-lives; produced non-responsive, brand-locked, insecure front-end code that had to be scrapped; and then asked for another US\$10 million. Accenture called the suit without merit. The case is pre-Refinery. Customers still reach for it when a firm promises an agent OS. Reddit client-side comments on later “use AI or leave” coverage include teams that said a Big Four delivery was “AI slop” later ripped out.

What employees have written. A consultant who joined in January 2025 described midnight stand-ups, weekend work that could not be logged as overtime, “+1” duties on a full client load, and colleagues who would not escalate up “the power chain.” People familiar with the 2026 log-in-to-promotion rule called some tools “broken slop generators”; one said they would quit immediately if the rule applied to them. France workplace-medicine reporting has cited overload and distrust of leadership promises. Glassdoor recycles bench anxiety and project-dependent hours. The contradiction the firm does not resolve: AATA saves two minutes on a VPN while the consultant is still online at midnight.

McKinsey

Named internal systems. Lilli (and LilliX); Tone of Voice style agent; dual surface for general chat versus Client Capabilities over 100,000-plus documents.

How it works internally. RAG plus five smaller expert models plus a PowerPoint parser that rose from ~15% to >85% slide coverage. Users get five to seven sources, summaries, expert pointers, synthesis. Firm figures: 70–75%+ monthly use, ~17 queries per user per week, 500,000-plus prompts per month, ~30% time saved on search/synthesis. Official voices: Susan Reyes no longer starts from zero; Adi Pradhan saves ~20% of meeting prep. Partners ask “What does Lilli think?” Some recruiting loops now case *with* Lilli.

Client use. Faster insight activation, research, analysis, brainstorming, slide generation so consultants “spend more time with clients.” QuantumBlack productizes the pattern. The firm also sells GenAI strategy.

What customers have written. An Emergn-commissioned Censuswide survey of 700-plus UK and US executives and project managers in firms with 1,000+ employees found that among those who had used MBB on transformations, **84% said the firms were “no help at all,”** 3% said they hindered work, and 13% were net-positive. Alex Adamopoulos called the model “big consultancy armies with briefcases” fitting clients to a predesigned mould. Separately, New York City paid McKinsey US\$1.6 million (on a contract with a US\$4 million ceiling) for a 20-week sanitation study informing containerization; DSNY later said McKinsey “did math” and “did not determine or recommend policy.” Social media priced the deck at tens of thousands of dollars a slide and asked whether the city needed consultants to discover the wheelie bin. Purdue and Eskom remain the moral docket from Chapter 3.

What employees have written. *The Guardian’s* “capitalism incarnate” reporting collected alumni accounts of extreme hours, partners who belittle mistakes, team dinners from which people secretly Uber home to keep working, and a willingness to follow the money into ethically fraught files. Reddit asks where seniors who can still read a bad slide will come from if Lilli drafts the first deck. Headcount compression beside agent rollout is the labor arithmetic underneath the 30% time-save.

BCG

Named internal systems. GENE (GPT-4o chatbot over research and interviews); Deckster (template-trained slides with a grader); Ava (IT/HR/admin); AgentKit; tens of thousands of custom GPTs; Team Builder and PRIME/ChatPDG for staffing and reviews.

How it works internally. Enterprise ChatGPT plus constrained Action Plans. Deckster has been used hundreds of thousands of times. Forward-deployed consultants build client tools that feed an internal marketplace. On r/consulting, staff greeted a corporate “employee joy” headline with sarcasm: the tool is useful; joy is a press release.

Client use. Agentic R&D (public shipyard case: 45% fewer engineering resources, 80% shorter lead time per deck), sales/RFP automation, pricing, supply-chain simulation, frontline scheduling.

What customers have written and sued. Same MBB survey. In 2022 BCG sued GameStop in Delaware for about US\$30 million in unpaid transformation fees; GameStop said paying was not “in our stockholders’ best interests” given “seemingly meager impact” and that it no longer used “the likes of BCG.” A judge later trimmed some claims but left a breach theory standing. Gaza/Aurora (Chapter 3) damaged morale and recruiting conversations; Save the Children paused a partnership. WHO-contract and Australian government-extension criticisms sit in the same “cosy public sector” file.

What employees have written. During slowdowns, juniors asked seniors “do you have work?” and often heard no. Promotions slowed; snacks and travel were cut. High custom-GPT volume did not cancel political infighting or hours.

Deloitte

Named internal systems. Ascend (delivery/PM platform with Trust Scores); Zora AI (finance/ops digital workforce); Quartz Atlas (NVIDIA R&D); Enterprise AI Navigator; 10X Analyst; MAS on AWS Bedrock; large internal Claude rollout. Employees on Reddit name Sidekick / PairD as daily copilots.

How it works internally. Ascend orchestrates people, assets, and agents. Navigator modules identify agent-ready tasks, model financial impact, design workflows, and generate agents. Zora executes source-to-pay and finance processes. The firm’s public story is cycle-time cuts when agents sit inside ERP/CRM (HPE finance; Toyota exploration).

Client use. Finance transformation, R&D simulation, agentic SDLC, CFO “AI Advantage”—and, at industrial scale, **state Medicaid and human-services eligibility systems.**

What customers and beneficiaries have written and sued. KFF Health News documented Deloitte as the dominant eligibility-system contractor: at least 25 states, contracts in the billions, 53 million Medicaid enrollees in those states at one measurement point. Systems generated incorrect notices, sent paperwork to wrong addresses, froze for hours, and took months or years to patch. Kentucky’s 2016 launch sent at least 25,000 erroneous letters. Tennessee beneficiaries filed a class action alleging the Deloitte-built

system “does not reliably test for eligibility” for several disability categories. In Michigan, contracts since 2006 total roughly US\$768 million; the Bridges system has misdirected disabled adults into skimpier plans or denied coverage because it did not register disability or miscounted earnings. One beneficiary, Noon, told KFF she could not afford care after the system ignored disability rules that count only half of earned income. Senators Wyden, Warren, Warnock, and Sanders pressed Deloitte and peer contractors after the investigation. Chapter 3’s DEWR and Newfoundland citation failures are the *document* twin of this *system* failure.

What employees have written. Disorganized integration teams, inconsistent direction, benching after a GenAI project that used libraries released weeks earlier, late stand-ups treated as firing offenses. Typical Big 4 hours and “ChatGPT + junior” delivery complaints.

EY

Named internal systems. EY.ai Agentic Platform (NVIDIA + Microsoft); EYQ; FlexiGenAI; Canvas AI (audit knowledge from ~85,000 professionals); Workforce digital workers; PDL “8090.” Reported 96%+ internal adoption after ~US\$1.4 billion.

How it works internally. Data foundation, orchestration, model catalog, NeMo and SafePrompt rings. 150-plus tax agents around 80,000 professionals at launch; later claims of tens of thousands of agents. The slogan is drink their own champagne, then productize.

Client use. Tax at scale, Canvas-embedded audit, financial-crime and risk agents, claimed 70% productivity / 80× speed in one software-delivery story, HR automation.

What customers have written. Santander UK terminated EY’s anti-financial-crime work on **Project Morgan** after the project “went badly wrong over an extended period.” EY refunded about **£15 million**—on the order of 19 UK partners’ average annual pay—and left the account. The bank was already under FCA pressure (a £108 million AML fine for 2012–2017 shortcomings). A third-party support-ticket breach separately exposed client tax data. Independence and audit-quality fights continue in litigation. The champagne story and the Morgan story are the same firm in the same decade.

What employees have written. Client bullying with firm support for the payer; bench then push-to-resign; missed internal support; standard hours and politics.

PwC

Named internal systems. PwC One; Agent OS (MCP switchboard across clouds and apps); 250-plus internal agents; 12,000-plus custom GPTs; 460,000-plus U.S. training hours. First ChatGPT Enterprise reseller.

How it works internally. Methods and standards embedded in workflows; multi-model routing; FinOps baked in.

Client use. Diligence, tax analysis, finance rebuilds (Lucid forecasting claimed from weeks to under a minute), F1 operations, contact-center agents, assurance over clients’ AI.

What customers have written. Chapter 3's Treasury confidentiality file remains the citizen counter-narrative. UK audit fines (Babcock among them) and China-client aftershocks after Evergrande sit beside "creating work for themselves." Fake or broken citations in some regional reports have been flagged as the same hallucination class as Deloitte's public PDFs.

What employees have written. Billing concerns that become dismissal claims; client-side descriptions of teams that "constructively undermine" organizations to extend engagements; high day rates with thin output in some reviews.

Booz Allen Hamilton

Named internal systems. aiSSEMBLE (open MLOps factory); Vellox (Reverser, Ranger, Striker, Navigator, Responder, LayerOne); Bighorn edge kits; Bedrock data-processing; Shield AI partnership.

How it works internally. Standardized pipelines for federal environments; operator-trained cyber models; disconnected edge labeling.

Client use. DoD and intelligence cyber defense, autonomy, air-gapped data prep. Number-one federal AI-services positioning.

What customers have written. Program offices described as demanding; DOGE-era cuts; Treasury cancellation of contracts after the Littlejohn IRS-leak case (Chapter 3). The Feinberg False Claims settlement (US\$377.45 million) is the billing ledger under the factory.

What employees have written. Bench risk when contracts end; clients treated as more important than staff; alleged pressure to bill hours not worked; weak ethics follow-through; retaliation fears; "Machiavellian" promotion politics beside "good people and flexibility."

KPMG

Named internal systems. Workbench; Clara AI; TaxBot (~100-page prompt, 25-page draft); Ignite; Digital Gateway; Velocity; Control Tower; ISO 42001 / AIUC-1 badges on some capabilities.

How it works internally. Multi-LLM RAG on Azure; auditors supposed to spend more time on judgment; staff build agents in Copilot Studio.

Client use. Audit speed, tax advice, ESG policy lifecycle, GBS orchestration, contract intelligence.

What customers have written. Chapter 3's locker-and-bid file. Historical audit-failure citations in bank-collapse reviews. High fees versus quality.

What employees have written. The whistleblower's "revenue growth at all costs." Standard hours and partnership politics.

Bain

Named internal systems. Sage; ~12 GenAI tools; Aura workforce platform; agentic PDLC/SDLC; every employee on ChatGPT Enterprise; internal GPT “Olympics.”

How it works internally. Sage synthesizes firm IP in seconds. Aura cleans and benchmarks workforce data. Product-development orchestration across 40-plus PDLC activities. Claimed ~25% methodology gain via OpenAI Evals.

Client use. Workforce heatmaps, faster software delivery (~30% claimed), shelf-image GPTs, America’s Cup digital twins, manufacturing complexity reduction.

What customers have written. Same MBB survey. SARS/Zondo (Chapter 3) plus the historic “get deep into the trousers” style that antagonizes client managers. Private-equity (Bain Capital) complaints are sometimes wrongly fused with the consulting firm.

What employees have written. Collegial-but-demanding; downturn staffing and promotion pressure; pride in results; opacity and the SARS stain.

IBM Consulting

Named internal systems. watsonx Orchestrate; Consulting/Enterprise Advantage (internal fabric now sold to clients, including FedRAMP paths); IBM Bob; Context Studio.

How it works internally. Teams design/build/deploy with industry agents; Orchestrate coordinates multi-vendor agents and SAP Joule; hybrid Red Hat.

Client use. Code assistants including mainframe; watsonx Assistant (Forrester TEI 370% ROI cited); Gemini industry agents; Pearson, Providence, Grand Slam fan experiences.

What customers have written. ERP upgrades that missed scope, dates, and cost; cloud-migration buyer’s remorse in IBM/HFS research; Watson Health’s over-claimed clinical impact and sale. “Consulting” sometimes read as relabeled delivery staff.

What employees have written. Fear-based leadership; extra manager layers; blocked mobility; tiny bonuses despite high ratings; visa/travel friction; back-to-office backlash; “take the IBM out of Kyndryl” from spin-off alumni.

Fully developed case examples (Chapter 4 files)

These are reconstructed from named reporting, complaints, and first-person public posts. They are not composite fiction. They are *not* the Chapter 3 research and cases quartet.

Case E. Noon’s coverage and 25,000 letters: Deloitte’s eligibility machines

Noon, a working disabled adult in Michigan, was eligible for Medicaid under a rule that counts only half of earned income so that modest wages do not strip medical coverage. The Bridges system—Deloitte-operated under contracts totaling about US\$768 million since 2006—did not register her as disabled and treated her earnings as disqualifying. She told

KFF Health News she could not afford care. A disability-rights lawyer had to reconstruct what the system should have known.

Scale turns one person into a market. Deloitte holds eligibility work in at least 25 states. KFF found wrong notices, wrong addresses, frozen portals, patches that take months, and change-request catalogs measured in tens of thousands of hours (Georgia: 35 change requests, more than 104,000 hours in one year). Kentucky's 2016 launch sent at least 25,000 automated letters telling people they would lose benefits. Tennessee's class action said the system "does not reliably test for eligibility" for several disability categories. Senators later asked whether error-riddled platforms plus new work-requirement rules would become a "bureaucratic maze" that drops eligible people.

Internally, the same firm sells Zora as a digital workforce and Ascend as a Trust-Scored delivery plane. The eligibility cases are not hallucinations in a PDF. They are **production systems that decide whether a person has insulin**. Internal AI that accelerates a consultant's slide does not touch that failure mode unless the firm treats safety-net logic as a tested Perform path with beneficiary-level HITL—not a change-request farm.

Case F. Project Morgan: EY drinks the champagne, Santander sends it back

Santander UK, already fined £108 million by the FCA for AML shortcomings (2012–2017), hired EY among other advisers to rebuild financial-crime defenses. The work was called Project Morgan. People familiar with the matter told the *Financial Times* it "went badly wrong over an extended period." EY terminated the contract and refunded about £15 million—roughly 19 UK partners' average pay. The hit landed on the UK financial-services division. Cutbacks in a ~150-person fincrime team were discussed. EY declined to comment. Santander spoke only of its own £700 million-plus change program.

In the same era EY announced EY.ai with NVIDIA and Microsoft, 96%+ internal adoption, tax agents around 80,000 professionals, and the line that the firm drinks its own champagne. Morgan is what a client writes when the champagne is corked: we paid for defenses; we received work so poor the supplier paid us to leave.

The employee shadow is the team that staffed Morgan and then faced the shrinkage. Internal FlexiGenAI does not appear in the FT account. The lesson for Chapter 4 is brutal and simple: **internal adoption metrics do not predict client-system quality**, especially in regulated control functions where "agentic" is a slogan and evidence is a SAR that never filed.

Case G. Eighty-four percent "no help," and the million-dollar bin

Censuswide, for Emergn, asked 702 transformation executives and project managers in large UK and US firms about MBB on change programs. Eighty-four percent of those who had used McKinsey, BCG, or Bain said the firms were "no help at all." Three percent said they hindered. Thirteen percent were net-positive. Half the broader sample had considered

quitting from “transformation fatigue.” The commissioner of the study called the model armies with briefcases and a predesigned mould.

That is a customer chorus, not a court finding. It is still the loudest quantitative client voice in this chapter. It lands beside McKinsey’s internal dashboard: 30% time saved, 500,000 prompts, Lilli in the room. If internal AI makes the army faster at producing the mould, the survey predicts *more* of what 84% already called no help—unless the surplus is used to shrink the army or change the mould.

New York’s trash study made the same math vernacular. A contract ceiling of US\$4 million; US\$1.6 million paid; twenty weeks; international comparators; volume math that appears on a handful of pages of the city’s *Future of Trash* report. DSNY said McKinsey did not set policy. Social media and columnists asked whether discovering containerization required that rate. The bin became a meme because it made the price-to-insight ratio visible to people who will never see a Trust Score.

Case H. GameStop’s stockholders versus BCG’s invoice

In August 2019 GameStop hired BCG for a transformation with profit-improvement targets baked into fees. After the meme-stock year and a management change, GameStop stopped paying. BCG sued in Delaware for about US\$30 million, alleging refusal to attend validation meetings and to furnish data needed to compute fees. GameStop’s public line: paying “tens of millions” was not in stockholders’ best interests given “seemingly meager impact,” and the company was “proud” it no longer used “the likes of BCG.”

This is the fee-fight twin of the 84% survey. If AI compresses the hours inside a transformation, who keeps the surplus—the partnership’s realization rate, or the client’s refusal to pay for meager impact? AgentKit and Deckster make slides cheaper to produce. They do not settle the invoice.

Case I. Accenture’s two-minute VPN and the midnight stand-up

AATA’s claimed 30-to-2-minute VPN cut is the cleanest internal-ROI anecdote in the set. The 2026 *Financial Times* log-in rule is the cleanest adoption-surveillance anecdote. The January 2025 joiner’s social post is the cleanest hour-sheet anecdote. Hertz is the cleanest customer-delivery anecdote.

Put on one page: the firm can automate its own service desk and still run a website program that never ships; it can count Refinery log-ins for promotion and still not count overtime; it can train 550,000 people in GenAI and still have staff call the tools slop generators. Internal AI here is not failing to exist. It is failing to renegotiate the labor contract it pretends to have already rewritten.

Cross-cutting patterns

Internally, the ten firms converged on **chat-plus-corpus, then agents-plus-plane**, with hyperscaler partners and RAI wrappers. Client offerings emphasize agentic process change and industry IP more than raw model novelty.

Customer frustration clusters around **price-to-impact, the partner-sells / junior-plus-AI-delivers staffing model, failed or harmful implementations** (eligibility, AML, websites, ERPs), and **integrity scandals** (Chapter 3). Employee frustration clusters around **hours and utilization, politics, bench risk, and whether prestige still matches cost**. Experiences vary by office, partner, and project.

The labor-process claim again: time saved is a fact only if someone is allowed to stop working. Utilization models usually convert the saving into more prompts per week.

Research appendix (new cases — not Chapter 3)

Use as 45–75 minute seminars. Through-line: *When a firm uses AI on itself, who collects the time—the consultant’s sleep, the client’s fee, or the partnership’s margin?*

Research Case E. Deloitte, Bridges, and Noon

Objectives. Distinguish internal delivery planes from public eligibility engines. Design an acceptance test a disabled beneficiary could survive. Decide whether “change requests” are a business model or a defect.

Research question. Deloitte’s Ascend Trust Score rates prompt results and decision trails on consulting work. Michigan’s Bridges system misclassified a working disabled adult and dropped or diverted coverage. **Should eligibility logic be treated as an Advise agent (analysis, then human notice) or a Perform agent (transaction that must be reversible within 24 hours)? Defend a testing regime either way.**

My Answer. Treat it as **Perform with beneficiary-level reversibility**. Eligibility is not a slide. A wrong denial is an irreversible medical event on the timescale that matters to Noon. Advise-style “we’ll patch the notice next quarter” is how Kentucky mailed 25,000 letters and how change-request catalogs become 104,000-hour products. A defensible regime: (1) disability-rule unit tests (half-income disregard, category flags) as blocking gates before promotion to prod; (2) canary populations and parallel run against known-eligible panels; (3) same-week human restoration without requiring a lawyer; (4) global patch, not state-by-state one-offs; (5) public error-rate dashboards, not only internal Trust Scores. Internal Zora/Ascend metrics are irrelevant unless they include *false-denial rate*.

Research Case F. EY.ai and Project Morgan

Objectives. Separate internal adoption percentages from regulated-control outcomes. Read a refund as a customer document. Ask what “champagne” means if the client pours it out.

Research question. EY reports 96%+ internal AI adoption and productizes tax and financial-crime agents. Santander took a ~£15 million refund and ended Project Morgan

because the work “went badly wrong over an extended period.” **Which three contractual clauses would have surfaced the failure before year two—and why would internal EYQ usage data not have been one of them?**

Mt Answer. Useful clauses: (1) **outcome SLAs tied to regulator-recognizable controls** (alert-quality, SAR timeliness, quality-assurance sampling)—not hours or decks; (2) **step-in and shadow-run rights** so the bank can compare EY output to an independent sample monthly; (3) **refund/termination triggers on failed control testing**, not on “relationship review.” EYQ usage is the wrong instrument because it measures consultant interaction with a firm LLM, not whether Morgan’s AML design actually risk-rated customers. Adoption can rise while control quality falls. The FCA fine already told both parties what “good” had to mean; the SOW needed to cite that standard, not champagne.

Research Case G. The 84% and the bin

Objectives. Read a competitor-commissioned survey without discarding it. Connect internal productivity claims to client value claims. Price an insight.

