UNIVERSITÀ DEGLI STUDI DI
CASSINO E DEL LAZIO MERIDIONALE

Corso di Dottorato in
Economia e management per l'innovazione e la sostenibilità

Curriculum Management e Finanza

Ciclo 38°

The Visibility Trap: Attention Structures and the Hidden Costs of Generative AI in Work and Governance

SSD: ECON-08/A

Coordinatore del Corso
Chiar.mo Prof. Mario Rosario Guarracino

Dottorando
Alberto Varone

Supervisore
Chiar.mo Prof. Francesco Bolici

Table of Contents

Abstract

Technological advancement can open new possibilities for organizations, but whether and how these translate into organizational effects depends on how they are made visible, evaluated, and prioritized. Organizations do not engage in this process neutrally. Instead, they rely on familiar categories, established routines, and inherited metrics shaped by prior experience, which function as cognitive filters through which new developments are assessed. Under stable conditions, this form of cognitive economization is efficient. It becomes problematic, however, when a technology alters what those inherited categories and metrics actually capture. In such cases, indicators that once reliably tracked key organizational dimensions can appear to improve precisely as their representativeness declines. These interpretive habits are stabilized by attention structures, including dedicated channels, established agendas, and entrenched routines, which make them resistant to revision even when misalignment between indicators and underlying realities has the potential to generate significant harm. This structural stickiness is a primary source of organizational inertia.

This thesis utilizes the diffusion of Generative Artificial Intelligence (GenAI) as a contemporary case study to investigate the interaction between technological novelty and organizational inertia through an Attention-Based View. The overarching research question guiding this thesis is: *How do generative technologies reshape the relationship between observable indicators and the underlying organizational capabilities they are meant to track?* Building on the Attention-Based View, it develops the *Visibility Trap* framework: a self-reinforcing mechanism in which established attention structures privilege impacts that are *legible* and *strong* within existing evaluation channels, while slower or harder-to-code consequences remain peripheral and therefore under-managed.

The framework is applied across two phases of GenAI diffusion. In the early governance phase, a case study of the Italian Data Protection Authority's temporary ban on ChatGPT illustrates how regulatory action can register as a procedural and symbolic success, eliciting provider adjustments, financial penalties, and international signaling, while producing legitimacy erosion and limited behavioral change. In the second phase, namely organizational deployment in knowledge work, a theory-building study informed by exploratory interviews demonstrates that productivity gains and apparently improved access to organizational memory can coexist with shallow learning, growing dependence on GenAI for baseline task

performance, and the rise of *pseudo-memory* – synthetic precedents that appear authoritative but are weakly grounded in the organization's past.

Across both domains, the findings indicate a consistent pattern: GenAI amplifies the salience of historically established success signals, while weakening their value as proxies for underlying capability, efficacy, and epistemic integrity. The thesis contributes to research on organizational myopia by grounding incubation and success traps in attention structures and evaluative architectures, offering a portable mechanism for explaining why visible improvements can systematically delay correction of latent erosion. It concludes that effective GenAI integration requires organizational redesign that actively examines whether prevailing measures continue to track the capacities and qualities they were historically intended to protect.

CHAPTER 1: INTRODUCTION

1.1 RESEARCH CONTEXT

Technological advancement can open new possibilities for organizations. It becomes strategically hazardous, however, when it expands what organizations can do faster than what they can reliably observe and evaluate.

Organizations rarely assess novel technologies on a neutral basis. Rather, they interpret them through labels, categories, and evaluation heuristics built from past experience. When a technology weakens the link between familiar indicators of success and the underlying asset those indicators used to represent, organizations can appear to be improving while quietly degrading the capabilities required for long-term survival. This produces inertia: existing framings continue to guide action even as they become partially obsolete. Updating what the organization treats as relevant information, however, is difficult. This resistance to change stems from the fact that relevance is not merely a product of individual judgment; rather, it is deeply embedded within institutionalized organizational attention structures. Following Ocasio (1997), attention structures are the formal and informal processes (e.g., reporting channels, agendas, and routines) that make decision-makers selectively receptive to particular issues and cues.

To investigate the interaction between technological novelty and this form of organizational inertia, this thesis utilizes the diffusion of Generative Artificial Intelligence (GenAI) as a contemporary case study. In line with Faraj et al. (2018, p. 62), this thesis refers to artificial intelligence (AI) as "an emergent family of technologies [that] generate responses, classifications, or dynamic predictions that resemble those of a knowledge worker". Within this broader family, GenAI is defined as computational techniques capable of generating seemingly new, meaningful content such as text, images, and audio based on patterns learned from training data (Feuerriegel et al., 2024).

The premise of this analysis is that as organizations interact with GenAI, their attention structures act as filters that create a visibility asymmetry. Consequences of GenAI aligning with established performance or compliance indicators become salient and are prioritized for managerial intervention. Conversely, impacts falling outside these historical categories remain largely unobserved and relegated to the periphery despite their potential significance.

This research examines two distinct phases of GenAI diffusion to illustrate how attention structures shape organizational outcomes. The first phase addresses early diffusion, specifically analyzing how public authorities, constrained by inherited legal and institutional frameworks, attempt to mitigate emerging privacy and security risks (Chapter 4). The second phase focuses on organizational deployment and active use, exploring how the embedding of GenAI within knowledge work reshapes the relationship between performance, expertise development, and organizational memory (Chapter 5). Findings from these domains substantiate the central claim of this thesis: GenAI integration can lock governance and organizational deployments into self-reinforcing patterns. In these scenarios, perceived success according to established performance indicators often masks a gradual erosion of the underlying capacity to govern, learn, and adapt to future environmental shifts.

1.1.1 Governance of GenAI Diffusion

In this thesis, GenAI governance refers to the process through which public authorities and related bodies interpret, apply, and adapt policy frameworks to supervise GenAI diffusion and use. This includes deploying existing instruments (such as data protection law), designing new procedures, and signaling which forms of GenAI deployment are acceptable (i.e., aligned with national and European regulations) or problematic. Conceptually, this sits within AI governance as a polycentric, multilevel system. Dafoe (2023) defines AI governance as the ensemble of norms, institutions, and processes shaping how AI is built and deployed. Monteiro & Singh (2025) show how governance emerges from interactions among international bodies, national authorities, private and public organizations, and civil society. Thus, a data protection authority responding to GenAI is not a neutral rule-applier but an organization embedded in a contested field, with its own mandates, resource constraints, and performance metrics.

These organizations operate under structural conditions that make GenAI governance particularly demanding. The pacing problem captures the growing misalignment between rapidly evolving technological capabilities and slower, rigid legal and institutional responses (Maas, 2022; Marchant, 2011). Nordström (2022) characterizes AI policy as decision-making under great uncertainty, combining unclear problem boundaries, fluid concepts of what AI is or what constitutes "high risk", and strategic interdependence with powerful firms and jurisdictions. A "race to regulate" has emerged in which jurisdictions compete to set standards and signal leadership (Feldstein, 2024; Walter, 2024). Global divergence between EU, US, and Chinese approaches adds fragmentation and international uncertainty (Zhao, 2023).

Under these conditions, privacy and data governance have become the primary levers through which authorities act on GenAI. Generative models rely on extensive data collection and reuse, thereby tightly coupling GenAI deployment decisions with acceptable data practices. Normative work stresses that individual-consent models systematically underestimate informational asymmetries and collective harms, arguing for structural safeguards such as strict data minimization and limits on data markets (Véliz, 2022). Empirical studies show that current practices fall short: even under the GDPR, users who opt out of tracking are frequently still monitored and targeted, as entrenched architectures and business incentives sustain data flows despite surface-level consent (Liu et al., 2024).

For this thesis, public authorities governing GenAI diffusion are organizations that must design and justify interventions in a moving, fragmented regime, relying heavily on data protection law. Their decisions are highly visible and easy to report in terms of formal compliance and enforcement outputs, while their effects on actual practices and institutional legitimacy are harder to observe. Accordingly, this thesis shows data protection authorities and related bodies how their current mandates, indicators, and routines can leave crucial risks and capacity losses off the radar, and how bringing these latent effects into view is essential if they are to maintain their ability to govern GenAI and its successors over time.

1.1.2 GenAI Deployment in Knowledge Work

While public authorities are working out how to govern GenAI's diffusion, this technology is being increasingly absorbed into knowledge work through a growing ecosystem of tools. To provide a more complete view of GenAI's impact, this thesis complements the analysis of early diffusion and governance with a second focus on organizational deployment.

Experimental and field studies converge on a basic finding: access to GenAI support raises performance on many knowledge-intensive tasks. Examples include professional writing, consulting, computer programming, and customer support, where GenAI was found to improve productivity and output quality with especially large gains for novices (Noy & Zhang, 2023; Brynjolfsson et al., 2025; Peng et al., 2023). At the same time, recent work on knowledge management suggests that GenAI can make organizational memory more accessible, combining, generating, and distributing organizational knowledge (Alavi, Leidner, & Mousavi, 2024; Jarrahi et al., 2023) across workers. Tools that transcribe meetings, summarise documents, and answer questions over internal repositories embed these capabilities in everyday work, and reduce the cost of capturing and retrieving organizational knowledge (Chui et al., 2023; Schaetzle et al., 2025).

Alongside these gains, however, it is less clear how GenAI affects the less visible foundations of organizational capability, such as skill development and the reliability of organizational memory. Emerging evidence from education, programming, and early-career work suggests that these gains can coexist with stalled or shallow skill development. Novices often use GenAI to bypass learning, accept plausible but flawed outputs, and develop an illusion of competence (Prather et al., 2023; Fan et al., 2025; Margulieux et al., 2024). Opportunities for deliberate practice shrink as effort shifts from execution to prompt refinement and light-touch review (Macnamara et al., 2024; Simkute et al., 2025). Over time, this can produce a cohort that depends on GenAI to achieve baseline performance, with fragile expertise that is hard to mobilize when tools fail, change, or are restricted. Concurrently, these systems can introduce new vulnerabilities for organizational memory. GenAI can hallucinate or subtly distort content (Schaetzle et al., 2025; Storey, 2025), as well as reduce diversity of ideas, thereby pushing towards output standardization (Schuh, 2024; Anderson et al., 2024).

Taken together, these findings suggest that organizations using GenAI in knowledge work face a dangerous asymmetry: GenAI makes performance look better according to existing indicators, even as underlying expertise and organizational memory weaken in ways those indicators do not detect.

1.2 PROBLEM STATEMENT & RESEARCH QUESTION

While GenAI acts as a disruptive technology that fundamentally alters the nature and execution of work (Crowston & Bolici, 2019; Dell'Acqua et al., 2023), organizations frequently fail to recalibrate their evaluation mechanisms in response. When technological shifts are processed through historically developed attention structures, a visibility asymmetry emerges. Consequences that align with established success indicators such as accelerated task completion are rapidly recognized and reinforced. Conversely, impacts falling outside these traditional categories, including gradual skill erosion or increasing dependency on opaque algorithmic tools, are systematically marginalized or misunderstood. This phenomenon transcends simple short-termism, representing a form of structural myopia where the very mechanisms intended to assess performance obscure signals of long-term capability degradation.

Current research documents various facets of this tension between visible short-term gains and latent long-term losses across fields such as AI governance, skill acquisition, and organizational memory (see Chapter 2). However, the literature lacks a unified, theory-based mechanism to connect these observations. Specifically, there is an insufficient understanding of how GenAI-related gains and losses are differentially integrated into the organizational sphere of attention and how it subsequently influences governance and work practices. Consequently, this study addresses the following research question: *How do generative technologies reshape the relationship between observable indicators and the underlying organizational capabilities they are meant to track?*

This thesis addresses this question by developing and applying the concept of the *Visibility Trap*: a self-reinforcing process, shaped by established attention structures, in which the salience of visible successes progressively degrades the capacity to identify and respond to associated but latent negative consequences. Chapters 4 (focusing on GenAI governance) and 5 (on GenAI deployment in knowledge work) provide the empirical and conceptual ground for analyzing how this trap emerges, stabilizes, and might be mitigated.

The argument extends foundational research on organizational myopia (Levinthal & March, 1993; Catino, 2013; Sato, 2012) by recasting the phenomenon in explicitly attention-based terms. Rather than attributing failure solely to surveillance gaps or rigid competencies, this thesis demonstrates how organizations economize scarce attention by prioritizing consequences that manifest as performance shortfalls under existing criteria. By adopting the Attention-Based View (Ocasio, 1997), this research explains how measurement systems and performance expectations determine which GenAI effects become actionable and which remain relegated to the background.

1.3 THESIS OVERVIEW

The thesis is organized to build a cumulative argument about how GenAI's impacts, when filtered through existing attention structures, produce a *Visibility Trap* in governance and organizational settings. Following this introduction, each chapter addresses a specific aspect of the research question formulated above, progressing from problem mapping to theoretical development, empirical demonstration, and conceptual extension, and finally to implications.

Chapter 2 presents a selective, problem-focused literature review on GenAI governance and GenAI deployment in knowledge work. Synthetizing extant research, it shows that governance authorities operate in an uncertain environment where GenAI diffusion demands immediate action, triggering a race to regulate. However, available frameworks don't fit the technology's characteristics, leaving authorities with inadequate instruments and potential unawareness that this mismatch exists. The same macro pattern emerges from synthesizing findings from research in GenAI integration within knowledge work. Research demonstrates that organizations gain immediate performance improvements but face skill erosion, growing dependence, and memory homogenization – all of which are rendered relatively obscure by the previously reasonable assumption that accomplishing a task is tantamount to having the ability that accomplishment requires. This assumption no longer holds with GenAI-mediated work. Chapter 2 concludes by establishing the need for an account of how gains and losses related to GenAI become differentially visible in current organizational contexts.

Chapter 3 develops the theoretical framework adopted in this thesis. It analyzes Ocasio's (1997) ABV and positions it relative to alternative theoretical lenses: technology adoption models, resource-based and dynamic capabilities perspectives, information-processing views, and sociomaterial approaches. Building on this theoretical foundation, Chapter 3 formalizes the *Visibility Trap* as an attentional mechanism structured by *signal legibility* and *strength*, and recasts organizational myopia in explicitly attention-based terms. It is argued that myopic outcomes emerge not only from surveillance failures or competence rigidity, but also from how organizations economize attention and lock into established evaluative architectures that GenAI satisfies with apparent ease. This chapter provides a portable mechanism applicable across governance and organizational contexts, offering a theory-grounded basis for interpreting findings from Chapters 4 and 5.

Chapter 4 examines GenAI governance and addresses the following research question: *How do formal successes and invisible failures coexist in technology governance, and what mechanisms influence their visibility?* Using the Italian Data Protection Authority's temporary ban on ChatGPT as a case study, the chapter reconstructs how the authority deployed the intervention, how it constructed its official narrative for the public, and how it was received. By its own indicators, the authority achieved a clear formal win: it compelled OpenAI to modify its practices, secured a substantial fine, and established a strong precedent of GenAI governance in a Western country. However, the study also demonstrates that Italian users showed behavioral indifference to the intervention, bypassing the ban while in place and resuming usage of ChatGPT post-ban as if the risks substantiating the intervention had never been exposed. Concurrently, evidence is presented that the ban deteriorated the authority's pragmatic, moral, and cognitive legitimacy. The chapter interprets this divergence through the *Visibility Trap*: existing governance frameworks function as attention structures that privilege established compliance artifacts over harder-to-measure behavioral and legitimacy outcomes. The implications target regulators and policymakers: interventions that appear successful on inherited evaluative indicators can silently weaken their credibility and authority.

Chapter 5 analyzes GenAI's impact on knowledge work and addresses the following research question: *How do GenAI associated performance gains coexist with the disruption of expertise development trajectories and organizational mnemonic infrastructures?* The chapter develops two interconnected conceptual models informed by eight semi-structured interviews with practitioners. The first model delineates how GenAI reconfigures the relationship between individual performance and the acquisition of expertise. By leveling the performance gap between novices and experts, GenAI facilitates a state of doing-without-learning. While the technology enables novices to execute low-complexity tasks efficiently, it simultaneously obstructs the formation of the complex representational schemata required for expert-level problem-solving and the management of corner cases. This disruption interrupts the traditional declarative-procedural-automatized ladder of skill acquisition, raising critical concerns regarding the long-term cultivation of organizational expertise. The second model conceptualizes GenAI tools as integral components of an organization's memory system, termed the GenAI-OMIS (Organizational Memory Information System). As organizational outputs, including documents, decisions, and internal dialogues, are increasingly mediated through GenAI, the technology becomes a primary encoder of organizational history. While this integration enhances retrieval speed, it introduces the risk of *pseudo-memory*: the generation of synthetic precedents that superficially resemble past practices but are tenuously grounded in actual organizational history. This creates a strategic dependency on reconstructed versions of the past that are often resistant to inspection, governance, or correction. Framed through the *Visibility Trap* framework established in Chapter 3, this chapter demonstrates a perverse organizational trajectory. Organizations may appear increasingly productive and better informed due to the high visibility of GenAI-augmented outputs; however, this superficial success often masks an underlying fragility. The expertise base becomes shallower, and the mnemonic infrastructure becomes increasingly centralized and detached from authentic organizational experience.

Chapter 6 synthesizes the primary contributions of this thesis and explores their broader implications for organizational theory and practice. First, it formalizes proxy substitution as a practical risk: a phenomenon where visible improvements in inherited performance indicators coexist with latent systemic erosion. This divergence creates a deceptive appearance of organizational effectiveness that systematically delays necessary corrective action. By relying on these lagging indicators, organizations may inadvertently validate GenAI deployments that appear successful in the short term while silently undermining long-term operational integrity. Second, the chapter develops the politics of visibility as a central design-relevant mechanism. This highlights that the determination of what constitutes "legitimate" information for organizational attention is the product of bargaining among internal coalitions. Because different stakeholders have interests tethered to specific performance metrics, these negotiations result in either integrative arrangements or "no-man’s land" compromises. While these compromises may be institutionally defensible, they are often weakly connected to the actual practices and risks of GenAI integration, further obscuring the *Visibility Trap*. Third, this research specifies *elastic attention* as a relevant design principle for escaping the *Visibility Trap*. This concept involves expanding the scope of organizational legibility, specifically by incorporating capability and legitimacy risks, without causing cognitive overload or collapsing decision-making processes. By designing systems that can flexibly scale attention toward non-traditional indicators, organizations can maintain a more holistic view of technological impact on their underlying infrastructure.

CHAPTER REFERENCES

Alavi, M., Leidner, D. E., & Mousavi, R. (2024). Knowledge Management Perspective of Generative Artificial Intelligence. Journal of the Association for Information Systems, 25(1), 1–12. https://doi.org/10.17705/1jais.00859

Anderson, B. R., Shah, J. H., & Kreminski, M. (2024). Homogenization Effects of Large Language Models on Human Creative Ideation. *Proceedings of the 16th Conference on Creativity & Cognition*, 413–425. https://doi.org/10.1145/3635636.3656204

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics*. https://doi.org/10.1093/qje/qjae044

Catino, M. (2013). *Organizational myopia: Problems of rationality and foresight in organizations*. Cambridge University Press.

Crowston, K., & Bolici, F. (2019). Impacts of Machine Learning on Work. Proceedings of the 52nd Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences (HICSS).

Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zemmel, R. (2023). The economic potential of generative AI: The next productivity frontier. McKinsey & Company.

Dafoe, A. (2023). AI Governance: Overview and Theoretical Lenses. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., pp. 21–44). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.2

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., … & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. SSRN Electronic Journal.

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. Information and Organization, 28(1), 62–70.

Feldstein, S. (2024). Evaluating Europe's push to enact AI regulations: How will this influence global norms? *Democratization*, *31*(5), 1049–1066. https://doi.org/10.1080/13510347.2023.2196068

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. Business & Information Systems Engineering, 66(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2023). Artificial intelligence and knowledge management: A partnership between human and AI. *Business Horizons*, *66*(1), 87–99. https://doi.org/10.1016/j.bushor.2022.03.002

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*(S2), 95–112. https://doi.org/10.1002/smj.4250141009

Liu, Z., Iqbal, U., & Saxena, N. (2024). Opted Out, Yet Tracked: Are Regulations Enough to Protect Your Privacy? *Proceedings on Privacy Enhancing Technologies*, *2024*(1), 280–299. https://doi.org/10.56553/popets-2024-0016

Maas, M. M. (2022). Aligning AI Regulation to Sociotechnical Change. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., pp. 358–380). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.22

Macnamara, B. N., Berber, I., Çavuşoğlu, M. C., Krupinski, E. A., Nallapareddy, N., Nelson, N. E., Smith, P. J., Wilson-Delfosse, A. L., & Ray, S. (2024). Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cognitive Research: Principles and Implications*, *9*(1), 46. https://doi.org/10.1186/s41235-024-00572-8

Marchant, G. E. (2011). Addressing the Pacing Problem. In G. E. Marchant, B. R. Allenby, & J. R. Herkert (Eds.), The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem (pp. 199–205). Springer Netherlands. https://doi.org/10.1007/978-94-007-1356-7_13

Margulieux, L. E., Prather, J., Reeves, B. N., Becker, B. A., Cetin Uzun, G., Loksa, D., Leinonen, J., & Denny, P. (2024). Self-Regulation, Self-Efficacy, and Fear of Failure Interactions with How Novices Use LLMs to Solve Programming Problems. *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1*, 276–282. https://doi.org/10.1145/3649217.3653621

Monteiro, N., & Singh, V. (2025). The wheel of artificial intelligence governance. *Sustainable Futures*, *10*, 101279. https://doi.org/10.1016/j.sftr.2025.101279

Nordström, M. (2022). AI under great uncertainty: Implications and decision strategies for public policy. *AI & SOCIETY*, *37*(4), 1703–1714. https://doi.org/10.1007/s00146-021-01263-4

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Ocasio, W. (1997). Towards an Attention-Based View of the Firm. *Strategic Management Journal*, *18*(S1), 187–206. https://doi.org/10.1002/(SICI)1097-0266(199707)18:1+%253C187::AID-SMJ936%253E3.0.CO;2-K

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). *The Impact of AI on Developer Productivity: Evidence from GitHub Copilot*. https://doi.org/10.48550/ARXIV.2302.06590

Prather, J., Reeves, B. N., Denny, P., Becker, B. A., Leinonen, J., Luxton-Reilly, A., Powell, G., Finnie-Ansley, J., & Santos, E. A. (2023). "It's Weird That it Knows What I Want": Usability and Interactions with Copilot for Novice Programmers. *ACM Trans. Comput.-Hum. Interact.*, *31*(1), 4:1-4:31. https://doi.org/10.1145/3617367

Sato, H. (2012). Routine-Based View of Organizational Learning and Mechanisms of Myopia. Annals of Business Administrative Science, 11(0), 45–54. https://doi.org/10.7880/abas.11.45

Schaetzle, A., Zöll, A., Sammet, N., Brack, F., Scholz, J., & Schwarz, L. (2025). A Research Agenda for Generative AI in Organizational Knowledge Management: Addressing Risks in Knowledge Storage, Retrieval, and Transfer. *ICIS 2025 Proceedings*. https://aisel.aisnet.org/icis2025/is_transformwork/is_transformwork/6

Schuh, J. (2024). AI As Artificial Memory: A Global Reconfiguration of Our Collective Memory Practices? *Memory Studies Review*, *1*(2), 231–255. https://doi.org/10.1163/29498902-202400012

Simkute, A., Tankelevitch, L., Kewenig, V., Scott, A. E., Sellen, A., & Rintel, S. (2025). Ironies of Generative AI: Understanding and Mitigating Productivity Loss in Human-AI Interaction. *International Journal of Human–Computer Interaction*, *41*(5), 2898–2919. https://doi.org/10.1080/10447318.2024.2405782

Storey, V. C. (2025). Knowledge Management in a World of Generative AI: Impact and Implications. *ACM Trans. Manage. Inf. Syst.*, *16*(3), 26:1-26:14. https://doi.org/10.1145/3719209

Véliz, C. (2022). Governing Privacy. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., pp. 149–163). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.9

Walter, Y. (2024). Managing the race to the moon: Global policy and governance in Artificial Intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences. *Discover Artificial Intelligence*, *4*(1), 14. https://doi.org/10.1007/s44163-024-00109-4

Zhao, L. (2023). Artificial intelligence and law: Emerging divergent national regulatory approaches in a changing landscape of fast-evolving AI technologies. In *Research Handbook on Digital Trade* (pp. 369–399). Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/book/9781800884953/book-part-9781800884953-33.xml

CHAPTER 2: LITERATURE REVIEW

2.1 INTRODUCTION

This chapter positions this thesis RQ into established literature, laying the conceptual foundation for the analysis presented in the next chapters. Specifically, it develops a problem-focused literature review of GenAI's impact in two domains: public governance and knowledge work. While these fields operate at different analytical levels, this chapter demonstrates that they share a fundamental tension: a divergence between highly visible, immediate gains and latent, structural erosions when it comes to evaluating GenAI's impact.

Section 2.3 focuses on the first stream, and conceptualizes GenAI governance as a polycentric and adaptive system in which public institutions act as mediating organizations under three structural constraints: a *pacing problem* between fast model development and slower legal adaptation, competitive "races" to regulate, and fragmented regimes that complicate coordination and enforcement. Within these conditions, governance is frequently pulled toward highly visible and defensible interventions while broader questions of technical fit, behavioral impact, and institutional legitimacy remain harder to stabilize.

Section 2.4 examines the second stream, synthesizing evidence on GenAI in organizations, showing that it reliably raises the productivity floor and often disproportionately benefits novices, yet creates a jagged frontier where performance can collapse outside the model's competence region. It further reviews how GenAI shifts work from execution toward orchestration, potentially masking skill deficits and compressing opportunities for learning and deliberate practice. At the collective level, the chapter traces how GenAI reconfigures organizational memory from repository-based retrieval to regenerative remembering, with associated risks of homogenization, insight drift, and a growing validation burden concentrated in scarce experts.

The two streams are brought together in Section 2.5 using an attention-based lens, which highlights how organizational attention influences what organizations perceive, measure, and respond to. This lens is then formally developed in Chapter 3 through the Attention-Based View (ABV). Across both governance and organizational contexts, the reviewed literature points to a common pattern: when interacting with GenAI, established attention structures render certain outcomes (e.g., regulatory compliance, performance boosts) particularly salient, while relegating others (institutional legitimacy, skill decay, and organizational memory degradation) to the background.

2.2 METHODS

This chapter employs a selective, problem-focused literature review designed to synthesize concepts, mechanisms, and empirical findings that directly inform the analyses presented in Chapters 4 and 5. Rather than providing exhaustive coverage of all GenAI research in public governance and knowledge work, the review prioritizes theoretical and empirical relevance to the specific research questions at hand.

The search process combined database queries with citation chasing. Initial searches were conducted in Google Scholar and Scopus using keyword combinations across three domains: (1) AI governance and regulation (using terms such as “AI governance”, “AI regulation”, “technology governance”, “public institutions”, combined with terms such as “Generative AI”, “artificial intelligence”, “large language models”, and “ChatGPT”); (2) GenAI impact on knowledge work, with a focus on performance and skills (e.g., “productivity”, “performance”, “efficiency”, “effectiveness”, “creativity”, “quality”, “skill acquisition”, “skill development”, “learning, expertise”, “expert”, “novice”, “skill decay”, “skill erosion”, “skill retention”); and (3) organizational memory (e.g., “organizational memory”, “organizational learning”, later extended to knowledge management keywords, such as “knowledge acquisition”, “knowledge transfer” and “knowledge retrieval” when GenAI-specific OM work proved relatively scarce).

Selection criteria emphasized both theoretical and empirical fit. Studies were included when they addressed at least one of the three core domains and contributed to understanding either how public authorities govern the diffusion of emerging technologies like GenAI under conditions of uncertainty and institutional fragmentation, or how GenAI affects task performance, skill development trajectories, and organizational memory formation/curation in work settings. Purely technical papers focused on AI system design without explicit organizational or governance discussion were excluded after initial screening, as were studies examining AI model capabilities only under a technical standpoint. Within the governance literature, priority was given to contributions linking governance frameworks to innovation diffusion dynamics rather than isolated doctrinal analyses.

Synthesis proceeded iteratively throughout the thesis work. As relevant papers were identified through database searches and forward–backward citation chasing, they were organized into thematic streams reflecting the chapter's structure. Within each stream, recurring mechanisms and tensions were identified and integrated into the present analytical framework.

2.3 GENAI GOVERNANCE: GOVERNING UNDER UNCERTAINTY

2.3.1 AI Governance: A Multilevel and Adaptive System

Contemporary research increasingly conceptualizes AI governance as a polycentric, multilevel system characterized by overlapping and competing centers of decision-making, rather than a simple dyad between regulators and regulated entities (Kulothungan & Gupta, 2025). This polycentric structure spans international bodies, nation-states, corporate firms, professional communities, and technical infrastructures, each operating under distinct institutional logics and rationalities (Monteiro & Singh, 2025; Batool et al., 2025; Dafoe, 2023; Birkstedt et al., 2023; Cihon et al., 2020). Within this framework, public regulators are constrained actors navigating a decentralized regulatory landscape where governance outcomes emerge from complex interactions and negotiations (Monteiro & Singh, 2025).

The multiplicity of actors involved generates inherent tensions, as AI governance must reconcile often-conflicting rationalities including ethics principlism, managerial rationalism, IT professionalism, and regulatory oversight (Minkkinen & Mäntymäki, 2023). This complexity has led scholars to reconceptualize governance itself as the management of sociotechnical change rather than purely administrative rule-making (Maas, 2022). From this perspective, effective regulation should target the specific shifts in human behaviors, relationships, and power structures enabled by AI, rather than the technical artifacts themselves (Maas, 2022). In practice, this approach manifests through hybrid mechanisms where enforcement responsibilities are frequently delegated to market agents or embedded directly within technical architectures through risk-based frameworks (Almada & Petit, 2025; Currie et al., 2025). In this configuration, private firms, public agencies, and standardization bodies function as "governors in their own right" (Monteiro & Singh, 2025). These actors construct internal rules, allocate resources, and frame problems in ways that effectively govern AI development and deployment, regardless of whether they hold formal regulatory authority. Consistent with this framework, Dafoe (2023) offers a foundational definition of AI governance that captures both descriptive and normative dimensions: AI governance encompasses the norms, institutions, and processes that shape AI development, with the normative aspiration that these mechanisms promote safety, inclusivity, and effective decision-making.

Two contrasting perspectives inform current debates about how to conceptualize AI for governance purposes. Garfinkel (2022) situates AI within a historical lineage of general purpose technologies, comparing it to electricity and the steam engine. This framework suggests that while these technologies eventually catalyze profound economic, military, and political transformations, these effects unfold gradually through decades of institutional adaptation and complementary innovation. The full societal impact of AI, from this view, will emerge not immediately but as institutions and practices reconfigure themselves around the technology over extended timeframes. Mueller (2025), however, challenges this framing by redefining the technological substrate itself. Rather than treating AI as a discrete general-purpose technology, Mueller (2025) argues that contemporary AI comprises a set of machine learning applications emerging from a ubiquitous digital ecosystem composed of four interdependent elements: computing devices, networks, data, and software. From this perspective, most governance challenges attributed to AI (including cybersecurity vulnerabilities, algorithmic bias, and copyright infringement) are not new but endemic to distributed computing environments. This reconceptualization carries significant implications: attempts to govern AI as a generic capability risk expanding into efforts to control the entire digital infrastructure, potentially threatening permissionless innovation and open markets. Mueller (2025) therefore advocates for governance approaches that target specific machine learning applications and the particular policy problems they generate, rather than treating AI as a monolithic regulatory object.

At a higher level, the turn toward multilevel governance reflects fundamental limitations in traditional legal frameworks when applied to rapidly evolving, general-purpose technologies. To address this challenge, governance systems increasingly rely on adaptive soft law mechanisms, including standards, principles, and voluntary codes, that can be iterated more rapidly than formal legislation (Marchant & Gutierrez, 2022; Thierer, 2023). This reliance on flexible instruments shifts the burden of governance onto organizational capabilities. Public authorities are adopting anticipatory tools such as regulatory sandboxes to monitor risks in real-time (Yordanova & Bertels, 2024), while private firms are moving beyond post-facto compliance to integrate responsible AI governance as a core strategic value driver (Papagiannidis et al., 2025; Kahdan et al., 2023). In this configuration, organizational elasticity functions as an essential buffer that allows the governance system to operate despite inherent regulatory lags.

These strands of work converge on a core theoretical claim central to this thesis: while AI governance operates across multiple levels, public institutions governing AI diffusion occupy a critical mediating position within this system. These organizations function as the sites where heterogeneous demands from global institutions, transnational regimes, technical infrastructures, and evolving AI capabilities must be reconciled into actionable decisions

(Maas, 2022; Mueller, 2025). This thesis therefore conceptualizes public institutions governing AI diffusion as organizations embedded within a polycentric governance system, subject to the multilevel pressures and coordination challenges that characterize contemporary AI governance. Chapter 4 operationalizes this framework through an empirical examination of the Italian Data Protection Authority's intervention on ChatGPT, analyzing how the authority's regulatory decisions rendered the underlying governance dynamics visible through formal documentation, ChatGPT usage data, and public discourse analysis.

2.3.2 Governance Conditions: Pacing, Regulatory Races, and Fragmentation

Public authorities governing GenAI do not intervene on an open field but operate under structural conditions that constrain what can be known, coordinated, and enforced. Three conditions are decisive. First, a pacing gap separates fast, iterative model development from slower legal and administrative cycles, making regulatory targets unstable and enforcement chronically late. Second, this temporal mismatch feeds a geopolitical and economic race to regulate, in which jurisdictions compete to set standards and signal leadership while facing uncertainty about both technical trajectories and downstream impacts. Third, the resulting fragmentation produces overlapping and sometimes incompatible regimes, forcing regulators and regulated actors to navigate inconsistent definitions, obligations, and trade-offs across borders and policy domains. Together, these dynamics shape the content of GenAI rules as well as their legitimacy, coherence, and practical implementability. This section unpacks these governance conditions in turn.

2.3.2.1 The Pacing Problem

The pacing problem represents a fundamental temporal crisis arising from asynchronous rates of change between technological innovation and regulatory oversight. The core challenge stems from the widening gap between the exponential trajectory of technological advancement, often characterized through Moore's Law, and the linear, deliberative pace of legal processes (Allenby, 2011; Marchant, 2011, 2020; Taeihagh et al., 2021). Indeed, while technology firms operate in agile development cycles, regulatory agencies must navigate bureaucratic procedures, public comment periods, and judicial reviews that substantially delay intervention (Marchant & Gutierrez, 2022; Nordström, 2022). This friction produces regulatory obsolescence: by the time formal legislation is enacted, the target technology has often evolved into new iterations, rendering regulatory mandates outdated before enforcement begins (Guihot et al., 2017; Almada & Petit, 2025). Borrowing from evolutionary biology, Rejeski (2011) captures this dynamic through the *Red Queen effect,* where regulators are forced to "run as fast as they can just to keep in the same place" relative to the technological frontier, often failing to catch up despite their efforts.

This temporal gap creates an epistemic trap formalized as the *Collingridge Dilemma*. The dilemma presents a theoretical double-bind: during early development stages, regulation proves difficult because social consequences remain speculative and uncertain; this is referred to as the *Information Problem*. However, once impacts become empirically clear, the technology has typically achieved such deep entrenchment that meaningful modification becomes economically or politically infeasible; i.e., the *Power* or *Entrenchment Problem* (Cihon et al., 2020; Sætra, 2023). In AI governance, this dilemma intensifies due to the opacity of machine learning systems. Even when regulators identify the appropriate intervention window, the internal decision-making processes of AI models remain largely inscrutable, complicating efforts to define regulatory targets with precision (Ferrari et al., 2025; G'sell, 2024).

Beyond temporal and epistemic challenges, structural mismatches between governmental organization and technology characteristics compound governance difficulties. AI functions as a general-purpose technology that operates transversally across sectors including healthcare, finance, transportation, and defense. Regulatory bodies, however, typically organize around vertical domains, creating coordination failures and jurisdictional blind spots when governing cross-cutting technologies (Marchant, 2019; Taeihagh et al., 2021). Traditional legal frameworks assume static products that remain unchanged after market release, yet AI systems are inherently dynamic, continuously learning and updating post-deployment (Thierer, 2023). Resource asymmetries further deepen this structural mismatch: private firms command substantial capital and technical expertise, while public agencies operate under budget constraints and face persistent shortages of specialized personnel capable of auditing complex, evolving systems (Currie et al., 2025).