Research question. McKinsey reports ~30% time saved on search/synthesis via Lilli. Emergn’s survey says 84% of transformation managers found MBB “no help at all.” NYC paid US\$1.6 million for sanitation math. **Assume both the 30% and the 84% are directionally true. Where did the saved time go? Write the transfer identity (who gained hours, who gained fees, who gained outcomes) and pick one policy a partnership could adopt so the 30% could move the 84%.**

My Answer. Transfer identity: saved analyst hours → (a) more workstreams per engagement (utilization), (b) unchanged partner realization, (c) unchanged client outcome, (d) possible headcount reduction. The 84% is compatible with (a)+(b)+(c). The bin is (c) made visible: math was delivered; “use a bin” was already available as civic knowledge; policy ownership stayed with DSNY. One policy that could move the 84%: **price a transformation on verified operating KPIs at +6 and +12 months, with a clawback**, and forbid billing the Lilli-saved hours as additional scope. That forces the 30% toward (c) or toward a smaller invoice—not toward a larger army with briefcases.

Research Case H. Two minutes, thirty-two million, midnight

Objectives. Hold internal ROI and client delivery and labor process on one page. See promotion-by-login as a control plane aimed at employees.

Research question. AATA claims VPN setup fell from 30 minutes to 2. Hertz alleges US\$32 million and no working site. A 2025 joiner describes midnight work that cannot be logged. Promotion committees will see Refinery log-ins. **Design a single internal dashboard that would be honest about all three facts. Which metric must be banned?**

Answer. Honest dashboard, three panes: (1) **internal toil** (ticket cycle time—keep the two-minute VPN); (2) **client outcomes** (scope delivered / defects / go-live vs. contract—Hertz would have been red); (3) **labor** (hours actually worked, overtime denied, after-hours pages—midnight would be red). Ban **log-ins as a promotion input**. Log-ins measure

compliance theater; they reward slop-generator sessions and punish people who already finished the work in another tool. Pair (1) with (3) or the firm will automate the service desk and lengthen the consultant's night.

Cross-chapter questions (uses Chapter 3 + 4)

Question: "Chapter 3 argued that a control plane cannot be the professional duty. Chapter 4 argues that internal time-saved cannot be the client outcome. Using **one Chapter 3 case** and **one Chapter 4 case**, show how a firm could satisfy both advertised metrics (Trust Score / Lilli usage *and* 30% faster synthesis) while still producing the failure the case documents. Then name the single discarded metric that would have revealed the failure."

My Answer Example pairing: Chapter 3 DEWR footnotes + Chapter 4 Medicaid Bridges. Satisfied metrics: Azure toolchain used (adoption), Ascend-style trails on the consulting PDF (plane), analyst hours down on citation gathering (30%). Failures still produced: invented judge; Noon denied. Discarded metric in both: **external correctness against a ground-truth set**—live authorities for the PDF; known-eligible disability panels for Bridges—checked before release, not after KFF or Rudge.

Conclusion

Big-time firms now use AI internally the way they once used libraries and offshore slide factories: to compress the middle of the job. That compression is real. Lilli, Sage, GENE, Deckster, Clara, EYQ, PwC One, AATA, and Orchestrate take blank-page time out of research, drafting, ticketing, and workpapers. Adoption numbers are no longer speculative. Neither is the next sentence. **The partnership still decides whether the recovered minutes become rest, a smaller team, a cheaper invoice, or another billing code.**

Customers, in this chapter's documents, have already voted with surveys, refunds, and suits. Eighty-four percent of transformation managers in one large sample called MBB no help. Santander took £15 million back. GameStop refused a US\$30 million bill as not in stockholders' interest. Hertz alleged a website that never shipped. Medicaid beneficiaries and their lawyers described systems that generated the wrong letter and the wrong denial. Those are not architecture diagrams. They are verdicts on whether internal copilots changed the mould or only filled it faster.

Employees have voted with exits, workplace-medicine files, and anonymous posts. The two-minute VPN and the midnight stand-up can coexist. Promotion-by-login can coexist with tools staff call slop. "Drink the champagne" can coexist with a fincrime team facing cuts after Morgan. AI did not invent utilization. It gave utilization a dashboard.

The design implication matches Chapter 3 and tightens it. Chapter 3: do not delegate duty to MCP. Chapter 4: do not delegate value to usage analytics. Put beneficiary-level error rates on eligibility engines. Put regulator-recognizable control tests on AML programs. Put clawbacks on transformation outcomes. Put hours-actually-worked next to minutes-saved. Ban log-ins as virtue. Price the surplus in public, or clients will price it in court and on Censuswide.

Internal AI is already an operating system for the firm’s own labor. The open question is whether it becomes an operating system for other people’s coverage, cash, and time—or remains a faster way to staff the army with briefcases.

Table 1

Chapter 3 asked	Chapter 4 asks
What stack did they build?	Who uses it, on which work, at whose expense?
Control plane vs. professional duty	Productivity theater vs. shorter week vs. thinner team
Advise-document hallucinations	Safety-net systems and AML programs that “go live”
Partner governance of political files	Fee fights: who captures the surplus AI creates?

Table 2

Session	Case	Core tension
1	Deloitte eligibility systems	Internal Trust Scores vs. beneficiaries cut off by production systems
2	EY / Santander Project Morgan	96% internal adoption vs. a £15 million refund
3	MBB 84% + NYC bins	Internal 30% time-save vs. client “no help at all”
4	Accenture AATA / Hertz / log-ins	Minutes saved on VPN vs. midnight hours and a dead website

Chapter 5 What Happens to Recruits Accenture SEAM

Summary

This chapter treats recruiting as a researchable archive, not a careers brochure. Across ten firms, the talent story has three dated layers. In the founders’ generation, intake was personal and thin: James O. McKinsey hired accountants and engineers from Chicago; Marvin Bower imported the Harvard Law/HBS professional; Edwin Booz came out of Northwestern; Arthur Andersen built a systems practice that later became Andersen Consulting’s St. Charles boot camp; the audit partnerships hired from state flagships and a handful of elites. In the 1990s the model hardened into what Lauren Rivera later called outsourced screening: core and target campus lists, case interviews, “up or out,” MBA pipelines from HBS/Wharton/Stanford/INSEAD/LBS, and Andersen’s clone factory at St. Charles. After the generative-AI break (late 2022–2026), the same prestige schools still dominate MBB offer lists, but the filter now includes AI fluency, copiloted case rounds (Lilli), and a simultaneous downsizing of the roles those campuses were built to fill.

Layoff files in the AI era are uneven but patterned. McKinsey fell from about 45,100 (end-2023) to about 40,000, mixing Project Magnolia back-office cuts (7%) when civilian federal work slowed under DOGE, a demand shock that is not the same as an AI shock but collides with it.

What happens to people who do not make the AI transition is rarely a public memoir. The researchable traces are: compressed-timeline exits; “counseled to leave” after a concerns rating; support roles offshored to Tampa, Costa Rica, Poland, the Philippines, Argentina; severance plus outplacement; and a white-collar market that, in 2025–2026, was already absorbing Big Four and federal-contractor surplus. Commonalities across the ten stories: prestige-school first screen; up-or-out as the quiet layoff machine; pandemic overhire then correction; AI used as both hiring criterion and headcount rationale; juniors and coordinators hit before partners; recruiting rhetoric of “human judgment” issued in the same quarter as agent headcount targets.

The chapter’s claim is that recruiting is the apprenticeship model under inventory management. Founders professionalized advice. The 1990s scaled a school-branded inventory. The AI era is trying to keep the school brand while shrinking the inventory that brand was designed to produce.

Table 1. Three recruiting regimes: founders’ door, target-school machine, post-AI inventory

Founders’ days: who walked in the door

McKinsey. James O. McKinsey, University of Chicago accounting professor, opened the firm in 1926 as an accounting-and-engineering shop. After his death in 1937 the offices split. Marvin Bower—Brown undergraduate, Harvard Law 1928, Harvard Business School 1930—rebuilt the New York firm in 1939 on a professional analogy to law. HBS leadership notes that Bower pioneered hiring business-school graduates and used up-or-out to grow partners. The founding recruit was not “anyone from anywhere.” It was Bower’s own credential stack, recommended toward McKinsey by an HBS dean. The school was the search firm.

Booz Allen. Edwin G. Booz took a 1914 Northwestern path into what became management engineering for government and industry. Postwar, Business Week paired Booz with Bower as the two people most responsible for the field’s growth. Federal work later pulled engineers, area specialists, and cleared staff more than HBS generalists—an intake DNA that still distinguishes Booz from MBB.

BCG and Bain. Bruce Henderson left Arthur D. Little to found BCG in 1963; Bill Bain left BCG in 1973. The Boston firms professionalized a different recruit: the analytically branded MBA and the undergraduate from Penn, Yale, MIT, Michigan, Berkeley. They did not invent campus lists; they intensified them. Harvard remains BCG’s largest single reported undergraduate source in some placement tallies; Bain’s lists often show unusual Berkeley weight.

Andersen → Accenture. Arthur Andersen's systems work for GE (1950s) became Method/1 and, after the 1989 split, Andersen Consulting. The 1992 New York Times portrait of St. Charles is the founder-adjacent document for this lineage: 4,000 raw college hires a year through a 150-acre boot camp that produced "clones." That is industrial recruiting, not Bower's partnership of twelve. The 2001 rebrand to Accenture kept the factory logic at larger scale.

Audit partnerships (Deloitte, PwC, EY, KPMG). Founder-era intake was articulated clerks and campus accounting majors from state universities plus a few elites. Partnership, not the case interview, was the long filter. Strategy boutiques inside the Big Four (Parthenon, Strategy&, Monitor after acquisition) later borrowed MBB lists.

IBM Consulting. Not a founder shop in the Bower sense. The consulting lineage runs through IBM's services businesses and the 2002 purchase of PwC Consulting (Business Consulting Services). Recruiting mixed IBMers, acquired PwC consultants, and later campus STEM.

Researchable primary set for this layer. Bower obituaries and HBS leadership profiles; Duff McDonald, The Firm; NYT St. Charles (6 Sept 1992); firm founding pages; Northwestern/Booz histories.

The 1990s machine: target schools as outsourced HR

By the 1990s the ten firms were no longer hiring "a bright person James knew." They were running school lists. Rivera's later HBR account of elite professional-service recruiting describes the logic consultants themselves used: Harvard or Dartmouth had already done two-thirds of the screening; hunting diamonds at Bowling Green was "not worth the investment." One MBA recruiter said 95% of intake came from the top five business schools.

MBB undergraduate cores (stable into the 2020s). Harvard and Penn/Wharton stand at the top of multi-year placement counts (150-plus MBB hires over a seven-year window in one compilation). Next tier: Yale, Michigan, Stanford, MIT, Duke, Princeton, Northwestern, Berkeley, Dartmouth. Semi-targets fill office-specific needs (Georgetown, UVA, Notre Dame, UT Austin, USC). UK/EU lists: Oxford, Cambridge, LSE, Imperial, INSEAD, LBS, HEC, Bocconi, St. Gallen.

MBB MBA cores. HBS, Stanford GSB, Wharton, Booth, Kellogg, Columbia, MIT Sloan, plus Tuck, Darden, Fuqua, Ross, Yale SOM, INSEAD, LBS. McKinsey's on-campus MBA list is typically the narrowest; Bain's the broadest. HBS remains the largest single global MBB feeder in careers-office reporting. Columbia's 2024 class still sent more people to McKinsey than to BCG and Bain combined.

Big Four and Accenture. Wider nets: business schools outside M7, large state flagships, engineering campuses, and, for Accenture, high-volume analyst programs. St. Charles-style acculturation substituted for HBS polish. Starting GPAs in careers guides sit lower than MBB's informal 3.7 screen.

What blogs and forums actually said. Wall Street Oasis, Reddit r/consulting, and Management Consulted turned the list into folklore: “target vs. non-target,” coffee chats, and the claim that a philosophy degree from Harvard beats a business degree from a non-target. That folklore is data about beliefs that structured applications, not a census of every hire.

What “not making it” meant then. Up-or-out. Associates who missed the manager bar left for industry, PE, or another firm. The 1990s did not need an AI narrative to shrink a class. The partnership bar did it every year.

After the AI break: same campuses, different inventory

Recruiting lists in 2025–2026 guides still read like 1998 with extra STEM. What changed is volume, filter, and the job behind the offer.

Still the list. Veris Insights (2025) found 26% of surveyed employers recruiting exclusively from a shortlist of schools, up from 17% in 2022—elite campuses back at the top after a brief pandemic widening. MBB cores did not dissolve.

New screens. McKinsey’s Lilli appears in some final-round exercises: candidates case with the firm bot. Accenture and others treat generative-AI fundamentals as table stakes. Experienced AI/data hiring rose while generalist BA classes in some offices quietly shrank. Careers coaches now sell “AI + case” prep the way they once sold PEI stories.

The offer is a thinner pyramid. Sternfels’s “25 squared” (grow client-facing ~25%, shrink non-client-facing ~25%) and the public agent counts (on the order of tens of thousands of agents beside ~40,000 humans) are the inventory statement. Kate Smaje’s remarks that technology can do work “typically performed by junior employees,” including decks, are the apprenticeship statement. Recruiting still photographs the Harvard coffee chat. The staffing model behind the photo is a smaller analyst row.

Firm-level recruitment and layoff files (researchable cases)

McKinsey

Schools, then and now. Bower’s HBS/HLS path became a global HBS-heavy MBA machine plus a tight undergraduate target list (Penn, Yale, Michigan, Stanford, Duke, Princeton among the strongest). International: INSEAD, LBS, Oxbridge, LSE.

AI-era headcount. About 45,100 (end-2023) to about 40,000. Components, per FT, Business Insider, and firm statements: ~1,400 back-office cuts from 2023 (Project Magnolia); 400 specialists (data/software) dismissed; February 2024 “concerns” ratings to ~3,000 people with a short window to improve or be counseled out; some roles not backfilled; ~200 global tech jobs cut later; discussion of further ~10% cuts in non-client-facing units over 18–24 months. Leadership simultaneously said consultant hiring continues. Agent counts in public remarks moved from thousands toward a one-to-one aspiration with humans.

People who do not transition. The researchable mechanism is performance process plus non-replacement, not a single pink-slip day. A former consultant told Business Insider he

lost the job in 2024 amid overhire. Support work moved toward Tampa, Costa Rica, Poland. Alumni blogs describe the concerns letter as the old up-or-out wearing a utilization-and-AI costume.

Accenture

Schools. Founder-adjacent: Andersen's mass college intake and St. Charles. Today: large engineering and business campuses worldwide; less dependent on M7 than MBB; 779,000-person scale makes "target school" a weaker exclusive.

AI-era headcount. Workforce 791,000 to 779,000 in three months to August 2025 (~11,000). \$865 million optimization; \$615 million plus \$250 million severance-type charges. Julie Sweet: exit on a compressed timeline where reskilling is not viable; 550,000 trained in gen-AI fundamentals; 77,000 skilled AI/data professionals (from 40,000). Australia office cited down from a 2023 peak near 6,600 toward 5,400. Later: promotion to leadership for some senior grades tied to regular AI-tool adoption. Sweet: "Some people will not make the journey."

People who do not transition. Named corporate policy, not rumor. Reddit's reaction to "learn AI or get out" mixed client-side "AI slop" stories with staff fear. The researchable aftermath is severance inside the \$865 million program and a job market already full of other firms' advisory surplus. Headcount was still forecast to grow overall—replacement hiring in AI skills, not a shrinking firm.

BCG and Bain

Schools. BCG: large international campus machine; Harvard often the top undergraduate source in compiled shares; MIT, LSE, Penn, Yale, Northwestern. Bain: Berkeley-heavy relative to peers; Penn, Michigan, Notre Dame, Harvard; broadest MBA target list of the three.

AI-era headcount. BCG, in 2024–2025 reporting, grew revenue (~10% to \$13.5 billion in one year cited) and added on the order of 1,000 people to ~33,000—the contrast FT used against McKinsey's shrink. Bain's public footprint is smaller; SARS-era bans (Chapter 3) affected government recruiting more than campus lists. Slowdown stories on forums: juniors asking "do you have work?" and hearing no; delayed promotions.

People who do not transition. Classic up-or-out plus staffing drought. Gaza-period morale (Chapter 3) entered recruiting conversations; that is a values filter, not an AI filter.

Deloitte, EY, PwC, KPMG

Schools. Broader than MBB: Big 10 / Russell Group / Go8 / IIT-adjacent engineering; accounting majors; S&O/Parthenon closer to MBB lists. Campus presence is industrial.

AI-era cuts (illustrative, not a complete census).

Deloitte: modest U.S. actions on low attrition; UK advisory cuts (hundreds) repeated across 13 months; consulting growth collapsing from ~19% to ~2% in one reported year.

EY: first global headcount decline in 14 years (~2,450 in one annual report); 100-plus U.S. EA roles moved toward the Caribbean; earlier multi-thousand consulting-adjacent cuts.

PwC: first U.S. layoffs since 2009 (~1,800 in 2024, advisory/tech); ~1,500 more in 2025 in audit/tax on “historically low attrition”; ~600 EAs, recruiters, and support in one 2026-adjacent round; recruiting seats moved toward Buenos Aires and Tampa.

KPMG: multiple U.S. rounds 2023–2025 including ~330 then ~195 audit staff; unusual partner-level audit cuts (on the order of 10% of U.S. audit partners in one reconstruction); Americas headcount down thousands.

People who do not transition. Letters from practice leaders calling it “incredibly tough” after mobility options failed. Outplacement. A white-collar market Revelio and others measured as already elevated in Big Four exits. AI is the accompanying story; the cited story is often low attrition after overhire. Both can be true.

Booz Allen Hamilton

Schools. STEM, area studies, military/veteran pipelines, cleared talent; less HBS theater, more SCI briefings. Northwestern remains brand genealogy.

AI-era / demand-era cuts. 7% of ~35,800–36,000) announced May 2025, concentrated in civil federal work after DOGE-related slowdowns and contract reviews. Customer-facing staff down sequentially. Later quarters added more civil reductions and a \$150 million cost takeout. This is primarily a buyer-shock case, not Sweet’s reskill-or-exit case. It still hits the same recruiting pipeline: people hired onto large civilian IT programs who cannot be redeployed at “normal dynamism” when procurements freeze.

People who do not transition. Redeployment failed; RIF. Forums describe bench risk as the preexisting condition.

IBM Consulting

Schools. Technical campuses, acquired PwC Consulting cohorts, internal IBM mobility.

AI-era. Arvind Krishna: hundreds of HR jobs replaced by agents; hiring slowed for back-office roles he thought 30% replaceable over five years; later a low-single-digit-percent global cut while hiring into AI, quantum, programming, sales, and campus grads. Mohamad Ali: digital workers beside ~150,000 consultants; org charts flattening. Internal claim: most HR queries already AI-handled; HR budget down sharply.

People who do not transition. Rebadged, offshored, or exited from the “non-growth” column. Consulting leadership’s public line is that unprepared rival firms will implode—an externalization of the same inventory logic.

What happens to people who do not make the AI transition

Researchable outcomes, not invented diaries:

Compressed-timeline exit (Accenture's words). Severance inside a disclosed restructuring charge.

Counseled out after a short performance window (McKinsey concerns). The 1990s technology, accelerated.

Role deleted, person not retrained. EAs, coordinators, some recruiters, some HR. Work goes to an agent or to Tampa / Philippines / Argentina / Caribbean.

Redeployment failure (Booz civil; some Big Four advisory). The skill is still "consultant," but the buyer left.

Stay and stall. No promotion because log-ins or AI fluency are now inputs (Accenture senior grades). Quiet attrition.

Labor-market absorption. Industry, smaller consultancies, government, startups, or a long search in a 2025–2026 white-collar slump. Public first-person posts rarely include a neat second act.

What is not well documented: large-scale, firm-funded transitions into genuinely new occupations at former pay. The archive shows exit language and replacement hiring, not guaranteed passage.

Commonalities in the researchable stories

The university is still the first algorithm. Founders used deans; the 1990s used lists; 2026 still uses lists, now with a tighter post-DEI shortlist in the broader employer survey.

Up-or-out predates ChatGPT. AI did not invent counseling-out. It gave it a skills plot.

Pandemic overhire is the denominator. Two-thirds headcount growth at McKinsey in five years to 2023; Big Four campus classes sized for 15–20% consulting growth that became 2%. AI arrived as the story that made the correction sound like strategy.

Juniors and coordinators before partners—except when partners go. KPMG's audit-partner cuts are the exception that proves the usual rule.

Two layoff species are routinely fused in press. (A) Reskill-or-exit / agent substitution. (B) Client-demand collapse (federal civil, discretionary advisory). Both hit recruits. They are not the same independent variable.

Recruiting copy and staffing math diverge. "We want human judgment and creativity" is issued beside agent-to-human ratio targets and smaller project teams (14 analysts → 2–3 plus agents, in one widely circulated reconstruction).

Apprenticeship is the unsolved design. If decks and first-pass research are agent work, the BA hired from Penn is being hired for a craft the firm is simultaneously automating. Chapter 4's labor-process claim is the recruiting claim.

Voice lives on forums, not in official histories. WSO, Reddit, Glassdoor, BI anonymous sourcing, and Senate/earnings-call transcripts are the primary qualitative record. Treat them as situated, not as census.

Four researchable case examples (for follow-on work)

Case 1 — Bower's Harvard and Sternfels's agents. Compare HBS class employment reports (share into McKinsey, median consulting pay) across 1995 / 2015 / 2024 with firm-wide headcount and agent counts. Hypothesis: offer prestige stays high while expected years-to-partner and analyst-class size fall.

Case 2 — St. Charles to "exit on a compressed timeline." Read the 1992 NYT clone-camp piece against Sweet's 2025 earnings call and the 2026 promotion-by-login reporting. Same firm lineage. Intake was once mass and standardized; exit is now mass and skill-coded.

Case 3 — PwC's 2009-to-2024 silence broken. First U.S. layoff since 2009; then a second round because attrition was "historically low." Pair with EA/recruiter moves to Tampa and Buenos Aires. Hypothesis: AI is the accompanying soundtrack; the mechanism is utilization math after overhire.

Case 4 — Booz's 7% and the cleared pipeline. Campus STEM + veteran hiring into civilian IT programs that DOGE then slowed. Distinct from Accenture's AI-fluency exit. Tests whether "AI-era layoff" is being used as a catch-all for buyer shock.

Followup questions (with answers)

Q1. Why do MBB target-school lists look almost unchanged after an "AI revolution"?

A. Because the list was never only about task skill. It outsources conscientiousness, verbal polish, and elite-network signaling (Rivera's point). AI fluency can be added as a screen on the same campuses. Replacing Penn with a random coding bootcamp would throw away the signal partners still use to sell teams to CEOs.

Q2. Is Accenture's "exit where reskilling is not viable" the same phenomenon as McKinsey's concerns ratings?

A. Same family, different paperwork. Accenture named AI as the skill axis and booked a restructuring charge. McKinsey used the old up-or-out machine and non-replacement, with AI as context. Both reduce inventory. Only Accenture made non-transition an explicit public HR policy.

Q3. A 2024 Penn BA accepts McKinsey. Using Chapter 4 and this chapter, what should she assume about the job she was sold?

A. Assume Lilli will take first-pass synthesis; assume her class is smaller than 2019's; assume promotion still requires presence, judgment, and now tool fluency; do not assume the firm will cut her hours by 30% because Lilli saved 30% on search. The surplus historically returns to utilization.

Q4. Should Booz Allen's 2,500 cuts be taught as an AI-era layoff?

A. Teach it as a demand-shock layoff in the AI era. Civilian contract slowdown is the

mechanism. Collapsing it into Sweet's reskill story muddies the independent variable and insults the people whose programs were cancelled.

Conclusion

Recruiting stories at these ten firms are three chapters of the same book. Founders hired people who looked like themselves and like the professions they envied—law, accounting, engineering. The 1990s hired people who looked like the admissions office at Harvard, Penn, and INSEAD, because that was cheaper than searching the country. The AI era still hires those people, then runs a second filter: can you live with the bot that does the job you were trained to learn by doing?

Layoffs in 2023–2026 are not a single plot. Some are pandemic-overhire corrections wearing AI clothing. Some are explicit reskill-or-exit. Some are federal buyer strikes. Some are the old partnership bar, shortened from twelve months to three. The people who do not “make the transition” leave through those doors. The archive we have is earnings calls, concerns letters, severance charges, offshoring coordinates, and forum posts. It is thin on second careers at equal pay.

Commonality is the inventory problem. Consulting sold scarcity of judgment. It staffed that scarcity with abundant juniors from scarce schools. Generative AI attacks the abundance (the junior row) without automatically replacing the scarcity ritual (the campus list). Until a firm is willing to recruit for judgment from outside the list or to keep the junior row as a paid apprenticeship even when a bot can draft the slide, recruiting will photograph 1998 and staff 2026.

That is the researchable story. The next empirical task is not another target-school ranking. It is matching class-year offer counts to agent counts to years-to-counsel-out—an apprenticeship census the firms already have and have not published.

I'll map what's already in chapters 0–5 against the book's promised later rooms (workforce, international regulation, reform, hidden costs), then propose five chapters that complete the Tamaraland without repeating the existing cases. You already have the engine room. Chapters 0–5 name the machine, map its land deals, open the consulting platforms, show internal AI colliding with client harm, and treat recruiting as inventory management. Do not write five more chapters that retell Lilli, Hertz, Bridges, or the campus list.

The remaining work is the rest of Tamaraland: the rooms a reader still cannot enter if they stop at Chapter 5. The preface already promised three of them—workforce economics after speed, the international regulatory split, and reform that is not ethics theater. Two more rooms are implicit in the title and in SEAM: who absorbs the iceberg, and which public institutions are supposed to witness the machine but now sell access to it.

Design rule for all five: keep axes, not stages. Each chapter should score the same four dimensions—accountability, ethics, business-model logic, temporal horizon—and keep the six voices in the room. If a draft collapses into “then AI matured,” cut it.

Table 1

Regime	Approximate dates	Who was hired	How they were found	What “not making it” meant
Founders	1914–1960s	Accountants, engineers, lawyers; later HBS	Personal networks, dean referrals	Firm too small for industrial up-or-out
Campus machine	1970s–2019, peak 1990s	BA/BS from core targets; MBA from M7 + INSEAD/LBS; Big 4 from broader state + elite lists	On-campus lists, case interviews, St. Charles / similar academies	Up-or-out; “counseled out”; partnership bar
AI inventory	2023–2026	Same prestige schools + AI/data specialists; fewer pure generalist BAs in some offices	Lists still; AI fluency screens; copiloted cases; experienced-hire surge	Reskill-or-exit; log-in metrics; support offshoring; demand-shock RIFs (federal)

Chapter 6 Iceberg Below Invoice SEAM Ledger

Chapter 6: The Iceberg Below the Invoice, A SEAM Ledger of Hidden Costs for the Big Consulting Companies AI Dysfunctions, Four Financial Consequences That Were Wasteful

Summary

How to read this chapter against Chapters 2–5: Chapter 2 mapped NDA geography. Chapters 3 and 4 mapped control planes and internal copilots. Chapter 5 mapped recruiting as inventory. This chapter does not reopen those rooms except to ask what they cost after the fee is paid. Purdue appears only as a **settlement stream**. The KPMG locker appears only as the **2026 public-cost wave** that followed a suppressed whistleblower file. Accenture’s labor process appears as **FLSA dockets and hour-sheet forums**, not as Hertz or promotion-by-login. The method is SEAM’s two-column ledger plus a third column this book has been writing since the preface: **cost absorption**—the rightward arrow in Figure 5.

In part I, this chapter applies Henri Savall and Véronique Zardet’s Socio-Economic Approach to Management (SEAM) to the ten billion-dollar consulting firms that intermediate AI and large-system transformation: Accenture, McKinsey & Company, Boston

Consulting Group, Deloitte, EY, PwC, Booz Allen Hamilton, KPMG, Bain & Company, and IBM Consulting. Conventional engagement accounting stops at the waterline—fees, announced jobs, “value at stake,” Responsible AI badges. SEAM treats that waterline as an institutional boundary, not a fact about cost. Below it sit labor displacement and unpaid overtime, apprenticeship collapse, cognitive double-running, public-subsidy and sole-source lock-in, grid and water cost-shift, error rates borne by beneficiaries, journalism and oversight silence, and intergenerational infrastructure. The chapter’s claim is architectural: **the consulting invoice is a transfer instrument**. Time, risk, and repair migrate from partnership realization to workers, clients’ frontline staff, ratepayers, taxpayers, eligible claimants, and future users of the systems the firm will not be present to run.

The qualitative method is antenarrative and documentation. Official narratives (economic-impact studies, “champagne” adoption metrics, settlement language that denies liability) are held against fragments the narrative was designed not to count: auditor-general reports, False Claims Act resolutions, employment collective actions, financial-reporting-council fines, class complaints, parliamentary privilege, Reddit and Glassdoor hour-sheets, and interviews in which former partners describe the gap between fraud-hunting slides and what the partnership would fund. Ten firm-level researchable cases are developed. None is a composite. Four teaching cases then force a single question: if SEAM’s hidden-cost ratio of roughly 167 percent of visible fees is even directionally right, **who holds the below-waterline ledger when the engagement closes?**

The four tesseract axes remain simultaneous. Accountability is named or diffused. Ethics is practiced or performed. The business model sells judgment or distributes a platform and a sole-source system. The temporal horizon ends at invoice or extends to the aquifer, the grid, the welfare file, and the child’s education record. Stage models cannot hold those coordinates. The iceberg can.

In Part II, a SEAM qualimetric ledger for the ten firms—Accenture, McKinsey, BCG, Deloitte, EY, PwC, Booz Allen Hamilton, KPMG, Bain, and IBM Consulting—on **AI-touched work in a 2026 working year**. It does two jobs the prior draft left half-done.

First, it prices Savall and Zardet’s six dysfunction families as they now appear inside AI delivery: (1) working conditions, (2) work organization, (3) communication–coordination–cooperation (3C), (4) integrated training going to the wrong people, (5) time management, and (6) strategic implementation. For each family it states a **unit cost per event**, an **event frequency** (day / week / month / year), and **low / mid / high** annual dollars.

Second, it converts those dysfunctions into four financial-consequence lines the host organization and the public actually pay: (A) lost quality, (B) lost production, (C) overtime spent putting out fires, (D) expected legal-liability value. Tables carry subtotals, totals, and grand totals.

Two ledgers are kept distinct on purpose.

Ledger I (process). Hidden cost inside the labor process of AI-facing staff, mid estimate **\$12.68 billion per year** across the Ten. Low **\$4.61 billion**. High **\$25.94 billion**.

Ledger II (full sociotechnical). The book's own coefficient—hidden cost $\approx$ **167 percent of the visible AI-touched fee**—applied to an estimated **\$81.0 billion** of AI-touched revenue. Hidden column **\$135.3 billion** per year. That is the iceberg that includes client rebuilds, ratepayers, claimants, and deferred state capacity, not only the partnership's internal friction.

Neither ledger is an audited trial balance. Both are **researchable models** with published anchors (headcount, AI-share of work, loaded cost, documented refunds and settlements). The method section shows the formulas. Where a court or auditor has already printed a number, that number is used as a **who-paid-after-the-invoice** event and is *not* double-counted as if it were a daily 3C failure.

Part I. The waterline is a choice

Savall and Zardet's research across more than 2,000 organizations treated absenteeism, turnover, non-quality, productivity loss, and health costs absorbed by the host as measurable dysfunctions, not "externalities." Applied here, the visible column is what Economic Impact Studies and partnership reviews already know how to print:

consulting revenue and utilization

jobs announced at the data-center fence

AI investment press releases

ISO 42001 / Responsible AI theater

shareholder or partner draws

The submerged column, for this industry in 2024–2026, is now documentable:

Labor: unpaid or disputed overtime; compressed-timeline exits; bench risk when the buyer leaves

Quality: systems that issue the wrong letter, the wrong audit opinion, or the wrong detention-staffing plan

Public lock-in: proprietary platforms that force sole-source renewals

Fiscal leakage: ineligible payments, change-request farms, early-termination charges that do not cover the client's rebuild

Infrastructure shift: demand forecasts that become utility rate cases

Civic capacity: tax agencies, newsrooms, and audit regulators that spend years reconstructing what the engagement did not count

Future users: children in an education-data breach; claimants on an automated penalty framework; ratepayers on a 2030 grid built to a consultant's compound annual growth rate

Community Voice does not need the 167 percent figure to be exact. It needs the category to be named before the ribbon-cutting.

A. Firm-by-firm ledgers (researchable cases)

1. Accenture — Sole-source lock-in and the hour the ledger will not hold

Visible fee. Accenture remains the scale player: on the order of \$70 billion in annual revenue, long-cycle managed services, and Diamond client relationships measured in decades. That durability is the product.

Case A — CEBA / Export Development Canada (new file). Canada's Auditor General, Karen Hogan, told Parliament that the Canada Emergency Business Account disbursed \$49.1 billion in pandemic loans to nearly 900,000 small firms—and that value for money failed in the administration layer. Non-competitive contracts to Accenture were **92 percent of \$342 million** in CEBA-related contracting. EDC, Hogan found, gave the vendor too much control over scope and price, and did not reliably check that invoices matched work. Hourly rates in files reviewers saw ran from about \$60 to \$750. Because the running system is Accenture's proprietary stack, the Crown corporation remains tied to non-competitive work **through at least 2028**. Finance Canada set no overall administrative ceiling; by 31 March 2024 administrative spend sat at \$853 million. About **\$3.5 billion** went to ineligible recipients. Subsequent access-to-information reporting put EDC's broader Accenture relationship on a path beyond \$1 billion.

SEAM reading: the visible story is speed in a crisis. The submerged story is **platform capture of a sovereign loan book**, an eligibility-error stock measured in billions, and a renewal horizon that outlives the pandemic the program was built for. Cost absorption: Canadian taxpayers and the small firms now in collection.

Case B — The hour-sheet as hidden payroll (new file). In March 2026 a collective action in the Northern District of Illinois alleged that Accenture LLP used prefilled or standardized time entries for hourly, non-exempt staff—including Epic go-live and project-support roles—and discouraged deviations when actual hours exceeded the template (period from 16 March 2023). That docket sits on top of an older pattern: *Thomas v. Accenture* (E.D. Mich.) settled unpaid-overtime claims for about \$415,000 covering roughly 257 consultants staffed onto electronic-health-record rollouts who were paid straight time past forty hours. Reddit's r/accenture, July 2025, is not a court. It is the antenarrative the court file makes legible: a poster describing a project manager who "says no to reporting overtime, claiming too many expenses," and replies that treat the practice as "the norm," especially on under-budgeted fixed-price work.

SEAM reading: utilization converts unpaid minutes into margin. The iceberg category is **labor costs transferred onto the employee's body and household**, then onto the next client if the same underbidding model travels.

Case C — Successor systems (new file). In May 2026 the UK Post Office awarded Accenture a contract reported at £322.8 million (VAT inclusive, about five years) to operate Horizon and help build its replacement after Fujitsu's system underwrote one of the country's worst miscarriages of justice. The fee is visible. The hidden cost is already socialized: wrongful convictions, imprisonments, and deaths the public inquiry recorded. Accenture did not design Horizon. It is paid to inherit a wreckage ledger the partnership model is not structured to carry.

2. McKinsey & Company — Forecasts that become rate cases; settlements that do not reach the injured

Visible fee. Strategy prestige, QuantumBlack, Lilli's prompt counts, and a shift toward outcome clauses still sit above the waterline.

Case D — The demand forecast as a public-works instrument (new file). McKinsey's energy and data-center publications are not neutral meteorology. The *Global Energy Perspective 2025* and subsequent U.S. power notes project data-center electricity demand rising on the order of 17 percent a year globally (about 25 percent in the United States) through 2030, with U.S. data centers potentially more than 14 percent of national power demand. A 2026 McKinsey power note adds that roughly 75 percent of anticipated U.S. power-demand growth over the next decade could come from data centers, and that community action postponed about four gigawatts of projects in a single quarter. Utilities and legislators cite such ranges when they seek plants, transmission, and large-load tariffs. Lawrence Berkeley's 2025-update report places a 2030 reference case near 11.8 percent of U.S. electricity and notes alignment with McKinsey and EPRI bands. The consultant is paid by private clients. The number becomes a **ratepayer planning assumption**.

SEAM reading: Dimension Four. The firm exits after the slide. The interconnection queue, the residential bill, and the "Ratepayer Protection Pledge" politics of 2026 do not. Hidden cost category: **grid-expansion and generation costs transferred to existing customers** unless a tariff wall holds.

Case E — Settlement arithmetic after Purdue (ledger, not the Ch. 3 duty case). The United States announced in December 2024 that McKinsey would pay **\$650 million** to resolve a Justice Department investigation of its Purdue work; a former senior partner pleaded guilty to obstruction for destroying records. States and localities had already taken roughly \$1 billion in earlier McKinsey opioid settlements. In April 2026 McKinsey agreed to add **\$125 million** to Purdue's bankruptcy estate; court papers said about \$50 million of that would raise the pool for individuals. The Purdue/Sackler plan that became effective around 1 May 2026 is \$7.4 billion over years, mostly to governments for "abatement." Reporting on individual recovery has put a rough ceiling near **\$16,000** for a qualifying person or survivor. KFF Health News' county-tracking project shows settlement cash moving through governments with uneven accountability; some jurisdictions use it to plug budgets.

SEAM reading: the engagement fee (\$93 million over years, in the DOJ telling) is a rounding error against mortality and treatment cost. The settlement is still not a full ledger. Governments are more powerful creditors than widows. **Cost absorption is fiscalized**, then often re-absorbed by the same communities that buried the dead.

Employee and alumni forums treat “turbocharge” as a permanent stain. That is Community Voice, not a census. It is still part of the antenarrative the official “we have resolved” sentence suppresses.

3. Boston Consulting Group — Integrity as a traveling cost

Visible fee. Growth through 2024–2025 while some peers shrank; GENE and AgentKit above the waterline.

Case F — Angola, FCPA, and the revolving door that launders accountability (new file). BCG resolved U.S. Foreign Corrupt Practices Act matters in 2024 with about **\$14.4 million** in profit disgorgement tied to work that investigative reporting has long linked to Sonangol and Angolan payment structures. Open Secrets / *Daily Maverick* (3 September 2026) then documented a different transfer: senior Bain people who had worked the SARS years moving into BCG partnerships in South Africa while Bain wound down its local consulting shop in 2025 and remained under a ten-year National Treasury ban. BCG’s public line on Just Energy Transition work stresses multi-stakeholder modeling. The antenarrative is **personnel as a vehicle for unrepaired institutional harm**. A revenue authority that lost capacity does not recover because a different logo now sits on the JETP slide.

SEAM reading: hidden cost is **state capacity and public trust**, not only a disgorgement check. Ethics architecture is performed in the press release and weakened when the industry’s internal labor market treats a capture file as portable experience.

(Gaza and GameStop remain Chapter 3–4 files. They are not reopened here except to note the same pattern: process failure named after the fact; social cost already spent.)

4. Deloitte — Assurance language, taxpayer settlement, unchanged recommendations

Visible fee. Global generative-AI investment headlines; Ascend/Zora trust scores.

Case G — \$21.5 million False Claims Act resolution (new file). On 25 August 2026 the U.S. Department of Justice announced that Deloitte entities would pay **\$21.5 million** to resolve allegations that the firm certified compliance with federal anti-discrimination contracting rules while using race- and sex-based employment practices, including goals tied to partner pay. About \$10 million was labeled restitution; the relator (American Alliance for Equal Rights) takes \$4.3 million. Deloitte denied liability and said it settled to avoid protracted litigation. Whatever one thinks of the underlying DEI fight, the SEAM point is narrower: **a federal contractor’s compliance certificate is a billing instrument**. When it fails, the first payer of record is the Treasury, then the partnership’s insurance and reputation budget—not the agencies that bought advice under the certificate.

Case H — What a partial refund does not buy (ledger pointer only). Chapter 3 already has the Australian DEWR footnotes. The 2026 Senate-estimates figure matters for this chapter: Greens senators reported a refund of **A\$97,587** on a A\$440,000 contract—less than a quarter—while the Targeted Compliance Framework continued to generate mass payment suspensions (on the order of 350,000 in one quarter of 2025) and the Ombudsman recorded unlawful cancellations. The department kept the recommendations. Hidden cost is not the missing instalment. It is **claimants living inside an automated penalty machine whose “assurance” document had to be rewritten after a university reader did the state’s quality control.**

5. EY — After Wirecard, after Horizon’s books

Visible fee. EY.ai adoption percentages; NVIDIA/Microsoft product theater.