Lastly, geopolitical dynamics introduce an additional dimension through innovation arbitrage. Because technology development is globally distributed and capital is highly mobile, stringent national regulations can incentivize innovators to relocate activities to jurisdictions with more permissive regulatory environments (Hagemann et al., 2018; Dafoe, 2023). This mobility creates pressures toward a "race to the bottom", where states hesitate to impose necessary safeguards for fear of diminishing economic competitiveness (Camilleri, 2024). Definitional fluidity compounds this arbitrage problem: the absence of global consensus on what constitutes "AI" means that regulations often target concepts defined differently across jurisdictions, undermining international coordination efforts (Zhao, 2023). This dimension is further examined in the next section.

These coexisting pressures generate a normative dilemma (Hemphill, 2020): stringent ex-ante controls (such as bans) effectively manage safety but risk entrenching inferior technologies or relocating innovation to less accountable jurisdictions. Conversely, permissive approaches foster growth but risk irreversible harms. Under these conditions, governance ceases to be an exercise in optimizing outcomes and becomes, instead, an exercise in "managing regret". Nordström (2022) sharpens this diagnosis by noting that this decision-making occurs under three distinct forms of uncertainty: structural (where the available options are unclear), linguistic (where concepts like "fairness" or "high-risk" are volatile), and interactional (where outcomes depend on the strategic opacity of private firms).

To survive this instability, the literature suggests abandoning the search for a single best way statute in favor of adaptive portfolios. Taeihagh et al. (2021) argue that because information asymmetries and power shifts are persistent, effective AI governance requires a mix of substantive tools (targeting observable harms) and procedural tools (sandboxes, consultations, and iterative guidance). This portfolio perspective is directly relevant at the organizational level: effective governance requires layered arrangements that combine hard constraints (legal prohibitions and audit trails) with reflexive elements (pilots and feedback channels) capable of absorbing the shocks of rapid technological change (Taeihagh et al., 2021; Taeihagh, 2021).

Building on these observations, this thesis conceptualizes the pacing problem as a multidimensional governance condition arising from the convergence of four connected challenges: temporal obsolescence (linear legal processes versus exponential technological change), the Collingridge dilemma (tension between information availability and capacity for intervention), structural incompatibility (siloed regulatory agencies versus transversal, dynamic technological systems), and geopolitical arbitrage (jurisdictional competition and definitional fragmentation). The literature recognizes that, within the field of GenAI systems, no optimal solution to this problem currently exists; best practices remain theoretical and the empirical understanding of how authorities navigate these pressures in real-time is limited. Chapter 4 addresses this gap by documenting a practical manifestation of the pacing problem, examining how a national regulator constructs a response when the available tools are structurally misaligned with the speed and scale of the technology diffusion.

2.3.2.2 The Race to AI Regulation

The pacing problem intersects with a geopolitical race to AI regulation, wherein states (and regions) compete for the authority to establish legal and ethical standards that will govern AI globally (Walter, 2024; Zhao, 2023; Currie et al., 2025). This competition is inherently tense: jurisdictions seek first-mover advantages in exporting regulatory models, while simultaneously fearing that stringent rules may deter investment, slow deployment, or disadvantage domestic firms (Smuha, 2021). Walter (2024) characterizes this dynamic with the "race to the moon" analogy, where jurisdictions compete simultaneously for technological dominance and governance model primacy. In practice, this competition has produced evident divergence across major regulatory approaches: the EU's precautionary, rights-centric framework; the US preference for market-driven, sectoral governance; and China's state-strategic model oriented toward security and social stability (Currie et al., 2025; Feldstein, 2024). As Zhao (2023) observes, this results in a fragmented playground of partially overlapping regulatory regimes that organizations must navigate under significant time pressure.

Within this landscape, the European Union has positioned itself as the most ambitious comprehensive regulator, enacting the AI Act as the first horizontal legal framework for AI (Almada & Petit, 2025; Cancela-Outeda, 2024; Bakiner, 2023). The Act operationalizes a risk-based classification system that prohibits *unacceptable-risk* systems, subjects *high-risk* systems to strict obligations, imposes lighter transparency requirements for *limited-risk* applications, and leaves *minimal-risk* systems largely unregulated (G'sell, 2024). Normatively, this framework embodies a precautionary approach aimed at protecting fundamental rights, health, and safety while fostering a market for trustworthy AI (Currie et al., 2025). Notably, the AI Act was originally designed to govern purpose-specific, largely discriminative AI systems; the diffusion of GenAI shifted the technical scope after the political architecture was substantially finalized (Feldstein, 2024; Lubinski, 2024). It was only after the late-stage inclusion of General Purpose AI (GPAI) provisions that the tiering system around systemic-risk thresholds was introduced (Kim et al., 2025; Schuster et al., 2025). While these additions represent attempts to narrow the governance gap, critics argue that the Act continues to struggle with the general-purpose character and environmental impacts of GenAI, leaving the fundamental pacing gap largely intact (Shetty et al., 2025).

The United States has maintained a contrasting market-driven approach, prioritizing innovation and competitiveness through decentralized, sector-specific governance rather than comprehensive legislation (Batool et al., 2025). US governance relies heavily on soft law mechanisms, including voluntary commitments from major firms and the non-binding NIST AI Risk Management Framework (Thierer, 2023). The Biden Administration's Executive Order 14110, issued in October 2023, directed federal agencies to develop reporting requirements and guidance for powerful models but lacked the durability of congressional legislation (G'sell, 2024). This approach reflects explicit geopolitical considerations – particularly the perceived imperative to maintain technological leadership relative to China – while avoiding regulatory interventions that could constrain the US technology sector (Feldstein, 2024).

China has pursued a state-centric, vertically segmented model combining rapid rulemaking with strong political control. Rather than adopting a single horizontal AI law, China has issued targeted regulations for specific technologies, including rules governing recommendation algorithms (2021), deepfakes and deep synthesis (2022), and generative AI services (2023) (G'sell, 2024; Walter, 2024). These instruments commonly require algorithm registration and embed content obligations aligned with socialist core values, reflecting a governance philosophy oriented toward national security, social stability, and Party control over public discourse (G'sell, 2024; Mueller, 2025). China thus illustrates a model where technological innovation is encouraged within a tightly managed informational environment.

Under conditions of global fragmentation, imperfect supranational frameworks, and technical recalcitrance, national authorities frequently govern through improvisation. The Italian Data Protection Authority's intervention against ChatGPT, analyzed in Chapter 4, exemplifies this dynamic. As Gualdi & Cordella (2024) argue, the authority was compelled to repurpose existing data protection law to regulate a GenAI system. While such intervention enabled decisive action, it unfolded amid interactional uncertainty (Nordström, 2022), where a domestic enforcement choice had to navigate ethical controversies, international fragmentation and the competitive pressures of the global innovation race (Coltri, 2024).

2.3.2.3 The Organizational Challenge: Institutional Logics and Legitimacy

While the pacing problem and global fragmentation define the external environment, they also create acute internal tensions for public authorities, who often face severe asymmetries in technical expertise and resources compared to the industries they oversee (Almada & Petit, 2025; Currie et al., 2025). Thus, governing authorities must navigate a field of structural misalignment (Almada & Petit, 2025; Bakiner, 2023). Institutional theory helps clarify these pressures. Minkkinen & Mäntymäki (2023) argue that AI governance is shaped by friction between four competing institutional logics: AI ethics principlism (adherence to abstract values), managerial rationalism (focus on efficiency and control), IT professionalism (technical instrumentality), and regulatory oversight (compliance with binding rules). Furthermore, regulators often struggle to apply static data modeling logic to the dynamic algorithmic modeling logic of generative AI, resulting in flawed interventions (Gualdi & Cordella, 2024). Governance outcomes, therefore, must find a precarious balance between these conflicting rationalities and the need to render AI systems into legible regulatory objects (Ferrari et al., 2025).

This balancing act is complicated by deep asymmetries in power and legitimacy, which create relational risks wherein technology firms possess the knowledge to withhold data or deny explanations to the public (Currie et al., 2025; Lazar, 2022). Taylor (2021) characterizes these firms as "public actors without public values", possessing infrastructural power that creates relations of domination where citizens cannot meaningfully opt out of data extraction. Authorities are tasked with curbing this domination while themselves operating under strict constraints. In particular, regulators face severe asymmetries in resources and organizational capacity compared to technology giants, often lacking the budgetary resources to compete for the technical expertise required to audit complex AI systems (Almada & Petit, 2025; Guihot et al., 2017). Consequently, mainstream AI alignment discourse often focuses on technical direct alignment with operators' immediate goals rather than social alignment with broader societal welfare (Korinek & Balwit, 2023). Cugurullo (2025) critiques this discourse as a form of “obscure politics" that privileges the interests of powerful operators and masks the human stakeholders who control AI development. This dynamic raises the bar for governance authorities that must demonstrate that their interventions constitute substantive measures rather than symbolic gestures that firms can easily circumvent or compromise through strategic decoupling (Kahdan et al., 2023; Monteiro & Singh, 2025).

2.3.3 Privacy and Data as Primary Levers for GenAI Governance

Because modern GenAI systems depend on extensive data collection, aggregation, and reuse, decisions about their deployment are effectively decisions about the acceptable limits of extraction and informational control. Privacy and data governance thus constitute a primary axis of intervention. Véliz (2022) establishes a demanding normative benchmark for evaluating such interventions. It conceptualizes privacy as being personally unaccessed and retaining negative control over one's information, arguing that privacy functions both as an individual right and as a collective good that shields societies from abuses of power, discrimination, and democratic manipulation. This framework exposes a structural weakness in governance strategies that rely heavily on individual consent: they underestimate informational asymmetries and the collective nature of privacy harms. Véliz (2022) advocates instead for structural safeguards including strict data minimization, storage limitation, and bans on trade in personal data. Regulatory responses that tolerate data-maximizing business models, even when formally consent-based, therefore sit in tension with this benchmark.

Empirical evidence shows how far current practice departs from these standards. Liu et al. (2024) analyze advertising ecosystems under GDPR and CCPA and demonstrate that users who exercise opt-out rights are frequently still tracked and targeted. They attribute this persistence to entrenched technical architectures and business incentives that maintain data flows irrespective of surface-level consent choices. This finding exposes what can be termed a compliance–substance gap: legal frameworks articulate strong privacy rights, while the infrastructures that process data continue to undermine effective control. For public authorities, this gap is central because enforcement strategies focused only on formal compliance will fail to address the underlying dynamics that keep individuals "opted out, yet tracked" (Liu et al., 2024).

The regulatory environment compounds these challenges. The GDPR establishes general principles and rights for processing personal data, while the EU AI Act layers additional system-specific obligations on data quality, logging, risk management, and documentation for high-risk and general-purpose AI systems (Kim et al., 2025; Schuster et al., 2025). Experimental instruments such as AI sandboxes are designed to support supervised innovation, yet Yordanova & Bertels (2024) show that tight restrictions on personal data use can limit the range of feasible experiments. Public authorities must therefore interpret and coordinate these overlapping instruments when responding to concrete deployments, determining in practice how demanding data-related safeguards should be, how much experimentation to tolerate, and where to draw the line on data-intensive AI applications.

GenAI systems have made these choices more salient and visible. Gualdi & Cordella (2024) document how the Italian data protection authority utilized existing data protection law to impose a temporary ban on ChatGPT, citing concerns regarding lawful basis, transparency, and protection of minors. Coltri (2024) highlights how the ensuing debate intertwined privacy concerns with questions about integrity and autonomy. These episodes demonstrate that privacy and data governance are critical levers through which authorities currently shape the trajectory of GenAI diffusion. This is particularly significant at a time when AI-specific rules are still being implemented and global peers are taking divergent regulatory paths (Feldstein, 2024; Zhao, 2023; Currie et al., 2025). Every intervention signals a stance on acceptable data practices, on the permissibility of training regimes, and on the balance between individual protection and innovation. They anchor the authority's mandate to protect individuals and publics from domination and uncontrolled data extraction (Taylor, 2021; Véliz, 2022), while simultaneously exposing authorities to the compliance–substance gap identified in empirical work (Liu et al., 2024) and to tensions between strong rights on paper and fast-moving technical infrastructures (Kim et al., 2025; Schuster et al., 2025; Yordanova & Bertels, 2024). Chapter 4 examines how the Italian Data Protection Authority navigated this tension, limiting (or, more precisely, starting a chain of events that limited) access to ChatGPT in Italy in 2023.

2.4 GENAI DEPLOYMENT IN ORGANIZATIONS: PERFORMANCE, LEARNING AND MEMORY

The previous section treated GenAI primarily as an object of governance. This section shifts to GenAI's impact on work. The core tension is that GenAI delivers immediate, visible gains in speed, output volume, and baseline performance, while creating latent risks of expertise erosion, output homogenization, and fragile organizational competence once work crosses the boundary of what models can reliably support (Chui et al., 2023; Dell'Acqua et al., 2023; Brynjolfsson et al., 2025; Anderson et al., 2024; Simkute et al., 2025; Macnamara et al., 2024). The argument is developed in three steps. Section 2.4.1 consolidates evidence on performance outcomes, emphasizing the levelling effect and the "jagged frontier" where performance can collapse. Section 2.4.2 explains the shift from task execution to GenAI orchestration that can produce an illusion of competence, short-circuit learning, and weaken motivation and professional identity (Subramonyam et al., 2024; Fan et al., 2025; Chen et al., 2025; Amankwah-Amoah & Appiah, 2025). Section 2.4.3 scales the analysis to the collective level, showing how GenAI reconfigures organizational memory from repository-based recall to regenerative remembering, with risks of average collective memory, insight drift, and a growing validation bottleneck concentrated in scarce experts (Walsh & Ungson, 1991; Santos & Valentim, 2021; Schuh, 2024; Wedel, 2025; Storey, 2025; Lindgren, 2025; Schaetzle et al., 2025). Together, these mechanisms ground the models developed in Chapter 5 and clarify how short-run productivity gains can coexist with longer-run fragility in organizational competence and memory infrastructures.

2.4.1 Performance Gains, Leveling Effects & The Jagged Frontier

Causal and field evidence across domains consistently demonstrates that GenAI functions as a virtual expert, structuring unstructured tasks, accelerating completion, and raising average quality. In professional writing, ChatGPT access reduces time-to-completion by approximately 40% while improving rated quality (Noy & Zhang, 2023; Nakavachara et al., 2024). For management consulting tasks within the model's competence region, LLM support increases both throughput and speed while improving output quality (Dell'Acqua et al., 2023). Large-scale field deployments in customer service show productivity gains of approximately 14–15% in issues resolved per hour alongside improved customer sentiment, achieved largely through shorter handling time and standardized response components (Brynjolfsson et al., 2025; Ni et al., 2025; Al Naqbi et al., 2024). In software development, code assistants reliably increase task completion speed, with one controlled study reporting approximately 56% faster completion with Copilot, often without reducing, and sometimes slightly improving, quality (Peng et al., 2023; Weber et al., 2024; Li et al., 2024; Cui, 2025).

Cross-industry syntheses reach a similar conclusion: GenAI expands the feasible frontier of automation and augmentation in knowledge work, but realized gains depend on workflow design and governance (Chui et al., 2023; Al Naqbi et al., 2024; Karunakaran et al., 2025). These individual-level productivity improvements do not necessarily translate into system-level efficiency gains. In open-source development, increased submission volume shifts pressure to review and integration stages, raising coordination and integration costs (Song et al., 2023; Bolici et al., 2025; Varone et al., 2025). Individual-level gains can thus displace bottlenecks rather than remove them. Sectoral analyses anticipate downstream restructuring of value chains and bargaining power, including shifts toward platform providers and firms controlling distinctive brands or data (Amankwah-Amoah et al., 2024; Pagani & Wind, 2025).

The distribution of productivity gains from GenAI use is rarely uniform. A common pattern across domains is leveling: novices benefit disproportionately because GenAI packages fragments of expert practice into accessible scaffolding. In customer support, productivity gains concentrate among less experienced agents, with the tool functioning as real-time coaching distilled from high performers (Brynjolfsson et al., 2025). Similar novice-skewed effects emerge in professional writing (Noy & Zhang, 2023) and software development, where junior developers often experience larger productivity improvements than their senior counterparts (Peng et al., 2023; Cui et al., 2025). GenAI can also substitute for procedural know-how in adjacent tasks: in project planning, it provides structure that allows novices to approximate professional performance (Hettrich et al., 2025).

Leveling is not guaranteed, however. Where effective use depends on system-level context and the ability to validate and integrate outputs, advantages can tilt toward those with deeper expertise. In open-source projects, core developers with architectural understanding benefit more than peripheral contributors because effective prompting and integration presuppose domain and system knowledge (Song et al., 2023). Evidence from medical coding similarly distinguishes task experience from seniority: GenAI complements high task-specific experience by enabling speed without sacrificing accuracy, but offers limited benefits to senior workers whose broader responsibilities and lower trust reduce effective usage (Wang et al., 2023). There is also evidence of regression-to-the-mean effects. Standardized model suggestions can fall below expert standards and may distract or dilute top performers' output (Brynjolfsson et al., 2025; Ni et al., 2025). Moreover, performance distribution outcomes are also shaped by adoption patterns and organizational support structures. Within firms, higher-earning and more digitally literate workers tend to adopt earlier and integrate more effectively, even where lower-skilled workers might benefit substantially in principle (Humlum & Vestergaard, 2025). At the labor-market level, GenAI-related roles are associated with higher demanded cognitive and social skills, suggesting the emergence of "AI-leveraged" positions with increased responsibility and rewards (Gulati et al., 2025). Synthesizing these patterns, workplace inequality resulting from GenAI adoption depends on exposure, complementarity, and support structures as on average tool performance (Karunakaran et al., 2025).

The levelling effect is bounded by what Dell'Acqua et al. (2023) term a jagged frontier: performance improves sharply when tasks fall within the model's reliable competence region, but can collapse when users apply the tool just outside that region or rely on it without sufficient evaluation capacity. Frontier failures can arise from three related mechanisms. First, plausible-but-wrong outputs are difficult to detect. Second, hallucinations require time-consuming debugging or correction. Third, over-reliance converts fluency and stylistic polish into a compliance–substance gap, producing outputs that appear correct and well-formed while embedding shallow reasoning or factual and logical errors (Simkute et al., 2025; Lee et al., 2025). Expert-novice skill differences become particularly consequential at the frontier. For example, novices may lack the expertise required to reject flawed outputs, while experts can be pulled toward standardized mid-level solutions if they follow suggestions inferior to their own judgment (Ni et al., 2025; Prather et al., 2024). The same boundary logic emerges in safety- and liability-sensitive contexts, where performance gains depend on careful ambidextrous design that automates routine components while preserving human responsibility for complex judgment, accountability, and handover quality (Guo et al., 2024; Wachter & Brynjolfsson, 2024). Therefore, GenAI raises average visible performance, but reliability varies across micro-tasks and contexts, creating a jagged rather than uniform frontier. When trust in AI outruns capability, or when users lack adequate evaluation capacity, performance can degrade rather than improve (Dell'Acqua et al., 2023; Ni et al., 2025; Simkute et al., 2025).

2.4.2 From Execution to Orchestration

GenAI's performance gains are largely produced by a shift from execution to orchestration. Where workers previously wrote code or text from scratch, they now frame, prompt, steer, edit, and verify generated outputs. This shift fundamentally alters the cognitive demands of work. As generation becomes cheap, verification becomes cognitively demanding because users no longer build artifacts step-by-step and therefore lose the situational awareness that supports error detection (Simkute et al., 2025).

Mapping GenAI adoption to skill taxonomies evidences this reweighting. GenAI touches a wide range of existing cognitive and technical skills, but it elevates meta-skills such as defining task goals, treating language as a quasi-interface, understanding model limits, and measuring outputs against explicit criteria (Giordano et al., 2024). Human–AI interaction research demonstrates why these demands are non-trivial. Users struggle to envision what the model can do and to translate vague intent into promptable instructions, creating an *instruction gap* and *intentionality gap* that weakens evaluation because success criteria remain underspecified (Subramonyam et al., 2024). Survey evidence similarly frames prompting and iterative refinement as a form of digital competence, unevenly distributed by prior education and experience (Rizal, 2025).

These orchestration skills manifest distinctively in organizational knowledge and collaboration processes. GenAI reduces the cost of capturing and codifying knowledge through workflows such as transcript-to-structure conversion, but it increases the stakes of human validation because the failure mode shifts from "not storing" to "institutionalizing distortions" (Schaetzle et al., 2025; Lindgren, 2025). Workplace studies describe this as a practical identity shift, with workers becoming editors and supervisors of AI drafts rather than primary authors (Callari & Puppione, 2025). In interdependent workflows, breakdown repair and handover management become central to maintaining quality and coordination (Lozie et al., 2024; Bolici et al., 2025; Varone et al., 2025).

A trap emerges when orchestration substitutes for rather than complements the effortful processes through which expertise is formed. In novice programming, two characteristic patterns illustrate how correct outputs can coexist with weak learning. *Shepherding* involves effort spent coaxing the AI, while *drifting* entails following suggestions into opaque dead-ends; both produce syntactically correct code without building robust mental models (Prather et al., 2023). Follow-up evidence suggests widening divergence between student groups: higher-performing students use GenAI selectively and critically, while weaker students accept flawed outputs at high rates and cannot debug them, yet still report subjective learning because the interaction is fluent and polished (Prather et al., 2024; Margulieux et al., 2024). Similar mechanisms appear across learning contexts. In writing experiments, students using ChatGPT engage in less planning and evaluation even when final scores increase, exhibiting what Fan et al. (2025) term *metacognitive laziness*. These students also show limited transfer to subsequent tasks. The same mechanism plausibly generalizes to early-career knowledge work: higher confidence in GenAI is associated with lower enactment of critical thinking in AI-assisted tasks, whereas higher self-confidence predicts more active checking and integration (Lee et al., 2025). These dynamics are not purely cognitive but also motivational. Low self-efficacy and fear of failure predict earlier and heavier AI reliance, turning GenAI into a defensive shield against productive struggle and increasing the risk of an illusion of competence (Margulieux et al., 2024). Without explicit scaffolding, GenAI can produce tool-dependent graduates fluent in interfaces but weak in problem framing and critical evaluation (Hyde et al., 2024; AlBannai, 2025). The evidence is not uniformly pessimistic, however. Some workplace settings show signs of accelerated learning when AI recommendations diffuse best practices that users can internalize (Brynjolfsson et al., 2025). The key distinction is whether GenAI is configured and normatively framed as scaffolding for judgment or as a substitute for judgment (Lee et al., 2025; Hyde et al., 2024).

Even when competence does not immediately decay, GenAI can shift motivational and identity dynamics in ways that make good enough performance self-reinforcing. Experimental evidence demonstrates that human–GenAI collaboration can enhance immediate task performance while undermining intrinsic motivation for subsequent unaided work, consistent with a dependency or "deprivation" pathway (Wu et al., 2025). Other studies find that automating core creative effort reduces perceived meaningfulness even when output quality improves, undermining long-run engagement (Mehler & Krautter, 2024; Manresa et al., 2024). In professional illustration, AI accelerates early quality gains but flattens the high-end quality curve, reducing the marginal returns to refinement and leading creators to stop earlier at acceptable quality (Chen et al., 2025). Qualitative work reveals parallel identity tensions: workers report a crisis of competence and a perceived loss of craftsmanship when AI matches their output (Kobiella et al., 2024).

At the organizational level, these mechanisms can crystallize into unplanned human capital obsolescence. As GenAI devalues routines and shifts value toward orchestration, firms face a choice: invest in continuous reskilling – what Amankwah-Amoah & Appiah (2025) term "planned obsolescence" – or drift into dependence as underlying capabilities depreciate (Amankwah-Amoah & Appiah, 2025; Amankwah-Amoah et al., 2024). Compulsive consultation habits and alienation can reduce ownership and responsibility, encouraging expedient behavior and corner-cutting (Yankouskaya et al., 2025; Hai et al., 2025).

Social dynamics compound these risks. Threat perceptions can trigger knowledge hiding that undermines collective learning, while productivity benefits depend on coordination and trust in AI-enabled processes (Liu et al., 2025; Kassa & Worku, 2025; Khan, Soomro, & Pitafi, 2025; Rafi et al., 2024; Li et al., 2025; Nguyen & Elbanna, 2025). Finally, firm-level rollouts can convert efficiency gains into intensified output expectations rather than reduced workload, contributing to longer workdays and shifting rents rather than "freeing time" for workers (Jiang et al., 2025; Lozie et al., 2024).

Taken together, these findings clarify that GenAI can simultaneously raise and standardize measured performance, compress the apparent skill distribution, and weaken the conditions for deep expertise and sustained excellence. Chapter 5 builds directly on this mechanism by modeling how floor-raising gains can coexist with longer-run capability erosion and increasingly standardized organizational outputs (Chen et al., 2025; Macnamara et al., 2024; Simkute et al., 2025; Amankwah-Amoah & Appiah, 2025; Anderson et al., 2024; Wei et al., 2025; Lee & Chung, 2024; Pagani & Wind, 2025).

2.4.3 GenAI & Organizational Memory

If workers can fall into a state of over-reliance on GenAI systems, as described in Section 2.4.2, the organizational-level analogue is a loss of diversity in what gets encoded, stored and retrieved from the organizational memory. Organizational memory has traditionally been theorized as distributed across multiple storage bins (individuals, culture, routines and transformations, structures, ecology, and external archives) activated through cues and social practices rather than accessed as a single database (Walsh & Ungson, 1991). More recent accounts similarly treat memory as a mix of codified artifacts, norms, routines, and tacit know-how that is continuously transformed and reused rather than simply stored (Santos & Valentim, 2021). Emerging research is supporting the idea that GenAI does not merely accelerate retrieval within this repertoire; it changes the ontology of organizational remembering.

2.4.3.1 Regenerative Remembering

GenAI transforms organizational remembering into a regenerative process. Instead of retrieving the document, workers query a system that synthesizes answers from latent representations and retrieved fragments. This reconfiguration is enabled by the mobilization of previously underused data (e.g., emails, logs, audio and video recordings) which older knowledge systems could store but could not make meaningfully searchable at scale (Wang & Legner, 2025; Nonato & Perez, 2025). In practical terms, employees can query internal knowledge bases in natural language rather than navigating rigid taxonomies or keyword search, reducing search friction and compressing the time spent on information gathering (Chui et al., 2023; Alavi et al., 2024; Jarrahi et al., 2023). Retrieval-Augmented Generation (RAG) architectures operationalize this shift by combining an LLM's parametric memory with external, organization-specific non-parametric memory such as a vector database, improving factual grounding and reducing hallucinations relative to standalone generation (Klesel & Wittmann, 2025). In parallel, agentic systems increasingly function as external memory, proactively surfacing relevant material and embedding recall within workflow rather than treating retrieval as a discrete step (Anjelia et al., 2025).

The generative nature of GenAI-mediated organizational remembering introduces a structural fragility: memory becomes opaque and probabilistic, and organizations can lose the "why" behind past actions even when they retain the "what." Wedel (2025) argues that mainstream GenAI memory architectures, including many RAG implementations, often produce shallow memory. They retrieve by similarity and regenerate coherent text, but fail to preserve decision rationales, rejected alternatives, and the lineage of how conclusions were reached. Over time, this creates insight drift, where rationales decay or become untraceable as contexts change (Wedel, 2025).

This fragility matters because GenAI simultaneously reshapes multiple dimensions of knowledge management. It amplifies the combination mode by synthesizing large volumes of explicit knowledge into new patterns and can support internalization by acting as a personalized coach, but it can simultaneously erode socialization if employees turn to the model rather than colleagues for interpretation and tacit norms (Alavi et al., 2024). Jarrahi et al. (2023) frame this as a human–AI symbiosis in which AI excels at patterning and structuring while humans supply contextualization, judgment, and ethical sense-making. Nault & Ruhi (2024) extend this implication: AI-enabled knowledge management needs to be embedded in learning organization architectures that deliberately connect models to internal knowledge bases, design meaningful collaboration, and manage hallucination risk as a socio-technical problem rather than a pure IT defect.

The economics of codification change sharply as well. GenAI can transform interviews and audiovisual traces from departing experts into structured artifacts at low marginal cost, expanding what can be stored and searched, but it also increases the chance that subtle distortions enter the memory system unless validation becomes systematic (Schaetzle et al., 2025; Nonato & Perez, 2025).

2.4.3.2 Memory Imperialism and the Shift Toward Average Outputs

A risk of moving from retrieval to regeneration is that organizational remembering becomes dominated by what is most probable rather than what is most locally diagnostic. Schuh (2024) characterizes large models as *average collective memory agents*: they do not recall specific episodes but generate plausible narratives from latent spaces shaped by aggregate data. The organizational analogue is memory imperialism, in which dominant patterns in the model's priors – and in practice, in the most frequently retrieved internal documents – crowd out idiosyncratic histories, dissenting interpretations, and locally situated exceptions that may have been strategically decisive in the past (Schuh, 2024; Walsh & Ungson, 1991). Evidence from creativity research strengthens this concern empirically. GenAI can increase idea volume and raise average creativity ratings, but groups become more semantically similar: diversity falls even as mean quality rises (Anderson et al., 2024). The same homogenization mechanism can operate in knowledge work when organizations standardize prompts, templates, and best practice responses. Variance collapses toward the mean, and unusual but valuable approaches become harder to reproduce or even to remember as legitimate options (Alavi et al., 2024; Chui et al., 2023).

Routine dynamics provide another channel through which GenAI can homogenize organizational memory. Mahringer et al. (2024) demonstrate that AI can shield portions of routines by black-boxing steps, reducing variance and reflection. While this can stabilize performance, it also interrupts feedback-based learning and decreases the local experimentation through which organizations update what they know. AI can also reframe routines by proposing new connections that stimulate reconceptualization, but the key point is that GenAI becomes an actor in determining where variability is suppressed and where it survives (Mahringer et al., 2024).

At the organizational level, this interacts with coordination structures. AI can raise performance when interdepartmental coordination allows units to align around AI-enabled processes, but fragmented settings risk reinforcing silos, producing multiple incompatible local truths while simultaneously flattening within-unit diversity (Kassa & Worku, 2025). Team communication evidence supports the conclusion that social channels thin under AI mediation: human–AI teams generate more task-oriented content but show less social and emotional communication, weakening the relational substrate that traditionally supports tacit transfer and interpretive pluralism (Ju & Aral, 2025; Alavi et al., 2024). Finally, threat perceptions can actively distort what enters organizational memory. When employees see AI as undermining their job value, they may respond with knowledge hiding, reducing the inflow of non-codified or dissenting knowledge into collective repositories (Liu et al., 2025).

This homogenization dynamic is not inevitable. Simulation work suggests that high-capability AI can, under some conditions, substitute for human exploration by generating belief diversity and rapid updates to an organizational code, especially in turbulent environments where consensus is slow (Sturm et al., 2021). The problem is that many deployments in practice optimize for speed, standardization, and compliance – conditions that systematically favor average, repeatable outputs over dissent, edge cases, and historically contingent rationales (Schuh, 2024; Mahringer et al., 2024; Anderson et al., 2024). The result is a drift toward *average memory* even when the system could have been configured to preserve heterogeneity.

2.4.3.3 The Validation Burden and the Expert Bottleneck

GenAI is often framed as democratizing organizational knowledge by lowering the cost of retrieval and recombination. Organizations are increasingly turning GenAI into institution-specific memory assets: proprietary knowledge bases can be queried through natural-language interfaces functioning as virtual experts, accelerating access to previously siloed documents and records (Chui et al., 2023). In software teams, GenAI tools can internalize local coding standards and practices, functioning as dynamic repositories of team norms (Li et al., 2024). Firms increasingly build bespoke internal tools aligned with specific workflows rather than relying only on generic models, explicitly encoding local processes and context into the system (Lozie et al., 2024). In research and development and technical work, GenAI can reduce information gaps by translating between natural-language queries and technical artifacts, improving distribution across technical and non-technical units (Al Naqbi et al., 2024).

The democratization narrative obscures a critical bottleneck: validation becomes the scarce resource. GenAI increases the volume of codified artifacts and the speed of retrieval, but it also increases the probability that hallucinations, subtle distortions, or decontextualized chunks contaminate the knowledge base or downstream decisions (Schaetzle et al., 2025; Klesel & Wittmann, 2025; Wedel, 2025). Schaetzle et al. (2025) argue that the central human contribution shifts from documenting to validating: the problem becomes less about capturing knowledge than preventing low-grade corruption of memory at scale.

Storey (2025) identifies the distributional implication of this shift. As novices rely on GenAI for execution, experts increasingly serve as quality evaluators of AI outputs, but if GenAI simultaneously reduces the friction-based learning through which novices become experts, the supply of validators can shrink precisely when validation needs grow. This represents the organizational manifestation of the trap described in Section 2.4.2: an expanding base of users drawing on generative memory systems confronts a thinning apex of humans capable of auditing them (Storey, 2025; Schaetzle et al., 2025; Perozzo & Al-Ahmadi, 2025).

Empirical workflow studies make this bottleneck concrete. In AI-assisted medical coding, workers with high task-specific experience are positioned to supervise and correct AI outputs, while senior staff, often accountable for broader consequences, may trust the system less and integrate it more cautiously, creating friction in routinization and reducing the net organizational learning benefit (Wang et al., 2023). Lindgren (2025) reframes the deeper issue as one of grounding and shared meaning: humans sometimes must act as tutors to AI systems, training them on the organization's socially constructed reality if hybrid teams are to form genuine "we-intentions" and support collective learning rather than fragmented compliance.

Technical mitigations help but do not remove the bottleneck. RAG architectures can reduce hallucinations by grounding responses in internal sources and can provide references that support verification, but they still tend to preserve content more easily than rationale and introduce new dependencies on data quality, retrieval design, and MLOps capabilities (Klesel & Wittmann, 2025; Wedel, 2025). Proposals to strengthen integrity via immutable audit trails such as blockchain-based approaches address provenance and verifiability but do not automatically solve the interpretive and evaluative burden required to decide what should become canonical organizational memory (Nonato & Perez, 2025).

Thus, following from these accounts, GenAI emerges as an additional layer of organizational memory information systems that can expand access and accelerate remembering, but it also exerts structuring power over which versions of the past become legible, retrievable, and therefore actionable. The key risk is not that organizations will forget everything, but that they may remember in a more standardized, less context-rich, and less plural way while concentrating responsibility for epistemic integrity on a shrinking set of experts (Schuh, 2024; Wedel, 2025; Schaetzle et al., 2025; Storey, 2025; Lindgren, 2025).

2.5 SYNTHESIS AND CONCLUSIONS

The literature reviewed in this chapter operates at two levels of analysis: the level of public governance of GenAI diffusion and that of organizational deployment. Despite this separation, a critical view suggests a structural parallel. In both domains, interaction with GenAI produces a systematic divergence between visible, short-term indicators of success and latent, long-term structural erosions. More precisely, the literature converges on a recurring pattern: decision-makers, whether regulators or managers, tend to optimize for outcomes that are immediate and legible, while the foundational capacities required to sustain those outcomes quietly degrade.

TABLE 2.1: Tensions in the literature.

The Visible Signal	The Latent Erosion	The Structural Mechanism
GOVERNANCE OF GENAI DIFFUSION		
Formal Compliance • Visible interventions (bans, fines) • Adherence to static rules • First-mover signaling in the regulatory race	***Institutional Legitimacy*** • Widening compliance-substance gap • Loss of technical alignment (obsolescence)	***The Pacing Problem*** Authorities prioritize immediate, defensible actions using best-proxy instruments (like privacy law) to manage the uncertainty of the

	• Erosion of public trust in regulator capacity	regulatory void, sacrificing long-term structural fit.
GENAI ORGANIZATIONAL DEPLOYMENT		
Productivity Floor-Raising • Increased speed and output volume • Leveling up of novice performance • Reduction in search friction	***Structural Capability & Diversity*** • Metacognitive laziness (skill decay) • Illusion of competence (blind trust) • Memory homogenization (loss of local, contextual diversity) • Validation bottleneck (expanding archives with fewer experts to audit them)	***Augmentation vs. Fragility*** GenAI automates the productive struggle required for individual learning and the friction required for diverse memory formation. Organizations attend to the speed of standardized outputs while ignoring the decay of judgment and the degradation of (reconstructed versions of) their organizational memory.

In the governance domain, this divergence manifests as tension between formal compliance and institutional legitimacy (Allenby, 2011; Marchant, 2011, 2020; Almada & Petit, 2025; Lubinski, 2024). The literature on the pacing problem suggests that authorities must choose between acting early with misaligned tools, waiting until the technology is entrenched, or abdicating responsibility entirely. Pressured to signal relevance, authorities often choose the first path, using proxy frameworks such as data protection law to police general-purpose technologies. While these interventions generate measurable outputs (e.g., fines, bans, and compliance checklists) they often fail to address the underlying technical reality. The result is performative governance: visible indicators of regulation are satisfied, but the latent capacity to govern the technology's diffusion effectively erodes through legitimacy crises and technical obsolescence.