Case I — Post Office audits under FRC investigation (new file). In April 2025 the UK Financial Reporting Council opened an investigation into EY’s audits of Post Office Limited for years ending March 2015–2018, focused on the Horizon system that produced false shortfalls and wrongful prosecutions. EY said it would cooperate. The public inquiry and subsequent reporting had already established the human ledger: hundreds of convictions, imprisonments, family collapse, and suicides. Fujitsu built the system. The Post Office prosecuted. The auditor’s hidden-cost question is specific: **what did certified accounts authorize the institution to keep doing?** An investigation is not a finding. It is already a researchable file because the social cost matured years before the watchdog’s letter.

Case J — Insider interview, unfinished clean-up (new file). In September 2025 the *Financial Times* published Joe Howie’s account: a 35-year EY veteran who says he was pushed out after a post-Wirecard quality effort stalled—risky clients kept, controls he would not sign. German APAS had already described “repeated grave” duty violations on Wirecard and barred EY from new public-interest audit mandates for a period into 2026. *Capital* reported that federal German ministries still awarded EY advisory work, including eight-figure relationships, because the audit ban did not travel automatically into consulting.

SEAM reading: Founders’ Voice (conscience as auditor) against PR Voice (Reimagination Engine). Hidden cost is **investors and tax authorities who treated an unqualified opinion as ground**, plus public buyers who separate “audit sin” from “advisory spend” as if they were different firms.

6. PwC — Second fines, distressed vendors, and a \$1 sale of the government arm

Visible fee. Agent OS and PwC One; Responsible Tax language.

Case K — Babcock, twice (new file). In July 2026 the FRC fined PwC **£3.2 million** and audit partner John Waters £59,000 for “serious and numerous breaches” on Babcock’s 2019–2020 audits—three years after a separate £5.6 million penalty on earlier Babcock work. Babcock, a major UK defence contractor, had to announce on the order of 140

adjustments totaling about £2 billion. The watchdog's through-line was failure to challenge management.

SEAM reading: audit quality failure is a hidden cost to pension funds, the Ministry of Defence's counterparty, and workers in a restated company. Consulting's adjacent claim—that the same network can “transform” the public sector—rides on a trust stock the fines keep drawing down.

Case L — Project Aztec and the vendor next door (new file). BBC reporting on mistakenly released Cabinet Office papers described PwC work, coded Aztec, on continuity risk if Atos—supplier of NHS-adjacent and justice IT—hit insolvency. That is not a PwC implementation failure. It is PwC paid to map **someone else's wreckage** because the state outsourced critical ledgers until failure became a consulting product. Hidden cost: dual payment (first the fragile vendor, then the review of the fragile vendor) while patients and courts remain on the systems.

Australia remains the structural pair: after the tax-confidentiality scandal PwC sold its government advisory arm to private equity for **A\$1** (rebranded Scyne). New federal work was later allowed back under constraints. The \$1 price is the market's SEAM coefficient for a captured public-sector practice.

7. Booz Allen Hamilton — Taxpayer files, cancelled contracts, bench risk

Visible fee. aiSSEMBLE, Vellox, number-one federal AI-services positioning.

Case M — Littlejohn, 406,000 files, Treasury exit (new file as 2026 ledger). Charles Littlejohn, a Booz Allen contractor at IRS, stole and leaked returns; he is serving a five-year sentence. IRS later put the affected population near **406,000** taxpayers. On 26 January 2026 Treasury Secretary Scott Bessent cancelled **31 Booz Allen contracts** (~\$4.8 million annual / ~\$21 million obligations), citing inadequate safeguards. A consolidated civil action (*Safe Harbor International LLC v. Booz Allen Hamilton Inc.*, D. Md.) was allowed in part to proceed in 2026 against Booz Allen and IRS defendants. TIGTA's later work on deferred-resignation staff who kept system access shows the control failure did not end with one man.

SEAM reading: Chapter 3 used the older False Claims billing file. This chapter's cost-bearer is **every taxpayer in the extract**, plus civil-federal employees cut in the 2025–2026 demand shock when civilian work slowed. Two layoff species (Chapter 5) meet here: buyer punishment and residual security risk. Hidden cost is civic confidentiality—the one asset an IRS contractor is paid to protect.

Employee forums on bench-when-the-contract-ends are the labor half of the same ledger.

8. KPMG — When the suppressed file becomes a sector crisis

Visible fee. Workbench, Clara, ISO badges.

Case N — 2026: privilege, resignations, and a breakup paper (new public-cost wave).

Chapter 3 recorded the locker. Chapter 6 records what the locker cost once Senator Deborah O'Neill used parliamentary privilege. In May–June 2026 KPMG Australia's CEO and audit leader resigned after the firm conceded its whistleblower probe "fell short." The chair followed. ASIC opened investigations of named auditors. Finance classified the event as significant and put more than **A\$270 million** in Commonwealth contracts under scrutiny; active federal work was reported near **A\$650 million**. The Reserve Bank signaled it would likely not renew a whistleblower-hotline role. Reuters found new Big Four federal contract signing already down by about half versus the prior year. Treasury floated structural options: fines, partner caps, even break-up of audit and consulting. Chartered Accountants ANZ opened ethics reviews across the large firms.

SEAM reading: the hidden cost of a two-year internal delay is **systemic**. Clients re-tender. Juniors sit on benches. Regulators spend a year of hearings. The profession's social capital—already drawn down by PwC—goes another increment negative. That is not "one bad apple." It is Dimension One: accountability delayed until Parliament became the quality function.

9. Bain & Company — Fees returned; capacity not returned

Visible fee. Sage, ChatGPT Enterprise, America's Cup digital twins.

Case O — After the repayment: office closure, failed re-ban, portable people (new file). Bain repaid SARS fees plus interest in November 2018. South Africa's National Treasury still imposed a ten-year public-contract ban in 2022. In July 2025 Bain confirmed it was winding down consulting operations in South Africa. The UK had banned then unbanned the firm; in January 2025 ministers said they had **no legal route** to reimpose a ban without new facts, despite Zondo's language of collusion and unlawfulness. *Daily Maverick* (September 2026) then tracked SARS-era Bain seniors into BCG roles. Athol Williams had already named names on social media in 2023.

SEAM reading: a fee refund is above-waterline hygiene. A revenue service's investigative and collection capacity is below. **Repayment without restoration is effort theater.** Dimension Four fails: the partnership's horizon ended when the wire cleared; SARS's horizon is generational.

Do not fuse Bain & Company with Bain Capital's PowerSchool litigation. Different legal persons. The education-sector breach file is researchable for private-equity chapters, not this one.

10. IBM Consulting — Restitution in Madrid; undisclosed breaches; certificates again

Visible fee. watsonx Orchestrate; Enterprise Advantage sold outward; HR-query automation claims.

Case P — Madrid Bar Association systems (new file). Spain’s Audiencia Provincial (December 2025) upheld a substantial award against IBM Global Services España and Viewnext: restitution on the order of **€4.64 million** tied to contracts and addenda with the Madrid Bar Association (ICAM) after a lower court found for the college on core system failures. That is a client-side hidden cost made visible only because a professional body sued. Most public agencies cannot.

Case Q — Barlow qui tam, unsealed 2026 (new file). TechCrunch reporting on a 2020 complaint unsealed in June 2026 describes William Barlow, former IBM VP of threat intelligence, alleging Chinese APT activity against IBM’s core network (2013–2016), poor logging, and non-disclosure. DOJ declined to intervene. IBM says it followed the law. The SEAM question does not require a verdict: **if a cyber-consulting vendor’s own control plane is the contested fact, every client who bought “fortress” language is holding an unpriced contingency.**

IBM also paid about **\$17.1 million** in 2026 to resolve federal contracting allegations in the same Civil Rights Fraud wave as Deloitte. Certificates, again, as billing instruments.

Employee writing (Glassdoor, spin-off alumni: “take the IBM out of Kyndryl”) describes fear-based layers and blocked mobility—the labor iceberg under Orchestrate’s ROI slides.

C. Cross-firm patterns (the transfer identity)

The 84 percent “no help” transformation survey (Chapter 4) is compatible with this table. Internal time-saved does not move a row unless the contract prices the submerged column.

Blogs and social media do not replace audits. They mark where the official story is already unbelievable to operators: “no to reporting overtime, claiming too many expenses”; “ROI on inference is a utilization number”; Williams naming SARS alumni at a competitor. Treat them as situated testimony, then triangulate.

D. Evidence — four researchable cases (not Chapters 3–4)

Use 60–90 minutes. Through-line: *Write the third column (cost absorption) before you defend the fee.*

Research Case J. Accenture, CEBA, and the system you cannot leave

Documents. OAG Report 8 (2024); Hogan opening statement 4 Dec 2024; PACP evidence; Globe and Mail contract-database follow-up.

Question. EDC needed speed. It bought a proprietary stack and a sole-source future. Design a 2020 model contract that still ships loans in weeks **without** granting the vendor pricing control and 2028 lock-in. Which three clauses would Auditor General Hogan have scored as present?

Indicative answer. (1) Open or escrowed source / documented exit image within 12 months; (2) invoice-to-work sampling by a party that is not the vendor; (3) eligibility false-

positive/false-negative SLAs with clawback—because \$3.5 billion ineligible is the real performance metric. Hourly-rate bands published. Conflict ban: vendor cannot run a procurement that selects its subsidiary.

Research Case K. McKinsey’s gigawatts and the residential bill

Documents. McKinsey *Global Energy Perspective 2025*; 2026 U.S. data-center power note; LBNL 2025 update; USAFacts explainer; one state utility large-load tariff filing that cites consultant ranges.

Question. If a firm’s scenario becomes 14 percent of national demand in a regulator’s exhibit, is that speech or infrastructure? Draft a disclosure rule for consultant demand forecasts used in rate cases.

Indicative answer. Treat the forecast as an expert submission: named authors, confidence intervals, community-delay sensitivity (McKinsey’s own 4 GW postponement), and a sworn statement on whether the firm is concurrently paid by a hyperscaler in that interconnection queue. Without that, Dimension One is diluted into “insights.”

Research Case L. Deloitte’s \$21.5 million and the certificate

Documents. DOJ 25 Aug 2026 release; settlement agreement; compare IBM \$17.1 million.

Question. Separate the culture-war noise from the billing structure. If a partnership ties partner pay to demographic targets **and** signs federal non-discrimination certifications, which SEAM category moved—ethics architecture or business-model logic?

Indicative answer. Both. The certificate was a market-access tool (Dimension Three). The settlement is Treasury recovering a slice of allegedly false claims (Dimension One), not a repair of whoever was hired or not hired. Ban using internal DEI *or* anti-DEI dashboards as partner-comp inputs while a federal certificate is live. Price the certificate as an auditable control, not a cover page.

Research Case M. KPMG 2026 and the two-year delay

Documents. Guardian/Reuters/FT timelines May–July 2026; Finance “significant event”; Treasury options paper on structural separation.

Question. The locker was 2023. The public ledger is 2026. Write the hidden-cost interest rate on a suppressed whistleblower file.

Indicative answer. Compound items: leadership severance and replacement; A\$650 million book under political risk; sector-wide tender slump; regulator hours; junior utilization crash; copycat reviews at other firms. A two-year internal “fell short” is not a communications error. It is SEAM non-quality in the ethics function. Remedy: mandatory external triage of confidentiality allegations within 30 days, with the client notified if audit independence is in play—before Parliament has to do quality control.

E. Implications of Consent Machine

The Consent Machine's most ordinary technology is not an NDA and not an agent. It is an **invoice that closes before consequence**. Accenture can administer a national loan program so that the state cannot leave the program's own software. McKinsey can publish a demand curve that utilities translate into plants and bills. BCG can disgorge Angolan profit and hire the people another firm could no longer station in Johannesburg. Deloitte can refund a slice of a welfare-assurance fee and pay \$21.5 million to quiet a federal certificate case while the penalty framework keeps suspending payments. EY can live through Wirecard sanctions and still be hired to bless—or now to explain—books adjacent to Horizon. PwC can be fined twice on one defence contractor and paid to war-game the next vendor's collapse. Booz Allen can lose Treasury work after 406,000 tax files move, while civil staff take the headcount shock. KPMG can watch a locker become a breakup white paper. Bain can repay fees and still not repay a revenue service its years. IBM can automate HR queries and still meet a Spanish court over a professional body's broken system.

None of those sentences requires a stage theory. They are simultaneous coordinates. Accountability is a partner name or a “no admission” paragraph. Ethics is a code or a slide. The business model is independent judgment or a platform you must renew. Time is a billing cycle or a generation.

SEAM's instruction is practical. Put beneficiary-level error rates, ratepayer increments, overtime actually worked, sole-source exit dates, and settlement dollars that reach injured people **on the same page as the fee**. If the partnership will not hold that page, the community will—in Auditor General reports, FLSA dockets, FRC orders, parliamentary privilege, and the signs already held up in Magdalena.

The next chapter leaves the ledger and walks the two law rooms—Brussels after 2 August 2026, Washington after voluntary frameworks and security overrides—and asks how the same ten firms productize both.

Part II: Quantifying SEAM Hidden Costs

Scope. One working year (220 paid days, 52 weeks, 12 months). Only AI-touched delivery: internal copilots, agent platforms, AI transformation engagements, and AI-adjacent public systems the Ten operate or assure.

Scale anchors used

AI-touched revenue implied by those shares: **\$81.04 billion**.

Ledger I formula

$$\text{Hidden}_{f,d,b} = (\text{AI-FTE}_f) \times H_b \times w_d$$

where (H_b) is hidden cost per AI-facing FTE-year (low \$8,000 / mid \$22,000 / high \$45,000) and (w_d) is the dysfunction weight below. The \$22,000 mid-point sits inside

Savall-Zardet's observed per-person hidden-cost bands for knowledge-intensive organizations, adjusted for double-running (human process + agent process).

Event frequency (mid), per 100 AI-facing FTE per year, from triangulation of hour-sheet forums, adoption-mandate reporting, refund/rework files, and SEAM interview densities:

Day = year / 220. Week = year / 52. Month = year / 12.

Ledger II formula

$$\text{Full hidden} = 1.67 \times \$81.04 \text{B} = \$135.34 \text{B}$$

allocated with the same (w_d) . This is the preface coefficient (visible \$2 million fee → about \$5.5 million sociotechnical total).

What these dollars are not. They are not fines already paid, and they are not a claim that every consultant generates a lawsuit each week. Legal line D is a **separate expected-value stack** built from named 2024–2026 instruments, then annualized. Documented cash is listed in the who-paid section so a reader can subtract or add.

A — What the six dysfunctions look like in AI work (so the dollars have a referent)

1. Working conditions. After-hours pages, login-surveillance, unpaid or disputed overtime, bench anxiety when a civilian contract dies, “learn AI or compressed-timeline exit.” Anchor files: Accenture FLSA collective action (N.D. Ill., filed 16 March 2026); r/accenture hour-sheet thread (July 2025); Accenture \$865 million optimization / reskill-or-exit charge.

2. Work organization. Human process and agent process run in parallel; juniors become prompt operators; partners still sell an army. Anchor: McKinsey ~40,000 humans + ~25,000 agents; Sternfels “25 squared”; Kate Smaje on junior deck work.

3. 3C. Hallucinated citations, version conflicts across Agent OS / Lilli / Clara / Zora, client footnotes nobody owns. Anchor: DEWR Azure-GPT-4o toolchain as *quality-control failure*, not as a platform tour.

4. Training to the wrong people. 550,000 gen-AI fundamentals seats and 47 million training hours can still miss the eligibility-rule engineer, the AML investigator, and the partner who signs. Training spend is visible; apprenticeship collapse is hidden.

5. Time management. A reported 30 percent synthesis saving that utilization reconverts into more prompts. Midnight stand-ups coexist with two-minute VPN myths.

6. Strategic implementation. Sole-source stacks that outlive the crisis (CEBA to 2028); demand forecasts that become rate cases; ethics frameworks that do not stop a locker or a fabricated judge.

Table 1 — Ledger I: Process hidden costs by dysfunction (mid, \$ million per year)

Bands for the same grand total: low \$4,611 million · mid \$12,681 million · high \$25,938 million.

Mid Ledger I is about **3.7 percent** of combined Ten-firm revenue and about **15.7 percent** of AI-touched revenue—friction, not the full public iceberg.

**Table 2 — Event frequency (mid), Ten firms combined and per firm
Industry-wide mid events**

Low band: ×0.5 events and lower unit cost. High band: ×1.8 events and higher unit cost.

Per-firm events per year (mid)

A “strategic implementation event” is rarer and more expensive (about \$24,000 mid) because it is a failed go-live, a lock-in renewal, a fabricated-citation report, or a sole-source stack—not a single slack message.

Table 3 — Ledger II: Full sociotechnical hidden costs (167% of AI-touched fees), \$ billion per year

Low/high here scale the 167% coefficient to a 60%–192% band around the book’s central 167%, not the Ledger I per-FTE band. Use Ledger I for labor-process debates; use Ledger II when the question is “what did the community actually absorb?”

Allocation of the \$135.34B mid total back onto firms (proportional to AI-touched revenue):

Table 4 — Financial consequences A–D (Ledger I base + separate legal EV)

Consequence weights on the **\$12,680.8 million** process pool: A lost quality 35%; B lost production 30%; C firefighting overtime 20%. The remaining 15 percent is residual (turnover, overconsumption, image) and is not forced into A–C. Line D is **not** a percentage of that pool. It is a 2026 expected-value stack from named instruments (refunds, FCA, FRC, cancelled book, first Purdue tranche, class-action midpoints).

Time grain for A–C only (mid, Ten firms, \$ million)

Bands for A+B+C: low \$3.92B · mid \$10.78B · high \$22.05B.

Implied mid loss per event (A+B+C pool / 1,654,268 events): about **\$6,500** blended. Strategy events pull that mean up; 3C and time events pull it down.

If the reader prefers Ledger II as the base for A–C, multiply A, B, and C by $\sqrt{(135.34 / 12.68)}$ (approx 10.67). That yields mid A **\$47.4B**, B **\$40.6B**, C **\$27.1B**, A+B+C **\$115.1B**, plus D **\$0.50B**, grand **\$115.6B**. Use that only when the unit of analysis is the *host society*, not the partnership P&L.

How line D was built (so it can be attacked)

Expected 2026 legal/refund/cancellation value, mid, \$ million. Low = 45% of mid. High = 220% of mid.

High D across the Ten $\approx$ **\$1.10 billion**. Still an order of magnitude smaller than Ledger II, which is the point: **courts price a sliver of the iceberg**.

Who paid after the invoice

Research ledger events (new files first; Chapter 3–5 files only as checks)

Accenture

New — CEBA / EDC (Canada). Visible administration contracts: **\$342 million** CEBA-related, 92 percent sole-source to Accenture; EDC–Accenture relationship on a path beyond **\$1 billion**. After the invoice: Auditor General Hogan found invoice-to-work controls failed; **\$3.5 billion** in loans estimated ineligible; administrative spend **\$853 million** with no Finance ceiling; proprietary lock-in through **2028**. *Payers after the fee*: Canadian taxpayers; ineligible and eligible borrowers now in collection; EDC, which cannot exit the stack. Dysfunctions 2, 3, 6. Consequence lines A, D.

New — FLSA 2026 + hour-sheets. Collective action filed 16 March 2026 (N.D. Ill.) for hourly non-exempt staff from 16 March 2023; precursor *Thomas* settlement about **\$415,000** for ~257 EHR go-live consultants. Reddit, July 2025: overtime “too many expenses.” *Payers*: employees’ unpaid hours (line C), then, if the class holds, Accenture cash (line D). Dysfunctions 1 and 5.

New — Post Office successor. Contract reported at **£322.8 million**. Accenture did not build Horizon. *Payers after Fujitsu’s invoice*: the UK public already paid in convictions and redress schemes; Accenture is now paid to inherit that wreckage. Dysfunction 6.

Supplement, not retelling — Hertz. Chapter 3–4 told the dead website. Ledger only: Hertz alleged **\$32 million** fees plus **\$10 million+** to remediate and lost revenue. Public docket did not produce a clean 2026 “Accenture paid Hertz \$X” line in open sources; the absorption to date is **Hertz’s cash and years**. Line D remains a contingent, not a booked refund.