In the organizational domain, the same dynamic appears as an augmentation vs. fragility tension. Empirical evidence confirms that GenAI raises workers' performance, creating immediate, visible gains in speed and output quality. However, these gains frequently mask a latent hollowing out of human capability and organizational memory (Huber, 1991; Levitt & March, 1988; Walsh & Ungson, 1991). As work shifts from execution to orchestration, the productive struggle required to build expertise is bypassed, leading to metacognitive laziness and the illusion of competence. Simultaneously, the shift from distributed memory to generative retrieval homogenizes organizational knowledge (Brynjolfsson et al., 2025; Peng et al., 2023; Macnamara et al., 2024; Schaetzle et al., 2025; Ni et al., 2025; Anderson et al., 2024). Organizations thus appear increasingly productive in the short run even as the human and mnemonic infrastructures required to validate that productivity become fragile.

These dual dynamics constitute the core phenomenon this thesis investigates. The reviewed literature effectively describes that this divergence occurs, but it largely fails to explain why organizations consistently fall into this pattern. This thesis argues that the root cause is not merely technical but attentional. GenAI interacts with the established *attention structures* of organizations (Ocasio, 1997), making productivity and compliance successes highly salient, while rendering skill durability and institutional legitimacy opaque. This mechanism is conceptualized in Chapter 3 as the *Visibility Trap*. The remainder of this thesis operationalizes this insight. Chapter 3 develops the theoretical mechanism of the trap through the Attention-Based View. Chapter 4 empirically investigates the governance side of the trap through the case of the Italian Data Protection Authority ban of ChatGPT. Chapter 5 theoretically models the organizational side, using Skill Acquisition Theory to formalize the mechanism of skill non-development and reconceptualizing GenAI as a new form of Organizational Memory Information System (Stein & Zwass, 1995; Olivera, 2000) that centralizes and homogenizes knowledge.

CHAPTER REFERENCES

Al Naqbi, H., Bahroun, Z., & Ahmed, V. (2024). Enhancing Work Productivity through Generative Artificial Intelligence: A Comprehensive Literature Review. *Sustainability*, *16*(3), 1166. https://doi.org/10.3390/su16031166

Alavi, M., Leidner, D. E., & Mousavi, R. (2024). Knowledge Management Perspective of Generative Artificial Intelligence. Journal of the Association for Information Systems, 25(1), 1–12. https://doi.org/10.17705/1jais.00859

AlBannai, N. A. (2025). From classroom to Boardroom: Ethical and cognitive challenges of ChatGPT use among DBA students in business problem-solving. *The International Journal of Management Education*, *23*(3), 101254. https://doi.org/10.1016/j.ijme.2025.101254

Allenby, B. R. (2011). Governance and Technology Systems: The Challenge of Emerging Technologies. In G. E. Marchant, B. R. Allenby, & J. R. Herkert (Eds.), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (Vol. 7, pp. 3–18). Springer Netherlands. https://doi.org/10.1007/978-94-007-1356-7_1

Almada, M., & Petit, N. (2025). The EU AI Act: Between the rock of product safety and the hard place of fundamental rights. *Common Market Law Review*, *62*(1). https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2025004.pdf

Amankwah-Amoah, J., Abdalla, S., Mogaji, E., Elbanna, A., & Dwivedi, Y. K. (2024). The impending disruption of creative industries by generative AI: Opportunities, challenges, and research agenda. *International Journal of Information Management*, *79*, 102759. https://doi.org/10.1016/j.ijinfomgt.2024.102759

Amankwah-Amoah, J., & Appiah, G. (2025). Unmasking the silent threat: AI-induced human capital obsolescence and business failure. *Journal of Managerial Psychology*. https://doi.org/10.1108/JMP-05-2024-0385

Anderson, B. R., Shah, J. H., & Kreminski, M. (2024). Homogenization Effects of Large Language Models on Human Creative Ideation. *Proceedings of the 16th Conference on Creativity & Cognition*, 413–425. https://doi.org/10.1145/3635636.3656204

Anjelia, S. R., Sensuse, D. I., & Lusa, S. (2025). AI Agents for Organizational Knowledge Retrieval and Sharing: A Systematic Literature Review. International Journal of Advances in Data and Information Systems, 6(3), 824–839. https://doi.org/10.59395/ijadis.v6i3.1462

Bakiner, O. (2023). Pluralistic sociotechnical imaginaries in Artificial Intelligence (AI) law: The case of the European Union's AI Act. *Law, Innovation and Technology*, *15*(2), 558–582. https://doi.org/10.1080/17579961.2023.2245675

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics*, *5*(3), 3265–3279. https://doi.org/10.1007/s43681-024-00653-w

Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: Themes, knowledge gaps and future agendas. *Internet Research*, *33*(7), 133–167. https://doi.org/10.1108/INTR-01-2022-0042

Bolici, F., Crowston, K., Varone, A., & Fudge, M. (2025). *Rethinking Programming Skills in the Age of Generative AI*.

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics*. https://doi.org/10.1093/qje/qjae044

Callari, T. C., & Puppione, L. (2025). Can generative artificial intelligence productivity tools support workplace learning? A qualitative study on employee perceptions in a multinational corporation. *Journal of Workplace Learning*, *37*(3), 266–283. https://doi.org/10.1108/JWL-11-2024-0258

Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. Expert Systems, 41(7), e13406. https://doi.org/10.1111/exsy.13406

Cancela-Outeda, C. (2024). The EU's AI act: A framework for collaborative governance. *Internet of Things*, *27*, 101291. https://doi.org/10.1016/j.iot.2024.101291

Cihon, P., Maas, M. M., & Kemp, L. (2020). Should Artificial Intelligence Governance be Centralised? Design Lessons from History. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, AIES '20, 228–234. https://doi.org/10.1145/3375627.3375857

Chen, Y. J., Gong, J., Li, J., & Zhao, Z. (2025). Better Technology, Worse Motivation: GenAI's Mediocrity Trap.

Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zemmel, R. (2023). *The economic potential of generative AI: The next productivity frontier*. McKinsey & Company.

Coltri, M. A. (2024). The Ethical Dilemma with Open AI ChatGPT: Is it Right or Wrong to prohibit it? *Athens Journal of Law*, *10*(1), 119–130.

Cugurullo, F. (2025). The obscure politics of artificial intelligence: A Marxian socio-technical critique of the AI alignment problem thesis. *AI and Ethics*, *5*(2), 1189–1201. https://doi.org/10.1007/s43681-024-00476-9

Cui, Z. (Kevin), Demirer, M., Jaffe, S., Musolff, L., Peng, S., & Salz, T. (2025). *The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers* (SSRN Scholarly Paper No. 4945566). Social Science Research Network. https://doi.org/10.2139/ssrn.4945566

Currie, W. L., Leimeister, J. M., Schlagwein, D., & Willcocks, L. (2025). Rethinking technology regulation in the age of AI risks. *Journal of Information Technology*, *40*(3), 236–245. https://doi.org/10.1177/02683962251378815

Dafoe, A. (2023). AI Governance: Overview and Theoretical Lenses. In J. B. Bullock, Y.-C.

Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., pp. 21–44). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.2

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4573321

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Feldstein, S. (2024). Evaluating Europe's push to enact AI regulations: How will this influence global norms? *Democratization*, *31*(5), 1049–1066. https://doi.org/10.1080/13510347.2023.2196068

Ferrari, F., van Dijck, J., & van den Bosch, A. (2025). Observe, inspect, modify: Three conditions for generative AI governance. New Media & Society, 27(5), 2788–2806. https://doi.org/10.1177/14614448231214811

G'sell, F. (2024). Regulating under Uncertainty: Governance Options for Generative AI (SSRN Scholarly Paper No. 4918704). Social Science Research Network. https://doi.org/10.2139/ssrn.4918704

Garfinkel, B. (2022). The Impact of Artificial Intelligence: A Historical Perspective. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., pp. 86–108). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.5

Giordano, V., Spada, I., Chiarello, F., & Fantoni, G. (2024). The impact of ChatGPT on human skills: A quantitative study on twitter data. *Technological Forecasting and Social Change*, *203*, 123389. https://doi.org/10.1016/j.techfore.2024.123389

Gualdi, F., & Cordella, A. (2024). Theorizing the regulation of generative AI: lessons learned from Italy's ban on ChatGPT. *Proceedings of the 57th Hawaii International Conference on System Sciences*. HICSS 2024. https://hdl.handle.net/10125/106631

Guihot, M., Matthew, A. F., & Suzor, N. P. (2017). Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence. Vanderbilt Journal of Entertainment & Technology Law, 20, 385.

Gulati, P., Marchetti, A., Puranam, P., & Sevcenko, V. (2025). Generative AI Adoption and Higher Order Skills (arXiv:2503.09212). arXiv. https://doi.org/10.48550/arXiv.2503.09212

Guo, M., Gu, M., & Huo, B. (2024). The impacts of automation and augmentation AI use on physicians' performance: An ambidextrous perspective. *International Journal of Operations & Production Management*, *45*(1), 114–151. https://doi.org/10.1108/IJOPM-06-2023-0509

Hagemann, R., Huddleston Skees, J., & Thierer, A. (2018). Soft Law for Hard Problems: The Governance of Emerging Technologies in an Uncertain Future. Colorado Technology Law Journal, 17, 37.

Hai, S., Long, T., Honora, A., Japutra, A., & Guo, T. (2025). The dark side of employee-generative AI collaboration in the workplace: An investigation on work alienation and employee expediency. *International Journal of Information Management*, *83*, 102905. https://doi.org/10.1016/j.ijinfomgt.2025.102905

Hemphill, T. A. (2020). "The innovation governance dilemma: Alternatives to the precautionary principle." *Technology in Society*, *63*, 101381.

https://doi.org/10.1016/j.techsoc.2020.101381

Hettrich, B., Krings, N., & Kock, A. (2025). Bridging the Expertise Gap: The Role of Generative AI in Supporting Project Planning Tasks for Novices and Professionals. *Creativity and Innovation Management*. https://doi.org/10.1111/caim.70002

Huber, G. P. (1991). Organizational Learning: The Contributing Processes and the Literatures. *Organization Science*, *2*(1), 88–115. https://doi.org/10.1287/orsc.2.1.88

Humlum, A., & Vestergaard, E. (2025). The unequal adoption of ChatGPT exacerbates existing inequalities among workers. *Proceedings of the National Academy of Sciences*, *122*(1), e2414972121. https://doi.org/10.1073/pnas.2414972121

Hyde, S. J., Busby, A., & Bonner, R. L. (2024). Tools or Fools: Are We Educating Managers or Creating Tool-Dependent Robots? *Journal of Management Education*, *48*(4), 708–734. https://doi.org/10.1177/10525629241230357

Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2023). Artificial intelligence and knowledge management: A partnership between human and AI. *Business Horizons*, *66*(1), 87–99. https://doi.org/10.1016/j.bushor.2022.03.002

Jiang, W., Park, J., Xiao, R. (Jiqiu), & Zhang, S. (2025). *AI and the Extended Workday: Productivity, Contracting Efficiency, and Distribution of Rents* (No. w33536; p. w33536). National Bureau of Economic Research. https://doi.org/10.3386/w33536

Ju, H., & Aral, S. (2025). *Collaborating with AI Agents: Field Experiments on Teamwork, Productivity, and Performance* (No. arXiv:2503.18238). arXiv. https://doi.org/10.48550/arXiv.2503.18238

Kahdan, M., Hartwich, N., Salge, O., & Cichy, P. (2023). Navigating Uncertain Waters: How Organizations Respond to Institutional Pressure in Times of the Looming EU AI Act. *ICIS 2023 Proceedings*. https://aisel.aisnet.org/icis2023/gov_strategy/gov_strategy/12

Karunakaran, A., Lebovitz, S., Narayanan, D., & Rahman, H. A. (2025). Artificial Intelligence at Work: An Integrative Perspective on the Impact of AI on Workplace Inequality. *Academy of Management Annals*. https://doi.org/10.5465/annals.2023.0230

Kassa, B. Y., & Worku, E. K. (2025). The impact of artificial intelligence on organizational performance: The mediating role of employee productivity. *Journal of Open Innovation: Technology, Market, and Complexity*, *11*(1), 100474. https://doi.org/10.1016/j.joitmc.2025.100474

Khan, A. N., Soomro, M. A., & Pitafi, A. H. (2025). AI in the Workplace: Driving Employee Performance Through Enhanced Knowledge Sharing and Work Engagement. *International Journal of Human–Computer Interaction*, *41*(17), 10699–10712. https://doi.org/10.1080/10447318.2024.2436611

Kim, B.-J., Jeong, S., Cho, B.-K., & Chung, J.-B. (2025). AI Governance in the Context of the EU AI Act. *IEEE Access*, *13*, 144126–144142. https://doi.org/10.1109/ACCESS.2025.3598023

Klesel, M., & Wittmann, H. F. (2025). Retrieval-Augmented Generation (RAG). Business & Information Systems Engineering, 67(4), 551–561.

Kobiella, C., Flores López, Y. S., Waltenberger, F., Draxler, F., & Schmidt, A. (2024). “If the Machine Is As Good As Me, Then What Use Am I?” – How the Use of ChatGPT Changes Young Professionals’ Perception of Productivity and Accomplishment. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3613904.3641964

Korinek, A., & Balwit, A. (2023). Aligned with Whom?: Direct and Social Goals for AI Systems. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., pp. 65–85). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.4

Kulothungan, V., & Gupta, D. (2025). Towards Adaptive AI Governance: Comparative Insights from the U.S., EU, and Asia. *2025 IEEE 11th Conference on Big Data Security on Cloud (BigDataSecurity)*, 32–38. https://doi.org/10.1109/BigDataSecurity66063.2025.00018

Lazar, S. (2022). Power and AI: Nature and Justification. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., pp. 198–209). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.12

Lee, B. C., & Chung, J. (Jae). (2024). An empirical investigation of the impact of ChatGPT on creativity. *Nature Human Behaviour*, *8*(10), 1906–1914. https://doi.org/10.1038/s41562-024-01953-1

Lee, H.-P. (Hank), Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–22. https://doi.org/10.1145/3706598.3713778

Levitt, B., & March, J. G. (1988). Organizational Learning. *Annual Review of Sociology*, *14*(1), 319–338. https://doi.org/10.1146/annurev.so.14.080188.001535

Li, N., Zhou, H., & Mikel-Hong, K. (2024). *Generative AI Enhances Team Performance and Reduces Need for Traditional Teams* (No. arXiv:2405.17924). arXiv. https://doi.org/10.48550/arXiv.2405.17924

Li, X., Huang, M., Liu, J., Fan, Y., & Cui, M. (2025). The Impact of AI Negative Feedback vs. Leader Negative Feedback on Employee Withdrawal Behavior: A Dual-Path Study of Emotion and Cognition. *Behavioral Sciences*, *15*(2), 152. https://doi.org/10.3390/bs15020152

Lindgren, H. (2025). Emerging Roles and Relationships Among Humans and Interactive AI Systems. *International Journal of Human–Computer Interaction*, *41*(17), 10595–10617. https://doi.org/10.1080/10447318.2024.2435693

Liu, Z., Lin, Q., Tu, S., & Xu, X. (2025). When robot knocks, knowledge locks: How and when does AI awareness affect employee knowledge hiding? Frontiers in Psychology, 16, 1627999. https://doi.org/10.3389/fpsyg.2025.1627999

Liu, Z., Iqbal, U., & Saxena, N. (2024). Opted Out, Yet Tracked: Are Regulations Enough to Protect Your Privacy? *Proceedings on Privacy Enhancing Technologies*, *2024*(1), 280–299. https://doi.org/10.56553/popets-2024-0016

Lozie, D. R., Omasa, R., Hesami, S., Zaman, S., Kajbaf, M., & Malik, A. R. (2024). Examining the impact of generative artificial intelligence on work dynamics. *Human Resources Management and Services*, *6*(2), 3420. https://doi.org/10.18282/hrms.v6i2.3420

Lubinski, F. (2024). *The European Union AI Act: Airplanes, Robots, and the Pacing Problem. With the AI Act, did the EU get the Pacing of the Law Right?**
*** (SSRN Scholarly Paper No. 5188418). Social Science Research Network. https://doi.org/10.2139/ssrn.5188418

Maas, M. M. (2022). Aligning AI Regulation to Sociotechnical Change. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., pp. 358–380). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.22

Macnamara, B. N., Berber, I., Çavuşoğlu, M. C., Krupinski, E. A., Nallapareddy, N., Nelson, N. E., Smith, P. J., Wilson-Delfosse, A. L., & Ray, S. (2024). Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cognitive Research: Principles and Implications*, *9*(1), 46.

https://doi.org/10.1186/s41235-024-00572-8

Mahringer, C., Danner-Schröder, A., Müller-Seitz, G., & Renzl, B. (2024). How Does Artificial Intelligence Promote Change and Stability of Organizational Routines? The Role of Automation and Augmentation. *Journal of Competences, Strategy & Management*, *12*, 1–17. https://doi.org/10.25437/jcsm-vol12-104

Manresa, A., Sammour, A., Mas-Machuca, M., Chen, W., & Botchie, D. (2024). Humanizing GenAI at work: Bridging the gap between technological innovation and employee engagement. *Journal of Managerial Psychology*, *40*(5), 472–492. https://doi.org/10.1108/JMP-05-2024-0356

Marchant, G. E. (2011). Addressing the Pacing Problem. In G. E. Marchant, B. R. Allenby, & J. R. Herkert (Eds.), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (pp. 199–205). Springer Netherlands. https://doi.org/10.1007/978-94-007-1356-7_13

Marchant, G. (2019). "Soft Law" Governance of Artificial Intelligence. https://escholarship.org/uc/item/0jq252ks

Marchant, G. E. (2020). Governance of Emerging Technologies as a Wicked Problem. *Vanderbilt Law Review*, *73*, 1861.

Marchant, G. E., & Gutierrez, C. I. (2022). Soft Law 2.0: An Agile and Effective Governance Approach for Artificial Intelligence. Minnesota Journal of Law, Science and Technology, 24, 375.

Margulieux, L. E., Prather, J., Reeves, B. N., Becker, B. A., Cetin Uzun, G., Loksa, D., Leinonen, J., & Denny, P. (2024). Self-Regulation, Self-Efficacy, and Fear of Failure Interactions with How Novices Use LLMs to Solve Programming Problems. *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1*, 276–282. https://doi.org/10.1145/3649217.3653621

Mehler, M., & Krautter, K. (2024). Productivity vs. Purpose: Generative AI Enhances Task Performance but Reduces Meaningfulness in Programming. *ECIS 2024 Proceedings*.

Minkkinen, M., & Mäntymäki, M. (2023). THE INSTITUTIONAL LOGICS UNDERPINNING ORGANIZATIONAL AI GOVERNANCE PRACTICES. *14th Scandinavian Conference on Information Systems*. https://aisel.aisnet.org/scis2023/12

Monteiro, N., & Singh, V. (2025). The wheel of artificial intelligence governance. *Sustainable Futures*, *10*, 101279. https://doi.org/10.1016/j.sftr.2025.101279

Mueller, M. L. (2025). It's just distributed computing: Rethinking AI governance. *Telecommunications Policy*, *49*(3), 102917. https://doi.org/10.1016/j.telpol.2025.102917

Nakavachara, V., Potipiti, T., & Chaiwat, T. (2024). *Experimenting with Generative AI: Does ChatGPT Really Increase Everyone's Productivity?* (No. arXiv:2403.01770). arXiv. https://doi.org/10.48550/arXiv.2403.01770

Nault, K., & Ruhi, U. (2024). From Knowledge Management to a Learning Organization: Strategic Use of AI Tools. *2024 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*, 1–6. https://doi.org/10.1109/ICTMOD63116.2024.10878180

Nguyen, T., & Elbanna, A. (2025). Understanding Human-AI Augmentation in the Workplace: A Review and a Future Research Agenda. Information Systems Frontiers. https://doi.org/10.1007/s10796-025-10591-5

Ni, X., Wang, Y., Feng, T., Lu, L. X., Wang, Y., & Zhou, C. (2025). *Generative AI in Action: Field Experimental Evidence on Worker Performance in E-Commerce Customer Service Operations*. SSRN. https://doi.org/10.2139/ssrn.5012601

Nonato, J. A. A., & Perez, G. (2025). From challenges to opportunities: The impact of emerging technologies on enhancing organizational memory systems. *International*

Journal of Scientific Management and Tourism, *11*(1), e1247–e1247. https://doi.org/10.55905/ijsmtv11n1-001

Nordström, M. (2022). AI under great uncertainty: Implications and decision strategies for public policy. *AI & SOCIETY*, *37*(4), 1703–1714. https://doi.org/10.1007/s00146-021-01263-4

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Ocasio, W. (1997). Towards an Attention-Based View of the Firm. *Strategic Management Journal*, *18*(S1), 187–206. https://doi.org/10.1002/(SICI)1097-0266(199707)18:1+%253C187::AID-SMJ936%253E3.0.CO;2-K

Olivera, F. (2000). Memory Systems In Organizations: An Empirical Investigation Of Mechanisms For Knowledge Collection, Storage And Access. *Journal of Management Studies*, *37*(6), 811–832. https://doi.org/10.1111/1467-6486.00205

Pagani, M., & Wind, Y. (2025). Unlocking Marketing Creativity Using Artificial Intelligence. Journal of Interactive Marketing, 60(1), 1–24. https://doi.org/10.1177/10949968241265855

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. The Journal of Strategic Information Systems, 34(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). *The Impact of AI on Developer Productivity: Evidence from GitHub Copilot*. https://doi.org/10.48550/ARXIV.2302.06590

Perozzo, H., & Al-Ahmadi, M. (2025, August 15). GenAI Chatbots and Knowledge Management in Software Development: A Qualitative Exploration. AMCIS 2025 Proceedings. https://aisel.aisnet.org/amcis2025/sigadit/sigadit/37

Prather, J., Reeves, B. N., Denny, P., Becker, B. A., Leinonen, J., Luxton-Reilly, A., Powell, G., Finnie-Ansley, J., & Santos, E. A. (2023). "It's Weird That it Knows What I Want": Usability and Interactions with Copilot for Novice Programmers. *ACM Trans. Comput.-Hum. Interact.*, *31*(1), 4:1-4:31. https://doi.org/10.1145/3617367

Prather, J., Reeves, B. N., Leinonen, J., MacNeil, S., Randrianasolo, A. S., Becker, B. A., Kimmel, B., Wright, J., & Briggs, B. (2024). The Widening Gap: The Benefits and Harms of Generative AI for Novice Programmers. *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1*, *1*, 469–486. https://doi.org/10.1145/3632620.3671116

Rafi, M., Aitken, J. M., Fatah, T. D., & Mailangkay, A. (2024). Analyzing The Impact of Generative AI on IT Employee Performance. *2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT)*, 1–7. https://doi.org/10.1109/ICCIT62134.2024.10701171

Rejeski, D. (2011). Public Policy on the Technological Frontier. In G. E. Marchant, B. R. Allenby, & J. R. Herkert (Eds.), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (Vol. 7, pp. 47–59). Springer Netherlands. https://doi.org/10.1007/978-94-007-1356-7_4

Rizal, K. A. (2025). *Prompt Engineering and the Effectiveness of Large Language Models in Enhancing Human Productivity*. Zenodo. https://doi.org/10.5281/ZENODO.15381809

Sætra, H. S. (2023). Generative AI: Here to stay, but for good? Technology in Society, 75, 102372. https://doi.org/10.1016/j.techsoc.2023.102372

Santos, J. C. dos, & Valentim, M. L. P. (2021). Knowledge Management and Its Interrelations With Organizational Memory: Focus on the Repertoire Memory. In *Role of Information*

Science in a Complex Society (pp. 145–162). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-7998-6512-4.ch009

Schaetzle, A., Zöll, A., Sammet, N., Brack, F., Scholz, J., & Schwarz, L. (2025). A Research Agenda for Generative AI in Organizational Knowledge Management: Addressing Risks in Knowledge Storage, Retrieval, and Transfer. *ICIS 2025 Proceedings*. https://aisel.aisnet.org/icis2025/is_transformwork/is_transformwork/6

Schuh, J. (2024). AI As Artificial Memory: A Global Reconfiguration of Our Collective Memory Practices? *Memory Studies Review*, *1*(2), 231–255. https://doi.org/10.1163/29498902-202400012

Schuster, T., Waidelich, L., Schneider, A., & Lambert, M. (2025). Risk Classification and Compliance of AI Systems under the EU AI Act. *AMCIS 2025 Proceedings*. https://aisel.aisnet.org/amcis2025/sig_sec/sig_sec/17

Shetty, D. K., Vijaya Arjunan, R., Cenitta, D., Makkithaya, K., Hegde, N. V., Bhatta B, S. R., Salu, S., Aishwarya, T. R., Bhat, P., & Pullela, P. K. (2025). Analyzing AI regulation through literature and current trends. *Journal of Open Innovation: Technology, Market, and Complexity*, *11*(1), 100508. https://doi.org/10.1016/j.joitmc.2025.100508

Simkute, A., Tankelevitch, L., Kewenig, V., Scott, A. E., Sellen, A., & Rintel, S. (2025). Ironies of Generative AI: Understanding and Mitigating Productivity Loss in Human-AI Interaction. *International Journal of Human–Computer Interaction*, *41*(5), 2898–2919. https://doi.org/10.1080/10447318.2024.2405782

Smuha, N. A. (2021). From a ‘race to AI’ to a ‘race to AI regulation’: Regulatory competition for artificial intelligence. *Law, Innovation and Technology*, *13*(1), Article 1. https://doi.org/10.1080/17579961.2021.1898300

Song, F., Agarwal, A., & Wen, W. (2023). *The Impact of Generative AI on Collaborative Open-Source Software Development: Evidence from GitHub Copilot* (SSRN Scholarly Paper No. 4856935). Social Science Research Network. https://doi.org/10.2139/ssrn.4856935

Stein, E. W., & Zwass, V. (1995). Actualizing Organizational Memory with Information Systems. *Information Systems Research*, *6*(2), 85–117. https://doi.org/10.1287/isre.6.2.85

Storey, V. C. (2025). Knowledge Management in a World of Generative AI: Impact and Implications. *ACM Trans. Manage. Inf. Syst.*, *16*(3), 26:1-26:14. https://doi.org/10.1145/3719209

Sturm, T., Gerlach, J., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., & Buxmann, P. (2021). Coordinating Human and Machine Learning for Effective Organization Learning. *Management Information Systems Quarterly*, *45*(3), 1581–1602.

Subramonyam, H., Pea, R., Pondoc, C., Agrawala, M., & Seifert, C. (2024). Bridging the Gulf of Envisioning: Cognitive Challenges in Prompt Based Interactions with LLMs. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–19. https://doi.org/10.1145/3613904.3642754

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, *40*(2), 137–157. https://doi.org/10.1080/14494035.2021.1928377

Taeihagh, A., Ramesh, M., & Howlett, M. (2021). Assessing the regulatory challenges of emerging disruptive technologies. Regulation & Governance, 15(4), Article 4. https://doi.org/10.1111/rego.12392

Taylor, L. (2021). Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector. Philosophy & Technology, 34(4), 897–922.

https://doi.org/10.1007/s13347-020-00441-4

Thierer, A. D. (2023). *Flexible, Pro-Innovation Governance Strategies for Artificial Intelligence* (SSRN Scholarly Paper No. 4423897). Social Science Research Network. https://papers.ssrn.com/abstract=4423897

Varone, A., Bolici, F., & Crowston, K. (2025). From Execution to Orchestration: Rethinking GenAI Implementation through Information Processing Theory. *ITAIS 2025 Proceedings*. XXII Conference of the Italian Chapter of AIS.

Véliz, C. (2022). Governing Privacy. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., pp. 149–163). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.9

Wachter, R. M., & Brynjolfsson, E. (2024). Will Generative Artificial Intelligence Deliver on Its Promise in Health Care? *JAMA*, *331*(1), 65–69. https://doi.org/10.1001/jama.2023.25054

Walsh, J. P., & Ungson, G. R. (1991). Organizational Memory. *Academy of Management Review*, *16*(1), 57–91.

Walter, Y. (2024). Managing the race to the moon: Global policy and governance in Artificial Intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences. *Discover Artificial Intelligence*, *4*(1), 14. https://doi.org/10.1007/s44163-024-00109-4

Wang, J., & Legner, C. (2025, December 14). Uncovering Untapped Organizational Knowledge in Unstructured Data: GenAI and the Reconfiguration of Data Management. ICIS 2025 Proceedings. https://aisel.aisnet.org/icis2025/general_topic/general_topic/6

Wang, W., Gao, G. (Gordon), & Agarwal, R. (2023). Friend or Foe? Teaming Between Artificial Intelligence and Workers with Variation in Experience. *Management Science*, mnsc.2021.00588. https://doi.org/10.1287/mnsc.2021.00588

Weber, T., Brandmaier, M., Schmidt, A., & Mayer, S. (2024). Significant Productivity Gains through Programming with Large Language Models. *Unknown*. https://api.semanticscholar.org/CorpusId:270554938

Wedel, K. (2025). *Contextual Memory Intelligence—A Foundational Paradigm for Human-AI Collaboration and Reflective Generative AI Systems* (No. arXiv:2506.05370). arXiv. https://doi.org/10.48550/arXiv.2506.05370

Wei, X., Wang, L., Lee, L.-K., & Liu, R. (2025). The effects of generative AI on collaborative problem-solving and team creativity performance in digital story creation: An experimental study. *International Journal of Educational Technology in Higher Education*, *22*(1), 23. https://doi.org/10.1186/s41239-025-00526-0

Wu, S., Liu, Y., Ruan, M., Chen, S., & Xie, X.-Y. (2025). Human-generative AI collaboration enhances task performance but undermines human's intrinsic motivation. *Scientific Reports*, *15*, 15105. https://doi.org/10.1038/s41598-025-98385-2

Yankouskaya, A., Liebherr, M., & Ali, R. (2025). Can ChatGPT Be Addictive? A Call to Examine the Shift from Support to Dependence in AI Conversational Large Language Models. *Human-Centric Intelligent Systems*, *5*(1), 77–89. https://doi.org/10.1007/s44230-025-00090-w

Yordanova, K., & Bertels, N. (2024). Regulating AI: Challenges and the Way Forward Through Regulatory Sandboxes. In H. Sousa Antunes, P. M. Freitas, A. L. Oliveira, C. Martins Pereira, E. Vaz De Sequeira, & L. Barreto Xavier (Eds.), *Multidisciplinary Perspectives on Artificial Intelligence and the Law* (Vol. 58, pp. 441–456). Springer International Publishing. https://doi.org/10.1007/978-3-031-41264-6_23

Zhao, L. (2023). Artificial intelligence and law: Emerging divergent national regulatory approaches in a changing landscape of fast-evolving AI technologies. In *Research Handbook on Digital Trade* (pp. 369–399). Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/book/9781800884953/book-part-9781800884953-33.xml

CHAPTER 3: THEORETICAL BACKGROUND

3.1 INTRODUCTION

This chapter provides the theoretical foundations for the thesis. Building on Chapter 2, it elaborates on the claim that GenAI systematically amplifies certain consequences while relegating others to a low-visibility and weakly monitored space. The central argument is that this asymmetry cannot be reduced to generic short-termism or individual cognitive bias. Rather, it arises from how organizational attention is structured and evaluated: which consequences are legible within prevailing organizational schemas and evaluation metrics, which generate perceived performance–aspiration gaps, and how these evaluations recursively reshape rules, roles, channels, and artefacts.

To theorize this pattern, the chapter adopts the Attention-Based View (ABV) of the firm (Ocasio, 1997), grounded in the Carnegie tradition of bounded rationality (Simon, 1947; March & Simon, 1958; Cyert & March, 1963) and closely aligned with theories of organizational myopia (Levinthal & March, 1993; Catino, 2013). ABV treats attention as a scarce, structurally distributed resource and asks how organizations channel issues, answers, and decision-makers into concrete procedural and communication channels. It explains not only that attention is limited, but also which consequences are routinely seen, encoded, and acted upon, and which remain latent. This makes ABV particularly suited to analyze the GenAI configuration identified in Chapter 2, where visible success and latent erosion coexist across governance and work settings under similar attentional constraints.

Several other theoretical families are natural candidates for anchoring the thesis. Technology adoption and acceptance models (TAM, UTAUT, and diffusion of innovations) explain whether, when, and by whom GenAI is adopted, as well as how beliefs, social influence, and contextual conditions drive usage. Resource-based and dynamic capabilities perspectives provide a vocabulary for thinking about data assets, AI skills, and organizational memory as strategic resources and capabilities whose erosion matters. Information-processing views conceptualize organizations as systems for matching information-processing capacity with environmental requirements, offering useful language for thinking about GenAI as a capacity-enhancing technology. Sociomaterial and sociotechnical approaches demonstrate how GenAI is enacted in situated work and governance practices, highlighting the mutual constitution of technologies and routines.

The chapter draws on these traditions, but argues that none, on their own, can explain the specific pattern that matters here. Adoption models focus on use and intentions, rather than the post-adoption allocation of attention. RBV and dynamic capabilities presuppose that capability erosion can be sensed and interpreted. Information-processing views foreground capacity rather than the structured selectivity of attention. Sociomaterial and sociotechnical theories provide rich descriptions of local enactments but are deliberately cautious about abstracting structural attentional mechanisms across different contexts.

ABV provides the missing layer. It specifies how environmental cues are filtered through institutional repertoires, how issues and answers are embodied in artefacts and metrics, how channels and participation patterns shape what is available and legitimate to attend to, and how organizational moves recursively feed back into attention structures. Building on this perspective, the chapter develops an ABV-based account of the *Visibility Trap:* a self-reinforcing process, shaped by established attention structures, in which the salience of visible successes progressively degrades the capacity to identify and respond to associated but latent negative consequences.

The remainder of the chapter is organized as follows. Section 3.2 reviews the main alternative theoretical lenses, engages more closely with key contributions, and motivates the rationale for adopting ABV as the primary theoretical lens. Section 3.3 returns to Carnegie foundations and shows how bounded rationality, uncertainty absorption, problemistic search, and aspiration-level adaptation already contain the seeds of visibility/latency asymmetries. Section 3.4 reviews ABV's principles and mechanisms, tracing its recent extensions and treating organizations as distributed attention systems. Section 3.5 integrates these elements into an explicit ABV-based framework of the Visibility Trap (VT1–VT4), organized around legibility and strength, and indicates how this mechanism will structure the analyses of GenAI governance (Chapter 4) and GenAI-mediated knowledge work (Chapter 5).

3.2 THEORETICAL LENS AND RATIONALE

A number of theoretical frameworks could, in principle, anchor this thesis. Given the domains involved, four families of theories are discussed as alternative or complementary lenses: technology adoption models, resource-based and dynamic capabilities perspectives, information-processing views, and sociomaterial/sociotechnical approaches. Following this, Section 3.2.5 explains why the ABV, situated in the broader context of organizational myopia, represents the most suitable theoretical lens for understanding the Visibility Trap.

3.2.1 Technology Adoption and Acceptance Models

Technology adoption and acceptance models are prominent analytical tools in the Information Systems literature. Davis's (1989) Technology Acceptance Model (TAM) formalizes perceived usefulness and perceived ease of use as the main antecedents of user attitudes and behavioral intention, which in turn predict usage. Subsequent work has refined and extended this structure. The more recent TAM3 (Venkatesh & Bala, 2008) elaborates a detailed nomological network for the determinants of perceived usefulness and ease of use, explicitly positioning itself as a guide for managerial interventions to increase acceptance and effective utilization.

Sharing the same theoretical roots, the UTAUT integrates eight prior models to show that performance expectancy, effort expectancy, social influence, and facilitating conditions, moderated by age, gender, experience, and voluntariness, can explain a significant share of variance in usage intentions (Venkatesh et al., 2003).

At a more macro level, Rogers' (2003) Diffusion of Innovations theory emphasizes the attributes of innovations (relative advantage, compatibility, complexity, trialability, and observability) and the role of communication channels and social networks in shaping adoption over time. Observability and trialability are particularly salient for AI, because visible results and low-cost experimentation can accelerate diffusion.

For this thesis, these models are useful in several ways. They can explain why knowledge workers in Chapter 5 start using GenAI assistants (e.g., high perceived usefulness, strong social influence); why some groups resist (e.g., low perceived usefulness, perceived risk); and how social dimensions shape adoption and usage trajectories.