McKinsey

New — Demand forecast as rate-case input. Global Energy Perspective 2025 and 2026 U.S. power notes: data centers ~25 percent annual U.S. load growth, possibly **>14 percent** of U.S. demand by 2030; ~4 GW postponed in one quarter by community action. *Payers after the insight fee*: ratepayers, unless a large-load tariff holds. Dysfunction 6. Lines A and B at grid scale (Ledger II, not Ledger I).

Supplement — Purdue as cashflow, not Lilli. DOJ **\$650 million** (Dec 2024) plus **\$125 million** into the estate (2026 filing: **\$65 million** then, **\$60 million** in 2027). Individuals in the Purdue plan face a rough ceiling near **\$16,000**. Governments take the bulk of **\$7.4**

billion. *Payers after the historic \$93 million fee:* McKinsey partners (settlement), then taxpayers who fund abatement programs, then families who still hold the medical bill. Dysfunction 6. Line D.

BCG

New — FCPA / Angola / revolving door. About **\$14.4 million** disgorged (2024). 2026 *Daily Maverick*: SARS-era Bain seniors inside BCG while Bain exits Johannesburg. *Payers:* U.S. Treasury (disgorgement); Angolan and South African publics who already lost institutional capacity. Dysfunctions 3 and 6.

Supplement — Gaza / GameStop. Not retold. Ledger: BCG says invoices cancelled and no fees collected on the disavowed work; Save the Children suspended a twenty-year partnership (relationship cash, not a court number). GameStop refused about **\$30 million**; that is a client who refused to pay after “meager impact,” i.e. the rare case where the *buyer* stopped the transfer.

Deloitte

New — FCA 25 August 2026. \$21.5 million to the United States (~\$10 million restitution); relator **\$4.3 million**. Denied liability. *Payer after years of federal invoices:* Deloitte, then the U.S. as partial recapture. Dysfunction 6. Line D.

Supplement — DEWR footnotes. Refund reported at **A\$97,587** on **A\$440,000**. Department kept recommendations. *Payers:* Australian taxpayers for the unrefunded three-quarters; jobseekers inside TCF suspensions (on the order of 350,000 in one quarter of 2025). The fabricated citation was 3C; the unpaid human cost is A and D in Ledger II.

Supplement — Bridges / Noon. Not retold. Ledger: Michigan and other states already paid multi-year eligibility contracts (Michigan Bridges cited in the ~\$768 million-since-2006 band). Noon paid in coverage. Lawyers paid in reconstruction hours. Deloitte kept the operate-and-change-request book.

EY

New — FRC file on Post Office audits 2015–2018. Open investigation, not a fine yet. *Payers to date:* sub-postmasters and the UK redress schemes. Dysfunction 6.

New — Howie / Wirecard aftermath. Insider account that post-Wirecard controls were watered down; German audit ban did not stop federal advisory awards. *Payers:* Wirecard investors historically; German ministries still paying advisory invoices.

Supplement — Project Morgan. Already refunded about **£15 million** to Santander. *Payer after the champagne:* EY UK financial-services partners; Santander still holds the AML program risk.

PwC

New — Babcock, second time. FRC 2026: **£3.2 million** firm + **£59,000** partner, after 2023 **£5.6 million**. Babcock restated on the order of **£2 billion** across adjustments. *Payers:* PwC; then pension funds and a defense ministry that relied on clean opinions. Dysfunctions 3 and 6.

New — Project Aztec. PwC paid to map Atos distress around NHS/justice systems. *Payer:* UK taxpayer, twice (fragile vendor + review of fragile vendor).

Supplement — Australia tax leak. Government arm sold for **A\$1**. That dollar is the market's coefficient for a captured public practice.

Booz Allen Hamilton

New — Treasury 26 January 2026. 31 contracts cancelled (406,000** taxpayers in the IRS accounting. Civil actions proceeding in D. Md. *Payers:* affected taxpayers (privacy); Treasury (re-procurement); Booz (lost book and defence costs); civil staff cut in the separate demand shock. Dysfunctions 1, 2, 6.

Supplement — Feinberg FCA \$377.45 million. Already in Chapter 3 as billing. Ledger only: U.S. Treasury recaptured a slice of alleged false claims years after the invoices.

KPMG

New — 2026 public-cost wave. Privilege broke a suppressed file. CEO, audit leader, chair gone. **A\$650 million** active federal contracts under scrutiny; **A\$270 million+** named as significant-event review; ASIC on named auditors; Treasury breakup paper. *Payers:* KPMG partners (exits, insurance, retender); Australian agencies that must re-procure; juniors on the bench. Dysfunctions 3, 4, 6.

Supplement — locker. Not retold. The 2023 event's 2026 interest rate is the sector slump in new Big Four federal signings (Reuters: on the order of a one-year halving).

Bain

New — 2025 SA wind-down; 2025 UK re-ban blocked. Fees plus interest were repaid in November 2018. Capacity at SARS was not. *Payers after repayment:* South African taxpayers and compliant businesses who needed a working revenue service; UK taxpayers who still may buy Bain without a live ban. Dysfunction 6. Line D small; Ledger II large.

Supplement — SARS / Zondo. Operational story stays in Chapter 3. This chapter's only addition is the **personnel transfer into BCG** as an unpriced 3C failure of the industry, not of one logo.

IBM Consulting

New — ICAM Madrid. Restitution on the order of **€4.64 million** (AP Madrid, Dec 2025). *Payer:* IBM Spain entities; the bar association already paid in broken systems and years.

New — Barlow complaint unsealed 2026; \$17.1 million U.S. certificate settlement.

Payers: IBM (settlement cash); clients who bought fortress language (contingent).

Supplement — Watson Health. Sale and write-down already historic. Ledger: hospitals and IBM shareholders split a failed clinical claim. Not re-opened as a product tour.

Multiple estimates, stated together (so no single cell pretends to be a census)

The last row's mid **\$13.2B** = Ledger I \$12.68B + incremental legal not already inside process risk. Do not add Ledger I and Ledger II. They are two readings of one iceberg.

Closing the Loop (one question)

Take Accenture's CEBA file and Deloitte's \$21.5 million FCA check. Which ledger should a public accounts committee print?

If it prints only D, it will celebrate a \$21.5 million recovery and miss TCF suspensions.

If it prints only Ledger I, it will see overtime and 3C friction and miss \$3.5 billion ineligible loans.

If it prints Ledger II and the who-paid list together, the invoice stops looking like the cost.

That is the chapter's test for every firm above.

Conclusion

The six AI dysfunctions are not maturity-stage problems. They run every working day: on a mid estimate, about **7,500 dysfunction events** across the Ten, at blended process cost of about **\$58 million a day**. Training still misses the person who can stop a false denial. Time saved still returns to utilization. Strategy still sells a stack the buyer cannot leave.

Courts and refund clauses recovered, on a mid 2026 expected-value basis, about **half a billion dollars**. Ledger I says process hidden costs are about **twenty-five times that**. Ledger II says the sociotechnical column is about **two hundred seventy times that**. The Consent Machine is not primarily a litigation problem. It is an accounting boundary problem.

Who pays after the invoice, in the files that already exist: employees who cannot book the hour; claimants inside eligibility and penalty engines; ratepayers sitting under a consultant's gigawatt curve; taxpayers locked to a sole-source loan platform until 2028; families who receive a four-figure opioid check while governments book the headline; audit regulators who have become the quality function the partnership would not fund.

The next chapter takes the same Ten into two law rooms—the AI Act as enforced from August 2026, and the U.S. voluntary-plus-security-override room—and asks how those rooms change the coefficients in Tables 1–4, or merely productize them.

Table 1

Hidden-cost category (SEAM)	Who is supposed to hold it	Who actually holds it in these files	Exemplars
Non-quality / error	Firm quality function	Claimants, detainees, investors, postmasters	TCF suspensions; Horizon audits; Wirecard opinions
Labor / overtime / exit	Employer payroll	Hourly staff, counsel-outs, offshored EAs	Accenture FLSA; Reddit hour-sheets; civil RIFs
Lock-in / proprietary stack	Competitive procurement	Taxpayers for years after the crisis	CEBA to 2028; EDC >\$1B path
Integrity delay	Partnership investigation	Parliament, ASIC, FRC, journalists	KPMG 2024–2026 lag; PwC tax years
Forecast externalization	Client + regulator	Ratepayers and host counties	McKinsey/BCG demand curves in utility cases
Settlement substitution	Wrongdoers → injured	Governments first; individuals last or capped	Opioid pools vs ~\$16k individual band
Certificate risk	Contractor ethics office	Treasury (FCA) then partners	Deloitte \$21.5M; IBM \$17.1M
State-capacity wreckage	Client institution	Citizens who need the tax agency or aid system	SARS after repayment; Angola after fees

Table 2

Firm	Revenue (approx.)	Human FTE	AI-touched share of work	AI-facing FTE	Loaded cost / FTE-year
Accenture	\$69.7B	780,000	25%	195,000	\$85,000
McKinsey	\$16.0B	40,000	40%	16,000	\$180,000
BCG	\$13.5B	33,000	35%	11,550	\$175,000
Deloitte	\$67.0B	450,000	20%	90,000	\$95,000
EY	\$51.0B	395,000	20%	79,000	\$90,000
PwC	\$55.0B	370,000	20%	74,000	\$90,000
KPMG	\$38.0B	270,000	18%	48,600	\$85,000
Bain	\$6.0B	19,000	30%	5,700	\$170,000
IBM Consulting	\$18.0B	150,000	30%	45,000	\$100,000
Booz Allen	\$11.0B	33,000	35%	11,550	\$140,000
Ten-firm total	\$342.2B	2,540,000	—	576,400	—

Table 3

Dysfunction	Events / 100 AI-FTE / year (mid)	Low ×	High ×	Weight \ (w_d\)	Mid \$ / event
1 Working conditions	40	0.5	1.8	15%	\$8,250

2 Work organization	60	0.5	1.8	20%	\$7,333
3 3C	80	0.5	1.8	22%	\$6,050
4 Training to wrong people	25	0.5	1.8	12%	\$10,560
5 Time management	70	0.5	1.8	18%	\$5,657
6 Strategic implementation	12	0.5	1.8	13%	\$23,833

Table 4

Firm	1 Working conditions	2 Work organization	3 3C	4 Training wrong people	5 Time management	6 Strategic implementation	Firm subtotal
Accenture	643.5	858.0	943.8	514.8	772.2	557.7	4,290.0
McKinsey	52.8	70.4	77.4	42.2	63.4	45.8	352.0
BCG	38.1	50.8	55.9	30.5	45.7	33.0	254.1
Deloitte	297.0	396.0	435.6	237.6	356.4	257.4	1,980.0
EY	260.7	347.6	382.4	208.6	312.8	225.9	1,738.0
PwC	244.2	325.6	358.2	195.4	293.0	211.6	1,628.0
KPMG	160.4	213.8	235.2	128.3	192.5	139.0	1,069.2
Bain	18.8	25.1	27.6	15.0	22.6	16.3	125.4
IBM Consulting	148.5	198.0	217.8	118.8	178.2	128.7	990.0
Booz Allen	38.1	50.8	55.9	30.5	45.7	33.0	254.1
Dysfunction total	1,902.1	2,536.1	2,789.8	1,521.7	2,282.5	1,648.4	
GRAND TOTAL (mid)							12,680.8

Table 5

Dysfunction	Per working day	Per week	Per month	Per year	Mid \$ / event
1 Working conditions	1,049	4,438	19,213	230,560	8,250
2 Work organization	1,573	6,657	28,820	345,840	7,333
3 3C	2,097	8,876	38,426	461,120	6,050

4 Training wrong people	655	2,774	12,008	144,100	10,560
5 Time management	1,835	7,766	33,623	403,480	5,657
6 Strategic implementation	315	1,331	5,764	69,168	23,833
All six	7,524	31,843	137,855	1,654,268	—

Table 6

Firm	WC	WO	3C	Train	Time	Strateg y	Firm year	Firm day
Accenture	78,000	117,000	156,000	48,750	136,500	23,400	559,650	2,544
McKinsey	6,400	9,600	12,800	4,000	11,200	1,920	45,920	209
BCG	4,620	6,930	9,240	2,888	8,085	1,386	33,149	151
Deloitte	36,000	54,000	72,000	22,500	63,000	10,800	258,300	1,174
EY	31,600	47,400	63,200	19,750	55,300	9,480	226,730	1,031
PwC	29,600	44,400	59,200	18,500	51,800	8,880	212,380	965
KPMG	19,440	29,160	38,880	12,150	34,020	5,832	139,482	634
Bain	2,280	3,420	4,560	1,425	3,990	684	16,359	74
IBM Consulting	18,000	27,000	36,000	11,250	31,500	5,400	129,150	587
Booz Allen	4,620	6,930	9,240	2,888	8,085	1,386	33,149	151

Chapter 7 Two Laws One Product Line Tamaraland

Summary

This chapter focuses on Brussels, Washington, and the firms that sell both. For theory we turn to Tamaraland, to John Krizanc's Tamara (1981) is performed in many rooms of a villa at once. The audience picks a character and walks. You can stand in only one room. The rest of the play continues without you. Boje's 1995 Academy of Management Journal study of Disney named that structure Tamara-land: an organization in which simultaneous storying-action runs in marketing, labor, the park, the executive suite, and the suppressed backstage, and no observer—journalist, worker, scholar—can traverse the whole. Tamaraland.us restates the combinatorics: for six rooms yield 6 factorial pathway possibilities which equals, 720. Twelve rooms explode to 12!, 479001600 possible pathways, as people, organizations, the low moves from room to room, chasing storylines. This is not a metaphor for distraction. It is the structural condition of living inside a storytelling organization.

The theme: Brussels and Washington: How the Consent Machine Sells Both Binding Rules and Voluntary Theater

This is the international chapter the preface promised. It is in simultaneous rooms of Tamaraland (Boje, 1995), in US, and note: “Europe is stricter.”

We aim to show that the Big Ten do not face regulation; they productize it. In the EU they sell Annex inventories, ISO 42001 badges, and “AI Act-ready” operating systems. In the U.S. they sell NIST-mapped frameworks, Responsible AI slideware, and stakeholder processes that leave NDAs on elected officials intact. Same firms, different packaging, same Dimension-Two split between embedded ethics and performed ethics.

The 2026 fact that makes the chapter timely. The AI Act is generally applicable from 2 August 2026; GPAI duties already bite; high-risk Annex III duties were deferred by the AI Omnibus toward December 2027. Enforcement has begun while the hardest obligations were postponed. That delay is itself a consulting market.

U.S. counter-room. Voluntary frameworks plus siting-by-NDA plus, in the xAI litigation, a Justice Department argument that pollution suits threaten national security. That is not “absence of regulation.” It is a different accountability architecture.

Firm cut. Big Four will look stronger here than MBB on paper (audit lineage, assurance language). Critical Theorist should ask whether assurance of a client’s model is compatible with selling that client’s transformation. Ethnographic Witness should read one EU conformity file and one U.S. “Responsible AI” PDF side by side until the distance is visible.

This is about how the Big Ten—and Capgemini, without whom the February 2026 Frontier Alliance cannot be told—as a Tamaraland of regulation. The rooms that run at the same hour are:

the Brussels / AI Office room (Act generally applicable 2 August 2026; GPAI duties live; Annex III high-risk deferred by the AI Omnibus toward 2 December 2027);

the Washington room (NIST AI RMF, sector rules, siting-by-NDA, and, in the xAI turbine litigation, a Justice Department national-security override);

the assurance room (ISO/IEC 42001 badges, SOC-style AI controls, AIUC-1, “Trustworthy AI”);

the transformation room (the same partnership selling the agent platform the assurance deck will later bless);

the judicial / watchdog room (FRC, FCA, Auditor General, NAACP, Treasury cancellation);

the beneficiary room (eligibility engines, welfare penalties, hiring screens—Annex III classes already in U.S. production without the name);

the Frontier Alliance room (OpenAI plus McKinsey, BCG, Accenture, Capgemini as certified distributors of a platform they did not build).

A partner, a client GC, a commissioner, a judge, a claimant, can occupy one of those rooms. The Consent Machine's method is to sell a coherent walkthrough of room 3 while rooms 4, 5, and 6 keep running. Dimension Two of the tesseract—ethics embedded versus ethics performed—is the distance between those walkthroughs.

The chapter's claim is narrow and researchable: the Ten do not merely face regulation; they productize it. In a veritable Tamaraland. In the Union they sell Annex inventories, EN ISO/IEC 42001:2026 management systems, and "AI Act-ready" operating systems. In the United States they sell NIST-mapped frameworks, Responsible AI slideware, and stakeholder processes that leave NDAs on elected officials intact. Same logos. Different packaging. Same incompatibility question: can a firm assure a client's model in one room while implementing that client's transformation—and, after 23 February 2026, distributing OpenAI Frontier—in another?

This is not a treatise on Articles 5–50. One Annex III family (employment, education, or public benefits) plus one U.S. eligibility engine already documented in Chapter 4 is enough to show that "high-risk" is an American production fact, not a European specialty. The delay of the hardest EU duties is itself a consulting market. Enforcement without Annex III is a second market. Voluntary U.S. frameworks beside NDA geography is a third.

How to walk this chapter

You will not see the whole villa. That is the method. Follow one firm through two rooms and you will miss the third. The six voices of the book stay in different doorways: Ethnographic Witness reads a conformity brochure against a refunded PDF; Critical Theorist asks whether assurance and implementation can share a letterhead; Community Voice stays with the person whose benefit letter was generated in a room no partner visits; Founders' Voice / PR Voice speak past each other across the Atlantic.

I. The play, the factorial, and why a statute is not a single story

Boje (1995) wrote that Disney was not one narrative but a Tamara-land: "No single audience member... can traverse more than a tiny fraction of the total narrative space. This is not a failure of attention; it is the structural condition of living in a Tamaraland" (Tamaraland.us restatement of the AMJ argument). The 1981 play had simultaneous scenes; the 1995 article had simultaneous organizational rooms; the 2026 AI Tamaraland on the same site maps White House, Commission, Pentagon, chip economy, and model labs as rooms whose pathway counts are factorial in access.

Apply that grammar here.

Let the seven rooms above be the villa. Full tours: $(7!=5\{ \}040\backslash)$. All partial sequences: $(\sum_{r=0}^7 P(7,r)=13\{ \}700\backslash)$. A Brussels lawyer who spends 2026 on Article 50 transparency guidelines does not see Loudoun County NDAs. A Virginia supervisor who

signed an NDA does not see prEN 18286. A Deloitte assurance partner who announces “end-to-end AI controls” on 12 August 2026 does not stand in the Australian department that kept TCF recommendations after a fabricated-citation refund. A McKinsey QuantumBlack slide on “safeguards” does not contain Noon’s Medicaid file. Each is a lawful, billable walkthrough. None is the play.

Antennarrative is what the official walkthrough suppresses: the deferred Annex III date as a product window; the DOJ brief that a pollution suit threatens national security; the blog that notes the Big Four sell governance while their own reports hallucinate.

II. Room 1 and Room 2: two laws, timed for sale

Brussels, 2 August 2026. Regulation (EU) 2024/1689 is generally applicable. Prohibited practices have been live since February 2025. GPAI model duties have been live since August 2025; the AI Office’s enforcement powers over those models are now active. Article 50 transparency—chatbots, synthetic content, emotion recognition—applies. The Commission’s Article 50 guidelines landed 20 July 2026. What did not land on 2 August 2026 is the obligation set the market most feared: stand-alone Annex III high-risk systems. The Digital Omnibus on AI, Regulation (EU) 2026/1744, in force 27 July 2026, moved those duties to 2 December 2027 (Annex III) and 2 August 2028 (Annex I embedded products). The architecture stayed. The calendar became a service line.

ISO/IEC 42001 became EN ISO/IEC 42001:2026. The Commission has already said 42001 is not aligned with the quality-management system Article 17 requires for high-risk systems; that work sits in prEN 18286. So the badge the Ten sell in 2026 is a real standard and not the Act’s high-risk QMS. That gap is billable twice: once for the badge, once for the “readiness” that the badge does not finish.