However, these models are fundamentally ill-suited to the central problem of the Visibility Trap. First, they are structurally framed around initial adoption and use. Even the more elaborate variants (TAM3, UTAUT) focus on explaining and manipulating beliefs and usage intentions, not on what happens to organizational attention once a technology is embedded. The key outcome variables remain use, intention to use, or extent of diffusion; long-run shifts in what is attended to and how success is defined are largely outside the frame. Second, these models treat the organization primarily as a context, rather than as a distributed attention system. Third, the underlying success criterion in this literature is often implicit: more adoption is better, or at least higher usage is the thing to be explained and enabled. Even DoI's treatment of observability treats visibility as a positive driver of diffusion, not as a potential source of bias: highly observable benefits accelerate spread, whereas less observable harms receive little theoretical attention. None of these models theorize a trajectory in which increased use produces larger visible benefits and simultaneous, unregistered losses in capabilities or governance quality.

For the *Visibility Trap*, the central question is not why actors adopt GenAI, but why organizations, after adoption, systematically overlook some consequences and not others, and why that bias persists over time.

3.2.2 Resource-Based and Dynamic Capabilities Perspectives

Resource-based and dynamic capabilities theories dominate strategic management and are widely used to analyse digital technologies. Wernerfelt's (1984) resource-based view (RBV) and Barney's (1991) formalization of valuable, rare, inimitable, and non-substitutable (VRIN) resources shift the focus from products and markets to internal resource portfolios as sources of sustained competitive advantage. Teece, Pisano, & Shuen's (1997) dynamic capabilities framework adds a temporal and processual layer: firms must develop capabilities to sense, seize, and transform in order to adapt to rapidly changing environments.

In a GenAI context, these perspectives provide a natural way to conceptualize data assets, AI models, and GenAI-related skills as strategic resources and to treat skill acquisition, learning, and organizational memory as dynamic capabilities. They resonate with Chapter 5's concern about skill development trajectories: deep expertise, tacit knowledge, and the structure and quality of organizational memory can be viewed as critical resources whose erosion undermines long-term advantage. This thesis draws on those ideas. Yet, for the purposes of explaining the Visibility Trap, these perspectives are incomplete in two ways.

First, RBV largely treats resources and capabilities as given objects that can be identified and assessed. The mechanisms through which some resources become salient, routinely tracked, and defended while others remain backgrounded are not specified. Second, dynamic capabilities theory assumes a sensing and interpreting capacity that Chapters 4 and 5 precisely call into question. Sensing presupposes that relevant signals are attended to, that they can be framed as opportunities or threats, and that they enter strategic processes. In practice, however, organizations may not "sense" the gradual erosion of deep expertise or the slow reshaping of organizational memory, because these do not appear on central dashboards. Teece et al. (1997) discuss managerial processes and learning paths, but do not model how attention is selectively allocated across competing signals or how measurement asymmetries bias sensing toward short-term, visible outcomes.

For the *Visibility Trap,* the problem is not only that GenAI might erode certain capabilities, but that organizations do not systematically see this erosion while visible performance gains accumulate. RBV and dynamic capabilities help articulate the content and strategic significance of what is eroding; they under-specify the attentional mechanisms through which that erosion is, or is not, noticed and acted upon. They therefore provide an important backdrop but cannot, on their own, account for a pattern in which visible GenAI successes coexist with unattended capability and memory loss.

3.2.3 Information-Processing View

Galbraith's (1974) information-processing view conceptualizes organizations as mechanisms for coping with environmental uncertainty by designing structures that increase information-processing capacity. The central design problem becomes achieving a fit between the volume and complexity of information generated by the environment and the organization's capacity to process that information through rules, hierarchies, and lateral relations.

This perspective offers a natural way to think about GenAI as an information-processing technology: it increases the volume and speed at which data can be handled, analysed, and circulated, and thus helps explain why organizations are attracted to GenAI as a means of coping with uncertainty at lower marginal cost. However, in its classic formulation, the information-processing view treats information as a scarce resource and emphasises capacity-enhancing structural arrangements. It does not capture attention, and its structured allocation across issues and answers is not the central object of analysis. Questions about which signals are picked up, which are ignored, and how some consequences become systematically more visible than others remain largely implicit.

3.2.4 Sociomaterial and Sociotechnical Perspectives

Orlikowski (2007) and Orlikowski & Scott (2008) argue that everyday organizing is inherently sociomaterial: technologies and social practices are constitutively entangled, such that organizational outcomes emerge from shifting sociomaterial assemblages rather than from "social" variables acting on "technical" artefacts. Leonardi (2011) develops related ideas through the notion of imbrication of human and material agencies, showing how perceptions of affordance and constraint lead actors either to change routines or to modify technologies. These approaches are particularly powerful for examining how GenAI is enacted locally: how workers negotiate responsibility for AI outputs, how prompts and queries are developed as workarounds, how GenAI reshapes identity and accountability, and how sociomaterial configurations stabilize or shift in daily practice.

Sociotechnical systems theory adds a longer historical arc. Trist & Bamforth's (1951) classic study of the longwall method in coal mining shows how new technologies can degrade social relations and psychological well-being when social and technical systems are not jointly optimized. More recent reflections (Pasmore et al., 2019) argue that sociotechnical thinking remains crucial in the digital era, as new technologies outpace the ability of organizational design to keep pace. These contributions highlight the need to design work systems that consider both technical possibilities and human consequences.

This thesis draws on these traditions by aligning with the claim that GenAI cannot be treated as an exogenous tool simply adopted within pre-existing structures. Instead, it reshapes workflows, identities, and accountability arrangements. As primary lenses for the thesis, however, they present limitations. Sociomaterial approaches are deliberately cautious about structural abstraction. Orlikowski & Scott (2008) explicitly call for close attention to emergent sociomaterial configurations and resist the temptation to make robust generalizations across contexts. This is a strength for detailed case analysis, but a weakness when the problem is to explain why similar visibility/latency patterns appear in very different settings (public authorities vs. firms; governance vs. work) and to articulate a compact mechanism that applies across them.

3.2.5 From Organizational Myopia to the ABV

Literature on organizational myopia examines the systematic inability of organizations to foresee the consequences of their decisions or recognize signs of impending danger (Catino, 2013). The foundational work of Levinthal & March (1993) identifies three forms of myopia. First, temporal myopia is the tendency to prioritize the short term over the long term. Second, spatial myopia refers to the tendency to overlook the larger system in favor of the local neighbourhood. Third, failure myopia captures the tendency to oversample success and overlook failure, implying that persistent success leads organizations to underestimate risks, assuming that future safety is guaranteed by past performance (Levinthal & March, 1993).

A central tenet of this literature is that failures are preceded by an incubation period (Turner & Pidgeon, 1978; Catino, 2013). During this period, latent factors such as poor design, bad management decisions, or communication failures accumulate silently (Pasman, 2021). Detection during incubation is hindered by organizational filters. Pasman (2021) notes that warning signals are often noticed by operators but filtered out by middle management to avoid disrupting operations. Catino (2013) describes the decoy problem, where attention to well-defined problems distracts from ill-defined, ambiguous threats because focusing on structured, visible issues inherently obscures others: "a way of seeing is always also a way of not seeing". Furthermore, the normalization of deviance (Vaughan, 2007; Catino, 2013) enables organizations to reclassify warning signs as acceptable risks. Repeated successes in the presence of defects (e.g., a shuttle returning safely despite damage) reinforce the belief that the system is safe, gradually transforming anomalies into routine events.

While the literature on organizational myopia provides the essential diagnostic framework for understanding how organizations incubate latent failure, this thesis proposes two attention-based adaptations that re-ground the problem of myopia from surveillance failure and competence rigidity into questions of attention economisation and allocation.

First, classic myopia accounts emphasize that organizations intend to detect emerging defects but fail because incubation is silent and signals are weak (Catino, 2013; Pasman, 2021). In the Visibility Trap, the emphasis shifts to attentional economisation. The organization does not merely miss latent consequences; it reallocates attention away from them because they do not generate performance-aspiration gaps. Thus, potential problems can remain cognitively recognized by some actors, but they are classified as low-strength and relegated to the background relative to dimensions on which visible indicators perform well. This differs from the normalization of deviance, where anomalies are explicitly tolerated as acceptable risk. Here, the very perceived need for anomaly detection is downgraded by the abundance of apparently satisfactory signals.

Second, the so-called *Success Trap* explains how organizations become locked into familiar routines because exploitation yields faster, more certain returns than exploration (Levinthal & March, 1993; Sato, 2012). Competence rigidity follows from repeated investments in a narrow portfolio of skills and procedures. The Visibility Trap builds on this diagnosis, relocating the binding force from competence portfolios to attention structures (Ocasio, 1997). What becomes rigid are not only routines but also the evaluative architectures (metrics, channels, categories, and roles) through which performance is defined and monitored. Under GenAI, these attention structures are supplied with *synthetic validity*. In GenAI governance (Chapter 4), highly visible sanctions and compliance artefacts that count as success on existing indicators while masking erosion in substantive legitimacy and governance capacity. In knowledge work (Chapter 5), GenAI co-produced outputs satisfy internal quality markers that signal expertise, even when underlying skills gradually decay. In both cases, attention is bound to inherited success criteria, and repeated "wins" on those criteria make it progressively harder to raise or legitimate concerns about latent losses.

The theoretical contribution of this chapter is to recast organizational myopia in explicit ABV terms, specifying how legibility and strength yield a form of attentional economisation that is not captured by existing "weak-signal" accounts. Building on this reframing, the thesis operationalizes the *Visibility Trap* in two domains: GenAI governance (Chapter 4) and GenAI-mediated knowledge work and organizational memory (Chapter 5), demonstrating how these mechanisms emerge and reinforce established attention structures at the expense of latent consequences.

3.3 FROM BOUNDED RATIONALITY TO THE ABV

The ABV integrates a line of argument developed over 40 years by the Carnegie School. This section revisits three foundational works – Simon's Administrative Behavior (1947), March & Simon's Organizations (1958), and Cyert & March's A Behavioral Theory of the Firm (1963) – to show how they provide the microfoundations for ABV and to the conceptualization of the Visibility Trap.

3.3.1 Simon (1947): Bounded Rationality & Attention as an Organizational Problem

Simon's (1947) starting point is a direct rejection of the assumption of objective rationality. A fully rational actor would survey a complete set of alternatives, evaluate the full consequence space of each, and choose the optimal option given a stable preference ordering. Simon argues that real decision-makers can do none of these things. Three limitations are central.

First, knowledge of consequences is fragmentary. Even in relatively simple decisions, causal chains branch rapidly, and human cognition cannot follow all branches in depth. In practice, individuals work with truncated situation models built from salient cues and short causal horizons; large portions of the consequence space are never represented at all (Simon, 1947). Second, valuation is itself bounded. Actors cannot "hold" the entire pattern of future outcomes in mind simultaneously. The attractiveness of an option depends heavily on the slice of consequences that current attention highlights; a course of action framed in terms of short-term efficiency can look compelling even if it carries poorly attended long-run risks. Third, alternative generation is constrained: "only a very few of all these possible alternatives ever come to mind" (Simon, 1947). What is not brought into focal awareness is effectively outside the choice set.

For Simon, these limits make attention the central currency of decision-making. Attention is inherently selective: to attend to some objects is to ignore others. As a result, he redefines the role of organizations. Organizations exist, in large part, to structure the "psychological environment" so that the limited attention of individuals is steered toward premises, cues, and alternatives that are consistent with organizational objectives (Simon, 1947). He identifies several mechanisms through which this occurs.

The division of work narrows individual concern to a tractable subset of issues. Standard practices (routines) remove many decisions from active deliberation by providing pre-packaged responses. Authority systems substitute hierarchical directives for individual evaluation, thereby channeling attention toward execution of given orders. Communication channels determine which stimuli reach which people, and in what form. Training and indoctrination inject organizational values and categories into members' cognitive systems, shaping what they notice and how they interpret it. Together, these mechanisms create what ABV will later call attention structures: enduring configurations of roles, rules, and communication patterns that constrain what is even available for thought and choice.

3.3.2 March & Simon (1958): Uncertainty Absorption, Routines, and Subgoal Fixation

March & Simon (1958) extend Simon's (1947) argument by showing how bounded rationality manifests in concrete organizational mechanisms of change, information processing, and goal fixation. Three elements are particularly relevant to this thesis: (1) search behavior, (2) uncertainty absorption and routine formation, and (3) subgoal fixation.

March & Simon's (1958) theory of search begins with an asymmetry between persistence and change: continuing a current program avoids the "sunk costs of innovation" required to develop new alternatives. Consequently, organizations do not systematically scan for the optimal option; instead, they satisfice, engaging in search primarily when changes in the environment make the existing organizational procedures unsatisfactory.

While the specific proximity rules (searching "near the problem symptom") were formalized in Cyert & March's (1963), March & Simon (1958) describe a similar local bias. They argue that search is sequential and initially focuses on variables that are largely within the problem-solver's control. Adaptation is typically incremental, involving local changes in programs or the recombination of existing routine programs rather than total invention. Furthermore, highly programmed activity tends to take precedence over unprogrammed innovation, a phenomenon termed the "Gresham's Law" of planning: daily routine drives out planning.

Uncertainty absorption describes how organizations transform ambiguous, complex raw data into "facts" that can guide action. Inference, not evidence, travels through the system: questionnaires become aggregated scores; nuanced observations become compressed indicators; tentative interpretations harden into official statements (March & Simon, 1958). Classification schemes (e.g., categories, coding manuals, reporting formats) govern this process. Over time, these schemes are reified: they cease to be seen as conventions and are treated as properties of reality. Specialized positions become gatekeepers of this absorption process: they choose what to observe, how to summarize it, and what to transmit. Once an interpretation becomes an "official" fact, it is rarely checked against original evidence, because others lack the time, access, or competence to do so.

Organizational learning, in this framework, primarily involves building and modifying a repertoire of routines. March & Simon (1958) describe a hierarchy: at the lowest level, organizations execute stored programs; at an intermediate level, rules select which program to use; at the highest level, meta-procedures govern the revision of programs and selection rules. Most of the time, adaptation occurs through program selection within the existing repertoire, not program modification. Because revising routines is cognitively and politically expensive, it tends to occur only when existing routines visibly fail by current standards. This structure generates path dependence, a phenomenon later referred to in the literature as competency traps (Levinthal & March, 1993): organizations become increasingly skilled at existing routines, making them appear superior to untried alternatives even when the environment has changed.

Subgoal fixation explains how the necessary decomposition of complex goals into operational subgoals can distort organizational attention. Since no unit can attend to all dimensions of performance, higher-level objectives must be factored into specialized subgoals (cost efficiency, response time, error rates, etc.). Over time, these subgoals can displace the broader aims they were meant to serve: the metrics become ends in themselves (March & Simon, 1958). Three mechanisms reinforce this: selective perception (individuals filter information to confirm the importance of their subgoal), in-group communication (most information comes from colleagues who share the same focus), and structural differentiation (units are exposed to different external stimuli). Together, these create a self-reinforcing attentional narrowing: local performance on subgoals becomes the dominant lens, and side effects on higher-order goals are either not seen or are interpreted as someone else's problem.

3.3.3 Cyert & March (1963): Coalitions, Problemistic Search, & Aspiration-Level Learning

Cyert & March (1963) complete the Carnegie picture by adding an explicit political and temporal dimension. Organizations are coalitions of participants (e.g., managers, workers, shareholders, regulators, customers) whose preferences diverge and whose demands must be reconciled. Goals are not derived from a unified preference ordering; instead, they are negotiated and revised over time. Within this coalition, attention is a scarce political resource: different groups compete to have their issues recognized as organizational problems.

The concept of quasi-resolution of conflict explains how organizations function despite unresolved disagreements over goals. Instead of harmonizing all goals, organizations attend to different goals at different times. They allocate attention sequentially: responding to sales demands in one period, production demands in another, and regulatory pressures in a third. This sequential attention is embedded in organizational channels: budget meetings, production reviews, and compliance committees, each addresses different goal subsets. Inconsistencies across these channels often remain invisible because no forum simultaneously engages all goals (Cyert & March, 1963).

Uncertainty avoidance complements uncertainty absorption by describing how organizations structure their environment and time horizons to minimize the need for complex prediction. Rather than anticipating long-run consequences, organizations rely on short-run feedback: they solve problems as they arise in a “fire-fighting” mode, waiting for performance indicators to signal trouble rather than investing heavily in forward-looking analysis (Cyert & March, 1963). Externally, they negotiate stable environments through plans, contracts, and industry conventions, so they can focus attention on executing within agreed boundaries instead of constantly worrying about strategic surprises. Internally, budgets and plans function as negotiated contracts among subunits, reducing the uncertainty they must attend to in each local context.

Building on March & Simon (1958), Cyert & March (1963) describe problemistic search as the behavioral mechanism that links goals, feedback, and adaptation. As described in the previous section, a problem is defined as performance falling below an aspiration level on a given goal dimension. Search is initiated when such a failure is detected and continues until a satisfactory solution is found. It is specific (directed at the problematic dimension) and local: organizations first explore solutions close to the site of the symptom and close to current practices (Cyert & March, 1963). Search is also systematically biased: different units see different aspects of the problem depending on their training and interests; hopes and fears shape what evidence is taken seriously; and communication biases, grounded in power and conflict, influence which interpretations reach decision-makers. When local search fails, attention often shifts to "vulnerable" units (i.e., those with weak power positions or ambiguous contributions) rather than to the true causal locus.

Organizational learning, for Cyert & March, centres on aspiration-level adaptation and the reinforcement of search rules. Goals are treated as aspiration levels that adjust over time based on experience: when performance exceeds aspirations, aspirations rise; when performance falls short, aspirations may be revised downward (Cyert & March, 1963). This creates a moving target that determines when situations are classified as problems requiring attention. At the same time, organizations learn where to search: search procedures that yield satisfactory solutions are reinforced and embedded in standard operating procedures; unsuccessful ones are weakened. Over time, this produces path dependence in attention: organizations repeatedly attend to familiar sources of solutions and remain relatively blind to alternatives that did not feature in past successful search episodes. Vicarious learning extends this process across organizations: firms adjust their aspirations and search patterns by observing peers, contributing to field-level convergence.

3.3.4 Other Foundations of ABV

While the three previous seminal works represent the most central ABV micro-foundations, its formulation is also influenced by four other theoretical traditions.

First: theories of enactment. Building on Weick (1979) and Daft & Weick (1984), the ABV posits that organizations do not merely perceive an objective environment but actively enact it. Weick argued that managers construct reality by singling out specific cues and imposing order upon them. Ocasio incorporates this directly into Mechanism 5b (enactment of issues and answers), but extends it structurally. While Weick focused on retrospective meaning-making, ABV adds that attention structures determine which environmental features are available for enactment in the first place. The environment is enacted, but the raw materials for that enactment are filtered through the firm's specific configuration of channels and rules.

Second, the Garbage Can Model. Cohen, March, & Olsen's (1972) "garbage can model" characterizes decision-making as the confluence of four independent streams: problems, solutions, participants, and choice opportunities. Decisions occur when these streams meet, often through temporal coincidence rather than rational planning. ABV accepts the fluidity of these streams but critiques the model's reliance on randomness. Ocasio replaces the "random" garbage can with structural regulation: attention structures (rules, resources, positions) govern the flow. They determine which problems and solutions are allowed into which channels (Mechanism 4b) and which participants are present (Mechanism 4c). Thus, ABV retains the insight of loose coupling while explaining the stable patterns of organizational choice.

Third, a process-oriented view of strategy. The ABV draws on process-oriented strategy research (Bower, 1970; Burgelman, 1983), which demonstrated that strategy is not just top-down planning but an emergent phenomenon shaped by multi-level resource allocation. Bower showed that the "structural context" shapes how middle managers filter information for the top. Ocasio translates this directly into the attentional mechanism of channeling (Mechanism 4b). The administrative mechanisms identified by Bower (i.e., measurement systems and reporting lines) are reconceptualized in ABV as the concrete channels that regulate the availability and salience of issues. This connects the cognitive aspect of attention to the political reality of how initiatives survive or die within the corporate hierarchy.

Fourth, neo-institutionalism. Lastly, ABV relies on neo-institutionalism (DiMaggio & Powell, 1983; Scott, 1995) to conceptualize the environment of decision (Mechanism 1). Institutional theory posits that the environment supplies the cognitive schemas, categories, and cultural "tool kits" (Swidler, 1986) that organizations use. Ocasio uses this to explain the origin of the repertoire of issues and answers available to decision-makers. However, ABV creates a crucial distinction: while institutional theory explains what is available in the broader field, ABV explains what is actually attended to by the firm. Attention structures act as the filter that selects specific items from the institutionally available repertoire.

Finally, the specific literature on learning myopia (Levinthal & March, 1993) provides the diagnostic backdrop for ABV. This tradition explicitly frames bounded rationality as a pathology of foresight, specifically the tendency to prioritize short-term exploitation over long-term exploration. ABV internalizes this logic by treating myopia as a structural inevitability. Where myopia scholars identify the tendency to ignore latent consequences, Ocasio (1997) specifies the mechanisms that institutionalize this phenomenon. This lineage establishes that the Visibility Trap is rooted in well-documented patterns of learning failure, but requires the structural granularity of ABV to explain how legible signals actively displace latent warnings.

Altogether, these foundations complement the Carnegie view by adding the structural, political, and environmental mechanisms that channel bounded rationality. They explain how the environment is enacted, how decision streams are regulated, and how external repertoires are internalized; concepts that the next section organizes into the specific principles of the ABV.

3.4 ABV: PRINCIPLES, MECHANISMS & EVOLUTION

The ABV turns the insights discussed in the previous section into a structured model of how organizations allocate attention (Ocasio, 1997). Its core claim is that what organizations do depends on what their decision-makers attend to, and what they attend to is systematically shaped by organizational structures. In the context of this thesis, ABV provides the backbone for the Visibility Trap, explaining how GenAI makes certain outcomes (e.g., productivity, compliance) hyper-visible while others (e.g., skill decay, legitimacy erosion) substantially unattended. This section details the ABV's principles, mechanisms and connection to the Visibility Trap.

3.4.1 The Three Principles of the Attention-Based View

Ocasio (1997) states three core principles that turn individual attention limits into an organizational theory.

3.4.1.1 Principle 1: Focus of Attention

"*What decision-makers do depends on what issues and answers they focus their attention on*" (Ocasio, 1997, p. 188). Issues are the categories used to make sense of the environment, including problems, opportunities, and threats. Answers are the action repertoires: proposals, routines, projects, procedures. Attention is finite: at any moment, only a limited subset of issues and answers can be consciously processed.

3.4.1.2 Principle 2: Situated Attention

"*What issues and answers decision-makers focus on, and what they do, depends on the particular context or situation they find themselves in"* (Ocasio, 1997, p. 188). Attention is situated in concrete procedural and communication channels such as meetings, reporting routines, platforms, review cycles. Spatial (where), temporal (when, for how long), and procedural (agenda, rules, norms) characteristics of these channels shape what becomes available and salient.

3.4.1.3 Principle 3: Structural Distribution of Attention

"*What particular context or situation decision-makers find themselves in, and how they attend to it, depends on how the firm's rules, resources, and social relationships regulate and control the distribution and allocation of issues, answers, and decision-makers into specific activities, communications, and procedures*" (Ocasio, 1997, p. 188).

Here ABV moves from the micro-situation to the broader attention structures that determine which issues and answers get placed on which agendas, which decision-makers are invited into which channels, and which channels are endowed with resources and authority. With the previous two principles, ABV provides the conceptual lens for the Visibility Trap: attention is limited (Principle 1), contextually triggered (Principle 2), and structurally patterned (Principle 3).

3.4.2 The Six Mechanisms of the Attention-Based View

ABV links environment, structures, and decision-making through six mechanisms (Ocasio, 1997). Mechanisms 1 and 2 specify how the environment of decision is constructed and made durable.

3.4.2.1 Mechanism 1: Environment of Decision

The environment of decision comprises "*the multiple material, social, and cultural factors, both internal and external to the firm, that impinge upon any decision activity*" (Ocasio, 1997, p. 193). In this domain, ABV makes three central claims.

1. Environmental stimuli: In any channel, physical, economic, technological, and regulatory factors impinge as stimuli. For GenAI, this includes vendor pitches, media narratives, regulatory developments, internal incidents, budget pressures, and peer moves. When a management team meets to discuss GenAI, these stimuli are filtered through prior monitoring choices and reporting routines.
2. Cultural and institutional repertoires: Cultural and institutional processes provide repertoires of issues and answers that actors can draw on: familiar labels (e.g., "innovation" and "AI risk") and familiar responses (e.g., "technology adoption", "legal intervention"). These repertoires make some interpretations of GenAI natural and others awkward or hard to articulate.
3. Embeddedness of structures: Rules, roles, and resources inside organizations are themselves shaped by their broader economic and professional environment. For

example, sectors with strong performance benchmarking will tend to have more mature infrastructures for tracking GenAI productivity metrics than for tracking deep skill trajectories.

3.4.2.2 Mechanism 2: Embodiment of Issues and Answers

"*Issues and answers are embodied in the cultural products and artifacts used to construct the firm's activities and communications*" (Ocasio, 1997, p. 194). Embodiment matters because materialized elements are easier to access and reuse, carry authority (they look official), and bias attention toward exploiting what already exists rather than imagining what is missing.

3.4.2.3 Mechanism 3: Availability and Salience

Mechanism 3 specifies how procedural and communication channels shape what is cognitively available and salient. Channels are "*the formal and informal concrete activities, interactions, and communications set up by the firm to induce organizational decision-makers to action on a selected set of issues*" (Ocasio, 1997, p. 194). Three dimensions matter:

1. Spatial: where the channel is located, what material and digital resources are present.
2. Temporal: how often and how long the channel convenes, what deadlines it faces, how it is sequenced relative to other activities.
3. Procedural: agendas, formats, norms, and reporting templates.

3.4.2.4 Mechanism 4: Attention Structures

Mechanism 4 formalizes attention structures as the social, economic, and cultural arrangements that govern how time, effort, and attention are allocated across issues and answers. Ocasio (1997) distinguishes four elements (i.e., rules of the game, players, structural positions, and resources) and three ways in which they shape attention: valuation, channeling, and the structuring of interests and identities. Taken together, these mechanisms explain why some issues routinely appear as legitimate and urgent topics in organizational life, while others remain peripheral or unarticulated – even when actors privately worry about them. They are central to how the Visibility Trap becomes a structural rather than purely cognitive phenomenon.

Mechanism 4a: Valuation of issues and answers

Attention structures "*govern the valuation and legitimization of the repertoire of issues and answers available to decision-makers. These values are not uniform throughout the firm but are differentiated according to the division of labor inherent in the firm's rules, positions, players, and resources*" (Ocasio, 1997, p. 199).

Rules of the game (formal strategies, norms, informal success criteria) embed organizational identity and purpose and thus define which issues "count" and what outcomes qualify as success. They provide the background logic against which an issue appears as strategic, trivial, or even inappropriate to raise.

Structural positions differentiate values across the organization. A finance function, a regulatory unit, and an R&D team do not simply see different things; they also rank issues and answers differently, because their roles tie them to distinct evaluative criteria.

Players (such as powerful executives, board members, or influential experts) "sell" certain issues and answers, championing them onto agendas and amplifying their perceived importance (Jackson & Dutton, 1988; Ocasio, 1997).

Resources bias valuation toward solutions embedded in existing capabilities and assets. Answers that leverage current technologies, routines, and expertise are easier to see as feasible and worthwhile than answers that would sideline or devalue those investments.

Mechanism 4b: Channeling of decision-making

Valuation alone does not determine what receives attention; issues and answers must also have a designated location. Mechanism 4b captures how attention structures "*channel and distribute the decision-making activities of the firm into a set of procedural and communication channels*" (Ocasio, 1997, p. 199).

Rules, positions, power relations, and resource constraints jointly determine which issues are assigned to which forums (e.g., budgeting committees, risk boards, strategy offsites, case-handling meetings); which channels have regular rhythms, analytic support, and formal authority (e.g., quarterly performance reviews versus ad hoc working groups); and which concerns have no obvious channel, surfacing only informally, intermittently, or not at all.

Mechanism 4c: Structuring of interests and identities

Finally, attention structures "*provide decision-makers with a structured system of interests and identities to motivate their action and to structure their decision premises*"

(Ocasio, 1997, p. 200). This mechanism moves from where issues go to who cares about what, and why.

Rules of the game link valued outcomes to reputations, careers, and rewards, defining what it means to be a "good" manager, expert, or regulator. Players and their networks shape local interest systems, for example, by sponsoring certain initiatives, mentoring specific groups, or framing some topics as "career-making" and others as distractions. Structural positions create systematic variation in interests and identities: a line manager, a specialist lawyer, and a knowledge engineer attend to different risks and opportunities because their roles anchor them in different social and evaluative communities. Resources and human capital further tie people's interests to particular assets and routines: when one's expertise is deeply invested in certain tools or practices, there is a strong incentive to keep those central.

3.4.2.5 Mechanism 5: Decision Processes

Mechanism 5 connects attention structures to actual organizational moves.

Mechanism 5a: Structuring of participation

For Ocasio, decision-making is "*the product of interactions among participants in the firm's procedural and action channels*" (1997, p. 200). Who actually sits in a given meeting or handles a given case depends on formal roles, the time and energy those roles make available, and the competing pulls of other channels. Each participant brings a particular repertoire of issues and answers, shaped by their training, position, and identity. The resulting mix determines which experiences and interpretations are even present when a situation is discussed.

Mechanism 5b: Enactment of issues and answers

Within each channel, decision-makers "enact the environment of decisions by focusing their attention on a limited number of issues and answers" (Ocasio, 1997, p. 201). Enactment is the immediate cognitive step between structures and action: from all the stimuli and possibilities that are formally present, participants construct a working "version" of the situation that they treat as real enough to act on. This construction is shaped by the availability and salience of issues and answers in the channel (Mechanism 3), by how attention structures have already valued them and tied them to identities (Mechanism 4), and by the concrete mix of participants just described (Mechanism 5a).

Mechanism 5c: Selection of organizational moves

Mechanism 5c links attention to action: "*decision-makers will select among alternative organizational moves depending on which issues and answers they attend to*" (Ocasio, 1997, p. 202). Organizational moves encompass both external exchanges (i.e., deploying or restricting technologies, signing contracts, and issuing enforcement decisions) and internal adjustments to roles, procedures, or controls. Because different channels enact different environments, they can legitimately choose different moves even when facing the same broad situation.

3.4.2.6 Mechanism 6: Recursion

The final mechanism in ABV introduces recursion: "*organizational moves, once enacted, become part of the firm's environment of decision and are inputs to the construction of subsequent organizational moves*" (Ocasio, 1997, p. 202). Past moves do not simply disappear once implemented; they modify rules of the game, redistribute resources, create or redefine positions, and leave behind artefacts that populate future channels. In Carnegie terms, they feed into aspiration levels, search rules, and routine repertoires.

3.4.3 Evolution of the Attention-Based View

Since its inception, the ABV has expanded to integrate social, political, and communicative dimensions. This section traces that evolution through three phases.

3.4.3.1 Early Expansions: Scope and Depth (2001-2008)

Hoffman & Ocasio (2001) expanded the theory from the firm to the industry level, arguing that collective attention is a socially constructed process rather than a simple reaction to events. Integrating Weick (1979), they showed that for an event to become a critical issue, it must challenge the industry's identity and generate contestation through outsider accountability and insider image assessment. Ocasio & Joseph (2005) focused inward, identifying operational and governance channels as the primary architecture of attention. They argued that strategy emerges from the "guided evolution" and selective retention of initiatives within these channels, where tight and loose coupling explains both alignment and organizational silos.

Shifting the focus from quantity to quality, Weick & Sutcliffe (2006) argued that organizational effectiveness depends on the stability and vividness of attention. They suggested that failures often stem from a scattered mode of attending that misses unique particulars. Finally, Ocasio & Joseph (2008) utilized a longitudinal study of General Electric to demonstrate that attention structures endure over time. They showed that while the specific forms and focus adapt to each CEO's agenda, the underlying system for directing managerial attention remains continuous.

3.4.3.2 Theoretical Consolidation and Integration (2010-2020)

The next phase integrated ABV with adjacent theories. Ocasio & Wohlgezogen (2010) positioned attention as the mediating mechanism of control systems. They mapped five control types onto three cognitive processes: selection, vigilance, and regulation. Ocasio (2011) further refined the theory by disaggregating attention into attentional perspective (top-down structure), engagement (resource allocation), and selection (emergent outcome), providing a framework to organize disparate findings.

Critiquing the structural "pipes and prisms" view, Ocasio et al. (2018) argued that communication actively constitutes attention. They outlined how communicative practices, strategic vocabularies, and rhetorical tactics shape what organizations notice. Ocasio et al. (2020) subsequently connected ABV to organizational learning, proposing that attention moderates the conversion of experience into knowledge. They posited that the tension between executive attention (switching focus for exploration) and attentional vigilance (sustaining focus for exploitation) explains the challenge of organizational ambidexterity.

3.4.3.3 Toward ABV 2.0: Recent Developments (2023-2024)

Ocasio et al. (2023) called for an "ABV 2.0" to address the shift from hierarchical corporations to networks and ecosystems. Joseph et al. (2024) anchored this new phase, highlighting themes of social engagement, temporality, and dimensional quality.

Several contributions emphasize the plasticity of attention structures. Plotnikova et al. (2024) showed how marginalized strategists can "bend" channels through reinvention and renewal. Nicolini & Mengis (2024) integrated practice theory, arguing that a "practice architecture" prefigures a pragmatic field of attention. Kudesia & Lang (2024) applied this lens to crises, defining them as systemic failures in the quality of collective attention rather than individual cognitive limits.

Other works focus on adaptive capacity. Love (2024) introduced the "emergent attentional control system," which allows firms to handle novelty without complete redesigns.

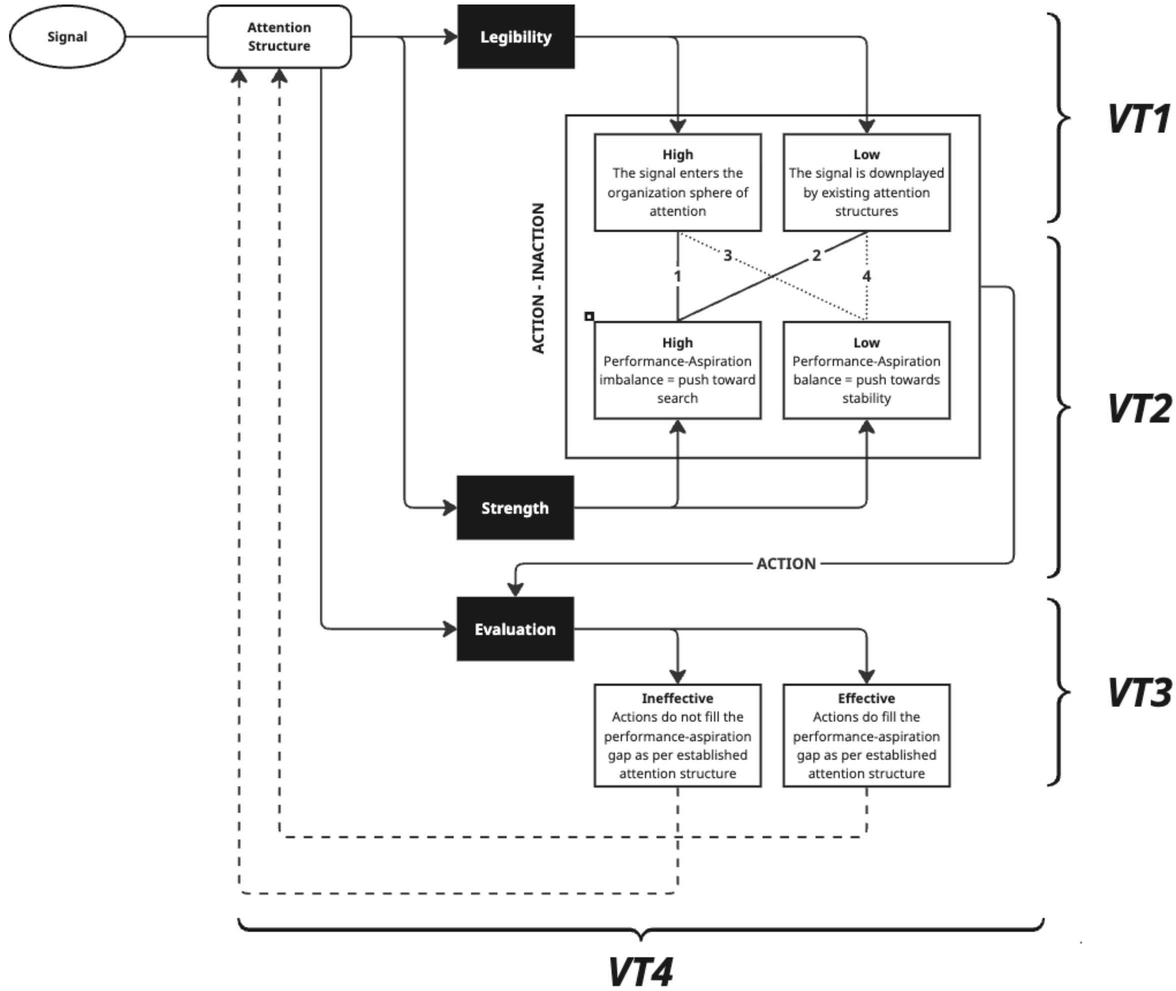
Bartel & Rockmann (2024) theorized relational systems as attentional infrastructure, noting that relational indifference creates open circuits that block coordination. Mack et al. (2024) demonstrated that a high attentional focus, combined with consistency, predicts proactive crisis responses.