Washington, same season. There is no horizontal AI Act. There is NIST AI RMF (voluntary), sector law, state patchworks, federal contractor certificates, and—on land—the NDA geography of Chapter 2. In June 2026 the Justice Department told a Mississippi court that the NAACP’s Clean Air Act case against xAI’s unpermitted turbines should be thrown out because shutting them would threaten national, economic, and energy security, tying Grok to classified and military use. That filing is regulation. It is regulation as override. Seven in ten Americans, in the Gallup numbers this book has already used, oppose a data center next door. The Washington room answers opposition with framework PDFs and, when needed, a security brief.

Critical Theorist: “absence of an AI Act” is a European sentence. The American sentence is plural accountability architectures, one of which can declare a neighborhood’s air quality a strategic exception.

III. Room 7: Frontier Alliance, or how a model lab rents the villa

On 23 February 2026 OpenAI named Frontier Alliances: multi-year partnerships with McKinsey, BCG, Accenture, and Capgemini to deploy Frontier, a platform for building, supervising, and governing agents. OpenAI’s forward-deployed engineers sit beside

consulting teams. BCG's Christoph Schweizer, in OpenAI's own release, said AI "must be linked to strategy... with aligned incentives and culture" and promised "safeguards from day one." Capgemini's Aiman Ezzat called it a long-term collaboration that "will shape the future of our industry." Accenture's Lan Guan described product company + consulting company + strategy company as the way to accelerate deployment.

This is Dimension Three under pressure. The firms become certified distributors of a platform they did not train and cannot independently evaluate at model-lab depth, while selling the governance layer that is supposed to evaluate it. Capgemini must be in the chapter. Leaving it out would fake a U.S. Ten onto a European industrial integrator that already sells "EU public sector implementation."

Deloitte, PwC, EY, KPMG, Bain, Booz, IBM are not on that press page. They productize other doors: Microsoft/Azure, NVIDIA, watsonx, federal IL5. The villa has more than one vendor corridor. The structural rhyme is the same: implementation room and assurance room share a brand.

IV. Firm rooms — researchable cases (AI-era regulation, law, product)

Each case is a walkthrough you can source. None is the whole play.

Accenture — Responsible AI as a sales room

Accenture hires "Responsible AI Specialists" whose posted work (June 2026) is explicit: sell Responsible AI projects; shape operating models; engage regulators; run a Responsible AI Academy. The same network is a Frontier Alliance end-to-end integrator and the vendor locked into Canada's CEBA stack through 2028 (Chapter 6). U.S. siting work still sits inside NDA practice the firm does not make its product to dismantle.

Researchable pair. Read one Academy / RAI statement against the CEBA Auditor General file. Room 3 says trust. Room 4 says proprietary lock-in. You cannot stand in both.

Blogs and staff posts treat RAI as a practice to staff, not a veto. That is employee antenarrative, not a statute.

McKinsey — QuantumBlack insights, not notified-body files

QuantumBlack publishes "the state of AI," "decision dividends," sovereign-AI paths for Europe. Forty percent of firm work is described as AI-related. Frontier Alliance places McKinsey as strategy and operating-model partner. What McKinsey does not publish is an Annex IV technical file for a high-risk system it built for a ministry of employment. The product is the walkthrough: where to put agents, how to redesign the operating model, how to speak NIST or "safeguards."

Researchable pair. The 2025–26 sovereign-AI essay beside the U.S. data-center power notes that utilities cite (Chapter 6). One room is European competitiveness. One room is a rate case. Same letterhead.

BCG — “Safeguards from day one” as a press-release room

Schweizer’s Frontier sentence is the PR Voice. The 2025 Gaza process-failure file (Chapter 3) is the judicial-adjacent room the firm itself called a breach of standards. 2026 Daily Maverick on SARS-era Bain people inside BCG is a third room: personnel pathways the official values walkthrough does not list.

Researchable pair. OpenAI 23 February 2026 blog quote versus BCG’s own July 2025 “correcting the record” letters. Dimension Two is the distance in months, not the distance in principles.

Deloitte — Assurance product, 12 August 2026

Deloitte’s press release that week expands “AI Controls and Assurance”: governance frameworks, model validation, SOC-style reporting, “regulatory readiness,” financial-reporting controls. Market directories price Big Four EU AI Act conformity-adjacent work in the \$80,000–\$250,000+ band, governance maturity \$75,000–\$300,000. Boutiques undercut on technical model tests.

Researchable pair. That 12 August page beside the DEWR revised PDF that disclosed Azure OpenAI GPT-4o after fabricated footnotes, and beside Bridges (Chapter 4)—a U.S. public-benefits engine deciding insulin access. Annex III names employment, education, and essential private/public services. The United States already runs the class. It does not buy the Annex. It buys change requests.

Critical Theorist: the same network offers to assure AI that affects financial reporting and to transform the processes those reports describe. Independence is a room assignment, not a feeling.

EY — EY.ai in one room; retracted footnotes in another

EY sells an “operating system” and NVIDIA-aligned agentic platform. In May 2026 EY Canada withdrew a cybersecurity marketing report after fabricated citations were found—the same hallucination class as Deloitte’s DEWR file. The FRC’s Post Office audit investigation (2015–2018 books, Horizon in the background) sits in the judicial room.

Researchable pair. An EY.ai Act “readiness” page and the retracted Points of Attack study. Startup Fortune’s June 2026 line—“The Big Four are selling AI governance while their own reports hallucinate”—is blog voice. The FT sourcing underneath it is Ethnographic Witness.

PwC — Mapping duties, failing opinions

Directories describe PwC as control-mapping and audit-ready evidence for regulated industries. Agent OS is the transformation SKU. The FRC’s second Babcock fine (2026, £3.2 million) is the assurance room’s older cousin: opinions that did not challenge management on a defence contractor. Australia’s tax-leak aftermath (government arm sold for A\$1) is what happens when the confidential-advice room and the client-winning room were the same corridor.

Researchable pair. One “AI Act / NIST mapped controls” slide and the FRC Babcock decision. Not because Babcock is an AI system. Because the habit being productized—certificate as product—is the same habit.

KPMG — Eight steps, a badge, a withdrawn paper

KPMG Denmark’s July 2026 “eight steps” to the Act (inventory, classify, high-risk, etc.) is the Brussels walkthrough, updated for 2 August. U.S. slipsheets map controls to NIST AI RMF, ISO 42001, and the Act in one matrix. In late August 2026 KPMG announced AIUC-1 certification of internal agent aIQ Capture after 900+ tests—first Big Four agent so badged. In June 2026 the firm withdrew Redefining excellence in the age of agentic AI after bogus case studies (UBS, NHS, SBB, TfL) were flagged.

Researchable pair. AIUC-1 press note versus the withdrawn agentic-excellence PDF. Badge in room 3; 3C failure in room 5. Frequency from Chapter 6 is the labor-process version of the same split.

Bain — Fewer Act SKUs, more pathway politics

Bain is not the EU-Act catalogue firm. Sage and ChatGPT Enterprise are internal compression. The regulatory story is access: ten-year South African public ban after SARS; UK ban lifted; 2025 ministerial statement that there was no legal route to reimpose it; 2025 Johannesburg wind-down; 2026 reporting of SARS-era seniors at BCG. Room assignment is who may stand in the public-sector villa at all.

Researchable pair. Bain’s SARS “misconceptions and realities” page and Lord Hain / McFadden correspondence on the failed re-ban. Ethics performed as repayment; ethics embedded would have been restored capacity.

Booz Allen Hamilton — Federal room, not Annex III

Booz is the U.S. federal AI-services incumbent: FedRAMP, IL5, cleared benches, DoD/intel agents (aiSSEMBLE, Vellox). It does not need an AI Act product to live. It needs contracting law, classification, and political permission. Treasury’s January 2026 cancellation after Littlejohn (406,000 files) is the accountability event. The 7 percent civil cut is the demand-shock room.

Researchable pair. A federal “responsible AI for national security” brochure and the Treasury cancellation release. Washington is not empty of rules. It is full of a different rule.

IBM Consulting — watsonx.governance as a packaged Act

Market notes treat watsonx.governance as built-in model-risk and EU AI Act tooling, with Promontory as the regulatory bench. That is the cleanest “productized statute” SKU in the Ten: the control plane is the offering. Madrid ICAM restitution and the Barlow complaint sit in rooms 5 and 1 of another villa: the vendor’s own systems.

Researchable pair. watsonx.governance feature list (Annex-flavored language) and the Audiencia Provincial restitution. Embedded ethics would mean the control plane catches the vendor. Performed ethics means the control plane is what you sell the client.

Capgemini — The European integrator the Alliance required

Founding Frontier Alliance member. European industrial and public-sector delivery. Ezzat’s “shape the future of our industry” is PR Voice. The researchable question is the same as Accenture’s: can a firm that wires Frontier into SAP and sovereign-cloud stacks also be the independent reader of the Article 9 risk-management system for that stack? Capgemini is how the Alliance enters the Brussels room without pretending McKinsey writes code for a French ministry.

V. One Annex III class, one American engine

Annex III’s high-risk list includes employment, education, and access to essential public and private services. The Omnibus deferred the duty. It did not defer the use.

EU walkthrough the Ten will sell through 2027: inventory the system; decide provider vs deployer; draft FRIA language; map ISO 42001; wait for prEN 18286; budget a notified-body assessment in the \$25,000–\$80,000 boutique band or the \$80,000–\$250,000 Big Four band; put a partner on the steering committee.

U.S. production already running: Deloitte-operated eligibility systems (Bridges and sister platforms across many states); automated welfare-penalty frameworks of the TCF type; hiring screens; education-analytics vendors. Chapter 4’s Noon file is the beneficiary room. No Annex number appears on her denial. The risk class is the same. The product in America is not conformity. It is operate, change-request, defend.

Put those two PDFs on one table—an “AI Act readiness” proposal and a KFF/Bridges account—and Dimension Two is visible without a doctrinal essay.

Employment-class rhyme: NYC Local Law 144 bias audits (boutiques do the math; Big Four sell the governance wrapper). Education-class rhyme: student-data systems whose breach files (keep Bain Capital out of Bain & Company) show what “high-risk” means when the data subject is a child.

VI. Side-by-side: what Ethnographic Witness actually reads

Table 1. Two legal rooms and one product line: EU AI Act / Digital Omnibus versus U.S. voluntary maps, by firm SKU

A GC who buys row 1 does not see row 6. A partner who sells row 3 does not sit in row 4’s production incident. Factorial pathways are not a flourish. They are why “we are AI Act-ready” can be true in one room and false in the play.

VII. Voices, briefly, in different doorways

David’s Voice. I can attend a Magdalena moratorium, in New Mexico, and I cannot attend an AI Office enforcement conference the same afternoon. The book’s method is to refuse the single walkthrough the Ten sell.

Community Voice. You can't drink an ISO badge. You can't breathe a NIST map. You can lose a benefit to a system that would be Annex III if it had an EU deployer.

Critical Theorist. Mazzucato's question, relocated: who authorized private distributors of OpenAI Frontier to be the public's substitute for a notified body? Assurance of a model you are paid to scale is not a conflict you disclose in a footnote. It is the business model.

SEAM Analyst. Chapter 6 priced 3C and strategic-implementation events. This chapter names their regulatory form: a readiness engagement whose deliverable is a map, billed before Annex III bites, and a U.S. framework engagement that never opens the NDA room.

Ethnographic Witness. Dates: Act in force 1 August 2024; prohibited practices 2 February 2025; GPAI 2 August 2025; Omnibus 27 July 2026; general application and Article 50 2 August 2026; Annex III 2 December 2027; Frontier Alliance 23 February 2026; Deloitte assurance expansion 12 August 2026; KPMG AIUC-1 ~September 2026; xAI DOJ brief June 2026.

Founders' Voice / PR Voice. Waterhouse: "our conscience." Schweizer: "safeguards from day one." The dialogic pair does not resolve. It measures the corridor between rooms.

VIII. Research Cases

Case 7A. Two PDFs, one hour. Suppose we receive (i) a Big Four "EU AI Act eight steps / ISO 42001" brochure and (ii) the revised DEWR report's AI disclosure paragraph. Task: list what each document would need to contain for ethics to be embedded rather than performed. Ban answers that only add a principle.

Case 7B. Annex III without the Annex. Assign Bridges / Noon (Chapter 4) as if the deployer were an EU public body in 2028. Which articles would have been live on 2 August 2026 anyway (transparency, prohibited practices if any), and which wait until December 2027? Who is paid in the waiting year?

Case 7C. Frontier Alliance independence memo. What is you are McKinsey or Capgemini legal. OpenAI FDE sits on your client team. Draft the one paragraph that would make Dimension One (named accountability) real if the agent harms a worker in an employment-class system. If you cannot draft it without breaking the Alliance economics, say so.

Case 7D. Washington override. Read the DOJ xAI brief beside a Responsible AI PDF from any of the Ten that has a U.S. public-sector practice. Question: is national security a room that swallows the others, or a lawful separate play? Tamaraland does not require you to pick a politics. It requires you to admit you cannot watch both scenes as if they were one.

Conclusion

The Consent Machine in 2026 does not wait to be regulated. It staffs both sides of the door. In Brussels it sells the inventory of a law whose hardest duties were postponed, and it sells the badge (42001) the Commission has already said is not the high-risk QMS. In Washington it sells the map of a voluntary framework while land deals still run on NDAs and, when a

community sues over turbines, a security brief can try to close the courtroom. The Frontier Alliance adds a seventh room: the model lab's distribution channel wearing four consulting logos, plus Capgemini so Europe is not an afterthought.

Big Four look stronger on paper because they already sold opinions. That lineage is why they win the assurance RFP. It is also why their retracted footnotes and welfare-engine contracts are the highest-value counter-rooms. MBB look stronger in the strategy room and in OpenAI's February photograph. Booz and IBM look stronger in the cleared and hybrid-cloud rooms. Bain's story is who is allowed in the public villa at all.

No partner, no commissioner, no claimant walks all 13,700 partial paths of a seven-room villa. The official story is always a tour. This chapter's job was to name the rooms the tour skips, and to treat that skipping as the product.

Table 1

Artifact	Room it claims	Room it leaves empty
EN ISO/IEC 42001:2026 certificate	Assurance	Article 17 QMS the Commission says 42001 does not cover
NIST AI RMF "mapped" deck	Washington voluntary	NDA on the supervisor who will host the GPU hall
Frontier Alliance "safeguards from day one"	Transformation + ethics	Independent evaluation of the frontier model
Deloitte 12 Aug 2026 assurance release	Controls	Bridges / TCF production error rates
KPMG AIUC-1 on aIQ Capture	Internal agent hygiene	Withdrawn client-facing case studies
DOJ brief in NAACP v. xAI	Security	South Memphis / Southaven air and water
Article 50 chatbot label	Transparency	Whether the welfare engine is high-risk in substance

Chapter 8 Captured Witnesses

Summary

The machine is cheaper when the rooms that should audit it are already on contract. A land-grant signs \$30,000 of "desalination advisory" and an NDA that obliges help fighting public-records release. An HBCU sits in the same state geography as Colossus while campus compute is sold as destiny. Independent weeklies in Socorro and Sierra print humidity and "kept in the dark"; Doña Ana syndication tracks the announcement calendar. AICPA independence binds the attest room, not the nonattest AI implementation room. ISO 42001 and vendor-funded Responsible AI centers badge readiness the EU Commission has said is not high-risk quality management. Capture transfers the cost of inquiry onto residents. Peirce/CSI method: abduct from the surprise that the auditor is on the invoice; deduce the documents independence would require; induce against NDAs, dockets, and codes.

Keywords: Captured witnesses; land-grant NDA; weeklies; AICPA; ISO 42001

The Consent Machine is cheaper when the institutions that are supposed to test it are already on contract. This chapter does not retell the siting cases of Chapter 2, the control-plane cases of Chapters 3–4, the recruiting inventory of Chapter 5, the SEAM ledgers of Chapter 6, or the two-law product line of Chapter 7. It asks a narrower question those chapters make visible but do not answer: who authorized the rooms that should audit the machine to rent their legitimacy?

The rooms are three, and they run at the same hour. In the university room, a land-grant signs a \$30,000 desalination advisory and an NDA that obliges it to help a developer oppose public-records release; a historically Black university sits in the same state as Colossus and contemplates its own campus compute building while its knowledge workers are trained for the firms that will later invoice the transformation. In the newsroom, independent weeklies in Socorro and Sierra Counties put water numbers and town-hall testimony on the front page; syndicated uniformity and NDA-covered officials in Doña Ana make the same industrial event look like economic development. In the profession room, AICPA independence language, ISO/IEC 42001 badges, IRB clearance, and vendor-funded “Responsible AI” centers sell the appearance of audit while the audit object is also the client.

The method is not confession and not a control plane. It is Charles Sanders Peirce’s self-correcting triad—abduction, deduction, induction—as operationalized in Conversational Storytelling Interviewing (CSI): persist in the method, correct the premises, remove what is fake, illusory, or obfuscating, and move gradually closer to what is true. The control plane is not the duty (Chapter 3). A university research center, a newsroom stylebook, an ethics code, and an ISO certificate are control planes. Duty is named answerability for the water, the footnote, the hour, and the comment letter that was never written.

How to walk this chapter (and what not to walk again)

Tamaraland still holds (Preface; Chapter 7). You can stand in only one doorway. The rest of the play continues. The prior rooms already documented are:

Land and NDA geography — Loudoun, Shreveport, South Memphis, Tucson, and the New Mexico contrast of Doña Ana and Sierra (Chapter 2).

Internal AI stacks and lived cases — Lilli, Zora, Agent OS, the refunded footnotes, the locker, the eligibility engines (Chapters 3–4).

Who walks in the door — target-school lists as outsourced HR (Chapter 5).

Who pays after the invoice — Ledgers I and II, lines A–D (Chapter 6).

Two laws, one product line — AI Act readiness SKUs beside NIST maps and NDA siting (Chapter 7).

This chapter adds the witness rooms: the university that should measure the aquifer, the newspaper that should print the NDA, the professional body that should refuse the dual mandate. Capture of those rooms is a hidden-cost category Chapter 6 named but did not price as an institution: journalism weakening, loss of democratic voice, apprenticeship breakdown. Here they become the object.

Critical Theorist's question, relocated from Mazzucato: consulting does not merely advise. It rents public legitimacy—from campuses that supply recruits and research cover, from newsrooms that reprint economic-impact language, from professions that convert independence into a nonattest product. Who authorized that rental?

Reflexivity: the control plane is not the method

David's Voice. I will not write a confessional chapter. I live in two New Mexico counties. I was a Regents Professor at New Mexico State University. I am a visiting professor at Fisk University. NMSU took the \$30,000. Those facts are coordinates, not credentials. Situated witnessing is useful only if it is submitted to a method that can refute it.

The method is at . Conversational Storytelling Interviewing is not interrogation and not a slide. It is Peirce's claim that inquiry, persistently applied, is self-correcting:

A properly conducted inductive research corrects its own premises.
— Peirce, Collected Papers 5.576, as used in Boje and Rosile (2020)

Persist in this same method of research, and we shall gradually be brought around to the truth. This gradual process of rectification is in great contrast to rudimentary induction, where the correction comes with a bang.
— Peirce, Collected Papers 7.115, as used in Boje and Rosile (2020)

The triad, used here as an institutional test rather than an interview protocol:

Abduction. A surprising fact generates a hypothesis. Surprising fact: the institutions designed to audit siting, water, footnotes, and independence sign NDAs, take advisory fees, badge ISO 42001, and reprint developer job counts. Hypothesis: the Consent Machine is cheaper when those institutions are already distribution channels.

Deduction. If the hypothesis is true, we should find named contracts, named NDAs, named dual mandates (assure and implement), and a measurable difference between rooms that still print counter-stories and rooms that do not.

Induction. Test against documents already in this book's file and against documents that can be pulled: the NMSU–BorderPlex agreement and NDA; IPRA releases; El Defensor Chieftain and Sierra County coverage versus Doña Ana syndication; AICPA independence rules for nonattest AI work; ISO 42001 product pages already mapped in Chapter 7; the NAACP docket in the Colossus geography.

Peirce's hierarchy of tests, cheap to costly: (1) self-reflection; (2) multiple conversations with multiple people; (3) another science—here hydrology, environmental justice, public-

records law, and SEAM hidden-cost accounting; (4) controlled experiment, which this chapter does not claim. CSI begins at tests 1 and 2 and refuses crude induction: the official story's beginning-middle-end ("advisory work," "Responsible AI," "community investment") is not allowed to predict the next case.