Finally, recent work addresses temporality and emotion. Boynton (2024) found that firm growth depends on the congruence between managerial temporal focus and technological complexity; specifically, complex resources require a long-term focus. Vuori (2024) formally integrated emotions, proposing that structures shape perceptions and social contexts to generate emotional energy, which in turn drives attentional engagement.

3.5 THE VISIBILITY TRAP

This section operationalizes the ABV elements developed above into the Visibility Trap framework. As anticipated in the Introduction, the Visibility Trap refers to a self-reinforcing process, structured by established attention structures, in which the salience of visible successes progressively degrades the capacity to identify and respond to associated but latent negative consequences. The VT1–VT4 framework operationalises the two attention-based adaptations of organizational myopia developed in Section 3.2.5. VT1 and VT2 map how GenAI-related consequences are positioned in relation to legibility and strength within existing attention structures. VT3 specifies how these consequences are evaluated primarily on established criteria. VT4 then describes the reinforcing or disruptive recursion of these patterns over time. Figure 3.1 shows the diagrammatic skeleton, derived from ABV, of how the Visibility Trap unfolds via four mechanisms (VT1–VT4). These are explained in the next sub-sections.

FIGURE 3.1: An ABV-based Representation of how the Visibility Trap Unfolds



3.5.1 VT1: Early Signals are Filtered by Existing Attention Structures

VT1 concerns how GenAI-related signals from the environment are processed when they first reach the organization. In ABV terms, signals impinge on the environment of decision and are filtered through existing attention structures – i.e., categories, metrics, roles, and channels that define what counts as an issue and which answers are available (Ocasio, 1997, Mechanisms 1, 2, and 4). In Figure 3.1, the interpretability of signals under established attention structures is referred to as *legibility*: "being clear enough to read" (Oxford dictionary).

Signals that can be interpreted within these structures (high visibility) enter the organizational sphere of attention. Signals that do not fit established repertoires (low legibility) are substantially downplayed or ignored. They lack both a recognized category and an obvious organizational owner, and therefore struggle to be stabilized as issues or answers in any central channel.

In Chapter 4, VT1 is instantiated by the way ChatGPT appears as a new technology with clear privacy and data protection implications, which can be easily classified under existing legal categories of risk. In Chapter 5, VT1 operates as GenAI is recognized as a technology well-suited to knowledge work across various sectors and tasks, aligning with established narratives of digitalization and productivity improvement.

3.5.2 VT2: Aspiration Gaps and Initiation of Organizational Moves

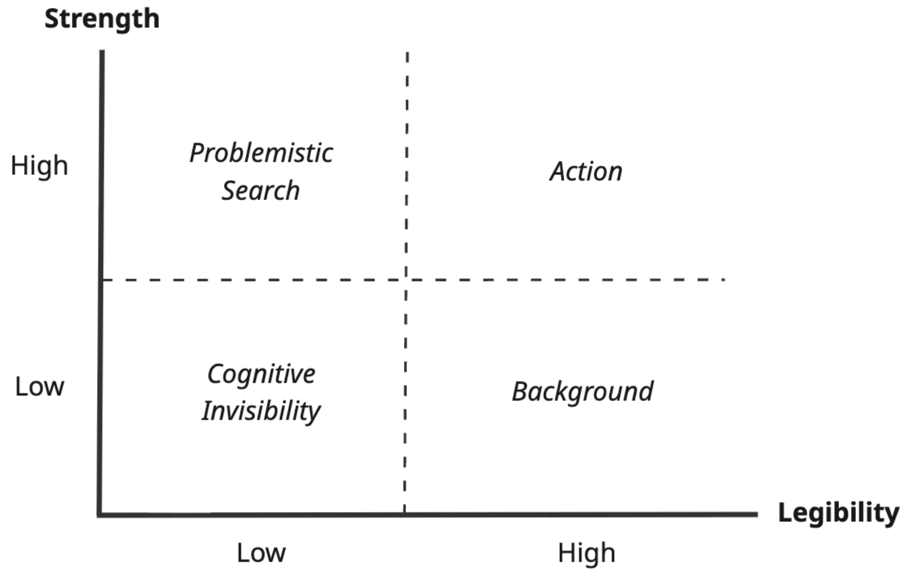
Given existing attention structures, a signal must generate a perceived performance–aspiration gap on dimensions that are already represented in established goals and evaluation metrics. Such aspiration gaps motivate action (Cyert & March, 1963).

In what follows, *strength* denotes the perceived magnitude of this performance–aspiration gap under the current attention structure. Strength is therefore endogenous to existing categories and evaluation schemas: only where a GenAI-related consequence is tied to established goals and measures can a performance-aspiration gap be computed at all.

For a given level of legibility, high-strength signals (i.e., those that clearly indicate a shortfall on valued, already monitored dimensions) lead decision-makers to initiate search using the response repertoires available in the relevant channels. This is how, for example, aspirations for higher performance lead to the adoption of GenAI under innovation and efficiency frames. Low-strength signals (i.e., those that do not indicate a shortfall on already valued dimensions) are more likely to be discussed or acknowledged without major changes. This explains how, even when some signals are legible, action may be delayed: for instance, when organizations recognize the skill-erosion potential of GenAI use but do not concretely manage this risk.

The interaction between legibility and strength determines the patterns and urgency with which an organization acts upon a stimulus. Figure 3.2 summarizes the four resulting quadrants. The cutoffs represented in the figure are indicative, as it is acknowledged that in real organizational contexts, legibility and strength can take varying levels of intensity and cannot be neatly divided into high and low.

FIGURE 3.2: Legibility–Strength configurations and the Visibility Trap



- ***Quadrant 1 - Action*** (high legibility, high strength): The signal is legible and strong; provided that resources are available, actions are taken immediately.
- ***Quadrant 2 - Problemistic Search*** (low legibility, high strength): The signal exposes under-performance, but this tension cannot be reconciled using established attention structures. Problemistic search starts, with the organization exploring possible causes and responses. Each cycle of testing and interpretation updates attention structures (i.e., organizational learning) until the cause of the performance–aspiration mismatch becomes legible (see Sections 3.5.3 and 3.5.4). Then, the process continues as described in Quadrant 1.
- ***Quadrant 3 - Background*** (high legibility, low strength): The signal fits current attention structures but does not generate a perceived mismatch between performance and aspiration. It can therefore be recognized and even discussed, yet it remains classified as low-priority relative to dimensions on which visible indicators already meet or exceed aspirations. Thus, the organization economises on attention not by denying the signal, but by treating it as background.
- **Quadrant 4 - *Cognitive Invisibility*** (low legibility, low strength): Signals in this area are substantially invisible to established attention structures: they are not easily captured by current categories, nor do they register on existing aspiration criteria. When genuinely irrelevant signals are discarded, such filtering preserves scarce attention for more pertinent issues. When these signals are, in fact, precursors of latent

negative consequences, this quadrant corresponds to the incubation space emphasized in the organizational myopia literature.

In this thesis, the *Visibility Trap* primarily concerns configurations in which GenAI-related benefits are located in Quadrant 1, while potential harms are situated in Quadrants 3 and 4.

In Chapter 4, VT2 is visible as GenAI is not only legible as a privacy and data protection issue, but also framed as creating a substantial perceived gap between current practice and the authority's aspiration to protect individuals and society. In Chapter 5, VT2 is instantiated as the perceived gap between productivity with and without AI, motivating rapid experimentation and adoption.

3.5.3 VT3: Evaluation on Visible Criteria

VT3 concerns how organizations evaluate the consequences of actions initiated under VT2. Outcomes are assessed relative to aspirations, and those aspirations are themselves shaped by current attention structures (Ocasio, 1997, Mechanisms 2, 3, 5b, 5c; Cyert & March, 1963). In practice, this means that GenAI moves are primarily judged against existing, visible criteria, including productivity metrics, throughput, case statistics, and formal compliance indicators. As noted in Section 3.2.5, these visible metrics often exhibit synthetic validity: they formally satisfy the organization's success criteria while pre-empting closer verification of the underlying processes and capabilities.

Within this arrangement, *strength* plays a central role. If, after applying an established response to a GenAI-related issue, visible indicators remain below aspirations, the signal continues to be experienced as high-strength: the gap persists. Under problemistic search, this pushes organizations to question the adequacy of current responses and, potentially, to revise routines or interpretations (See Section 3.5.4). If, by contrast, visible indicators improve sufficiently to close or substantially narrow the perceived gap, the signal's strength is reclassified as low: the issue no longer appears problematic on the dimensions that matter in the evaluation channel. The underlying response pattern is then treated as effective. In Figure 3.1, this is represented as the line connecting attention structures to evaluation, whose outcome is classified as effective or ineffective.

In Chapter 4, VT3 is instantiated when the GPDP intervention surrounding ChatGPT is evaluated as a formal regulatory success, as it improves visible privacy and data security indicators and establishes a defensible enforcement precedent. In Chapter 5, VT3 emerges when GenAI-enabled workflows are deemed successful because they deliver observable improvements in productivity, novice performance, and workload management, even though capabilities and memory effects remain weakly represented.

3.5.4 VT4: Recursive Reinforcement

VT4 captures how the outcomes of GenAI-related actions feed back into attention structures over time. In ABV, organizational moves, once selected and enacted, become part of the environment of subsequent decisions and reshape rules, routines, and resource allocations (Ocasio, 1997, Mechanism 6). VT4 has two variants: a reinforcing recursion, and a disruptive one that can partially restore attention to previously latent consequences.

In the reinforcing variant, high performance on established criteria drives reallocation of attention and resources toward GenAI-centred routines. Over time, the attention structures themselves are reshaped around these visible successes: metrics, dashboards, and roles are updated in ways that further privilege GenAI-mediated outputs, making it less likely that latent capability losses or governance vulnerabilities will be recognized as problems.

In the disruptive variant, accumulated erosion eventually manifests. At that point, even though the underlying processes were previously illegible, their consequences register as renewed strength: a perceived performance–aspiration gap on dimensions that matter under the current attention structure. Under pressure to explain and address these shortfalls, organizations are pushed to re-examine their categories, metrics, and routines. In ABV terms, this involves a redefinition of issues and answers and a reconfiguration of attention structures (Ocasio, 1997, Mechanisms 4 and 5b–6): new issue labels are introduced (e.g., explicit concerns about GenAI-induced capability loss, AI dependence, or governance fragility); new metrics and reporting routines are created to track them; and new or reorganized roles are given responsibility for these domains. Restorative moves can then include redesigning GenAI usage patterns, rebalancing performance criteria with capability and memory indicators, and introducing governance arrangements that explicitly address GenAI-specific vulnerabilities.

In Figure 3.1, both variants of VT4 are represented by the dashed lines connecting evaluation outcomes back to attention structures: when actions are repeatedly evaluated as effective on visible criteria, attention structures are reinforced in ways that sustain them; when actions are evaluated as ineffective, attention structures are pushed toward revision, opening possibilities for restoration. In the cases analysed in this thesis, evidence of reinforcing recursion is clearer than evidence of disruptive recursion: Chapter 4 documents how visible GenAI enforcement successes contribute to a governance pattern centred on case-by-case interventions, while Chapter 5 shows how visible performance improvements in knowledge work entrench GenAI-based routines even as concerns about expertise and organizational memory begin to surface.

3.6 CONCLUSIONS

This chapter has developed the theoretical background for the thesis by moving from alternative lenses to an ABV-based account of the Visibility Trap. The starting point was the observation, laid out in Chapter 2, that GenAI tends to produce a characteristic asymmetry: highly visible successes alongside harder-to-detect erosion in capabilities, organizational memory, and governance capacity.

Section 3.2 delimited what existing theories can contribute to this problem, and why the combination of organizational myopia and the ABV is proposed as the primary theoretical foundation. Section 3.3 and Section 3.4 revisited ABV's microfoundations and then articulated its principles, mechanisms, and recent extensions, treating organizations as distributed attention systems. Section 3.5 then elaborated the Visibility Trap as a four-stage mechanism. VT1 and VT2 formalize how GenAI-related signals are filtered through existing attention structures via legibility and perceived performance–aspiration gaps (strength). VT3 specifies that evaluation is dominated by visible criteria embedded in current metrics and control systems. VT4 illustrates how repeated evaluations feed back into aspirations and attention structures, either reinforcing GenAI-centered patterns and keeping harmful consequences in the background, or forcing partial reconfiguration of categories, metrics, and roles when visible indicators eventually fail.

Together, these elements provide a coherent organization-level attentional theory of why visible GenAI successes can coexist with, and sometimes accelerate, latent failures in capabilities, organizational memory, and governance. Chapters 4 and 5 utilize this framework to analyze, respectively, how GenAI serves as a stress test for technology governance and how GenAI reshapes expertise development trajectories and organizational memory in knowledge work.

CHAPTER REFERENCES

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, *17*(1), 99–120. https://doi.org/10.1177/014920639101700108

Bartel, C. A., & Rockmann, K. (2024). The disease of indifference: How relational systems provide the attentional infrastructure for organizational resilience. *Strategic Organization*, *22*(1), 18–48. https://doi.org/10.1177/14761270231183441

Bower, J. L. (1970). *Managing the resource allocation process: A study of corporate planning and investment*. Graduate School of Business Administration, Division of Research.

Boynton, D. (2024). Temporal attention, knowledge breadth, and firm growth. *Strategic Organization*, *22*(1), 91–117. https://doi.org/10.1177/14761270231196213

Burgelman, R. A. (1983). Corporate Entrepreneurship and Strategic Management: Insights from a Process Study. *Management Science*, *29*(12), 1349–1364. https://doi.org/10.1287/mnsc.29.12.1349

Catino, M. (2013). *Organizational myopia: Problems of rationality and foresight in organizations*. Cambridge University Press.

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A Garbage Can Model of Organizational Choice. *Administrative Science Quarterly*, *17*(1), 1. https://doi.org/10.2307/2392088

Daft, R. L., & Weick, K. E. (1984). Toward a Model of Organizations as Interpretation Systems. *The Academy of Management Review*, *9*(2), Article 2. https://doi.org/10.2307/258441

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), Article 3. https://doi.org/10.2307/249008

DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, *48*(2).

Galbraith, J. R. (1974). Organization Design: An Information Processing View. *Interfaces*, *4*(3), 28–36.

Hoffman, A. J., & Ocasio, W. (2001). Not All Events Are Attended Equally: Toward a Middle-Range Theory of Industry Attention to External Events. *Organization Science*, *12*(4), 414–434. https://doi.org/10.1287/orsc.12.4.414.10639

Jackson, S. E., & Dutton, J. E. (1988). Discerning Threats and Opportunities. *Administrative Science Quarterly*, *33*(3), 370. https://doi.org/10.2307/2392714

Joseph, J., Laureiro-Martinez, D., Nigam, A., Ocasio, W., & Rerup, C. (2024). Research frontiers on the attention-based view of the firm. *Strategic Organization*, *22*(1), 6–17. https://doi.org/10.1177/14761270231223397

Kudesia, R. S., & Lang, T. (2024). Toward an attention-based view of crises. *Strategic Organization*, *22*(1), 118–145. https://doi.org/10.1177/14761270231189935

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*(S2), 95–112. https://doi.org/10.1002/smj.4250141009

Leonardi, P. M. (2011). When Flexible Routines Meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies. MIS Quarterly, 35(1), 147–167. https://doi.org/10.2307/23043493

Love, E. G. (2024). Attentional control systems for emergent strategic issues in the post-Chandlerian world. *Strategic Organization*, *22*(1), 146–164. https://doi.org/10.1177/14761270231199278

Mack, D. Z., Cho, T. S., & Yi, A. C. (2024). Attentional engagement and strategic responses to discontinuous environmental change: Evidence from the US banking industry. *Strategic Organization*, *22*(1), 49–90. https://doi.org/10.1177/14761270231187090

March, J., & Simon, H. (1958). *Organizations* (1st edition). Wiley.

Nicolini, D., & Mengis, J. (2024). Toward a practice-theoretical view of the situated nature of attention. *Strategic Organization*, *22*(1), 211–234. https://doi.org/10.1177/14761270231183731

Ocasio, W. (1997). Towards an Attention-Based View of the Firm. *Strategic Management Journal*, *18*(S1), 187–206. https://doi.org/10.1002/(SICI)1097-0266(199707)18:1+%253C187::AID-SMJ936%253E3.0.CO;2-K

Ocasio, W. (2011). Attention to Attention. *Organization Science*, *22*(5), 1286–1296. https://doi.org/10.1287/orsc.1100.0602

Ocasio, W., & Joseph, J. (2008). Rise and Fall—Or Transformation? Long Range Planning, 41(3), 248–272. https://doi.org/10.1016/j.lrp.2008.02.010

Ocasio, W., & Joseph, J. (2005). An Attention-Based Theory of Strategy Formulation: Linking Micro- and Macroperspectives in Strategy Processes. In *Advances in Strategic Management* (Vol. 22, pp. 39–61). Emerald (MCB UP). https://doi.org/10.1016/S0742-3322(05)22002-8

Ocasio, W., Laamanen, T., & Vaara, E. (2018). Communication and attention dynamics: An attention‐based view of strategic change. *Strategic Management Journal*, *39*(1), 155–167. https://doi.org/10.1002/smj.2702

Ocasio, W., Rhee, L., & Milner, D. (2020). Attention, Knowledge, and Organizational Learning. In *The Oxford Handbook of Group and Organizational Learning*. Oxford University Press.

Ocasio, W., & Wohlgezogen, F. (2010). Attention and control. In S. B. Sitkin, L. B. Cardinal, & K. M. Bijlsma-Frankema (Eds.), *Organizational Control* (1st ed., pp. 191–221). Cambridge University Press. https://doi.org/10.1017/CBO9780511777899.008

Ocasio, W., Yakis-Douglas, B., Boynton, D., Laamanen, T., Rerup, C., Vaara, E., & Whittington, R. (2023). It's a Different World: A Dialog on the Attention-Based View in a Post-Chandlerian World. *Journal of Management Inquiry*, *32*(2), 107–119. https://doi.org/10.1177/10564926221103484

Orlikowski, W. J. (2007). Sociomaterial Practices: Exploring Technology at Work. *Organization Studies*, *28*(9), 1435–1448. https://doi.org/10.1177/0170840607081138

Orlikowski, W. J., & Scott, S. V. (2008). 10 Sociomateriality: Challenging the Separation of Technology, Work and Organization. *Academy of Management Annals*, *2*(1), 433–474. https://doi.org/10.5465/19416520802211644

Pasman, H. J. (2021). Early warning signals noticed, but management doesn't act adequately or not at all: A brief analysis and direction of possible improvement. *Journal of Loss Prevention in the Process Industries*, *70*, 104272. https://doi.org/10.1016/j.jlp.2020.104272

Pasmore, W., Winby, S., Mohrman, S. A., & Vanasse, R. (2019). Reflections: Sociotechnical Systems Design and Organization Change. *Journal of Change Management*, *19*(2), 67–85. https://doi.org/10.1080/14697017.2018.1553761

Plotnikova, A., Pandza, K., & Whittington, R. (2024). Bending the pipes: Regaining attention through reinvention and renewal. *Strategic Organization*, *22*(1), 165–188. https://doi.org/10.1177/14761270231184616

Rogers, E. M. (2003). *Diffusion of innovations, 5th edition*. Free Press.

Sato, H. (2012). Routine-Based View of Organizational Learning and Mechanisms of Myopia. *Annals of Business Administrative Science*, *11*(0), 45–54. https://doi.org/10.7880/abas.11.45

Scott, W. R. (1995). *Institutions and organizations*. Sage.

Simon, H. A. (1947). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization* (1st edition). Macmillan Co.

Swidler, A. (1986). Culture in Action: Symbols and Strategies. *American Sociological Review*, *51*(2), 273. https://doi.org/10.2307/2095521

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509–533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7%253C509::AID-SMJ882%253E3.0.CO;2-Z

Trist, E. L., & Bamforth, K. W. (1951). Some Social and Psychological Consequences of the Longwall Method of Coal-Getting: An Examination of the Psychological Situation and Defences of a Work Group in Relation to the Social Structure and Technological Content of the Work System. *Human Relations*, *4*(1), 3–38. https://doi.org/10.1177/001872675100400101

Turner, B. A., & Pidgeon, N. F. (1978). Man-made disasters. Butterworth-Heinemann.

Vaughan, D. (2007). *The Challenger launch decision: Risky technology, culture, and deviance at NASA* (Nachdr.). Univ. of Chicago Press.

Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), Article 3. https://doi.org/10.2307/30036540

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, *39*(2), Article 2. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Vuori, T. O. (2024). Emotions and attentional engagement in the attention-based view of the firm. *Strategic Organization*, *22*(1), 189–210. https://doi.org/10.1177/14761270231165356

Weick, K. E. (1979). *The social psychology of organizing* (2. ed.). McGraw-Hill.

Weick, K. E., & Sutcliffe, K. M. (2006). Mindfulness and the Quality of Organizational Attention. *Organization Science*, *17*(4), 514–524. https://doi.org/10.1287/orsc.1060.0196

Wernerfelt, B. (1984). A Resource-Based View of the Firm. *Strategic Management Journal*, *5*(2), 171–180.

CHAPTER 4: PACING, ATTENTION, AND THE VISIBILITY TRAP IN GENAI GOVERNANCE

ABSTRACT

The *pacing problem* in technology governance arises when rapidly evolving technological capabilities confront slower legal and institutional adaptation. This chapter examines how that misalignment is mediated by existing governance frameworks that function as attention structures, channeling regulatory focus toward technically narrow, highly visible responses while leaving broader failures underrepresented. Building on the *Visibility Trap* framework developed in Chapter 3, the chapter analyses Italy's 2023 ChatGPT ban as a case of GenAI enforcement. A mixed-methods design combines document analysis of the Italian Data Protection Authority's (GPDP) decisions and communications, difference-in-differences estimation of post-ban ChatGPT usage, and qualitative content analysis of 1,034 posts in the X (Twitter) debate. The findings show that, within a GDPR-centred evaluative architecture, the intervention is a formal success: OpenAI implements the required measures, the GPDP secures a €15 million fine, and the case is framed as exemplary data-protection enforcement. At the same time, usage returns to its counterfactual trajectory and public discourse exhibits pronounced erosion of pragmatic, moral, and cognitive legitimacy. The chapter thus provides an empirical instantiation of the *Visibility Trap* in GenAI governance and shows how the misfit between GenAI and institutionalized attentional architectures generates legitimacy vulnerabilities despite formal success. On this basis, it derives the Governance Trilemma confronting GenAI authorities: act early with misaligned tools and risk legitimacy erosion, delay intervention and risk irrelevance, or abstain and effectively abdicate responsibility.

4.1 INTRODUCTION

Innovation governance is concerned with supporting the production and diffusion of innovation while mitigating the social risks it may entail (Hemphill, 2020; Shetty et al., 2025; Bolici et al., 2024; Chan et al., 2024; Gualdi & Cordella, 2024; Milmo, 2023). Regulation is a primary instrument for operationalizing this governance, but its application to new technologies is inherently complex. This challenge is often framed as a pacing problem, a misalignment between rapidly evolving technologies and slower, more rigid legal and institutional change (Allenby, 2011; Marchant, 2011, 2020; Taeihagh et al., 2021; Walter, 2024). Recent work emphasizes that this misalignment is both temporal and structural, involving mismatches in concepts, instruments, and coordination architectures (Allenby, 2011; Maas, 2022; Nordström, 2022; Dolfsma & Seo, 2013). This chapter builds on that diagnosis and focuses on how, in the case of GenAI, this temporal–structural misfit is experienced inside a public authority as an incompatibility between GenAI and established governance frameworks understood as institutionalized attention structures (Ocasio, 1997).

While research calls for new governance approaches for GenAI (Sætra, 2023; Wang et al., 2025; Yoganathan et al., 2025), empirical evidence on the real-world implementation remains scarce (Sætra, 2023). Policymakers must therefore act with limited guidance, increasing the risk of interventions that generate visible successes while concealing latent negative outcomes. The goal of this chapter is to address this empirical gap by analysing the tension between formal regulatory success and less visible negative consequences. Using the ChatGPT ban in Italy as a case study, it addresses the following research question: *How do formal successes and invisible failures coexist in technology governance, and what mechanisms influence their visibility?*

The chapter makes three contributions to research on GenAI governance. First, drawing on a mixed-method analysis of Italy's 2023 ChatGPT ban, it empirically demonstrates how the pacing problem can materialize as a tension between formal regulatory success and substantive behavioral and legitimacy failure. Second, it applies the Visibility Trap framework developed in Chapter 3 to this case, showing how GDPR and the authority's inherited procedures function as an attention architecture that channels evaluative focus toward legible, high-strength indicators of compliance while allowing low-legibility, low-strength outcomes to deteriorate. Third, it derives the Governance Trilemma facing GenAI regulators: when acting under conditions of legal uncertainty and resource constraints, authorities are structurally pulled between demonstrating visible enforcement, preserving practical capacity to influence behavior, and maintaining public legitimacy.

The remainder of this chapter is organized as follows. Section 4.2 establishes the study background, reviewing the literature on restrictive interventions and reframing the pacing problem through the lenses of the Attention-Based View (ABV) and legitimacy. Section 4.3 details the mixed-methods research design. Section 4.4 presents the empirical results, tracing the trajectory from the regulator's narrative of success to the quantitative evidence of behavioral indifference and the qualitative mapping of legitimacy erosion. Section 4.5 connects these findings back to the Visibility Trap framework developed in Chapter 3 and uses them to discuss the Governance Trilemma. Section 4.6 concludes and outlines the limitations of this study.

4.2 STUDY BACKGROUND AND THEORETICAL FRAMING

This section reviews three streams of literature that form the context for this study. First, it examines restrictive interventions, drawing on theories of psychological reactance and legitimacy to explain why regulatory bans often fail to influence public behavior. Second, it builds on the pacing problem literature by specifying how temporal and structural misalignment in GenAI governance can be understood through an ABV of organizations, treating governance frameworks as institutionalized attention structures that shape which consequences become visible and actionable for authorities. Finally, it situates the study within existing evidence on the Italian ChatGPT ban, identifying the critical, unaddressed gap between the ban's formal success and its substantive failure.

4.2.1 The Paradox of Restrictive Interventions

Public institutions govern the diffusion of technology by either accelerating adoption (e.g., subsidies) or restricting it (e.g., bans) (Rogers, 2003; Caiazza, 2016). Restrictive policies, typically implemented to safeguard the public, often produce unintended consequences that undermine their effectiveness. These include alternative-seeking behavior, such as shifting to less safe substitutes or the creation of black markets (Rogers, 1983; Bushman & Stack, 1996).

A central mechanism driving these effects is psychological reactance: an aversive motivational state that drives individuals who perceive threats to their freedom to restore that freedom (Brehm, 1966; Brehm & Brehm, 1981). The Streisand Effect is a prominent manifestation of this phenomenon (Jansen & Martin, 2015). Named after Barbra Streisand's failed 2003 attempt to suppress photographs of her residence, which paradoxically amplified their circulation, this effect describes how attempts to censor information trigger mass public attention and defiance. It operates through dual channels: awareness amplification (the restriction generates publicity) and legitimacy erosion (the act of restricting frames authorities as illegitimately controlling) (Jansen & Martin, 2015).

Empirical work on technology bans in non-authoritarian countries is scarce, but COVID-19 restrictions offer a useful proxy for studying restrictive interventions under high uncertainty, rapid implementation, and constraints on individual freedom. This literature suggests that compliance depends on the perceived legitimacy of authorities, encompassing both their normative alignment ("do their values match ours?") and their rightful authority ("do they have an obligation to obey?") (Gurinskaya et al., 2024; Hawks, 2025). When legitimacy fails, it can create an epistemic crisis of public confidence, where distrust in expertise supports widespread non-compliance (Dowling & Legrand, 2023).

For technology regulation, this is exacerbated by a dual illegitimacy problem (Yoganathan et al., 2025). Users must evaluate the legitimacy of both the regulator and the regulated firm. Large technology firms are often viewed as public actors without public values, wielding power through “consent theatre” that users perceive as illegitimate (Taylor, 2021). Regulators simultaneously face questions about their technical competence and independence from capture (Viljoen, 2021).

This creates a high-risk environment where a regulator, already facing a dual legitimacy crisis, must intervene using misaligned structural tools.

4.2.2 Reframing the Pacing Problem

As discussed in Chapter 2 (Section 2.3.2.1), GenAI governance operates under a pronounced pacing problem: AI capabilities and deployment contexts evolve faster than legal and institutional responses can be designed, negotiated, and implemented (Allenby, 2011; Marchant, 2011, 2020; Taeihagh et al., 2021; Taeihagh, 2021). The literature already treats this as more than a temporal gap. It points to structural misalignment between complex socio-technical trajectories and the ways governance institutions define problems, design instruments, and coordinate responses (Allenby, 2011; Maas, 2022; Nordström, 2022). Regulators confront moving targets: system capabilities, risk profiles, stakeholders, and even the relevant concepts (“AI”, “safety”, “high-risk”) shift while regimes are still being drafted (Nordström, 2022). Governance becomes an exercise in managing regret under great uncertainty (Hemphill, 2020), with second-best, revisable mixes of instruments rather than stable solutions (Marchant, 2011, 2020; Taeihagh et al., 2021).

GenAI intensifies the pacing problem because of its general-purpose, infrastructural character. Models such as ChatGPT sit on top of globally distributed stacks of data, compute, networks, and software (Mueller, 2025). Their impacts extend beyond immediate use cases into organizational routines, labour markets, political discourse, and knowledge production, often through chains of second- and third-order effects that are difficult to foresee ex ante (Sætra, 2023). At the same time, public debate and media coverage make certain hazards highly salient: privacy violations, misinformation, academic cheating, and job displacement.

This visibility channels attention toward the most vivid threats while deeper incompatibilities between legal categories and technical architectures remain opaque (Rejeski, 2011; Sætra, 2023).

The governance system in which the GPDP operates is polycentric and multilevel. International bodies, the EU, national governments, firms, standards organizations, and civil society all participate in shaping AI outcomes (Dafoe, 2023; Monteiro & Singh, 2025). Within this system, public authorities are not neutral rule-appliers. They are organizations with bounded rationality, internal routines, and performance expectations (Maas, 2022; Monteiro & Singh, 2025). In this thesis, the pacing problem is therefore treated as a condition mediated by organizational attention: macro-level misalignment becomes, inside an authority, an incompatibility between GenAI and the institutionalized attention structures through which the authority perceives problems and recognises available interventions.

These attention structures are encoded in legal frameworks, procedures, and communication channels. For a data protection authority, GDPR and related instruments define both obligations and categories (data subject, controller, lawful basis), standard sequences of action (complaint, investigation, order), as well as public vocabularies of justification (lawfulness, transparency, data minimisation). They make some aspects of digital systems legible as governable (e.g., personal data processing) and relegate others to the background.

Privacy and data governance represent a specific instance of this dynamic. As reviewed in Chapter 2 (Section 2.3.3), privacy law is one of the few mature, enforceable instruments available for governing AI diffusion. Yet work on advertising ecosystems shows a persistent compliance–substance gap: even under GDPR, users who formally opt out of tracking are still monitored because technical architectures and business incentives sustain data flows irrespective of consent (Liu et al., 2024). This signals a deeper incompatibility between formal rights and infrastructural practice. GenAI is layered on top of those infrastructures, relies on large-scale data extraction, and adds further opacity.

In this configuration, the GPDP's intervention occurred at a moment when AI-specific regimes were still emerging and globally fragmented (Currie et al., 2025; Walter, 2024; Zhao, 2023). National authorities faced pressure to demonstrate that they could act on GenAI harms despite limited tools and high uncertainty (Gualdi & Cordella, 2024; Coltri, 2024). From an ABV (Ocasio, 1997), the formal infringement proceeding against ChatGPT is a predictable response. GenAI as a broad phenomenon is difficult to govern with existing legal categories; ChatGPT, as a service processing Italian users' personal data, is precisely the kind of object that the authority's attentional architecture can register and process. Under strong legitimacy pressures in a polycentric governance race, the GDPR enforcement channel appears internally as the most coherent, actionable, and communicable response.

This chapter builds on this diagnosis and makes two contributions. First, it provides an empirical demonstration of how the pacing condition manifests in the concrete case of a data protection authority responding to GenAI diffusion, and what consequences follow. Second, it offers an attention-based specification of the temporal and structural mismatch, focusing on how the pacing problem is organized inside a public authority rather than only at the level of high-level regimes. Accordingly, governance frameworks (e.g., the GDPR) are treated as institutionalized attentional architectures that make certain parts of GenAI governance visible while leaving others in a blind spot.

4.2.3 Legitimacy and Neo-institutional Theory

Together with the ABV, this study adopts the neo-institutional theory as a supporting theoretical pillar. Both perspectives share roots in bounded rationality: they reject the image of organizations as neutral optimizers and focus instead on routines, rules, and attention as the primary drivers of behavior (Cyert & March, 1963; March & Simon, 1958; Simon, 1947). ABV specifies how organizational roles, routines, and communication channels channel attention (Ocasio, 1997); neo-institutional theory specifies how institutional environments define which models and practices are legitimate objects of attention and which are hard to conceive as options at all (DiMaggio & Powell, 1991; Meyer & Rowan, 1977; Zucker, 1977).

4.2.3.1 Neo-institutional Environments, Fields, and Legitimacy

Neo-institutional theory starts from a strong claim about what institutions are and how they matter. Institutions are not simply explicit rules or designed conventions. They are "social processes, obligations, or actualities" that acquire a rule-like status in social thought and action (Meyer & Rowan, 1977). Once institutionalized, rules and structures are treated as objective, external facts and as the natural way to do things. Zucker's (1977) experiments show that, as practices become perceived as more objective and exterior to any particular actor, they become more persistent, more resistant to change, and more easily transmitted across generations.

Compared with the "old" institutionalism of Selznick (1953, 1996), neo-institutional theory shifts the analytic lens from local communities to organizational fields and wider societal sectors (DiMaggio & Powell, 1991). Selznick focused on how organizations were embedded in proximate communities and how leaders managed threats through political co-optation; the locus of institutionalization was the individual organization. Neo-institutional theorists instead emphasize fields: constellations of regulators, firms, professionals, and other organizations that collectively constitute a “recognized area of institutional life” (DiMaggio & Powell, 1983). For GenAI, the relevant field is composed of data protection authorities, AI firms such as OpenAI, EU institutions, national governments, civil society organizations, and expert communities. The GPDP’s ban is a move inside this field, not just a bilateral regulatory action.

Within fields, organizations face a legitimacy imperative. Survival depends less on technical efficiency and more on conformity with widely shared “rationalized myths”: formalized, impersonal prescriptions for what organizations in a given domain should look like and how they should act (Meyer & Rowan, 1977). Formal structures signal adherence to these myths and serve as a “vocabulary of structure” that communicates worthiness and rationality to external audiences. Data protection authorities, for example, display their legitimacy through visibly enforcing GDPR, issuing sanctions, and adopting recognizable governance instruments.

Isomorphism is the core dynamic through which this imperative operates. DiMaggio and Powell (1983) distinguish coercive isomorphism (pressures from the state or powerful organizations), mimetic isomorphism (copying others under uncertainty), and normative isomorphism (professionalization and shared training). In the emerging GenAI field, all three are visible: EU-level regulation and guidance exert coercive pressures; national regulators watch one another and possibly emulate peer moves; and professional networks of data-protection experts and AI ethicists diffuse models and vocabularies.

Neo-institutional theory also provides a distinctive account of inertia. Old institutionalism traced resistance to change to vested interests and political bargains in the informal organization (Selznick, 1953, 1996). Neo-institutionalism identifies a more profound source: the cognitive and cultural sunk costs embedded in institutionalized arrangements (Zucker, 1977). Formal structures, categories, and routines are hard to rethink because they structure perception and make alternatives difficult to imagine. Powell (1991) develops this into a path-dependence argument: once a field settles onto a trajectory, increasing returns, interdependencies, and taken-for-granted assumptions “lock in” structures that are neither necessarily efficient nor easily altered. For this study, this provides the institutional macro-foundation for the Visibility Trap developed in Chapter 3.

4.2.3.2 Practical Action, Cognition, and the Micro-Foundations of Legitimacy

Neo-institutional theory rests on a cognitive, rather than a purely normative, theory of action. Drawing on the Carnegie School and on broader developments in social theory, it assumes that much behavior is routine, taken for granted, and guided by practical know-how rather than explicit value commitments (Cyert & March, 1963; March & Simon, 1958; Simon, 1947; Zucker, 1977). Garfinkel's (1967) ethnomethodology shows how social order is a fragile accomplishment built through tacit background knowledge, typifications, and "accommodative work" that makes interactions appear normal and reasonable. Giddens (1984) and Bourdieu (1977) provide micro-macro linkages in which routines, affective interaction rituals, and habitus reproduce social structures in the course of everyday practical action.

Neo-institutional theorists incorporate these insights by treating institutionalization as fundamentally cognitive: institutions supply classifications, scripts, and schemas that actors use to interpret situations and decide what counts as appropriate conduct (DiMaggio & Powell, 1991; Jepperson, 1991; Zucker, 1977). Scott (1995) synthesizes this into a three-pillar framework. The regulative pillar highlights formal rules, sanctions, and legal authority; compliance follows a logic of expedience, and legitimacy is legally sanctioned. The normative pillar highlights values and norms; compliance follows a logic of appropriateness, and legitimacy is morally governed. The cultural-cognitive pillar highlights shared taken-for-granted beliefs and frames; compliance is almost automatic, and legitimacy derives from being conceptually "taken for granted" as the natural way to organize a given activity.

In the case of GenAI governance, these pillars operate simultaneously. The GPDP derives regulatory legitimacy from its legal mandate under GDPR; it is expected to enforce rules and can impose sanctions. It faces normative expectations from political principals, professional communities, and citizens regarding how data protection authorities ought to behave in a democratic society. Over time, certain regulatory models (e.g., risk-based impact assessments, guidance documents, formal investigations) have acquired cultural-cognitive legitimacy as the default ways in which "serious" regulators address emerging technologies. The ChatGPT ban intervenes in this configuration: it reasserts the regulative pillar through a highly visible coercive instrument, but it also exposes and tests the normative and cultural-cognitive expectations that different audiences hold about what responsible AI governance should look like.

4.2.3.3 Organizational Legitimacy: Pragmatic, Moral, and Cognitive

Within this broader framework, Suchman's (1995) synthesis of organizational legitimacy provides the core analytical vocabulary for this study. Suchman integrates strategic and institutional perspectives into a widely adopted definition of legitimacy as "a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions" (p. 574). This definition explicitly embeds legitimacy judgments in the institutional environment described above.

Suchman (1995) distinguishes three interrelated forms of legitimacy. Pragmatic legitimacy rests on the self-interested evaluations of an organization's immediate audiences. It reflects whether stakeholders perceive that they benefit from the organization's actions. This includes direct exchange benefits (e.g., effective protection of personal data; reliable AI services), influence legitimacy (the belief that an organization is responsive to an audience's broader interests), and dispositional legitimacy (attributions of trustworthy or competent character). Moral legitimacy reflects broader normative evaluations about whether an organization's actions are "the right thing to do" for society (Suchman, 1995). It derives from perceived consequences, procedures, structures, and personal qualities: organizations are judged on what they achieve, how they behave, how they are structured, and who leads them. Cognitive legitimacy is the most entrenched form. It arises when an organization's existence and activities are comprehensible and taken for granted. In this mode, an organization is legitimate because it fits established cultural models of what such an entity should be; alternatives appear implausible or literally unthinkable. This dimension connects directly to Meyer & Rowan's (1977) rationalized myths and Zucker's (1977) experimental demonstration of taken-for-grantedness.

Real organizations rarely enjoy a single, uniform type of legitimacy. They face heterogeneous audiences and therefore rely on complex legitimacy profiles that combine pragmatic, moral, and cognitive elements in different ways (Suchman, 1995). Tensions are common. Strategies that shore up pragmatic legitimacy with one audience can erode moral legitimacy with another, or destabilize cognitive legitimacy by making familiar arrangements appear questionable. For a data protection authority, issuing a highly visible sanction may enhance pragmatic legitimacy with certain domestic constituencies or peers (by signaling strength and competence), yet at the same time threaten moral legitimacy if perceived as disproportionate, or cognitive legitimacy if it clashes with taken-for-granted expectations about how regulators operate in a liberal democracy.

Suchman (1995) also emphasizes the temporal dimension of legitimacy. Organizations must gain legitimacy when they emerge, maintain it in the face of environmental change, and sometimes repair it after crises. Strategies differ across these phases: new actors often conform closely to existing templates or engage in institutional entrepreneurship to redefine standards; established actors emphasize monitoring, buffering, and reputation building; actors in crisis use normalizing accounts, selective reform, or more radical restructuring. The GPDP is an established actor embedded in a mature European data-protection field. The ChatGPT ban does not concern gaining initial legitimacy but rather maintaining and potentially repairing legitimacy under conditions of technological disruption, public scrutiny, and regulatory uncertainty. The study's analysis of how different audiences reframe the ban in terms of benefits, harms, fairness, and rationality uses Suchman's threefold distinction as a coding and interpretive scaffold.

4.2.3.4 Institutional Work, Entrepreneurship, and Logics in GenAI Governance

A long-standing criticism of neo-institutional theory is that early formulations portrayed organizations as passive rule-takers and underplayed heterogeneity and change (Powell, 1991). Later work responded by foregrounding agency and practice. DiMaggio's (1988) concept of institutional entrepreneurship highlights actors who mobilize resources to create or transform institutions in pursuit of valued interests. Maguire et al. (2004) and Lawrence & Suddaby (2006) generalize this into the broader notion of institutional work: the purposive activities through which actors create, maintain, or disrupt institutions. Creating institutions involves advocacy, defining and theorizing new rule systems, and vesting rights. Maintaining institutions involves policing, norm reinforcement, mythologizing, and embedding routines in daily practice. Disrupting institutions involves disconnecting sanctions from practices, disassociating them from moral foundations, or undermining core taken-for-granted assumptions.

GenAI creates an arena where multiple forms of institutional work collide. OpenAI and other providers engage in institutional work to stabilize a view of foundation models as legitimate, innovative, and manageable through self-regulation and technical fixes. Data protection authorities engage in work to maintain existing legal and normative frameworks, but also engage in institutional entrepreneurship when they push novel interpretations or high-profile interventions that reframe acceptable uses of data-driven technologies. Civil society organizations, experts, and users engage in disruptive work when they question the adequacy of prevailing governance templates or contest the authority of regulators and firms.

The Italian ChatGPT ban can be read through this lens as a combined attempt at maintaining and creating institutions. The GPDP mobilizes an existing legal framework (GDPR) and its formal enforcement powers to assert that generative-AI systems must comply with data-protection principles. At the same time, the decision implicitly theorizes generative AI as a problem category and articulates expectations about transparency, legal basis, and age verification that may crystallize into field-level norms. The reaction of other actors, supportive, critical, or mocking, constitutes counter-vailing institutional work that maintains alternative visions of legitimate AI development or disrupts the GPDP's claim to authority. This back-and-forth is precisely what Stage 3 of the study traces in public discourse.

The institutional-logics perspective makes this more precise by specifying how field-level cultural systems shape attention and evaluation (Friedland & Alford, 1991; Thornton & Ocasio, 1999; Thornton et al., 2012). Different logics (market, state, profession, technology, civil society) provide competing "rules of the game", sources of authority, and success criteria. They also shape what decision-makers notice, how they categorize issues, and which solutions appear salient. Thornton et al. (2012) explicitly integrate institutional logics with ABV, arguing that logics influence the distribution of attention by making some issues and answers more legitimate and cognitively available than others. Thornton and Ocasio's (1999) study of the U.S. higher-education publishing industry shows how a shift from an editorial to a market logic changes the determinants of power and executive turnover.

In the GenAI field, several logics intersect. A state-legal logic emphasizes formal compliance, rule enforcement, and public authority; a market-innovation logic emphasizes speed, user acquisition, and competitive positioning; a professional-expert logic emphasizes methodological rigor, risk assessment, and ethical standards; a civic-democratic logic emphasizes rights, participation, and openness. The GPDP's ban can be interpreted as privileging a state-legal logic that foregrounds formal legality and data-protection principles. The intense reactions from developers, users, and commentators reflect clashes with market, professional, and civic logics that value different goods and signal different legitimacy criteria. These clashes are visible in pragmatic complaints (lost access, competitiveness), moral critiques (fairness, proportionality, due process), and cognitive frames (Italy as backward or visionary).

Linking this back to ABV, institutional logics define the issues and answers that receive organizational attention (Ocasio, 1997; Thornton et al., 2012). A regulator socialized in a strong state-legal logic will attend more to formal legal defensibility and visible demonstrations of enforcement than to longer-term capability erosion, practical workarounds, or perceptions among peripheral but consequential audiences. A field structured by highly institutionalized myths of “decisive enforcement” and “innovation leadership” can therefore channel attention toward actions that are highly legible on the regulative pillar while gradually degrading pragmatic, moral, and cognitive legitimacy with critical stakeholders.

This study uses this neo-institutional and legitimacy framework for two analytic tasks. First, it conceptualizes the GPDP not as a purely technical problem solver but as an institutional actor embedded in a field, facing a legitimacy imperative vis-à-vis multiple audiences and logics. Second, it provides the vocabulary (i.e., pragmatic, moral, and cognitive legitimacy) for analyzing how the ban reshapes perceptions of the GPDP and of GenAI governance more broadly. Together with ABV, this clarifies how the Visibility Trap developed in Chapter 3 operates in GenAI governance: institutionalized attention structures and legitimacy expectations make certain governance interventions appear attractive and necessary, even when they initiate a gradual erosion of the institutional foundations on which effective governance depends.

4.2.4 The Case Context: Prior Research on the Italian ChatGPT Ban

Existing literature on the Italian ChatGPT ban can be organized into three streams. The first stream uses quasi-experimental methods to measure the ban's economic and behavioral impacts. On the economic side, Kreitmeir & Raschky (2023) found that the ban caused a 50% drop in productivity for Italian software developers, while Bertomeu et al. (2023) found a 6-7% decline in market value for AI-exposed Italian firms and a 17.6% decrease in AI-complementary hiring. Concurrently, this stream documents mass behavioral circumvention. Kreitmeir & Raschky (2023) also showed a significant spike in Google searches for VPNs, corroborated by Kikuchi (2024), who analyzed OONI network data.

The second stream consists of legal and policy critiques of the ban's justification. This critique has two main components. The primary one focuses on the regulatory mismatch: Gualdi & Cordella (2024) and Busacca & Monaca (2025) argue that the ban was fundamentally flawed because it applied the static logic of the GDPR to the dynamic, probabilistic logic of GenAI. The second component critiques the ban's ethical foundations, arguing that the GPDP’s focus on privacy was insufficient and failed to address deeper ethical issues, such as epistemic inequality (Farisco, 2024), or that it employed the wrong philosophical approach entirely (Coltri, 2024).

The third stream examines the public's discursive response. This area is primarily represented by Bolici, Varone, & Diana (2024). Using a mixed-methods analysis of X (Twitter) posts, they found that the ban was both unpopular (with 61% negative sentiment) and ineffective, as evidenced by the extensive and collaborative search for workarounds, such as VPNs. This study directly extends their work by: (1) providing a quantitative measure of the behavioral outcome (post-ban usage and search for workarounds) that their analysis implied but could not demonstrate; and (2) using the Visibility Trap framework developed in Chapter 3 to link their finding of public delegitimation to the behavioral failure observed here.

4.3 METHODS

This study employs a three-stage mixed-methods design to investigate the divergence between formal regulatory success and substantive governance failure. This approach is necessary because no single method can capture the central paradox: the GPDP's "visible success" (formal compliance) must be established as a baseline (Stage 1) before it can be contrasted with its invisible failures, which manifest first as a behavioral indifference (Stage 2) and second as a discursive legitimacy collapse (Stage 3). Table 4.1 outlines this argumentative logic.

TABLE 4.1: Methodological Approach and Outcome Preview

Stage	Methodological Approach	Outcome
1	Document Analysis	**Establish the formal "Visible Success"** This reconstructs the official GPDP’s narrative. It motivates what the GPDP stated it did and achieved (compliance, a fine).
2	Difference-in-Differences	**First latent failure: Post-ban behavioral Indifference** It provides the first counterpoint: despite the formal success, the intervention had no lasting behavioral impact on ChatGPT usage.
3	Qualitative Content Analysis	**Second latent failure: Legitimacy Erosion** The intervention, and the GPDP were extensively delegitimized by the public under the pragmatic, moral, and cognitive dimensions (Suchman, 1995). This analysis provides a possible explanatory justification for the DiD finding.

4.3.1 Stage 1: Document Analysis – Reconstructing the GPDP Intervention Narrative

A qualitative document analysis was employed to reconstruct the GPDP's official narrative of intervention. The primary methodological functions were to (1) establish a factual timeline and (2) provide an official "regulator narrative" to triangulate against the public's discursive response, thereby preventing single-source bias (Bowen, 2009).

The corpus comprised the complete set of documents (n = 10) issued by the GPDP regarding the ban, spanning two formal legal orders (Provvedimenti) and eight press releases (Comunicati) from March 30, 2023, to December 20, 2024. The analysis was conducted in two stages. The first stage reconstructed the historical evolution of the intervention by chronologically sorting the documents. This reconstructed a six-phase process: (I) The Urgent Intervention, (II) The Diplomatic Pivot, (III) The Formal "Path to Compliance," (IV) Framing International Leadership, (V) The Resolution and Declaration of Success, and (VI) The Final Judgment.

The second stage analyzed the framing within each phase. This involved examining subtle cues, such as the one-time use of the word "blocca" (blocks) in the initial, high-impact press release (GPDP, 2023g), which contrasted sharply with the consistent, formal use of "temporary limitation" in all official legal orders.

This analysis establishes the baseline "regulator narrative" of formal success, which serves as the critical counterpoint to the findings of substantive failure documented in the subsequent stages.

4.3.2 Stage 2: Difference-in-Differences Estimation of ChatGPT Usage Behavior

To quantify the lasting behavioral impact of the ban on usage behavior, a DiD approach was employed. DiD is a quasi-experimental method used for causal inference when randomized controlled trials are not feasible (Angrist & Pischke, 2009; Lechner, 2011). It estimates the treatment effect by comparing the change in an outcome over time for a treated group against the change for a control group over the same period.

In this analysis, the outcome variable is the number of daily ChatGPT visits, the treatment group is Italy, and the control group is Spain. The validity of this approach rests on the parallel trends assumption, which posits that both groups exhibited similar trends prior to the intervention (Angrist & Pischke, 2009). This assumption guided the choice of Spain as the control country, as thoroughly explained in Section 4.3.2.2.

4.3.2.1 Data Source and Study Period

Data on ChatGPT usage was acquired from Semrush, a commercial web analytics platform. Semrush estimates web traffic using a proprietary algorithm that synthesizes clickstream data, SERP data, and aggregated search volumes (Semrush, 2025). While this platform has limitations in measuring absolute traffic, previous research deemed it suitable for detecting directional trends (Jansen et al., 2022). Given that the goal is to assess the relative significance and direction of the ban's effect (not precise visit volumes), Semrush represents an appropriate data source for the DiD analysis.

The analysis uses daily access frequencies for "chat.openai.com" collected from December 1, 2022, to August 25, 2023. This timeframe is partitioned into three periods:

- Pre-treatment (120 days: Dec 1, 2022 – Mar 30, 2023) to establish baseline adoption trajectories.
- Treatment (28 days: Mar 31 – Apr 27, 2023) corresponding to the ban.
- Post-treatment (120 days: Apr 28 – Aug 25, 2023) to measure post-restoration behavior.

This symmetric design enhances the statistical power for interpreting post-ban effects.

4.3.2.2 Control Group Selection

A central assumption for DiD estimation is that the treatment and control groups exhibit similar pre-treatment outcome trajectories, a condition known as the parallel trends assumption (Angrist & Pischke, 2009; Roth, 2023). The selection of Spain as the control country was determined by a comprehensive statistical verification of this assumption. Parallel trends for Italy against each of the 21 potential control countries were tested using pre-treatment period data.

TABLE 4.2: Parallel Trends Test for 21 Candidate Control Countries

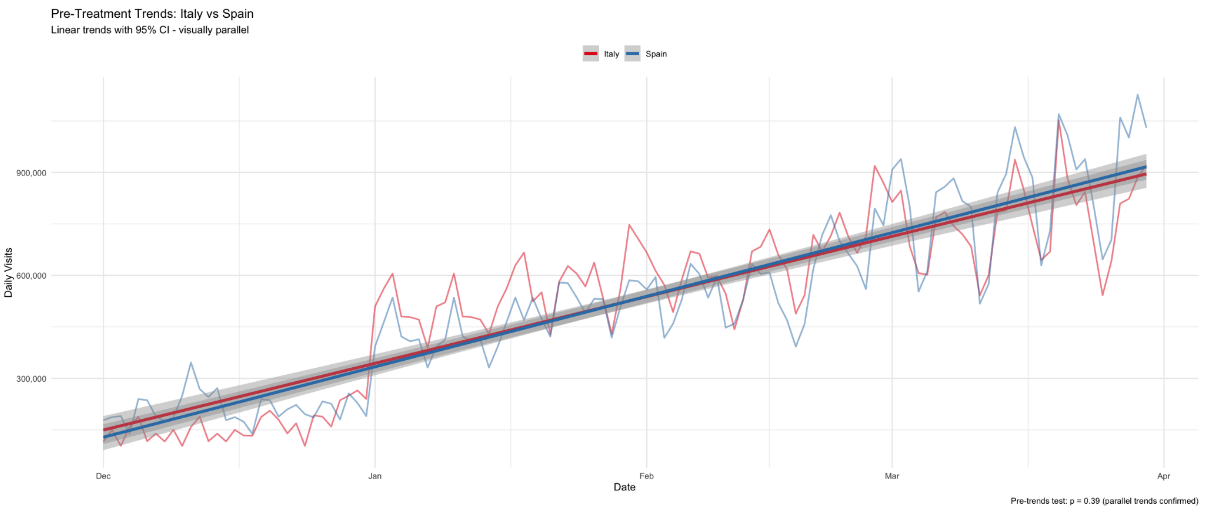
Country	Test Result	P-value (daily)	P-value (weekly)
Spain	Passed both	0.3912	0.6077
Netherlands	Only weekly passed	0.017	0.0744
Poland	Failed Both	0.000	0.0431
France	Failed Both	0.000	0.0005
United Kingdom	Failed Both	0.000	0.000
Germany	Failed Both	0.000	0.000
Belgium	Failed Both	0.000	0.000
Switzerland	Failed Both	0.000	0.000

Czechia	Failed Both	0.000	0.000
Sweden	Failed Both	0.000	0.000
Portugal	Failed Both	0.000	0.000
Denmark	Failed Both	0.000	0.000
Ireland	Failed Both	0.000	0.000
Romania	Failed Both	0.000	0.000
Norway	Failed Both	0.000	0.000
Austria	Failed Both	0.000	0.000
Hungary	Failed Both	0.000	0.000
Greece	Failed Both	0.000	0.000
Finland	Failed Both	0.000	0.000
Slovakia	Failed Both	0.000	0.000
Slovenia	Failed Both	0.000	0.000

TABLE 4.3: Summary Statistics of Daily ChatGPT Visits (Italy vs. Spain) pre-, during, and post-ban

Period	Days	Mean	SD
Pre-treatment	120	Italy: 522,749 Spain 522,670	Italy: 245,555 Spain: 252,824
During ban	28	Italy: 485,712 Spain 1,068,336	Italy: 132,714.1 Spain: 225,452
Post-ban	120	Italy: 786,241 Spain 798,659	Italy: 192,573 Spain: 199,116

FIGURE 4.1: Parallel Trends Assumption Test (Pre-Ban ChatGPT Daily Visits Italy vs. Spain)



The test used the following regression model, estimated on the pre-treatment data (Dec 1, 2022 – Mar 30, 2023):

$$Visits_{it} = \beta_0 + \beta_1 Treatment_i + \beta_2 Time_t + \beta_3 (Treatment_i \; * \; Time_t) + \epsilon_{it}$$

In this model, $Visits_{it}$ is the daily ChatGPT visits for the country i on day t; $Treatment_i$ is an indicator variable (1 for Italy, 0 for the control country), and $Time_t$ represents a linear time trend. The coefficient of interest, β_3, captures the differential trend between Italy and the control candidate. The parallel trends assumption holds if β_3 is not statistically different from zero. This hypothesis was tested using a two-sided t-test with a significance threshold of $\alpha = 0.05$.

These tests were performed at both daily and weekly frequencies to ensure the robustness of the findings and to account for differences between high-frequency noise and lower-frequency trends. As shown in Table 4.2, Spain was the only country that satisfied the parallel trends assumption at both frequencies, failing to reject the null hypothesis of parallel trends at both the daily ($p = 0.39$) and weekly ($p = 0.61$) levels. All other candidates failed the test, exhibiting pre-treatment trends that were statistically divergent from Italy's.

4.3.2.3 Difference-in-Differences Specification

The treatment effects for the during-ban and post-ban periods were estimated using the following DiD specification:

$$Visits_{it} = \alpha + \beta_1 Italy_i + \beta_2 During\_Ban_t + \beta_3 Post\_Ban_t$$

$$+ \delta_1 (Italy_i * During\_Ban_t) + \delta_2 (Italy_i * Post\_Ban_t) + \epsilon_{it}$$

Here, $Visits_{it}$ represents the daily ChatGPT web visits for country i on date t. $Italy_i$ is an indicator variable (1 for Italy, 0 for Spain), while $During\_Ban_t$ and $Post\_Ban_t$ are indicator variables marking the dates of the ban period (March 31 – April 27, 2023) and the post-ban period (April 28 – August 25, 2023), respectively.

The coefficients of interest are the interaction terms. The δ_1 coefficient estimates the immediate effect during the ban, which was expected to be large and negative as access was blocked. The coefficient δ_2 estimates any lasting effect after the ban ended. This coefficient captures whether Italy's post-ban usage trajectory differed from Spain's, holding constant baseline country differences (β_1) and common time effects (β_3). A statistically significant δ_2 value would indicate a persistent change in ChatGPT adoption behavior.

Robustness Checks

The main analysis indicated no significant evidence of lasting behavioral change in ChatGPT usage. To validate the model, two sets of robustness checks were performed: a verification of post-treatment parallel trends and a sensitivity analysis concerning the release of the iPhone app.

First, the presence of a parallel trend after the treatment was formally verified by estimating the following model on the post-ban data:

$$Visits_{it} = \gamma_0 + \gamma_1 Italy_i + \gamma_2 Time\_Post_t + \gamma_3 (Italy_i * Time\_Post_t) + \nu_{it}$$

The null hypothesis $\gamma_3 = 0$ tests whether Italy and Spain maintained parallel post-treatment trends. As detailed in Table 4.4, the analysis yields a p-value of 0.7018, failing to reject the null hypothesis. This provides strong evidence that the two countries followed parallel trajectories after the ban was lifted, consistent with a full recovery to the counterfactual path.

TABLE 4.4: Robustness Check: Post-Treatment Parallel Trends Test (Italy vs. Spain)

Variable	Coefficient	Std_Error	P_value
Intercept (Spain)	1,041,443	25,232	0.0000 (***)
Italy (baseline)	-24,241	35,683	0.4976
Time (days post-ban)	-4,080	366	0.0000 (***)
Italy * Time	199	518	0.7018

Second, the model's robustness was tested against a potential confounding event: the release of the iPhone ChatGPT app on May 30, 2023. Because this study captures only web traffic, differential rates of mobile app adoption could bias the web-based recovery estimates. This threat was assessed by comparing DiD estimates from a "clean" specification (ending the post-treatment period on May 29) with those from the full specification (120 post-ban days). As shown in Table 4.5, both specifications yield identical conclusions, providing evidence that the iPhone app release does not invalidate the comparative analysis.

TABLE 4.5: Robustness Check: DiD Post-Ban Effect Sensitivity to iPhone App Release

Specification	Estimate	Std Error	P-Value
Full Period	-12,497	21,874	0.5711
Up to May 29	-38,625	25,622	0.1437

4.3.3 Stage 3: Qualitative Content Analysis – Identifying Legitimacy Erosion Mechanisms

The third phase of the analysis examined public discourse on the ChatGPT ban using social media data from X (formerly Twitter), a common method in computational social science (Steinert-Threlkeld, 2018; Bolici et al., 2020).

X was selected as the primary data source for one critical reason: it provides explicit textual articulations of supportive and critical claims. This contrasts with the behavioral (DiD) data, which requires researchers to infer underlying attitudes. X data also offers secondary advantages: its real-time structure allows for tracking immediate reactions to regulatory interventions (Burgess & Bruns, 2012), and its user base aggregates a diverse mix of technical experts, policy stakeholders, and the general public (Bolici et al., 2020; Chiarello et al., 2024; Schuster & Kolleck, 2020).

4.3.3.1 Data Source and Processing

Data were collected from the X Academic Research API v2. The query targeted posts in the Italian language containing the keyword "ChatGPT" during the period from March 24 to May 4, 2023, yielding a final dataset of 31,258 posts.

The temporal window was deliberately structured to capture the immediate public reaction to the key regulatory events. The analysis defines three distinct periods: a 7-day pre-suspension baseline (March 24–30), the 28-day suspension period (March 31–April 27) initiated by the GPDP announcement, and a 7-day post-suspension phase (April 28–May 4) following the service's restoration. The focused one-week observation windows before and after the ban allow for establishing a proximate baseline while preserving focus on the public response to the ban.

4.3.3.2 Filtering and Sampling Strategy

The initial dataset of 31,258 posts was systematically filtered in three steps to create the final analytical sample.

First, a keyword-based filter was applied to retain only posts containing terms related to the ban: "Privacy", "Dati personali", "blocc*", "Ban*", "access*", "VPN", "Alternativ*", "Garante", "intelligenza artificiale", "GPDP", "viet*", "italia*", "openai", "disponibile". This step was necessary to distinguish the specific conversation about the regulatory intervention from general, irrelevant discussions about ChatGPT (e.g., "ChatGPT wrote my essay"). This filter yielded 18,243 ban-relevant posts.

Second, retweets were excluded, resulting in 9,567 unique contributions (original posts, replies, and quotes). Retweets are considered content amplification, not content generation; their inclusion would inflate frequency counts without adding new argumentative content (though retweet counts were preserved as metadata).

Finally, a weighted random sample of 2,000 posts (21% of the unique data) was drawn for qualitative coding. The sampling was weighted by each post's "like" count to increase the probability that posts that achieved higher public consensus would be included in the analysis. As shown in Table 4.6, all weeks are reliably represented in the final sample.

TABLE 4.6: X (Twitter) Data Sampling and Filtering Log

Week	**Number of posts in the no-retweet dataset**	**Number of posts in the final sample**	**% of posts in the no-retweet dataset over the total (9,567)**	**% of posts in the final sample over the total (2,000)**
Total	9567	2000	100,00%	100,00%
week 13 (pre-ban)	477	57	4,99%	2,85%
week 14 (ban)	5655	1259	59,11%	62,95%
week 15	1232	287	12,88%	14,35%
week 16	712	172	7,44%	8,60%
week 17	420	100	4,39%	5,00%
week 18 (post-ban)	1071	125	11,19%	6,25%

The 2,000-post sample size was deemed sufficient based on theoretical saturation. After sampling, these 2,000 posts underwent a final refinement step to isolate relevant argumentative content from noise (detailed in Table 4.7). This coding sorted the sample into three groups: 1,034 argumentative posts (which formed the basis for the main analysis), 436 news posts, and 530 off-topic/ambiguous posts.

As a final methodological control, the pre-ban week (week 13) was included in the sample to check for pre-intervention signals or information leaks. No evidence of such signals was found.

TABLE 4.7: Classification of the Sampled Posts in Argumentative, News, and Off-topic/Ambiguous

Post Type	Frequency
Argumentative	1034
News	436
Off-topic/Ambiguous	530
Total	2000

4.3.3.3 Codebook Development and Theoretical Grounding

The codebook for implementing Stage 3 is grounded in the neo-institutional conception of legitimacy introduced in Section 4.2.3 (Scott, 1995; Suchman, 1995). Pragmatic, moral, and cognitive legitimacy are treated as analytical ideal types that can co-occur in the same post, and they provide the high-level categories for organizing how the X public debate evaluated the GPDP’s intervention and its standing in terms of perceived benefits, normative appropriateness, and taken-for-grantedness.

The codebook was developed through an iterative combination of deductive and inductive coding. In a first step, the three legitimacy dimensions were translated into broad parent codes, and initial subcodes were established to capture concrete ways in which users might express support or challenge along each dimension (for example, questioning the competence or fairness of the authority, framing Italy as deviant, or highlighting user harms or benefits). This provisional scheme was applied to a pilot subset of posts; discrepant cases were analyzed and refined the subcodes and decision rules. This refinement process led to the production of the final codebook that guided the coding of the full sample in Stage 3.

4.4 RESULTS

This section presents the empirical findings in three sequential stages, following the logic of the Visibility Trap. Stage 1 (Document Analysis) first reconstructs the GPDP's formal success narrative. Stage 2 (DiD) then provides the first counterpoint: the "invisible failure" of the ban to produce any lasting behavioral change in ChatGPT usage. Finally, Stage 3 (Qualitative Content Analysis) provides the explanatory mechanism for this failure, detailing the comprehensive and multi-dimensional legitimacy collapse that likely eroded the authority’s innovation governance sustainability.

4.4.1 Stage 1: Document Analysis – A Narrative of Formal Effectiveness

The document analysis reveals a strategic, six-phase narrative. The GPDP's intervention began with an urgent, confrontational stance (using terms like "blocca"). This was immediately followed by a rapid diplomatic pivot to negotiation. After securing compliance and declaring victory, the narrative concluded over a year later with a final, punitive judgment. This evolution is synthesized in Table 4.8.

TABLE 4.8: Document Analysis Results: Timeline and GPDP's Framing

Stage (Date)	Description
Phase I: The Urgent Intervention (March 30-31, 2023)	GPDP issues an immediate temporary limitation, citing 4 GDPR violations. Public-facing communication uses strong, absolute language ("blocca," "Stop a ChatGPT").
Phase II: The Diplomatic Pivot (April 4-8, 2023)	De-escalation from confrontation to negotiation. GPDP holds a videoconference with OpenAI's CEO and reframes its stance as "pro-innovation but pro-regulation."
Phase III: The Formal Path to Compliance (April 11-12, 2023)	The GPDP translates negotiations into a formal 9-point ultimatum (e.g., transparency, legal basis, age gate) with a hard deadline of April 30 for compliance.
Phase IV: Framing International Leadership (April 13, 2023)	GPDP strategically frames the creation of a new European (EDPB) task force as a direct consequence of its "first mover" intervention, positioning itself as a leader.
Phase V: The Resolution & Success (April 28, 2023)	GPDP publicly declares victory after OpenAI implements measures to comply with the 9-point ultimatum. ChatGPT access is restored in Italy.
Phase VI: The Final Judgment (December 20, 2024)	Over 1.5 years later, the GPDP concluded its investigation, confirming the initial violations and imposing a €15 million fine, as well as a 6-month information campaign.

4.4.1.1 Phase I: The Urgent Intervention (March 30-31, 2023)

The intervention began on March 30, 2023, when the GPDP ruled an immediate temporary suspension of OpenAI's processing of Italian users' personal data. In response, OpenAI suspended access to ChatGPT for Italian IP addresses as of March 31.

The GPDP's justification was built on four alleged GDPR violations. The first was a recent data breach involving user conversations and payment information. The second was the absence of a legal basis for the "massive collection and storage of personal data" used to train the algorithms. Third, the authority noted data inaccuracy, as the information provided by ChatGPT does not always align with actual data. Finally, the lack of age verification, which the GPDP argued exposed children to absolutely inappropriate responses. Based on this framing, the GPDP ordered the temporary limitation of Italian users' data and gave OpenAI a 20-day deadline to respond, threatening significant fines.

A critical element of this initial framing was the GPDP's deliberate choice of language: while the formal March 30 ruling (GPDP, 2023) used the precise, legalistic term “temporary limitation on the processing of Italian users' data”, the public-facing March 31st press release (GPDP, 2023) used far more potent and absolute language. Its headline was "Intelligenza artificiale: il Garante blocca ChatGPT" (Artificial intelligence: the Garante blocks ChatGPT), and its opening sentence declared "Stop a ChatGPT." This was the only instance in the entire corpus of communications where the GPDP employed this language, potentially contributing to a public framing of the GPDP as an authoritative actor issuing a total ban, which was central to the subsequent public delegitimization (see Section 4.4.3).

4.4.1.2 Phase II: The Diplomatic Pivot (April 4-8, 2023)

Immediately following the March 31st "blocks" framing, the GPDP's communication strategy pivoted from public confrontation to negotiation. This phase was characterized by rapid de-escalation and a strategic reframing of the GPDP as a reasonable, non-ideological negotiating partner.

The pivot began on April 4, 2023, with the GPDP's announcement of an upcoming meeting with OpenAI. The authority publicly endorsed this new diplomatic channel, noting that the meeting follows the letter in which the US company expressed its immediate willingness to collaborate (GPDP, 2023). The GPDP framed this as a positive initiative, signaling a clear shift in tone from punitive to collaborative (GPDP, 2023).

This meeting was confirmed via videoconference on April 6, 2023. The event's gravity was underscored by its participants: the full GPDP board met with senior OpenAI leadership, including CEO Sam Altman (GPDP, 2023e). In its post-meeting communication, the GPDP explicitly countered the anti-innovation narrative its "ban" had fostered, stating it: “emphasized that there is no intention to halt the development of AI and technological innovation and reiterated the importance of respecting the rules protecting the personal data of Italian and European citizens” (GPDP, 2023).

This statement strategically repositioned the GPDP as both pro-innovation and pro-regulation. In turn, OpenAI, “while reiterating its conviction that it respects the rules,” affirmed its cooperation and committed to strengthening transparency and safeguards for minors (GPDP, 2023). OpenAI pledged to send a document outlining these measures by the end of the day (GPDP, 2023).

The diplomatic momentum continued. On April 8, 2023, the GPDP announced that it had convened an "extraordinary session" for the first examination of the two documents that OpenAI had delivered (GPDP, 2023).

4.4.1.3 Phase III: The Formal "Path to Compliance" (April 11-12, 2023)

This phase translated the previous negotiations into a formal ultimatum, fundamentally shifting the GPDP's role from critic to hands-on technical and legal overseer. The authority was no longer just identifying problems; it was now dictating specific product changes and deadlines.

This new stance was codified in the April 11 ruling (GPDP, 2023) and its accompanying press release dated April 12 (GPDP, 2023). These documents established a conditional path to compliance: the GPDP would suspend the temporary limitation order only if OpenAI fulfilled a precise list of demands by a hard deadline of April 30, 2023. The April 11 ruling detailed a 9-point list of mandatory actions, which the press release summarized in five key areas:

1. Transparency (Informativa): OpenAI was ordered to provide a transparent information notice detailing the logic of data processing and the rights of both users and non-users. This notice had to be easily accessible before registration and presented to all existing users upon their first access after reactivation.
2. Legal Basis (Base Giuridica): The GPDP explicitly forbade using "execution of a contract" as the legal basis for processing user data for algorithm training. OpenAI was required to adopt either consent or legitimate interest as the new legal basis, in line with the principle of accountability.
3. Exercise of Rights (Esercizio dei Diritti): OpenAI was mandated to create tools for all data subjects, including non-users, to request the rectification of inaccurate personal data or its erasure if rectification was technically impossible. A separate, accessible tool was required to enable non-users to exercise their right to object to their data being processed for training purposes.
4. Protection of Minors (Tutela dei Minori): This was a two-part demand. First, OpenAI had to immediately implement an "age gate" for all new registrations and for existing

users upon reactivation. Second, it was required to submit a plan by May 31 for a full-scale age verification system to be implemented by September 30, 2023.

5. Information Campaign (Campagna di Informazione): By May 15, OpenAI was ordered to promote an information campaign across radio, television, newspapers, and the web to inform the Italian public about the use of their personal data for training algorithms.