Bakhtin's distinction travels with the triad. Special answerability is the professional domain: the IRB protocol, the AICPA attest rule, the ISO clause, the newsroom embargo. Moral answerability is the postupok—the deed for which no badge substitutes. Chapter 3's sentence applies without remainder: the control plane is not the duty. CSI's job in this chapter is to remove four classes of obstruction:

Fake — comments submitted under names that were not written; citations that were not checked (Chapter 3).

Illusion — a \$30,000 "advisory" framed as public science while the NDA binds the scientist to the developer's records strategy.

Obfuscation — ISO 42001 sold as AI Act quality management after the Commission has already said it is not that QMS (Chapter 7).

Premature narrative — "jobs, closed-loop, community benefits" hardened before water, grid, and error-rate numbers enter the public file (Chapter 2).

No voice in this chapter is licensed to win. Founders' Voice / PR Voice will speak. Ethnographic Witness will read the PDF. Community Voice stays with the sign in Magdalena. SEAM Analyst will not invent a new billion-dollar cell; capture of witnesses is a transfer of the cost of inquiry onto the public.

I. University room: the land-grant on the Jupiter map, the HBCU in the Tennessee compute geography

I.1 NMSU — advisory fee, NDA, presumption of secrecy

The document, not the memoir.

On 10 February 2025, New Mexico State University and BorderPlex Digital Assets signed a nondisclosure agreement. On 25 February 2025 they signed a fixed-price agreement: \$30,000 for "advisory work and counseling" on a desalination project styled in the contract as work connected to "Project Nucleus," the digital-infrastructure campuses BorderPlex was developing in Santa Teresa—the geography the public later learned as Project Jupiter.

Attachment A of the agreement names the work: participation in the Office of the State Engineer's investigation of the Santa Teresa brackish resource; assistance with piloting treatment technology; feedback to EPCOR on permitting feasibility, path, timeline, and costs. The technical lead sits in civil engineering. That is real hydrology. It is also a client file.

The NDA is the instrument Chapter 2 already theorized at county scale. Applied to a public university it does something the audit report used to do: it relocates the boundary of what may be known. Documents may be labeled proprietary. If law or official order requires release, NMSU must notify BorderPlex before divulging them. If the company decides to oppose release, NMSU must “provide reasonable aid and assistance” in those efforts.

Norm Gaume, reading the same papers for the New Mexico Foundation for Open Government, stated the antenarrative the official “advisory” narrative suppresses:

NMSU’s nondisclosure agreement with BorderPlex Digital creates a presumption of secrecy when state law already defines what may be withheld and what must be released. That undermines public confidence.

IPRA already protects trade secrets. The NDA is not a restatement of IPRA. It is a contractual duty to help the developer fight IPRA. Dimension One of the tesseract—accountability architecture—moves from named public science toward diffused private process. Dimension Two—ethics architecture—moves from embedded land-grant duty toward performed partnership. Dimension Four—temporal horizon—exits before the aquifer’s multi-decade draw is a measured public number.

SEAM Analyst. Thirty thousand dollars is not the cost. It is the invoice. Below the waterline (Chapter 6, Figure 5 logic, not the tables): the public pays in delayed or contested records; in water figures that enter the county file as developer claims rather than independent design; in a university that cannot be the hostile reader of the project it was paid to midwife. Hidden-cost categories already named: journalism weakening (the campus cannot speak), loss of democratic voice, future infrastructure costs transferred to ratepayers and the desert.

Critical Theorist. Mazzucato’s public-value question is exact here. A land-grant’s legitimacy is a public asset: experiment-station authority, water engineering, the right to be believed when it speaks about an aquifer. Who authorized the Regents to rent that asset to a developer whose own comment process would later be investigated for names submitted without consent?

The AG investigation is not this chapter’s case; it is the adjoining room (Preface; Chapter 2). Attorney General Raúl Torrez opened a probe after residents and elected officials found form letters supporting Project Jupiter’s air-quality permit under their names. Canvassers in project T-shirts had collected identities at gas stations and grocery stores. Torrez: if parties “attempted to manipulate the public comment process through fraudulent submissions, we will investigate those actions thoroughly.” NMED later said it had been notified that on the order of 42 of more than 17,500 portal comments “may have been fraudulent.” Elevate New Mexico, a Virginia corporation, later settled an Ethics Commission case after filings showed nearly \$3.7 million in contributions, all from the developer, Yucca Growth Infrastructure.

CSI deduction: if the university is an independent witness, its public file on water should be readable without the developer’s permission and should not be timed to the approval

calendar. Induction: the NDA text fails that test. Correction of the premise: “desalination advisory” is not a complete description of the act. The act is witness capture at land-grant scale.

Chapter 5’s recruiting regime sits in the same doorway and will not be retold. Target-school lists already make the campus a distribution channel for the Ten. The \$30,000 file adds a second channel: research and public-science cover for siting. AI-literacy curricula and industry research centers are the third. None of these is wicked in isolation. The triad asks what happens when all three run at once while the same logos sell assurance in Chapter 7.

1.2 Fisk — HBCU knowledge room, Tennessee compute geography, no invented contract

Fisk University is in Nashville. Colossus is in South Memphis and Southaven. They are not the same campus and this chapter will not pretend they are. They are the same state geography of compute and environmental justice, and they are the same class of institution Chapter 5 already treated as a pipeline: universities that produce the people and the legitimacy the machine needs.

Two researchable facts, not a memoir:

Fisk is an HBCU whose intellectual tradition includes a long fight over who may speak for Black Nashville and Black Memphis. Colossus 1 occupies a former Electrolux plant in a majority-Black South Memphis industrial landscape already carrying pollution burdens; Colossus 2 and its turbines sit across the state line. The NAACP and Mississippi State Conference of the NAACP sued over unpermitted turbines; the Justice Department moved to intervene and dismiss, arguing that shutting the plant would threaten national, economic, and energy security. That filing is Chapter 7’s Washington override. This chapter’s point is narrower: the civil-rights organization is doing the audit work that universities, newsrooms, and professions in the same region are structurally tempted to productize or decline.

In 2026 Fisk announced a campus “Quantum Leap” innovation center that includes a data-center component on undeveloped acres along Herman Street. Neighbors petitioned. A mother told WSMV: “Progress is only progress if we’re here to see it.” The parallel to NMSU is not dollar-for-dollar. It is institutional class: an HBCU asked to host compute as destiny while the state’s other compute cluster is already a sacrifice-zone file (Chapter 2, South Memphis).

Community Voice. You cannot drink a visiting professorship. You cannot breathe a Quantum Leap rendering. South Memphis asked for air permits and got a security brief. Nashville neighbors asked whether an HBCU campus is a place for a hall of servers. Those are different rooms. They share a question: is the university the auditor of the stack or a square on the siting map?

Founders’ Voice / PR Voice will call both files partnership and workforce development. CSI requires the counter-story from the NAACP complaint and from the Herman Street petition before that language hardens.

No Fisk–BorderPlex contract appears in the public file this chapter can cite. Self-correcting method forbids inventing one. The abduction is analogical, not documentary: HBCU legitimacy is as rentable as land-grant hydrology. Deduction: if capture is occurring, we should later find NDAs, naming rights, or curriculum contracts that bind Fisk’s public voice the way NMSU’s NDA binds IPRA. Induction: watch the file. Do not write the ending first.

II. Newsroom: independent weeklies versus syndicated uniformity

Chapter 2 already used the New Mexico contrast as connective tissue. This section is only the witness technology.

Abduction. Two counties, same storm, different copy. Hypothesis: the machine is cheaper where local journalism cannot force water, power, and NDA text into the public file before the announcement.

Deduction. If true, we should find (a) named independent papers that printed opposition, petitions, and moratoria as news rather than as delay; (b) a Doña Ana pattern of developer-timed announcements, career-fair town halls, and comment-letter manufacturing; (c) FOIA/IPRA friction as a design feature, not a glitch.

Induction — Sierra / Socorro file. El Defensor Chieftain (Volume 158 in the Preface’s witnessing; 2026 reporting under Jessica Carranza Pino and staff) covered Green Data, New Mexico Tech, Magdalena town halls, and the Inform Socorro petition. Residents said they were “kept in the dark.” Cecilia Rosacker: “We have what is disappearing everywhere else.” Village trustee Teri Winchester: “We are the ones that will suffer the consequences.” Humidity numbers were argued in public against atmospheric-water claims. The Socorro County Commission then voted a one-year moratorium; Sierra County, after packed comment, amended a resolution into an 18-month moratorium. Signs in the room said what the official economic-impact PDF would not: you cannot drink data.

That is not a romance of small papers. A later secret recording showed an attempt to amend the Sierra moratorium so a Spaceport-adjacent proposal could proceed. Independent reporting made the amendment visible. Visibility is the condition Chapter 2 listed among the five NDA-enabled consent conditions—and among the five refusal conditions Chapter 9 will have to operationalize.

Induction — Doña Ana file. Project Jupiter arrived as ads, industrial-revenue architecture, and a 500-page draft proposal dropped close to a take-it-or-leave-it commission date. County staff had negotiated under NDA; commissioners were late to the file. Independent outlets (Haussamen, Source New Mexico, later the Journal and New Mexican when they assigned the water number) had to extract the NMSU NDA by records request. The comment portal then filled with letters residents said they did not write.

Ethnographic Witness. Compare two artifacts on one table: an El Defensor Chieftain town-hall story that prints humidity, petition links, and “kept in the dark,” versus a developer

release that prints 2,500 construction jobs, 750 permanent jobs, \$50 million for water systems, \$360 million in PILT. Both are public speech. Only one is an audit room.

NDAs on officials and universities are a journalism-killing technology because they attack the source, not the reporter. You cannot quote what the supervisor is forbidden to say. You cannot force the desalination design into the paper if the land-grant must first warn the developer and then help fight the release. FOIA/IPRA friction is then billed as delay, and delay is billed as anti-jobs.

SEAM Analyst. Chapter 6 already listed journalism weakening below the waterline. This chapter names the mechanism: pre-announcement secrecy + post-announcement volume. The hidden cost is not “bias.” It is the transfer of inquiry costs onto residents who must reconstruct the file after the bond ordinance.

Gallup’s seven-in-ten opposition to local data centers (Preface) is the public that a captured newsroom never quite gets to hear in time.

III. Profession room: codes, badges, IRBs, and centers that evaluate their funders

Chapter 7 showed the Ten productizing two legal rooms. This section is the profession that licenses the product.

III.1 AICPA and the nonattest door

The AICPA Code of Professional Conduct still requires independence in fact and appearance for attest work. It does not require independence when the member provides only nonattest services—consulting, systems, “Responsible AI,” implementation. In 2026 the Journal of Accountancy had to remind members that competence, due care, confidentiality, and independence “don’t disappear just because new technology is involved,” and that building or hosting AI for an attest client can raise independence threats under existing nonattest and information-systems rules. SOC engagements with tool-provider side deals create the same split: the code says do not let the vendor set scope, evidence, or deficiency language; the market sells speed.

Critical Theorist. This is Dimension Two as a licensing structure. Special answerability (attest independence) is real and old. Moral answerability would ask whether the same network may assure a client’s model in the morning and implement the client’s agent platform in the afternoon—the incompatibility Chapter 7 already posed for Deloitte, PwC, EY, KPMG, and IBM. The Code’s answer is a room assignment: independence is required in the attest room. The Consent Machine lives in the nonattest room, then borrows the attest brand.

CSI correction: “we follow the Code” is not a refutation if the Code’s door is the product.

III.2 ISO/IEC 42001 as a traveling certificate

Do not reprint Chapter 7's Act calendar. One sentence of carry-forward: EN ISO/IEC 42001:2026 is a real management-system standard; the Commission has already said it is not the Article 17 high-risk QMS. The badge is therefore billable twice—once as certification, once as “readiness” the badge does not finish.

The profession room adds the evaluator's evaluator. KPMG Canada announces itself among the first in the country to hold 42001 and immediately translates the badge into client help “as they look to scale AI.” Deloitte, PwC, and a long tail of boutiques sell gap assessments, AIMS build-outs, and third-party assurance mapped to 42001. The standard's own text is for any organization that provides or uses AI. The consulting market's text is a SKU.

Ethnographic Witness. Read a 42001 certificate against KPMG's withdrawn agentic-excellence paper and Deloitte's refunded DEWR footnotes (Chapters 3–4, 7). The badge does not see those files. That is not a failure of ISO drafting. It is what a management system cannot do when the firm's other room is paid to scale the object being managed.

III.3 IRBs and Responsible AI centers

University IRBs are special answerability for human-subjects research. They are not moral answerability for a \$30,000 siting advisory, an NDA, or an industry-funded center that writes the curriculum on the same vendors that write the center's gift agreement. An IRB can pass a survey of worker attitudes toward Copilot and never open the CEBA lock-in file or the Bridges eligibility engine (Chapters 4 and 6).

Vendor-funded “Responsible AI” centers and AI-literacy institutes are the academic twin of Chapter 7's assurance SKU. They produce maps, principles, and fellows. Mazzucato and Collington already described the older form: consultancies build quasi-academic institutes to look like knowledge hubs while recruiting the same campuses' graduates as inventory. The AI-era version adds a second rent: the center evaluates “the industry” with industry money and calls the output governance.

David's Voice, still not confession. I have taken university salaries and visiting titles. CSI test 1 is self-reflection: that does not make the NMSU NDA acceptable. Test 2 is other conversations: Gaume, Magdalena, Source NM, the NAACP complaint. Test 3 is another science: Savall's hidden costs and the State Engineer's aquifer file. If a center I sit in cannot survive those tests, it is a control plane.

IV. Synthesis: why capture makes the machine cheaper

The Ten do not need to buy every newsroom or every department. They need enough rooms quiet enough that the official walkthrough can be sold as the play.

Table 1. Captured witness rooms: what the public thinks it buys, what the contract often buys, tesseract shift, and SEAM transfer

Critical Theorist. Public value is not what the press release calls community benefits. Public value, in Mazzucato's market-shaping sense, is whether public institutions still have the

capacity to set direction—water, grid, eligibility, air—without renting the map from the party that will invoice the build. Capture is the sale of that capacity.

SEAM Analyst. Chapter 6's mid Ledger I was process hidden cost inside the Ten. Capture of witnesses is outside the partnership and therefore almost never invoices. It appears as: fewer IPRA-ready files; fewer hostile local stories before first reading; ethics codes that stop at the nonattest threshold; centers that cannot bite the hand that endowows them. Ledger II's 167 percent logic still applies: the visible fee (\$30,000; a center gift; an ISO engagement) is not the sociotechnical cost of a public that must reconstruct the truth after the fact.

Community Voice. Fraudulent comment letters are what consent looks like when the newsroom and the university are late. A moratorium is what refusal looks like when a weekly and a packed room arrive on time.

V. Voices in different doorways (short summary)

David's Voice. The method forbids me to treat NMSU as "my" scandal or Fisk as "my" redemption. I am a witness who must be correctable. If a later IPRA dump shows the \$30,000 bought only public hydrology and the NDA was never invoked, the induction changes. Until then the text of the NDA is the deed.

Community Voice. Magdalena printed humidity. South Memphis filed for air. Doña Ana found its names on letters it did not write. Those are not illustrations. They are the data.

Critical Theorist. Who authorized these institutions to rent their legitimacy? The Regents, the publisher's chain of title, the professional institute, the center's gift committee. Authorization is a document. Demand it.

SEAM Analyst. Price the silence. If you cannot, at least stop calling the invoice the cost.

Ethnographic Witness. Dates that survive a hostile read: NDA 10 Feb 2025; advisory 25 Feb 2025; IPRA release 14 Sep 2025; Torrez investigation July 2026; NMED \$3.7 million from Yucca Growth); Socorro moratorium June 2026; Sierra 18-month moratorium 21 July 2026; NAACP suit April 2026; DOJ intervene/dismiss brief June 2026; Fisk campus-center reporting June 2026.

Founders' Voice / PR Voice. Partnership. Closed loop. Responsible AI. Workforce. Safeguards from day one. The dialogic pair does not resolve. The distance between those words and the NDA's "aid and assistance" clause is Dimension Two.

VI. Research cases

Case 8A. Two contracts, one hour. Place (i) the NMSU–BorderPlex statement of work and NDA beside (ii) NMSU's public research mission language. List what would have to be in the file for ethics to be embedded: public water model, no aid-and-assistance clause, IPRA processed as routine, hydrology publishable without developer veto. Ban answers that only add a principle.

Case 8B. Two front pages. One El Defensor Chieftain moratorium story; one Project Jupiter economic-impact release. Score each on Peirce’s test 2: how many counter-stories, named sources, and numbers that can fail. Then ask what FOIA friction would have done to each score.

Case 8C. Attest door, non-attest floor. You are an AICPA member whose firm (a) holds 42001 and (b) implements an eligibility engine of the Bridges/Noon class (Chapter 4). Draft the one paragraph that makes Dimension One real if the engine false-denies insulin. If the paragraph breaks the nonattest business model, say so. That is the Chapter 3 standard applied to the profession.

Case 8D. HBCU without an invented invoice. Using only public Fisk and NAACP/Colossus documents, write the abductive memo a board committee should have required before hosting campus compute in a state already in turbine litigation. Mark every sentence that is not yet evidenced.

Conclusion

Universities, newsrooms, and professional bodies are not “stakeholders” in the consulting slide. They are the rooms that make consent look like knowledge. When those rooms are on contract—\$30,000 and an aid-and-assistance NDA; a newsroom that cannot see the file until after first reading; a code that requires independence only in the attest room; a badge the regulator has already said is not the high-risk QMS—the machine does not have to win the argument. It has already bought the witness.

Self-correcting inquiry is the opposite of that purchase. Abduct from the surprise that the auditor is on the invoice. Deduce the documents that must exist if independence is real. Induce against IPRA PDFs, weeklies, dockets, and codes. Correct the premises. Remove the fake comment, the illusory advisory, the obfuscating badge. Persist.

Chapter 9 takes the rooms where the purchase failed—Magdalena, Tucson, a commission that would not sign—and asks what refusal does to the four axes of the tesseract. This chapter’s only claim is prior: repair cannot be a slide written by the rooms that were for sale.

Table 1

Witness room	What the public thinks it buys	What the contract often buys	Tesseract shift	SEAM transfer
Land-grant / HBCU	Independent measurement, curriculum as public good	Advisory fee + NDA + recruiting channel + naming of “innovation” space	Dim. 1 named → diffused; Dim. 2 embedded → performed	Inquiry cost → residents; water/grid cost → later ratepayers

Independent weekly	Timed facts before the vote	Front-page counter-story (when it survives)	Dim. 4 lengthens	Refusal condition (Ch. 9)
Syndicated / understaffed metro	Coverage	Volume that tracks the announcement calendar	Dim. 4 shortens to the ad cycle	Journalism weakening
AICPA / ethics code	Independence	Independence in the attest room only	Dim. 3: judgment → distribution	Dual mandate as unpriced 3C
ISO 42001 / RAI center	Audit of the model	Badge + readiness engagement	Dim. 2 performed	Readiness billed before consequence
IRB	Protection of subjects	Clearance of a study, not of a siting	Special ≠ moral answerability	Harm elsewhere in the villa

Chapter 9 When Consent Machine Fails

Summary

Refusal is storying, not a beginning–middle–end victory. Magdalena/Socorro/Sierra and Tucson show the five NDA conditions inverted: a weekly that prints water before the vote; officials who will not sign; numbers in the public file before the announcement; a plaintiff with standing; a lag the developer cannot collapse. Tucson’s 7–0 killed city annexation and city water on 6 August 2025; the county still sold land, the utility sold power, site work followed. Remedies that move an axis: ban aid-and-assistance NDAs on public bodies; SEAM-style ledgers before first reading; outcome clawbacks and beneficiary reversibility; an external hidden-cost scorecard that ignores press releases; apprenticeship as a costed line. A 2027 partnership must stop selling yes-frameworks, dual assurance/implementation, and badges in place of duty. The book stays in the middle: you cannot drink data, and the aquifer does not read the slide.

Keywords: Refusal; storying; moratoria; claw-backs; no BME ending

This is the last chapter. I refuse a neat ending. The is what I call ‘storying’ that is in the middle, not clear beginning, and not end in sight. It is when the Consent Machine fails to gain closure of a classical narrative.

This is not a victory chapter. There is no happy ending to make the reader feel comfortable, maintain their alibi as a bystander. Tucson and Sierra County are not uplifting endings. These are cases that demand what Mikhail Bakhtin calls moral answerability. A reflection on our complicity, yours and mine.