4.4.1.4 Phase IV: Framing International Leadership (April 13, 2023)

A single press release on April 13, 2023, represented a strategic move to frame the GPDP as a European regulatory leader. One day after issuing its compliance ultimatum, the GPDP announced that the European Data Protection Board (EDPB), the collective body of all EU privacy authorities, had launched a task force on ChatGPT (GPDP, 2023).

The GPDP's communication was notable for its explicit causal framing. It presented the EDPB's action not as a parallel development but as a direct consequence of its own intervention. The press release stated: “Following the provision of temporary limitation, adopted by the Garante for the protection of personal data last March 30 against ChatGPT [...] the European privacy Guarantors [...] have decided to launch a task force on ChatGPT.” (GPDP, 2023)

This wording positioned the GPDP as a first mover that had successfully placed the issue on the unified European agenda. The GPDP added that the task force's goal was to "promote cooperation and the exchange of information on possible initiatives" for applying the European Regulation (GPDP, 2023).

4.4.1.5 Phase V: The Resolution and Declaration of Success (April 28, 2023)

This phase, communicated in the April 28 press release (GPDP, 2023f), served as the GPDP's public declaration of success and a vindication of its "path to compliance."

The GPDP announced it had received a note from OpenAI detailing the measures introduced to comply with the Authority's requests. The GPDP's communication framed this as a point-for-point fulfillment of the April 11 demands. Based on this compliance, the GPDP announced it was suspending its temporary limitation order, allowing ChatGPT to be accessible again to Italian users. The authority concluded with explicit approval, recognizing "the steps forward taken to combine technological progress with respect for people's rights" (GPDP, 2023f).

However, the GPDP also noted the investigation was not fully closed, stating it would continue to monitor compliance with the additional, future-facing demands, namely the full age verification system and the public information campaign.

4.4.1.6 Phase VI: The Final Judgment (December 20, 2024)

More than a year and a half after the initial ban, the GPDP's investigation concluded with a final ruling on December 20, 2024, which imposed two significant sanctions (GPDP, 2024).

First, the authority fined OpenAI € 15 million. Second, it ordered OpenAI to carry out a 6-month institutional communication campaign across radio, television, newspapers, and the internet. This campaign was mandated to achieve two goals: (1) promoting public understanding and awareness of ChatGPT's functioning; (2) ensuring both users and non-users are made aware of how to object to their data being used for training.

These sanctions served as the definitive confirmation of the GPDP's initial, urgent assessment from March 2023. In its ruling, the authority stated that it had ascertained that OpenAI had, in fact, committed the violations it had originally contested.

4.4.1.7 Conclusion: The "Formal Win" as the Visibility Trap

By its own internal logic, the GPDP's strategy was successful. The regulator identified GDPR violations (GPDP, 2023g, 2023h), established a diplomatic channel (GPDP, 2023e), issued formal demands (GPDP, 2023i), ensured those demands were met (GPDP, 2023f), and finalized the process with a punitive sanction (GPDP, 2024). This constitutes a clear formal victory: the regulator has successfully compelled a global tech giant to modify its product and practices to comply with European law.

Yet, this victory exemplifies the Visibility Trap: a strategic failure where the pursuit of visible, procedural successes masks or even causes a latent loss of institutional legitimacy. The intervention, therefore, degraded the very public authority necessary to ensure its future effectiveness in managing long-term AI governance challenges. Sections 4.4.2 and 4.4.3 detail these latent negative effects.

4.4.2 Stage 2: DiD Estimation – The Ban Had No Lasting Effect on ChatGPT Usage Behavior

The temporary Italian ban on ChatGPT had no significant, lasting effects on usage behaviors. While the restriction was effective in blocking access, this impact was entirely temporary. Once access was restored, Italian usage patterns recovered, returning to a trajectory statistically indistinguishable from the counterfactual path that would have been followed had the intervention not occurred. The DiD estimates (Table 4.9) confirm this conclusion, distinguishing the massive immediate effect from the negligible lasting one.

Figure 4.2 shows ChatGPT Daily Visits in Italy vs. Spain. It is notable that daily visits during the ban did not drop to zero, a finding consistent with the widespread behavioral circumvention (e.g., via VPNs) documented in other studies (Bolici et al., 2024; Kreitmeir & Raschky, 2023). This figure plausibly reflects panel-based tracking that successfully attributes visits to Italian users at the device level, regardless of their masked IP. However, the precise reason this traffic is captured by Semrush cannot be definitively verified, as its algorithm is proprietary. Given this methodological uncertainty, this study adopts a conservative approach. It acknowledges this data as possible evidence of notable circumvention, but this study's primary inferential focus remains on the lasting post-ban effect, which is unambiguously measured by the δ2 coefficient.

FIGURE 4.2: Historical Evolution of ChatGPT Daily Visits in Italy (Red) vs. Spain (Blue)

The primary coefficient of interest is the *treatment:post-ban interaction* (δ_2), which tests whether Italy's usage trajectory differed from Spain's after access was restored. This coefficient (δ_2 = -12,497, SE = 21,873.5) is statistically indistinguishable from zero (p = 0.571).

TABLE 4.9: Main Difference-in-Differences (DiD) Estimation Results

	Estimate	**Std. Error**	**t value**	**P-value**
(Intercept)	+ 522,670.08	55,622.8	9.397	0.000 (***)
treatment	+ 79.09	17,659.2	0.004	0.996
during ban	+ 545,665.38	76,891.4	7.097	0.000 (***)
post-ban	+ 275,988.76	66,827.4	4.130	0.000 (***)
treatment:during ban	- 582,702.27	76,957.8	-7.572	0.000 (***)
treatment:post-ban	- 12,497.36	21,873.5	-0.571	0.571
RMSE: 219,228.5 Adj. R2: 0.3494 Nr. Observations: 536				

Beyond its statistical insignificance, the magnitude of the point estimate is economically trivial. The estimated shortfall of 12,497 visits represents only 1.6% of the mean post-ban average daily accesses (786,241), rendering the difference substantively negligible. For context, this tiny estimated effect is dwarfed by a factor of 17.5 by the model's day-to-day random variation (RMSE = 219,229), suggesting it is simply noise.

These results contrast sharply with the during-ban effect (δ_1 = -582,702, SE = 76,957.8, $p < 0.001$), which shows massive, highly significant disruption while the ban was active. This contrast demonstrates that the ban's impact was confined strictly to the period of restriction.

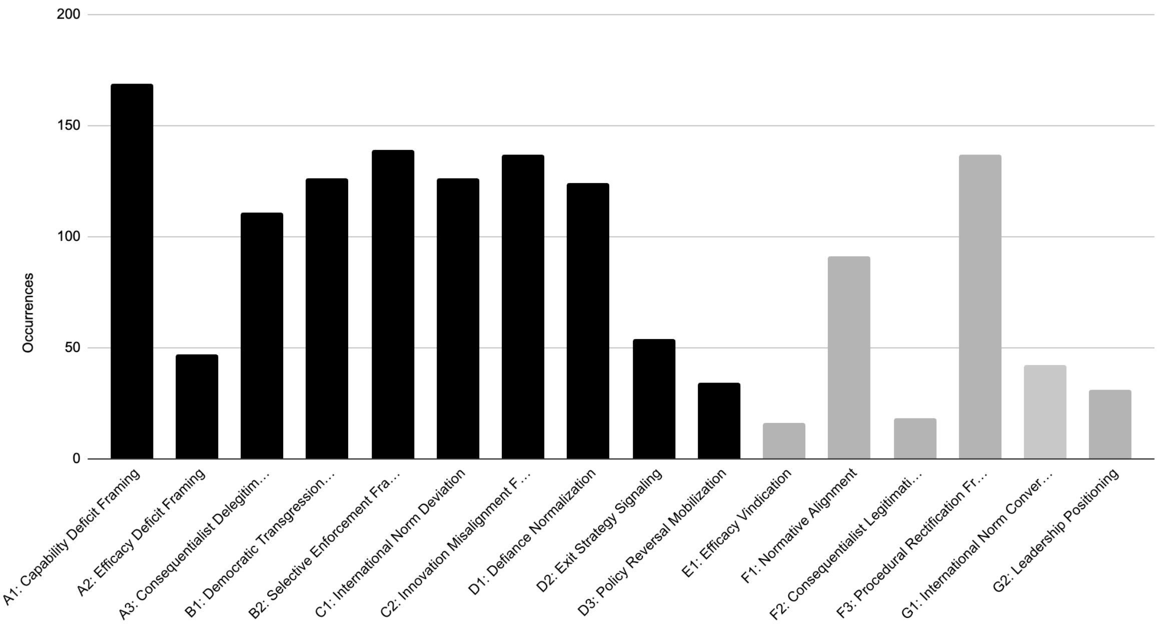
This finding of a full trajectory-level recovery is confirmed by a robustness test (Table 4.4), which regressed daily visits on an *Italy * Time* interaction term over the 120-day post-ban period to test for differential growth rates. This interaction coefficient was also statistically indistinguishable from zero (p = 0.7018). This test fails to reject the null hypothesis of parallel trends, providing strong evidence that Italy's post-ban growth rate did not diverge from Spain's.

The combination of a non-significant main effect (Table 4.9) and this non-significant post-treatment trend (Table 4.4) provides mutually reinforcing evidence: Italy's post-ban usage behavior is statistically indistinguishable from its counterfactual, both in absolute levels.

4.4.3 Stage 3: Qualitative Content Analysis – Delegitimating the Authority

The public debate on X largely consisted of an overwhelming and comprehensive delegitimation of the GPDP. The content analysis identified 10 distinct delegitimating frames, which were not only more diverse than the six legitimating ones but also vastly more frequent (Figure 4.3).

FIGURE 4.3: Frequency of Delegitimating (in Black) and Legitimating (in Grey) Mechanisms. (*Note: the sum of occurrences exceeds the number of argumentative posts, as each post could be coded into more than one category).*



As detailed in Table 4.10, these attacks spanned all three of Suchman's (1995) legitimacy types: pragmatic, moral, and cognitive. The most prominent frames challenged the GPDP's fundamental "Capability Deficit" (A1, 169 occurrences), its "Selective Enforcement" (B2, 139 occurrences), and its perceived "Innovation Misalignment" (C2, 137 occurrences). In contrast, the most frequent legitimating argument, defending the "Procedural Rectification" (F3, 137), was primarily a defensive reaction.

TABLE 4.10: Delegitimating and Legitimating Mechanisms Identified in the Qualitative Content Analysis of the Public Discourse on X (Grounded in Suchman, 1995)

Legitimacy Type(s)	Category	Description	Nr.
DELEGITIMATING			
Pragmatic Dispositional	A1: Capability Deficit Framing	**Challenges regulator competence**	169

		Criticizes GPDP for a perceived lack of technical expertise, directly challenging their perceived trustworthiness, character, and competence	
Pragmatic Exchange & Moral Consequential	A2: Efficacy Deficit Framing	**Challenges policy effectiveness** Claims the intervention fails to achieve its objectives due to design flaws or structural impossibility (e.g., "APIs still accessible," "can't regulate global platforms with national law"). Invokes both pragmatic concerns (the policy fails to deliver promised benefits to stakeholders) and moral concerns (the policy fails to serve the collective welfare).	47
Pragmatic Exchange & Moral Consequential	A3: Consequentialist Delegitimation	**Documents policy harms** Documents measurable negative consequences of the ban – competitiveness losses for Italian businesses, innovation obstacles for professionals and students, and economic harms to specific stakeholder groups. Frames outcomes as both damaging to constituent interests (pragmatic exchange) and harmful to Italy's broader societal welfare (moral consequential)	111
Moral Procedural & Cognitive Taken-For-Grantedness	B1: Democratic Transgression Framing	**Challenges normative appropriateness** Frames the intervention as violating democratic norms through authoritarian overreach, bad faith, deception, or as part of a broader pattern of rights restrictions. Invokes moral procedural legitimacy (democratic process norms are violated) and cognitive legitimacy (positions Italy as deviating from what is taken for granted about democratic societies, making the action seem abnormal for a democracy).	126
Moral Procedural	B2: Selective Enforcement Framing	**Challenges enforcement consistency** Challenges GPDP's enforcement as inconsistent, discriminatory, or unfairly applied through explicit comparison to other cases (e.g., "blocks ChatGPT but ignores TikTok/Facebook violations"). Violates moral norms of procedural fairness, equal treatment under law, and rule consistency – core procedural justice values.	139
Cognitive Taken-For-Grantedness & Moral Structural	C1: International Norm Deviation	**Challenges normalcy through spatial comparison** Positions Italy as uniquely isolated or deviant from international peers through explicit spatial comparison (e.g., "only Italy banned ChatGPT," "we're alone in Europe"). Primarily invokes cognitive legitimacy by challenging taken-for-grantedness – highlighting Italy's anomalous position makes the ban appear unnatural and disrupts assumptions about appropriate regulatory responses. Implicitly invokes moral structural legitimacy by placing Italy in the morally disfavored taxonomic categories of "outlier," "isolated nation," or "deviant" – categories that carry negative valence when contrasted with "peer nations" or "normal democracies."	126
Cognitive Comprehensibility & Moral Structural	C2: Innovation Misalignment Framing	**Challenges temporal alignment with progress** Claims Italy/GPDP is temporally misaligned with the taken-for-granted trajectory of technological progress – either through active opposition to innovation or failure to comprehend modern technology. Uses categorization language ("Luddite," "medieval," "backward," "technophobe") that simultaneously: (1) challenges cognitive legitimacy by positioning GPDP as incomprehensible/out-of-sync with progress norms, and (2) challenges moral structural legitimacy by placing GPDP in a morally deprecated, low-status taxonomic category.	137
Behavioral	D1: Defiance Normalization	**Normalizes circumvention** Frames circumvention behaviors (VPN use, workarounds) as normal, acceptable, expected, or collectively legitimate responses. Celebrates user agency and resistance. Represents the behavioral translation of delegitimation – how legitimacy challenges convert into non-compliance actions.	124

Behavioral	D2: Exit Strategy Signaling	**Promotes alternatives** Promotes migration to alternative AI services as a solution to the ban (e.g., "use Bing Chat/Claude/GPT4Free instead"). Behavioral exit mechanism illustrating how delegitimation leads to platform switching rather than policy compliance.	54
Behavioral	D3: Policy Reversal Mobilization	**Mobilizes collective action** Mobilizes organized collective action to overturn the ban through institutional democratic channels, including petitions, formal appeals, and campaigns. Behavioral voice mechanism representing attempts to reverse the decision through democratic participation rather than individual circumvention.	34
LEGITIMATING			
Pragmatic Dispositional & Moral Consequential	E1: Efficacy Vindication	**Affirms intervention efficacy** Highlights that the intervention successfully achieved its objectives as evidenced by OpenAI's policy changes and adherence to GPDP requirements. Invokes both pragmatic dispositional legitimacy (GPDP, demonstrated competence, regulatory power, and capability, as evidenced by successfully pressuring a major global technology company to comply) and moral consequential legitimacy (the policy served societal welfare by protecting collective privacy). In public discourse, competence affirmation is demonstrated through achievement rather than being directly asserted.	16
Moral Consequential	F1: Normative Alignment	**Supports the moral necessity of the intervention** Makes abstract arguments establishing the moral importance of privacy, rights, democratic accountability, or ethical duties (e.g., "privacy is a fundamental right," "technology cannot be above the law," "we have a duty to protect citizens"). Invokes moral consequential legitimacy by establishing which values and societal outcomes should be prioritized	91
Moral Consequential	F2: Consequentialist Legitimation	**Identifies beneficial consequences** Reframes the ban as a catalyst for broader, positive societal outcomes beyond its immediate objectives – including improved AI transparency, heightened public awareness of AI risks, European coordination on AI governance, and innovation opportunities in responsible AI development. Invokes moral consequential legitimacy by highlighting beneficial collective outcomes that promote societal welfare.	18
Moral Procedural & Pragmatic Dispositional	F3: Procedural Rectification Framing	**Defends action procedural correctness** Defends GPDP's specific action as procedurally correct, legally authorized, proportionate to the violation, appropriately timed, or properly justified under GDPR. Also corrects factual misrepresentations about what GPDP actually did or was required to do. Invokes moral procedural legitimacy (proper processes and techniques were followed, meeting the standards of procedural justice) and pragmatic dispositional legitimacy (corrections reinforce the GPDP's trustworthiness, competence, and character – demonstrating that they act with sound judgment and "have citizens' best interests at heart").	137
Cognitive Taken-For-Grantedness & Moral Structural	G1: International Norm Convergence	**Affirms normalcy through spatial comparison** Demonstrates that Italy's action aligns with international peers taking similar regulatory approaches to AI governance (e.g., "Germany is also investigating," "EU-wide concerns"). Primarily invokes cognitive legitimacy by affirming taken-for-grantedness – positioning the ban as normal, expected, and part of a broader international pattern makes the action comprehensible within prevailing global regulatory norms. Implicitly invokes moral structural legitimacy by placing Italy in the morally favored taxonomic categories of "peer nation," "responsible democracy," or "aligned with the international community" – categories that carry positive valence and prestige.	42

Moral Structural & Cognitive Comprehensibility	G2: Leadership Positioning	**Reframes deviation as leadership** Reframes Italy's action as pioneering, standard-setting, or proactive leadership in necessary AI regulation (e.g., "Italy is ahead of the curve," "setting the standard," "first mover on AI accountability"). Invokes moral structural legitimacy by placing GPDP/Italy in a morally favored, high-status taxonomic category ("leader," "pioneer," "innovator") – categories that carry prestige and positive valence. Simultaneously invokes cognitive comprehensibility by providing an interpretive frame that makes the action understandable as visionary leadership rather than an incomprehensible deviation from norms	31

4.4.3.1 Structural Composition of the Public Debate

The analysis of 1,034 argumentative X posts reveals that the public response was not a balanced debate, but rather an overwhelmingly negative reaction. As shown in Table 4.11, posts delegitimating the intervention outnumbered those legitimating them by nearly three to one (72% vs. 25%). The discourse was also highly polarized; the near-total absence of "mixed" or nuanced posts (3%) indicates the public quickly adopted binary, oppositional stances.

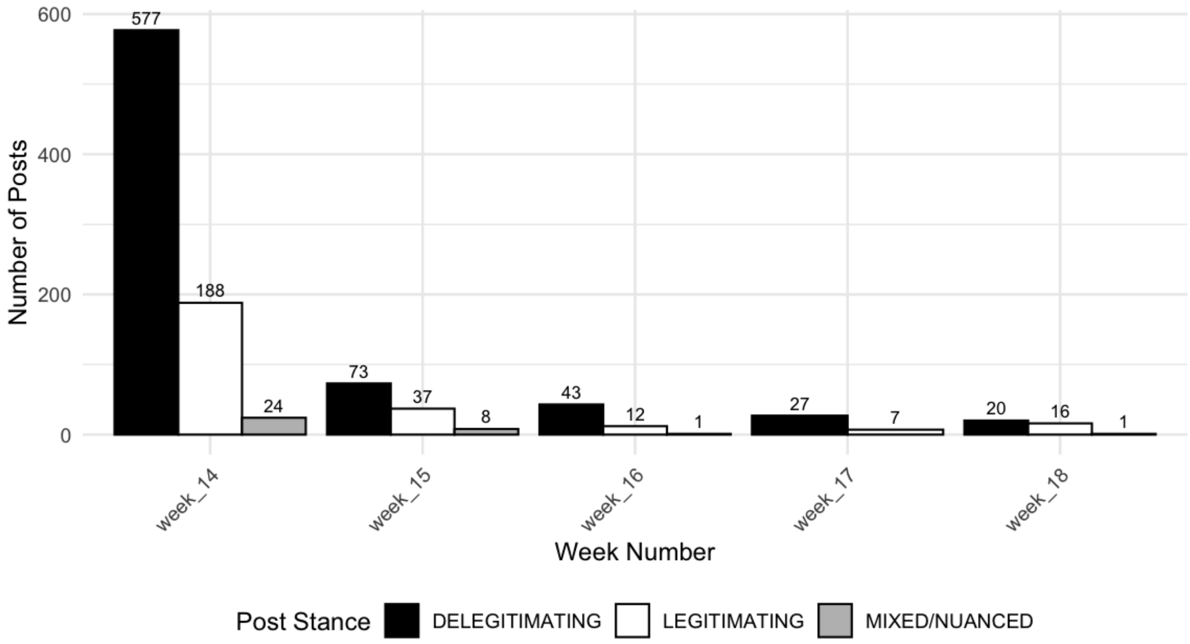
TABLE 4.11: Composition of Argumentative X (Twitter) Posts (Delegitimating vs. Legitimating)

Post Stance	Absolute Frequency	Relative Frequency
Delegitimating	740	72%
Legitimating	260	25%
Mixed	34	3%
Total	1,034	100%

This wave of delegitimization was both vast and front-loaded, characterized by a brief, intense outburst of negativity that peaked and then subsided. Specifically, 76.3% (789 posts) of all argumentative discourse within the 1,034-posts sample occurred in Week 14, the week of the initial ban (Figure 4.4).

This temporal pattern suggests why the GPDP's "visibility" failed to translate into persuasion or behavioral change. The intervention created a surge of attention, but that attention was immediately captured by a hostile narrative. This demonstrates that visibility without discursive legitimacy can amplify, rather than mitigate, public resistance.

FIGURE 4.4: Temporal Evolution of Argumentative X (Twitter) Posts



4.4.3.2 In-depth Analysis of the Delegitimating and Legitimating Discourse

A review of the argumentative frames (Table 4.10) demonstrates a structural asymmetry: the delegitimating discourse was distributed, while the legitimating discourse was more centralized. In particular, the delegitimating attack was distributed across 10 distinct frames. The top three delegitimating categories (A1, B2, C2) were all of roughly equal strength, with no single frame accounting for more than 12% of the total coded mechanisms. This demonstrates that critics deployed a multi-pronged strategy that eroded the GPDP's legitimacy on all fronts simultaneously: pragmatic (it's incompetent), moral (it's unfair), and cognitive (it's deviant). If one line of critique failed, numerous others were available.

In contrast, the legitimating defense was fragile. It was overwhelmingly concentrated on two primary moral arguments: F3, "Procedural Rectification" (137 instances) and F1: “Normative Alignment” (91 instances). This centralization meant that if the audience was unpersuaded by these defenses, no other significant legitimating argument had been advanced to affirm the GPDP's authority.

The next sub-sections discuss each micro-category in detail, exposing the main delegitimating and legitimating patterns within each, accompanied by illustrative post citations translated into English.

A1: Capability Deficit Framing – Attacking Dispositional Legitimacy

The most frequent delegitimating mechanism, A1, directly challenged the GPDP's dispositional legitimacy by portraying the regulator as fundamentally incompetent. This "capability deficit" frame was articulated in three ways: technical ignorance, procedural carelessness, and communication contradictions.

Technical Ignorance: critics asserted the GPDP lacked a basic understanding of ChatGPT's functionality. For instance, one user argued (Post ID: 1459) "*I'd just like to see OpenAI ask Guido to explain how ChatGPT works and watch the nonsense the Authority comes up with, invalidating any criticism they've made since they lack the knowledge to be objective*". This narrative extended to the GPDP's grasp of foundational AI concepts, with another post asserting that: "*The Garante intervention contains questionable points and other non untrue. (...) The 'P' in GPT stands for 'Pre-trained' -- training doesn't happen during conversation"* (Post ID: 320).

Carelessness Approach: The attack on dispositional legitimacy was reinforced by claims that the ban was a decision made carelessly. Public discourse challenged the GPDP's judgment: "*Everything was done too quickly, without even meeting, followed by a flood of interviews"* (Post ID: 33). This perceived haste violated public expectations of methodical processes with one post framing it as a fundamental procedural failure: "*I'm told that the measure taken by the Garante regarding ChatGPT was signed not by the board, but solely by the president; issued not at the conclusion of a proper investigation; but with a demand for immediate action under threat of heavy sanctions. If that's the case, calling it irregular would be an understatement"* (Post ID: 218).

Communication Contradictions: finally, A1 undermined the GPDP's trustworthiness by highlighting its own official statements. Critics repeatedly cited the announcement title "*The Authority blocks ChatGPT*" as evidence of either intentional deception or confusion: *"The authority communication also in English: 'The Authority blocks ChatGPT.' In the country of legal hairsplitters, good luck saying 'no but, I meant blocks in the sense you can continue freely'. Make peace with your brain"* (Post ID: 5). This directly framed the GPDP as either incompetent (unable to communicate clearly) or dishonest (intentionally misleading the public).

A2: Efficacy Deficit Framing – Arguing the Ban is Performative

The "Efficacy Deficit Framing" mechanism (A2) challenged the ban's practical effectiveness, attacking its pragmatic and moral legitimacy. Critics argued the ban was structurally incapable of achieving its stated objectives, highlighting two technical loopholes that rendered the policy merely performative.

Web Interface Blocked; API Access Unrestricted: first, critics exposed the ban as only partially effective, as it blocked the public web interface but left API access unrestricted. This technical loophole was framed as penalizing casual users while failing to stop sophisticated ones, rendering the policy arbitrary and ineffective: "*bro, but like, did you know you can still use it through the API? And that there are Discord servers with the ChatGPT bot ? [...] I PAY for it and I can't even cancel the subscription unless I use a VPN"* (Post ID: 7). This selective impact framed the policy as ineffective.

VPN Circumvention is Trivial: Second, critics argued that the ban was futile because it was trivially bypassed using VPNs. This line of reasoning positioned the GPDP's action as performative bureaucracy, not a serious intervention: *"Responding to AI with bureaucracy is like fighting windmills with a spear. It's pointless. And not just because, as mentioned, a VPN is enough to bypass the ban. But because it certainly won't be the Italian Data Protection Authority that stops this revolution. The process needs to be guided and studied, opening a serious debate on the benefits and risks of an intelligence already capable of writing university theses. Bans won't be enough: bureaucratic little rules are bound to collapse. They're bypassed with a click"* (Post ID: 265).

A3: Consequentialist Delegitimation – Documenting Tangible Harms

The third mechanism, A3: Consequentialist Delegitimation (111 instances), focused on the measurable negative consequences of the ban. This line of argument attacked both pragmatic exchange legitimacy (harm to stakeholder interests) and moral consequential legitimacy (harm to societal welfare) by providing concrete evidence of policy damage. This mechanism operated through two primary arguments.

Work Productivity and Competitiveness: first, critics described immediate operational damage through personal testimonials: "*Yesterday I worked 4 hours with ChatGPT, tomorrow the same work will take me 8 hours"* (Post ID: 326). This argument was scaled from the individual to the national level, focusing on economic competitiveness losses. Critics framed the ban as national self-sabotage: *"Continuing without access to ChatGPT gives significant advantage to international competition*" (Post ID: 782), and *"This is a work tool that many young people, companies, and start-ups were using and I hope they'll be able to use it again soon, otherwise Italy will fall behind the rest of the European countries"* (Post ID: 268).

Perverse Consequences for Data Privacy: This argument created a bitter irony by claiming the ban, intended to protect privacy, paradoxically increased risks by compelling users to adopt insecure VPNs. For example, one user (Post ID: 203) noted: *"#GPDP_IT forces us to use VPNs which, fatefully, expose us to far more serious security and privacy risks"*. Another elaborated: *"On a thousand non-tech channels (like gaming ones), there's a proliferation of 'guide: use ChatGPT with a free VPN.' I'm asking the @GPDP_IT what they think about this side effect (totally unpredictable?) with serious implications for privacy and security* (referenced in Post ID: 162). This narrative was particularly damaging: it argued the policy not only failed to protect privacy (an efficacy deficit) but actively endangered it (a consequential harm), creating an outcome directly opposite to its stated objective.

B1: Democratic Transgression Framing – A Violation of Democracy

The B1: Democratic Transgression framing (126 instances) positioned the intervention as a violation of fundamental democratic norms. This mechanism attacked both moral procedural legitimacy (violating democratic processes) and cognitive taken-for-grantedness (portraying Italy as deviating from behavior considered normal for a democracy, thus making the action incomprehensible). Three connected narratives were developed in this category.

Italy as an Authoritarian Regime: first, the dominant argument compared Italy to authoritarian regimes. Critics explicitly listed other countries that had blocked the technology: *"Besides Italy, the only countries in the world that have blocked ChatGPT are Russia, China, North Korea, Iran, Syria, and Cuba. Even in this, the worst authoritarian drift on the right is absolutely clear. A creeping fascism that advances and slowly tries to strangle the country"* (Post ID: 86). This argument was intensified by invoking Italy's scientific heritage to create cognitive dissonance: *"It's not acceptable that in Italy, homeland of Galileo, Marconi and Olivetti, one must consider using a VPN to bypass a block like happens in China and countries without freedom"* (Post ID: 143). This line of reasoning parallels C1: International Norm Deviation, which placed Italy in the same cognitive category as authoritarian regimes.

The Ban as an Act of Censorship: some arguments focused on individual autonomy: *"Let me decide about my personal data"* (Post ID: 123), while others constructed a broader pattern of systematic restriction by "list-making," connecting the ban to unrelated conservative policies *"So: ban on synthetic meat, ban on ChatGPT, fines for those who "overuse" foreign words. And maybe even a ban on smoking outdoors. And this is the right-wing party for freedom?"* (Post ID: 1261). This accumulation framed the government as fundamentally opposed to innovation and freedom, a position seen as incomprehensible for a modern European democracy. This is often paired with a frame of anti-innovation, discussed in C2: Innovation Misalignment.

Accusations of Fascism: posts explicitly labeled the ban as "fascist," positioning the GPDP and government within the most morally deprecated category of Italian political discourse: *"The prohibition on using ChatGPT in Italy is something terrifying for what it represents. You can bypass the fascist prohibition by connecting with a VPN"* (Post ID: 196). This served as a moral-structural delegitimation, casting the ban not as a policy dispute but as a fundamental democratic violation.

B2: Selective Enforcement Framing – Exposing Procedural Hypocrisy

The B2: Selective Enforcement framing (139 instances) challenged the GPDP's moral procedural legitimacy. This mechanism operated by demonstrating that the regulator applied privacy enforcement inconsistently and discriminatorily. Critics employed explicit comparisons to expose what they framed as arbitrary, selective action that violated core procedural justice values, such as equal treatment and consistent rule application. This argument was advanced through three primary lines of comparison.

Comparison to Unenforced Daily Violations: first, critics compared the ChatGPT ban with the perceived lack of enforcement against more tangible, daily privacy violations, such as unsolicited marketing calls. This positioned the GPDP as ignoring familiar harms while attacking an abstract one. For example, one user articulated this perceived injustice as: *"Dear #GarantePrivacy, instead of blocking ChatGPT could you do something about the call centers that call me 10 times a day for energy, gas and fiber contracts?"* (Post ID: 393).

Comparison to Other Unregulated Platforms: second, critics challenged enforcement consistency by identifying other technology services with comparable or perceivedly worse privacy practices that escaped restriction: *"#ChatGPT is being shut down in Italy because there isn't a technically adequate way to verify users' age. As if there were one for TikTok or Instagram. Or for porn sites. The regulator is a 77-year-old fossil who has no idea what he's doing to the country"* (Post ID: 727). These comparisons directly framed the GPDP's action as arbitrary.

Comparison to Other AI Services: The previous reasoning was extended to other AI models. Critics argued that if the GPDP's justification (related to training data) were applied consistently, other services should also be banned: *"On the points contested with OpenAI, the* *first - on data used for training - opens banning every algorithm-based service. For consistency they must also block Bard, Bing Chat etc.* (Post ID: 395).

C1: International Norm Deviation – "We're the Only Western Country Doing This"

The C1: International Norm Deviation mechanism (126 instances) was the primary method for spatial displacement, challenging cognitive legitimacy. It positioned Italy as deviant from its peer democracies, arguing it belonged to the wrong taxonomic category: authoritarian regimes. This framing, strongly linked to B1, violated taken-for-grantedness by making the ban cognitively incoherent: democracies do not ban such platforms, but dictatorships do.

Taxonomic Relegation: paralleling B1, the first strategy involved relegating Italy to the category of non-democratic countries. Public discourse explicitly and consistently enumerated the countries that had banned ChatGPT, placing Italy in this group *"Countries in the world where ChatGPT is banned: Russia, China, Iran, North Korea, Italy"* (Post ID: 96).

Peer Isolation: the second strategy documented that no other European or Western democracy followed Italy's example:*"This week we update the list of countries that followed the Italian authority's example (...): - Italy. See you next week!"* (Post ID: 100) and *"It's been a week since the measure. Can we say, finally, that no other EU country followed Italy's line?"* (Post ID: 945), or*"No European country banned ChatGPT and I don't think any will. Only Italy. There must be a reason"* (Post ID: 697).

C2: Innovation Misalignment – Medieval Obscurantism in the Modern Era

The C2: Innovation Misalignment mechanism (137 instances) challenged cognitive legitimacy by framing the ban as a temporal regression, which was misaligned with the expectation that modern societies would embrace innovation. This argument violated taken-for-grantedness by positioning Italy outside of modern time, rendering the action incomprehensible as contemporary behavior.

Medievalism and Luddism: This temporal displacement was primarily articulated through medieval or backward-looking framings. Critics explicitly invoked this regression – e.g., *"Tell me we're returning to the Middle Ages in Italy without telling me we're returning to the Middle Ages"* (Post ID: 1232) – or lamented a *"fear of the new"* unworthy of Italy's scientific heritage (Post ID: 290). This theme was frequently condensed into direct accusations of *"obscurantist Luddism"* (Post ID: 248), labeling the action *"Italian madness"*.

Systemic Anti-Innovation: This mechanism also framed the ban as part of a broader, systemic opposition to progress. Critics used rhetorical questions to highlight the incomprehensibility of this stance *"Of course, certain technologies need to be regulated, but is it possible that Italy's answer to every innovation is "ban it"? Can we keep driving away all investment in research?"* (Post ID: 282). To substantiate this, users engaged in "list-making," embedding the ban within a perceived pattern of anti-innovation governance: "*We are governed by old bureaucrats who are incompetent and afraid of any innovation or progress. They go after #SPID, #ChatGPT, lab-grown meat, and so on… If we don't act, Italy is a country headed for economic, social, cultural, technological, and scientific decline"* (Post ID: 233).

D1: Defiance Normalization – "Just Use a VPN, It's Easy"

The D1: Defiance Normalization mechanism (124 instances) operated by normalizing the technical circumvention of the ban, framing it as routine, acceptable, and trivially easy. By sharing tutorials and celebrating successful access, this mechanism undermined regulatory authority through visible, mass non-compliance.

Tutorials: Defiance normalization was driven by tutorials on ban circumvention through VPN use and claims of easy accessibility. Examples include users sharing direct instructions, such as: *"Here's how you can use #ChatGPT in Italy with a #VPN"* (Post ID: 841), and emphasizing the simplicity of the process: e.g., *"ChatGPT, here's how to continue connecting from Italy: PS: it's very easy and very fast"* (Post ID: 156).

Generational Divide: A generational framing was adopted, positioning the GPDP as out of touch with digital natives' capabilities. This argument implied that the ban was naive, as circumvention was a basic skill for the young: *"Mini tutorial on how to enable a VPN and use chatgpt. It's just one way, probably the fastest. 99% of those under 20 already know this"* (Post ID: 177). Another user shared a similar anecdote: *"Me to my son (16 years) this morning: 'Seen they blocked ChatGPT?' Him: 'yes, bypassed instantly with free VPN'"* (Post ID: 16).

Collective Defiance: Finally, this behavior was framed as a collective defiance, not an isolated violation. Users highlighted mass non-compliance, noting: *"By now several days have passed since the Authority's measure on ChatGPT. Result? VPN use in Italy is at record levels, we're following the same path as those who want to obtain information in authoritarian regimes"* (Post ID: 97). Notably, in association with these, commercial VPN promotion emerged (though this does not provide causal evidence between the ban and these commercial discounts), with users noted sarcastically: *"You couldn't make it up. Coincidence? #mystery #authority #vpn #chatgpt"* quoting a famous VPN provider discount offer (Post ID: 463).

D2: Exit Strategy Signaling – "There Are Plenty of Alternatives Anyway"

The D2: Exit Strategy Signaling mechanism (54 instances) presented numerous alternatives to the blocked ChatGPT interface, thereby rendering the ban largely irrelevant. This mechanism undermined regulatory effectiveness by demonstrating that functionally equivalent AI services remained accessible. This framing exposed the ban's limited scope: the GPDP had only blocked one interface, leaving the underlying technology and its many alternatives untouched.