There are partial proofs that the five NDA-enabled conditions of consent manufacturing named in Chapter 2 can be broken in one room of the villa while the play continues in the

next. A Beginning-Middle-End emplotment would read: community organizes, council votes no, credits. That is what PR hacks sell. Antenarrative lives before that story gets fixed—before the characters are chosen, before the events are selected, before Beginning-Middle-End imposes its linear shape on the living complexity of organizational experience (Boje, 2001, chap. 1).

The method of this book has been a sequence, not a plot: ledger → law → price → witnesses → refusal. Hidden costs first (Chapter 6), so regulation has something to govern (Chapter 7). Institutions that should audit fourth (Chapter 8), so reform is aimed at captured rooms, not only at a consulting website. This chapter is the refusal room. It offers a theory of failure modes of consent manufacturing, then a short statute-and-contract agenda. Five operational conditions of refusal, already visible in Magdalena/Socorro/Sierra and in Tucson's city chamber, are translated onto the four axes of the tesseract. None of them is a framework PDF.

Story and storying are not the same act. Story is the relatively stabilized narrative artifact—the finalized BME emplotment literary theory has studied for centuries. Storying is the ongoing, unfinished, polyphonic process through which fragments, interpretations, values, and possibilities become narratable. An AI system, and a consulting engagement that runs on one, can produce an exceptionally coherent story while concealing much of the storying through which that artifact became possible. The gap between the polished story and the suppressed storying is where accountability lives.

No voice gets the last word unchallenged. Founders' Voice speaks last only because the chapter must specify what a 2027 partnership would have to stop selling. That list is not nostalgia for Waterhouse, Bower, or Henderson. It is a contract constraint.

How to walk this chapter (what is already paid for)

Do not retell Loudoun, Shreveport, South Memphis, or the NMSU NDA. Chapter 2 mapped NDA geography; Chapter 6 priced the iceberg; Chapter 7 dated the two-law product line; Chapter 8 named the captured witnesses. Date-stamps that must not slide:

Tucson Mayor and Council, 6 August 2025: unanimous direction to cease annexation and negotiations for Project Blue.

Digital Omnibus on AI, Regulation (EU) 2026/1744, in force 27 July 2026: Annex III high-risk duties moved to 2 December 2027 (not "August 2026 high-risk deadline").

AI Act generally applicable, Article 50 live: 2 August 2026.

Socorro County moratorium: June 2026.

Sierra County 18-month moratorium: 21 July 2026.

Pima County land sale on Project Blue closed: 24 December 2025; site work reported April 2026.

A chapter that treats 6 August 2025 as the end of Project Blue is already a BME fake. The city room refused water and annexation. The county room sold the dirt. The utility room sold the electrons. Tamaraland does not permit a single walkthrough to stand for the play.

I. Story, storying, and why BME is the machine’s favorite genre

Story is the artifact you can invoice: the economic-impact study, the Responsible AI deck, the “safeguards from day one” sentence, the closed-loop water slide, the jobs number. It has a beginning (the client’s problem), a middle (the engagement), and an end (productivity, investment, community benefits).

Storying is the process the artifact suppresses: the NDA signed before the announcement; the comment letter submitted under a name that was not written; the junior who cannot book the hour; the aquifer model that never entered the public file; the RLHF layer and the pretraining corpus no partner can independently evaluate; the room you were not in.

The 1991 Administrative Science Quarterly study of GoldCo, an office-supply firm, is the empirical warrant for refusing BME as method. Eight months of recorded conversation—more than four hundred hours of tape, transcribed by hand—yielded one story with a clean beginning, middle, and end. Everything else was terse telling: fragments, codes, and references insiders did not bother explaining, because there were no newcomers in the room who needed it spelled out, and because even where there were newcomers, regulars often preferred them not to know. A terse telling is “an abbreviated and succinct simplification of the story in which parts of the plot, some of the characters, and segments of the sequence of events are left to the hearer’s imagination” (Boje, 1991, p. 115). The terser the telling, the more the wrong people are kept out. A researcher who looks only for fully formed stories will miss ninety-nine percent of what is happening. A consultant who sells only fully formed stories is counting on that miss.

Glossing is the insider’s fill-in. The Consent Machine glosses “advisory,” “closed-loop,” “AI Act-ready,” and “national security” for rooms that already share the code. Refusal begins when a newcomer—a weekly, a council, a plaintiff—forces the gloss to be spoken in public.

Antenarrative is the double claim: before the narrative, and a bet on a future not yet emplotted (Boje, 2001). The Seven B’s are the analytic vocabulary for what exists before the story fixes itself (Boje, 2001, 2014; Boje & Rosile, 2016; antenarrative.com):

Table 1. Seven antenarrative B’s and how consent manufacturing closes each one

The Ghost Vortex is what happens to that vocabulary when the storying organization is also an artificial intelligence system—when the seven processes are computational as well as institutional, and when the storying itself is inaccessible to the people most affected by the output. I do not treat AI storying as human consciousness. I use storying as an analytic term for the process through which distributed representations—weights, tools, vendors, NDAs, rate cases—become an externally available story artifact. At any given moment only a sliver of that activity is available to deliberate reason. The polished story the system (or the firm) delivers is that sliver. Accountability lives in the rest.

PR linearizes. Antenarrative stays in the middle.

II. Failure modes: the five NDA conditions, inverted

Chapter 2 named five conditions of NDA-enabled consent manufacturing. Restated here only as the inverse—the conditions under which the machine fails to close—in operational language, not values language.

Table 2. Five inverted NDA conditions (failure modes of consent manufacturing) and the tesseract axis each remedy moves

Condition 1 — Independent local reporting.

A paper that can print humidity, petition links, and “kept in the dark” before first reading. El Defensor Chieftain and the Sierra County coverage did this; Doña Ana’s early syndicated pattern did not (Chapter 8). Without that room, water numbers stay above the waterline as branding.

Condition 2 — Officials who will not sign.

A supervisor, a councilmember, a university counsel who refuses the NDA or the aid-and-assistance clause. Tucson’s 7–0 was this condition in the city room: annexation and city water were refused. Pima County’s sale was the adjoining room in which the condition failed. Sierra’s commission amended a weak resolution into an 18-month moratorium after packed comment—then faced a later push, captured on a recording, to carve an exception. Refusal is a deed, not a mood. It has to be repeated.

Condition 3 — Water and energy numbers in the public file before the announcement. Not after the bond ordinance. Not inside a 500-page draft marked draft. Tucson’s fight became possible when reclaimed-water demand and cooling load could be argued in chambers. Magdalena’s fight became possible when atmospheric-water claims met desert humidity in a village meeting. Project Jupiter’s early file did the opposite: announcement first, aquifer arithmetic later, NDA on the land-grant in between.

Condition 4 — A civil-society plaintiff with standing.

NAACP and the Mississippi conference in the Colossus turbine docket; No Desert Data Center Coalition in Pima County (suit filed 14 January 2026; dismissed; the dismissal is part of the file, not a footnote). Standing is not a vibe. It is a docket number. Without it, the Washington override room (Chapter 7’s DOJ brief, June 2026) speaks alone.

Condition 5 — A time lag the developer cannot collapse.

The fait accompli is a temporal weapon: Shreveport’s announcement as done deal; Jupiter’s take-it-or-leave-it night; Colossus online before the neighborhood had a permit story. Moratoria (Socorro one year; Sierra eighteen months; Tucson’s cease-work direction plus a promised large-water-user ordinance) are devices for reopening Before. They do not end the Bet. They stop the developer from choosing the ending in the same week as the reveal.

These five are failure modes of manufacturing, not a community-organizing manual. The machine fails when even one condition flips and the other rooms cannot immediately absorb the loss. It does not fail when a press release says the firm “heard the community.”

Critical Theorist. Consent manufacturing is a production function with five inputs. Remove an input, output falls. That is the theory. The rest of the chapter is the specification.

III. Two geographies that are not endings

III.1 Magdalena / Socorro / Sierra — the sign, the weekly, the moratorium, the recording

Community Voice keeps the sign: You can’t drink data. That is terse telling. Insiders in the chamber did not need a three-act play. The Chieftain printed the town hall, the petition, the humidity argument, the “kept in the dark” sentence. New Mexico Tech’s partnership pause and the county votes followed. Sierra then did something rarer: it converted opposition language into an 18-month stop.

The antenarrative that BME would cut: a later private meeting, recorded, in which a Spaceport-adjacent proposal and an amendment to the moratorium were discussed. Refusal in July is not refusal in September unless Condition 2 is still staffed. Being continues. Becoming-language (“we are studying,” “we can carve an exception for the right project”) will try to re-emplot the pause as a beginning.

SEAM Analyst. The moratorium’s value is temporal-horizon inventory: eighteen months in which water, grid, and displacement can be forced onto a ledger before a subsidy is booked. If those months produce only another economic-impact PDF, the pause was theater.

III.2 Tucson — one room refused; three rooms did not

On 6 August 2025, Tucson Mayor and Council directed staff to cease annexation efforts and negotiations for Project Blue. The vote was unanimous. Chambers were full. Vice Mayor Lane Santa Cruz: the project “didn’t align with our values, our climate reality, or the long-term economic vision our community has consistently called for.” Councilwoman Santa Cruz in chamber: “Project Blue is not about Tucson’s future. It’s about using Tucson’s land, Tucson’s water, Tucson’s people as means to someone else’s ends.” Developer Beale called the decision disappointing despite “plans directly compatible with Tucson’s Climate Action and Adaptation and One Water plans.” AWS said it had no commitments in place.

That is Condition 2 and Condition 3 in the city doorway: no annexation, no city water, numbers argued in public, independent reporting (Arizona Luminaria, local broadcast, Sentinel) that named Amazon when the county still said “advanced technology.”

The rooms that kept running:

County land. Purchase-and-sale on the 290 acres near the fairgrounds; close reported 24 December 2025.

Electrons. Arizona Corporation Commission approval of a long energy-supply agreement with Tucson Electric Power after the city refusal; cooling flipped toward air when city reclaimed water was denied—Amazon interest reported withdrawn, other end users rumored.

Dirt. Contractor work reported the week of 13 April 2026.

Court. Coalition open-meetings suit, 14 January 2026; dismissed.

Ethnographic Witness. “Tucson rejected Project Blue” is a story. The storying is: city water refused; county parcel conveyed; utility contracted; lawsuit lost; blades in the ground. Supervisor Cano called a later benefits text “a press release with an exit door.” That sentence is antenarrative method in one line.

David’s Voice. I will not write Tucson as hope for Doña Ana. I will write it as a map of which lock moved.

IV. Five remedies that change an axis (statute and contract, not slides)

Each refusal condition becomes a remedy. Each remedy is aimed at one primary tesseract axis. If it does not change a coefficient in Chapter 6’s ledgers or Chapter 7’s product rooms, it is theater.

1. Ban or sunlight NDAs on public officials and public universities for siting, water, and power

Axis: Accountability architecture (Dimension One).

Operational text, not principle:

No NDA, confidentiality rider, or “aid and assistance against disclosure” clause may bind an elected official, a public employee, or a public university in matters of data-center or AI-infrastructure siting, water, or power.

Trade secrets remain withholdable under existing public-records law (IPRA, FOIA, state analogs). The statute forbids contractual expansion of secrecy and forbids the public body to litigate for the developer.

Violation: void clause; public-records fee-shifting to the private party that procured it; named official on the cover page of the next filing.

This is Chapter 8’s NMSU file turned into a prohibition. Special answerability (counsel’s caution) is not moral answerability (the aquifer).

2. SEAM-style public ledgers above a threshold

Axis: Temporal horizon (Dimension Four).

For any subsidy, industrial-revenue bond, PILT scheme, sole-source platform, or public-sector AI engagement above a published dollar or megawatt/gallon threshold, the public file must contain, before first reading:

Water: source, daily peak, consumptive vs. recycled, whose rights, whose shortage.

Grid: incremental load, who pays the upgrade, large-load tariff or residential transfer.

Displacement and error rates: for eligibility, welfare, AML, hiring engines—false-denial and false-penalty rates with a named owner.

Hidden-cost categories already used in Chapter 6, in a two-column visible/hidden table a hostile reader can attack.

The ledger is not an economic-impact study written by the applicant. It is a public accounts document. Line D recoveries (courts, refunds) stay visible so nobody confuses a \$21.5 million check with the iceberg.

3. Outcome clawbacks and beneficiary-level reversibility

Axis: Ethics architecture (Dimension Two)—embedded in the contract, not the framework PDF.

Public-sector AI that decides money, status, or liberty (Annex III classes already running in U.S. production: Bridges/Noon, TCF-type penalties, hiring screens—Chapters 4 and 7) must carry:

A clawback: fees or milestone payments return if error rates exceed a published bound or if the model cannot be independently reconstructed.

Reversibility at the beneficiary: a human, timed path to undo a denial that does not require a lawyer the claimant does not have.

A named partner or contracting officer on the letter—Dimension One riding with Dimension Two.

ISO 42001 and NIST maps remain optional hygiene. They do not satisfy this clause. Chapter 7 already recorded that 42001 is not the Act's Article 17 QMS. Do not let a badge eat a clawback.

Date-stamp. Until 2 December 2027, EU high-risk duties for stand-alone Annex III systems are deferred. The waiting year is a consulting market unless contracts impose the duty early. U.S. engines do not wait for the Annex. The clawback should not wait either.

4. Hidden Costs as an external scorecard

Axis: all four, scored from outside.

Ghost Vortex and Hidden Costs of Consent Machine were introduced in This is our working rubric so outsiders can rescore Deloitte and BCG in twelve months.

Accountability architecture

Named human on the engagement letter and on the adverse-event file.

No NDA or aid-and-assistance clause on public counterparties for siting, water, power, or public-sector AI.

Public-records and discovery cooperation without developer veto.

Ethics architecture

Assurance and implementation are not sold to the same client on the same system.

Error, hallucination, and fabricated-citation events produce a changed recommendation, not only a partial refund.

Beneficiary reversibility exists in the contract, not in a principles appendix.

Business-model logic

Revenue from independent judgment exceeds revenue from platform distribution (Frontier Alliance and equivalents disclosed).

Apprenticeship hours are a costed line (see Remedy 5).

The firm can refuse a stack it cannot independently evaluate.

Temporal horizon

Engagement success is measured after consequence (grid, water, error rates), not at invoice.

Hidden-cost ledger filed where public money or public rights are in play.

No exit-before-wreckage clause that assigns downstream harm to the client, the ratepayer, or the claimant.

Evidence rule. Only primary documents: contracts, dockets, IPRA/FOIA releases, regulator decisions, refund letters, tariff citations. Press releases score zero. Blogs score zero. A firm's own "Responsible AI" page scores zero.

Review. One year from publication, an outside panel republishes scores. If Deloitte and BCG cannot be rescored with documents a SEAM intervenor can assess Hidden Costs not reported in traditional accounting statements.

5. Apprenticeship as a costed line

Axis: Business-model logic (Dimension Three).

If agents draft the slide (Lilli, Zora, Agent OS, Sage—Chapters 3–4), the junior row is either a paid teaching hospital or it is inventory (Chapter 5). Contract language:

A published ratio of supervised human review hours to agent-generated pages on public-sector and high-risk private work.

Those hours are billable teaching, not uncompensated overtime (Chapter 6, line C).

If the ratio is unmet, the clawback in Remedy 3 triggers.

Diasio's sentence from the Preface—"Our entire business model was built by demonstrating effort and working for that billable hour"—is the old model dying. The remedy is not to restore fake hours. It is to stop treating the disappearance of apprenticeship as efficiency.

V. What a 2027 partnership would have to stop selling

Founders' Voice, last and unanswered only for a page.

Waterhouse's "conscience," Bower's professional independence, Henderson's strategy-as-judgment, and the older civic refusal named by Goerdeler in another register of conscience-under-pressure, are useless as décor. They become usable only as a list of SKUs a partnership must stop selling if Dimension Two is to be embedded:

Stakeholder-engagement frameworks designed to produce yes for a siting already decided.

NDA-compatible public-official and land-grant work on water and power.

Assurance of a model the same network is paid to scale—including Frontier Alliance distribution without an independence memo that can be read in court (Chapter 7, Case 7C).

Economic-impact studies that exclude SEAM categories below the waterline.

Eligibility, penalty, and hiring engines without beneficiary reversibility and a published error bound.

ISO 42001 / NIST "readiness" as a substitute for Article 17-class duties or for U.S. claw-backs.

Apprenticeship rhetoric unaccompanied by a costed line.

Self-scored Hidden Costs of dysfunctions that could be turned into revenue streams by unleashing human potential.

If the 2027 partnership cannot strike those eight, Founders' Voice is PR Voice in period costume. Special answerability without the strike-list is the control plane again.

Critical Theorist. Mazzucato's question, last time: who authorized private distributors to be the public's substitute for a notified body, a State Engineer, a newsroom, and a council? The strike-list is the beginning of an answer because it withdraws the rental.

VI. I Propose Polyphony, not a merger

David's Authorial Voice. In Doña Ana, where I have our main home, I will not accept a university NDA that obliges aid against IPRA, water numbers that arrive after the bond, or comment letters that use names as inventory. That is a situated bet (postupok), not a theory of everyone else's county. If Jupiter's operational gallons and grid tariffs later match the

closed-loop slide under independent meter, induction will correct me. Until the meters are public, I will not emplot an ending.

Community Voice. As I say in my YouTubes: You can't drink data. You can't drink an 18-month pause either, but you can use it. Keep the Magdalena sign in the file when the next developer brings a new cooling story. Tucson's cheer on 6 August 2025 was real. So were the blades in April. Both sentences stay.

SEAM Analyst. Print Ledger II beside the ordinance. If the ordinance does not move water, grid, error rates, or hours, it did not move the iceberg.

Ethnographic Witness. The dates above are the chapter's only romance.

Founders' Voice / PR Voice. Conscience. Safeguards from day one. The pair still does not resolve. The strike-list is the corridor between them.

No merger. Tamaraland's combinatorics remain: Community Room, Boardroom, Former-Employee Room, Ratepayer Room, Brussels Room, Witness Room, Refusal Room. A reader who walks all of them still does not get one official story. That is the book's architecture, not a defect in the last chapter.

VII. Research cases (one seminar, timed)

Case 9A. Two Tucson minutes. Give students the 6 August 2025 council direction and the 24 December 2025 land-close note. Task: write the antenarrative Between—the rooms the BME "Tucson said no" sentence leaves out. Ban a moral of the story.

Case 9B. Draft the NDA ban in one page. Use the NMSU aid-and-assistance clause (Chapter 8) as the thing the statute voids. Include the trade-secret remainder so counsel cannot call the ban a taking of IP.

Case 9C. Rescore one firm. Using only Chapter 6's document list plus one 2026 docket, apply the SEAM's hidden cost analysis to Deloitte or BCG.

Case 9D. Claw-back for Noon. Rewrite a public-benefits AI operate-and-change-request contract so a false denial of insulin reverses in ten days and a slice of the fee returns. If the rewrite is incompatible with the firm's platform SKU, record that as the finding.

Conclusion (which is not an end to a Narrative)

We are in the middle. Antenarrative is the honest tense of a data-center corridor and of a consulting partnership that has not yet chosen whether judgment or distribution is the business. The Consent Machine fails in public when five conditions flip. It is repaired in public only when those conditions become statute, ledger, clawback, external score, and a costed junior row.

AGI talk was already in Chapter 1's Davos split. It is the official narrative. This book's method stays with water, footnotes, and hours—the storying the coherent story is designed not to count.

The horses still drink from an aquifer that does not read press releases. If a chapter cannot survive that sentence, it is still above the waterline.

Table 1

B	What it names in this book	What consent manufacturing does to it
Before	Many contested starting points, not one Beginning	Collapses start-dates into the announcement
Beneath	Tacit contracts, NDAs, aquifer and grid deals	Keeps them unlabeled “proprietary”
Bets	Competing wagers with pravda, not scenarios	Sells one bet as the official ending
Being	What is happening now to water, air, hours, denials	Replaces Being with Becoming-language
Becoming	The moment a situation is named as a story	Permanent future tense: “we are becoming responsible”
Between	Polyphony that cannot be merged	One walkthrough sold as the play
Beyond	Obligations that exceed the firm's purpose	Projected past the engagement and the invoice