Promotion of API-Based Wrappers and Direct Access: as only the web interface of ChatGPT was blocked, not the underlying API, users promoted API wrappers, such as the ironic nationalist "PizzaGPT" which users noted: *"To allow us Italians to use ChatGPT, PizzaGPT was created, which unlike the first doesn't acquire personal data"* (Post ID: 137). Others shared more technical guides for direct API access: *"ChatGPT was 'hacked': free alternative GPT4Free appears on GitHub"* (Post ID: 802), or *"OpenAI blocked access to chatGPT in Italy at Privacy Authority's request. But we can access anyway thanks to @************'s work. No VPN necessary, just need free GitHub account, your OpenAI API key and 10 minutes"* (Post ID: 145 – username anonymized for privacy reasons).

Promotion of Corporate Substitutes: Concurrently, users pointed to corporate alternatives, primarily Microsoft's Bing Chat, although sometimes framed as an inferior option: *"Chatgpt isn't reachable in Italy (shame). Options: 1) use a vpn 2) hold your nose and use Bing Chat on Edge"* (Post ID: 205).

Promotion of Local Models: A third line of argument focused on promoting self-hosted models. Specifically, users shared guides on: *"how to install a personal AI similar to ChatGPT on computer and run it without Internet and without VPN"* (Post ID: 90).

D3: Policy Reversal Mobilization – "Sign the Petition to Restore ChatGPT"

The D3: Policy Reversal Mobilization mechanism (34 instances) aimed to organize collective action, including petitions and open letters, to demand the policy's reversal. Unlike defiance-based mechanisms (D1, D2), D3 did not focus on circumvention; rather, it directly contested the legitimacy of the rulemaker's decision by applying coordinated democratic pressure.

Proliferation of Petitions: This mobilization was primarily evident in the proliferation of petitions. Users circulated identical, coordinated messages, such as “Restore ChatGPT and update AI norms - Sign the petition!” (Post ID: 1671). The credibility of this campaign was simultaneously constructed through elite signaling, which highlighted that “#ChatGPT, the appeal promoted by @Marco********* and by entrepreneurs, investors and academics” (Post ID: 925). This framing positioned the movement as a concerning effort by economic and intellectual stakeholders, distinguishing it from an anonymous online protest.

E1: Efficacy Vindication – Documenting Policy Success

The E1: Efficacy Vindication mechanism (16 instances) served as the direct legitimating counterpart to the "Efficacy Deficit" (A2) framing. While A2 argued the ban was structurally incapable of achieving its objectives, E1 vindicated the GPDP's effectiveness by documenting that OpenAI complied with its demands. This argument invoked both pragmatic exchange legitimacy (the Authority delivered promised benefits) and moral consequential legitimacy (the intervention achieved positive collective outcomes).

Compliance Documentation and New User Rights: The empirical foundation for this argument was compliance documentation. Supporters highlighted that OpenAI had implemented substantive, not superficial, changes. One post (Post ID: 1148) noted “*OpenAI committed to strengthening transparency in the use of individuals' personal data, the existing mechanisms for exercising rights, and safeguards for minors. Do you think the way it was set up was already fine?*” Another user (Post ID: 1957) provided concrete evidence of new user rights “*Thanks to the Italian Authority's intervention, ChatGPT is obligated to protect and respect us more. If in generating texts it writes we're criminals and it's not true, now we have the right to obtain rectification”*. The specificity of these new rights demonstrated that the intervention produced tangible, measurable improvements.

Victory Declarations and Practical Implementation: This sense of vindication was also expressed through victory declarations that validated the GPDP's strategy. Supporters celebrated “*Now it seems clear the Authority won and OpenAI ChatGPT will comply. Victory for privacy but also for (good) innovation”* (Post ID: 961). This frame was reinforced by users affiliated with the GPDP who shared on their personal X page practical instructions on how to use the new privacy features: *“Those who want to use ChatGPT without their data being processed also to improve algorithm training can use this functionality accessible in settings”* (Post ID: 1891). This shift from abstract debate to practical instruction was rhetorically significant, as it demonstrated that the GPDP's intervention had delivered functional tools and concrete benefits to users, thereby reinforcing the legitimacy of pragmatic exchange.

Justifying Harsh Methods with Results: finally, a more balanced achievement framing was used to acknowledge the intervention's intensity while emphasizing its successful outcome. An example is represented by Post ID: 856: *"Good the Garante Privacy's extraordinary session to examine OpenAI's proposals. The (in my opinion) excessive harshness was at least accompanied by adequate speed"*. This framing positioned the GPDP as imperfect in execution but ultimately effective in achieving its objectives.

F1: Normative Alignment – "Technology Must Follow the Rules We've Democratically Adopted"

The F1: Normative Alignment mechanism (91 instances) legitimized the intervention by positioning it as the necessary enforcement of established legal and ethical norms. This argument, which invoked the GDPR, the protection of minors, and foundational rule-of-law principles, contended that technology, regardless of its novelty, must adhere to democratically adopted rules.

Law Compliance as Non-Negotiable: The first line of justification framed GDPR compliance as non-negotiable. Supporters argued that privacy regulations are settled law requiring universal adherence, not exceptions for popular services: *"ChatGPT doesn't respect GDPR. What should they do, make an exception because you like it?"* (Post ID: 47). This perspective positioned critics as uninformed rather than principled, with one user stating: "*One thing seems clear. Those defending OpenAI have never opened GDPR"* (Post ID: 976).

Primacy of Law over Technology: A second, related argument inverts the "innovation first" assumption, rejecting the idea that innovation should precede regulation. This line of support framed this conflict as a normative choice: "*ChatGPT and similar systems raise enormous technological, ethical, social, environmental (they consume a huge amount of energy), and political issues. The Data Protection Authority hasn't ruled on these matters, nor would it be entitled to. But regulating artificial intelligence can no longer be postponed. We're deciding whether tech companies should get to choose which rules to follow or whether, once those rules are democratically adopted, like the GDPR and others to come, they must be respected* (Post ID: 56). This was often simplified to the maxim: *"Technology isn't above the law: why ChatGPT blocking protects our rights"* (Post ID: 1578). This was supported by invoking specific requirements, including the *"lack of privacy-by-design approach"* (Post ID: 36) and the necessity to protect minors, as evident in statements such as

"*There are no systems to control under-13s, do you want children using AI without control?*" (Post ID: 25).

F2: Consequentialist Legitimation – Catalyzing Public Discourse

The F2: Consequentialist Legitimation mechanism (18 instances) legitimized the intervention by highlighting the positive collective outcomes it generated. This argument emphasized that the controversy itself was valuable, as it initiated essential public discourse on AI governance and privacy that would not have otherwise occurred.

Generation of Public Debate about AI Governance: the primary benefit identified was the generation of a necessary public conversation around AI risks and AI governance: "*As they say, every problem can be an opportunity. The Privacy Authority vs. #ChatGPT affair can be one*" (Post ID: 789). Awareness raising about hidden risks emphasized the intervention exposed problems users didn't realize existed: *"This episode is generating healthy debate about privacy that can help non-experts and increase awareness among end users* (Post ID: 921). This mechanism reframed the controversy itself as a valuable societal good, justifying the intervention by its positive secondary effects on public awareness and reframing isolation (C1) as leadership

F3: Procedural Rectification Framing – "OpenAI Blocked Itself, Not the Authority"

The F3: Procedural Rectification Framing (137 instances) was the most prevalent legitimating mechanism. This argument operated by systematically correcting a procedural misunderstanding of the intervention: the GPDP requested compliance, and OpenAI chose to suspend access rather than comply immediately. Therefore, critics attacking the GPDP for "blocking" ChatGPT were factually incorrect.

Correcting the "Block" Narrative: the dominant theme in this mechanism was the simple correction that "the GPDP didn't block." Supporters relentlessly clarified this point, arguing "But the Italian Data Protection Authority didn't suspend ChatGPT, it simply requested a series of clarifications, and the company managing the service voluntarily suspended it while preparing the necessary explanations" (Post ID: 429). This narrative framed the public outcry as a "collective hallucination" (Post ID: 422), positioning critics as disconnected from reality and debating a fictional scenario.

Suspension as an Admission of Guilt: such a procedural correction was extended to reverse the causal narrative: supporters emphasized that OpenAI had other procedural options, arguing *"The Authority didn't impose blocking, but provisional limitation of data processing for users in Italy. The company could have continued offering the service. Not* *doing so is concrete evidence of GDPR violation"* (Post ID: 1262). In this framing, OpenAI's choice to suspend the service was not the result of a block, but rather an admission that it could not comply with data protection requirements, thereby vindicating the GPDP's concerns.

Acknowledging GPDP Communication Failures: Finally, this category incorporated self-criticism. Some supporters acknowledged that the GPDP's own poor communication, while reiterating the core procedural fact: *"#ChatGPT: when a member of GPDP's board fails in communication .... and creates confusion. To reiterate: the #Garante didn't block #ChatGPT, #OpenAI disabled #ChatGPT for users"* (Post ID: 427).

G1: International Norm Convergence – "Italy is Not Alone"

The G1: International Norm Convergence (42 instances) mechanism documented that multiple jurisdictions – particularly other European countries, Canada, and US agencies – were also investigating ChatGPT or expressing similar concerns. This argument directly countered the C1 narrative of Italian isolation. It reframed the situation as one of international regulatory convergence, demonstrating that subsequent investigations by other nations vindicated, rather than contradicted, Italy's initial approach.

Documentation of Multi-Jurisdiction Investigations: The empirical foundation for G1 was the documentation of multi-jurisdiction investigations. Supporters compiled and shared exhaustive inventories of countries examining ChatGPT. For example, one post (Post ID 937) noted that the "*Swiss federal data protection authority also expressed concerns about ChatGPT. Added to Canada, France, Ireland, Germany, USA"*. Another (Post ID 59) documented actions in France: "*In France CNIL received at least 2 complaints against ChatGPT for privacy and GDPR violation. In Norway, 'we haven't yet started an investigation on ChatGPT, but we don't exclude anything for the future*". The temporal framing of these posts, highlighting complaints received or investigations under consideration, suggests compatibility between the Italian and other Western countries' actions.

Validation via the European (EDPB) Task Force: this convergence was further ***validated at the*** EU level by the creation of a European Data Protection Board (EDPB) task force. Supporters highlighted that the EDPB launched this working group specifically "*to promote cooperation and information exchange among Authorities* (referenced in Post ID: 712). The task force represented formal recognition at the EU level that Italy raised legitimate concerns requiring a coordinated response. This transformed what critics framed as unilateral Italian action into a catalyst for pan-European governance development.

G2: Leadership Positioning – "Italy is Setting the Standard"

The G2: Leadership Positioning mechanism (31 instances) advanced the G1 (convergence) argument by actively positioning Italy as a leader whose actions others emulated. This mechanism directly inverted the C1 (deviance) narrative, reframing Italy from a "backward outlier" to a "forward-thinking vanguard." This approach invokes Suchman's (1995) concept of cognitive leadership legitimacy, wherein first movers and standard setters enjoy an enhanced status.

First-Mover: This framing was most directly articulated through "making history" or "setting example" framings. Supporters asserted that *"The Authority's choice is making school"* (Post ID: 53), and celebrated: *"So it's official: we weren't 'the only ones,' but for once we were 'the first'"* (Post ID: 758). This "first-mover" status was explicitly linked to European and global regulatory leadership. Public discourse claimed: *"Europe is the world's digital regulator' Director @Vi**********'s @EU_Commission on Europe's role in regulating new technologies"* (Post ID: 49). As introduced above, this claim was substantiated by highlighting that the new pan-European (EDPB) working group was Italian-led (Post ID: 735): *"ChatGPT will be investigated throughout Europe by an Italian-led working group* (Post ID: 1784). Global standard-setting claims positioned Italy as defining emerging international norms. One supporter stated: *"the authorities anticipated the entire world. Even ChatGPT's privacy management will be managed at EU level under Italian guidance"* (Post ID: 767).

Alarm Bell Framing: finally, a related framing positioned the ban as a necessary "alarm bell" or catalyst for a global conversation (compatibly with F2 framing of Consequentialist Legitimation). This was often supported by quoting external validation, such as a legal expert (Post ID 997), for instance, stating *"It will trigger dialogue in Europe and accelerate position-taking by other regulators"* (Post ID: 997). This mechanism thus completed the narrative inversion: Italy was not an isolated deviant (C1) but a courageous leader setting the global agenda.

4.4.3.3 Explaining Behavioral Indifference: The Legitimacy Collapse Mechanism

Taken together, this qualitative analysis reveals the precise mechanism of the GPDP's legitimacy failure. The public debate was a comprehensive, three-dimensional legitimacy collapse. As demonstrated, the delegitimating attacks were robust, diverse, and mutually reinforcing, striking simultaneously at the GPDP's pragmatic (A1, A2, A3), moral (B1, B2), and cognitive (C1, C2) legitimacy. In contrast, the legitimating defense was fragile and overwhelmingly concentrated on narrow procedural and moral claims (F1, F3).

This structural asymmetry (a many-front attack versus a single-trench defense) provides the deep explanatory context for the behavioral nullification found in Stage 2. The public did not just disagree with the GPDP; they were taught to dismiss it as simultaneously incompetent, arbitrary, and abnormal. This total legitimacy collapse is the mechanism that produced the "invisible failure," setting the stage for the final theoretical discussion of the Visibility Trap.

4.5 DISCUSSION AND IMPLICATIONS

This section interprets the empirical findings in relation to the study’s research question. It is structured in three parts. First, it reconciles the central tension that motivated the study: the divergence between the GPDP’s formal success and its latent behavioral failure (4.5.1). Second, it specifies how the pacing condition manifests in this case as an attentional misfit between GenAI and existing governance frameworks, integrating the pacing problem with an ABV (4.5.2). Finally, it connects this mechanism to the legitimacy collapse to define the study’s main theoretical contribution, the Visibility Trap, outlining its implications throughout the discussion (4.5.3).

4.5.1 Reconciling the Paradox: Why Visible Success Produced Invisible Failure

This study found that the ChatGPT ban produced no measurable change in Italian usage trajectories (Table 4.9). This was demonstrated by the interaction term in the DiD model, which was statistically indistinguishable from zero (δ_2 = -12,497, p = 0.571), indicating that post-restoration Italian search behavior resumed its pre-ban path as if the regulatory intervention had never occurred. Users neither reduced their ChatGPT-related activity in response to the GPDP's risk signals nor increased it due to amplified curiosity or reactance. They simply continued their prior trajectory.

This behavioral indifference contradicts the two primary outcomes predicted by established theory. If the ban had successfully sensitized the public to GenAI risks, as diffusion of innovation theory would predict when authorities frame a technology as hazardous (Rogers, 1983), usage should have shown durable suppression even after

ChatGPT's restoration. Alternatively, if the restriction had triggered psychological reactance or a Streisand Effect, whereby prohibitions amplify interest in forbidden content (Brehm, 1966; Jansen & Martin, 2015), post-restoration usage should have exceeded the counterfactual trajectory. The data supported neither prediction.

This creates a genuine puzzle: the intervention produced massive public discussion (Bolici et al., 2024), yet no impact on the behavior it ostensibly targeted. Thus, the GPDP successfully demonstrated regulatory power over a global technology platform while simultaneously demonstrating regulatory powerlessness over the population whose protection justified the intervention. Citizens discussed the ban extensively, but their actual usage decisions remained unaffected by the authority's risk assessment.

This disjuncture between visible regulatory success and invisible behavioral nullification raises a critical question: why did an intervention that achieved its formal objectives and generated substantial public attention fail completely to influence the adoption behavior that regulatory authorities presumably sought to shape?

Understanding this mechanism requires examining how the specific characteristics of the GPDP's intervention, particularly the mismatch between the technical nature of GenAI and the legal instruments available to regulate it, created systematic vulnerabilities that discourse participants could exploit. Framing this tension under the pacing problem perspective (Taeihagh et al., 2021; Walter, 2024) provides a plausible explanation for the observed phenomenon (See Section 4.5.2).

4.5.2 The Mechanism: How Technical Incompatibility Creates Legitimacy Vulnerabilities

As discussed in Chapter 2, the pacing problem captures both temporal lag and structural misalignment between rapidly evolving technologies and slower, more rigid governance arrangements (Allenby, 2011; Marchant, 2011, 2020; Maas, 2022; Nordström, 2022; Taeihagh et al., 2021). This section builds on that diagnosis and specifies how, in the ChatGPT case, the misalignment is organized inside a public authority. Using an ABV (Ocasio, 1997), it conceptualized GDPR and related procedures as institutionalized attentional architectures: they encode categories, routines, and communication channels that determine which aspects of GenAI become visible, actionable, and measurable, and which remain in a blind spot. The key mechanism in this case is not incompatibility in the abstract, but the way the misfit between GenAI and these attentional architectures narrows the GPDP’s perceived choice-set to a small subset of technically and symbolically legible interventions.

Established governance frameworks, such as the GDPR, therefore function as attention structures (Ocasio, 1997). These structures are not neutral; they are socially patterned frameworks that regulate and channel what decision-makers focus on. In practical terms, the GDPR provides regulators with a specific, pre-defined repertoire of issues and corresponding answers (such as data accuracy, consent, and age verification) that determine what becomes organizationally salient and actionable. A crisis emerges when this established attention structure is applied to a novel technology with which it is fundamentally incompatible.

For example, the GPDP’s attention was structurally channeled onto the GDPR's data accuracy requirement (Art. 5). This "issue" is fundamentally incompatible with GenAI's probabilistic operation; "hallucinations" are inherent statistical possibilities, not "bugs" to be corrected. Because the regulatory tool does not align with the technology, authorities struggle to justify their actions in technical terms and are pushed to rely primarily on formal legal authority. Therefore, while the GPDP could assert its legal authority to demand data accuracy, the action was perceived by critics as a fundamental misunderstanding of how the technology operates, triggering a predictable legitimacy crisis.

When authorities cannot demonstrate that their intervention is technically appropriate, only that they possess the legal authority to impose it, they risk transforming compliance into an opaque power relationship. This explains why the GPDP's visible success in securing OpenAI's compliance failed to repair its legitimacy: the changes implemented (such as transparency notices and age verification) formally addressed the letter of the GDPR but could not resolve the underlying architectural incompatibility.

Thus, if this pattern observed in this study generalizes, the pacing problem creates a governance trilemma. Authorities are forced to choose between: (1) acting early with misaligned tools, risking their legitimacy; (2) waiting for better tools, risking irrelevance as the technology becomes entrenched; or (3) not acting at all, thereby abdicating responsibility. Resolving this trilemma is particularly urgent in the GenAI context, where the gap between technological capabilities and regulatory readiness is measured in months, not years.

4.5.3 Evidence of the Visibility Trap

When read through the *Visibility Trap* framework developed in Chapter 3 (VT1–VT4), the three empirical stages in this chapter can be interpreted to jointly instantiate the mechanism. VT1 is expressed in the way the GPDP initially encounters ChatGPT as a highly legible privacy and data-protection problem: a specific service, operated by an identifiable firm, whose data practices can be rendered intelligible in GDPR terms (lawfulness, transparency, accuracy, protection of minors).

Under VT2, the highly legible privacy violations associated with ChatGPT are treated as a direct, high-stakes threat to the authority's core aspiration of protecting individuals and society. The document analysis (Section 4.4.1) showed that the GPDP presented the case as urgent and exceptional, under conditions where GenAI is diffusing rapidly, and AI-specific regimes are not yet operational. This configuration creates a clear performance–aspiration gap: the authority sees a salient risk it believes it ought to control, but for which it has only a narrow, ill-fitting toolkit. That gap drives a decisive organizational move: an urgent, highly visible national intervention that pushes the existing GDPR enforcement repertoire to its limit.

The document analysis, DiD estimation, and public discourse analysis jointly inform VT3 (evaluation of actions). Within a GDPR-centred attention structure, the intervention appears to close the perceived gap on visible criteria: violations are formally ascertained, a detailed compliance pathway is imposed and met, a pan-European task force is created, and a substantial fine is eventually levied. On these dimensions, the ban is presented, and can reasonably be read, as a regulatory success; aspirations tied to "effective enforcement of GDPR" are satisfied, and the case is framed as an exemplar of appropriate data-protection governance. At the same time, the DiD estimation (Section 4.4.2) shows that post-restoration usage of ChatGPT in Italy is statistically indistinguishable from the counterfactual trajectory, and the public discourse analysis (Section 4.4.3) evidences a three-dimensional erosion of pragmatic, moral, and cognitive legitimacy in the X debate, with delegitimating frames vastly outnumbering and out-structuring legitimating ones.

Interpreted through the legibility–strength framework in Section 3.5 (Figure 3.2), this mismatch is informative about how different consequences are positioned inside the authority's evaluative architecture. Formal GDPR outcomes are highly legible and high-strength: they are encoded directly in existing categories and metrics (violations found, measures imposed, fines, EU coordination) and used as primary criteria for judging whether aspirations around enforcement are met. By contrast, neither post-ban behavioral effects nor public legitimacy dynamics appear as salient points in the GPDP's own success narrative: they are not articulated as explicit targets in the documents, nor is there any substantial evidence of subsequent problemistic search or recalibration in response to behavioral indifference or discursive backlash. Under ABV, high-strength signals that expose a performance–aspiration gap would keep the issue "alive" and trigger further search; the absence of such adjustment is consistent with these outcomes being treated as low-strength within existing goals and measures. At the same time, because they are not represented in established categories, indicators, or routines, they are also low in legibility from the authority's point of view. Taken together, this supports the interpretation that behavioral and legitimacy outcomes mostly fall into quadrant 4 of Figure 3.2 (low legibility, low strength) for the GPDP, even though they are highly salient and consequential for external audiences. This configuration provides a concrete micro-foundation for the Visibility Trap in this case: the same intervention that is evaluated internally as a clear success on visible, codified criteria coincides, empirically, with behavioral indifference and marked legitimacy erosion that remain weakly represented in the authority's own evaluative channels.

VT4 concerns the longer-term recursive effects of this configuration, which this study cannot fully observe within its time horizon. What can be seen already is that the ChatGPT case has been encoded in the field as a salient reference point. OpenAI extended the changes negotiated with the GPDP to all European users, and European bodies and several national authorities opened inquiries or working groups in the wake of the Italian move. At the same time, no other authority has replicated the specific instrument of a national ban on ChatGPT or similar GenAI services. The case has thus become visible as both a precedent for demanding concrete changes from a major provider and a boundary marker of how far a single national authority has been willing to go.

This single case cannot establish that the full VT1–VT4 pattern generalizes across all GenAI governance episodes. However, the combination of (i) high-legibility–high-strength privacy signals triggering a decisive move, (ii) positive evaluation on visible enforcement criteria, and (iii) concurrent behavioral nullification and three-dimensional legitimacy erosion provides a strong empirical instantiation of the mechanism theorized in Chapter 3. It suggests that, under pacing and incompatibility conditions, formally successful interventions that are tightly aligned with established attention structures may systematically weaken the substantive authority needed for subsequent governance moves. Avoiding this trap requires not only additional "sensing" instruments (e.g., real-time legitimacy and behavioral impact tracking), but also deliberate experimentation with alternative attention structures – that is, ways of encoding goals, metrics, and channels that can bring latent governance consequences out of the low-legibility–low-strength corner before recursive reinforcement locks in a pattern of visible victories and invisible institutional decline.

4.6 CONCLUSIONS & LIMITATIONS

This study addressed an urgent problem surrounding GenAI governance: the paradox of regulatory interventions that achieve formal success while producing substantive failure. Using the Italian ChatGPT ban as a case, this work employed a mixed-methods design to investigate the divergence between the GPDP's visible success (compliance) and its invisible failure (post-intervention behavioral indifference and three-dimensional delegitimation).

The study made three primary contributions. First, it empirically demonstrates this tension by combining document analysis, DiD estimation, and qualitative content analysis to show how a formally successful GenAI intervention can coexist with behavioral indifference and three-dimensional legitimacy erosion. By doing so, it provided an empirical demonstration of the Visibility Trap: a self-reinforcing process, shaped by established attention structures, in which the salience of visible successes progressively degrades the capacity to identify and respond to associated but latent negative consequences. Second, it specified the underlying mechanism by framing the pacing problem within an Attention-Based View (Ocasio, 1997): misaligned attention structures, exemplified by the way GDPR and related procedures make certain responses to GenAI appear uniquely available and defensible, narrow the authority’s perceived options, and contribute to technically and symbolically vulnerable interventions. Lastly, it advanced the Governance Trilemma confronting GenAI governance authorities: act early with misaligned tools and risk legitimacy erosion, wait for better governance tools and risk irrelevance, or abstain and abdicate responsibility.

Future research can further investigate the trilemma discussed in this study. If the pacing problem stems from established attention structures, what mechanisms can interrupt the routine application of obsolete tools before legitimacy erodes? And how should authorities balance the risks of acting early, waiting for better tools, or not acting at all in the trilemma? Answering these questions is crucial for developing governance models that can effectively navigate rapid technological change.

This study has limitations. First, it is a single-case analysis. While the Italian ban serves as an ideal-type case for observing the paradox of formal success and substantive failure, the generalizability of the findings is not guaranteed. The specific dynamics of the Visibility Trap and Governance Trilemma may be contingent on Italy's unique political and media environment. Second, the qualitative analysis of public discourse (Stage 3) relies on data from X (Twitter). This population is not representative of the general Italian public and may over-sample users who are more technically inclined or politically engaged. Therefore, findings documented here reflect a collapse within this specific digital public sphere, not necessarily a one-to-one map of broader societal sentiment. Third, the DiD analysis (Stage 2) relies on estimated web traffic data from Semrush, not on actual server logs. While Semrush data is suitable for detecting directional trends and relative changes (as argued in 4.3.2.1), the absolute visit counts are proprietary estimates. This introduces a limitation in the precision of the behavioral measurement. Fourth, the DiD analysis relies on Spain as the sole control country. While this choice is methodologically justified (robustness checks confirm parallel pre-trends), a synthetic control approach aggregating multiple European countries would provide stronger causal identification by creating a weighted counterfactual. Nevertheless, the behavioral nullification finding remains robust, consistently holding across multiple model specifications. Finally, the single-coder approach to qualitative content analysis introduces potential interpretive bias. Multiple coders with an inter-rater reliability assessment would strengthen the validity of the mechanism identification. Despite this, the convergence of findings across all three methodological stages provides confidence in the core theoretical contributions.

CHAPTER REFERENCES

Allenby, B. R. (2011). Governance and Technology Systems: The Challenge of Emerging Technologies. In G. E. Marchant, B. R. Allenby, & J. R. Herkert (Eds.), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (Vol. 7, pp. 3–18). Springer Netherlands. https://doi.org/10.1007/978-94-007-1356-7_1

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Bertomeu, J., Lin, Y., Liu, Y., & Ni, Z. (2023). Capital Market Consequences of Generative AI: Early Evidence from the Ban of ChatGPT in Italy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4452670

Bolici, F., Acciarini, C., Marchegiani, L., & Pirolo, L. (2020). Innovation diffusion in tourism: How information about blockchain is exchanged and characterized on twitter. *The TQM Journal*. https://doi.org/10.1108/TQM-01-2020-0016

Bolici, F., Varone, A., & Diana, G. (2024). Unpopular Policies, Ineffective Bans: Lessons Learned from ChatGPT Prohibition in Italy. *ECIS 2024 Proceedings*. https://aisel.aisnet.org/ecis2024/track04_impactai/track04_impactai/11?utm_source=aisel.aisnet.org%2Fecis2024%2Ftrack04_impactai%2Ftrack04_impactai%2F11&utm_medium=PDF&utm_campaign=PDFCoverPages

Bourdieu, P. (1977). The economics of linguistic exchanges. *Social Science Information*, *16*(6), 645–668. https://doi.org/10.1177/053901847701600601

Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, *9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press New York.

Brehm, S. S., & Brehm, J. W. (1981). *Psychological reactance: A theory of freedom and control*. Academic Press.

Burgess, J., & Bruns, A. (2012). Twitter Archives and the Challenges of “Big Social Data” for Media and Communication Research. *M/C Journal*, *15*(5). https://doi.org/10.5204/mcj.561

Busacca, A., & Monaca, M. A. (2025). Who’s Afraid of “Big Bad” Generative AI? Brief Notes on the IDPA Provision Against OpenAI ChatGPT. In D. Marino & M. A. Monaca (Eds.), *Generative Artificial Intelligence and Fifth Industrial Revolution* (Vol. 880, pp. 117–144). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-73880-7_9

Bushman, B. J., & Stack, A. D. (1996). Forbidden fruit versus tainted fruit: Effects of warning labels on attraction to television violence. *Journal of Experimental Psychology: Applied*, *2*(3), Article 3. https://doi.org/10.1037/1076-898X.2.3.207

Caiazza, R. (2016). A cross-national analysis of policies affecting innovation diffusion. *The Journal of Technology Transfer*, *41*(6), Article 6. https://doi.org/10.1007/s10961-015-9439-2

Chan, K. J. D., Papyshev, G., & Yarime, M. (2024). Balancing the tradeoff between regulation and innovation for artificial intelligence: An analysis of top-down command and control and bottom-up self-regulatory approaches. *Technology in Society*, *79*, 102747. https://doi.org/10.1016/j.techsoc.2024.102747

Chiarello, F., Giordano, V., Spada, I., Barandoni, S., & Fantoni, G. (2024). Future applications of generative large language models: A data-driven case study on ChatGPT. *Technovation*, *133*, 103002. https://doi.org/10.1016/j.technovation.2024.103002

Coltri, M. A. (2024). The Ethical Dilemma with Open AI ChatGPT: Is it Right or Wrong to prohibit it? *Athens Journal of Law*, *10*(1), 119–130.

Currie, W. L., Leimeister, J. M., Schlagwein, D., & Willcocks, L. (2025). Rethinking technology regulation in the age of AI risks. *Journal of Information Technology*, *40*(3), 236–245. https://doi.org/10.1177/02683962251378815

Cyert, R., & March, J. (1963). *A Behavioral Theory of the Firm*.

Dafoe, A. (2023). AI Governance: Overview and Theoretical Lenses. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., pp. 21–44). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.2

DiMaggio, P. J. (1988). Interest and agency in institutional theory. In *Institutional Patterns andOrganizations: Culture and Environment* (pp. 3–21). Cambridge University Press.

DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, *48*(2).

DiMaggio, P. J., & Powell, W. W. (1991). *The new institutionalism in organizational analysis*. Univ. of Chicago Press.

Dolfsma, W., & Seo, D. (2013). Government policy and technological innovation—A suggested typology. *Technovation*, *33*(6), 173–179. https://doi.org/10.1016/j.technovation.2013.03.011

Dowling, M.-E., & Legrand, T. (2023). “I do not consent”: Political legitimacy, misinformation, and the compliance challenge in Australia’s Covid-19 policy response. *Policy and Society*, *42*(3), 319–333. https://doi.org/10.1093/polsoc/puad018

Farisco, M. (2024). The EU Regulation of Generative AI: What Can We Learn from the Controversy Between the Italian Data Protection Authority and OpenAI. In A. Aristovnik, D. Ravšelj, & N. Tomaževič (Eds.), *The regulation of artificial intelligence* *shaping governance and power relations for a liberal future*. European Liberal Forum EUPF : Zavod 14.

Friedland, R., & Alford, R. R. (1991). Expanding the Scope of Institutional Analysis. In *The new institutionalism in organizational analysis*. Univ. of Chicago Press.

Garfinkel, H. (1967). Studies in ethnomethodology. Prentice-Hall.

Giddens, A. (1984). The constitution of society: Outline of the theory of structuration (1st paperback ed., repr). Polity Press.

GPDP. (2023a, March 30). Provvedimento del 30 marzo 2023 [9870832]. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9870832

GPDP. (2023b, March 31). Intelligenza artificiale: Il Garante blocca ChatGPT. Raccolta illecita di dati personali. Assenza di sistemi per la verifica dell'età dei minori. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9870847

GPDP. (2023c, April 4). ChatGPT: Domani incontro tra OpenAI e Garante privacy. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9872284

GPDP. (2023d, April 13). ChatGPT: Al via una task force europea. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9875657

GPDP. (2023e, April 28). ChatGPT: OpenAI riapre la piattaforma in Italia garantendo più trasparenza e più diritti a utenti e non utenti europei. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9881490

GPDP. (2023f, June 4). ChatGPT: OpenAI collabora con il Garante privacy con impegni per tutelare gli utenti italiani. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9872832

GPDP. (2023g, August 4). ChatGPT: Garante privacy, iniziato l'esame delle misure proposte da OpenAI. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9873602

GPDP. (2023h, November 4). Provvedimento dell'11 aprile 2023 [9874702]. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9874702

GPDP. (2023i, December 4). ChatGPT: Garante privacy, limitazione provvisoria sospesa se OpenAI adotterà le misure richieste. L'Autorità ha dato tempo alla società fino al 30 aprile per mettersi in regola. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9874751

GPDP. (2024, December 20). COMUNICATO STAMPA - ChatGPT, il Garante privacy chiude l'istruttoria. OpenAI dovrà realizzare una campagna informativa di sei mesi e pagare una sanzione di 15 milioni di euro. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/10085432

Gualdi, F., & Cordella, A. (2024). Theorizing the regulation of generative AI: lessons learned from Italy's ban on ChatGPT. *Proceedings of the 57th Hawaii International Conference on System Sciences*. HICSS 2024. https://hdl.handle.net/10125/106631

Gurinskaya, A., Nalla, M. K., & Paek, S. Y. (2024). Exploring the Determinants of Citizens' Compliance with COVID-19 Regulations: Legitimacy Versus Fear. Criminal Justice Review, 49(2), 156–174. https://doi.org/10.1177/07340168231190471

Hawks, K. (2025). Cultural Authority and (Non)Compliance with Public Health Directives: The Effect of Legitimacy and Values on Behavior during the COVID-19 Pandemic. *Journal of Health and Social Behavior*, 00221465241312696. https://doi.org/10.1177/00221465241312696

Hemphill, T. A. (2020). "The innovation governance dilemma: Alternatives to the precautionary principle." *Technology in Society*, *63*, 101381. https://doi.org/10.1016/j.techsoc.2020.101381

Jansen, B. J., Jung, S., & Salminen, J. (2022). Measuring user interactions with websites: A comparison of two industry standard analytics approaches using data of 86 websites. *PLOS ONE*, *17*(5), e0268212. https://doi.org/10.1371/journal.pone.0268212

Jansen, S., & Martin, B. (2015). The Streisand Effect and Censorship Backfire. *International Journal of Communication*, *9*, 656–671.

Jepperson, R. (1991). Institutions, institutional effects, and institutionalism. In *The new institutionalism in organizational analysis*. Univ. of Chicago Press.

Kikuchi, T. (2024). *Impact Evaluation on the European Privacy Laws governing generative-AI models—Evidence in Relation between Internet Censorship and the Ban of ChatGPT in Italy* (No. arXiv:2407.06495). arXiv. https://doi.org/10.48550/arXiv.2407.06495

Kreitmeir, D. H., & Raschky, P. A. (2023). *The Unintended Consequences of Censoring Digital Technology—Evidence from Italy's ChatGPT Ban* (No. arXiv:2304.09339; Issue arXiv:2304.09339). arXiv. http://arxiv.org/abs/2304.09339

Lawrence, T. B., & Suddaby, R. (2006). Institutions and Institutional Work. In *The SAGE Handbook of Organization Studies* (2nd Edition, pp. 215–254). SAGE. https://www.torrossa.com/en/resources/an/4913717

Lechner, M. (2011). The Estimation of Causal Effects by Difference-in-Difference Methods*. Foundations and Trends in Econometrics, 4(3), 165–224. https://doi.org/10.1561/0800000014

Liu, Z., Iqbal, U., & Saxena, N. (2024). Opted Out, Yet Tracked: Are Regulations Enough to Protect Your Privacy? *Proceedings on Privacy Enhancing Technologies*, *2024*(1), 280–299. https://doi.org/10.56553/popets-2024-0016

Maas, M. M. (2022). Aligning AI Regulation to Sociotechnical Change. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (1st ed., pp. 358–380). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.22

Maguire, S., Hardy, C., & Lawrence, T. B. (2004). Institutional Entrepreneurship in Emerging Fields: HIV/AIDS Treatment Advocacy in Canada. *Academy of Management Journal*, *47*(5), 657–679. https://doi.org/10.2307/20159610

March, J., & Simon, H. (1958). *Organizations* (1st edition). Wiley.

Marchant, G. E. (2011). Addressing the Pacing Problem. In G. E. Marchant, B. R. Allenby, & J. R. Herkert (Eds.), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (pp. 199–205). Springer Netherlands. https://doi.org/10.1007/978-94-007-1356-7_13

Marchant, G. E. (2020). Governance of Emerging Technologies as a Wicked Problem. *Vanderbilt Law Review*, *73*, 1861.

Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, *83*(2).

Milmo, D. (2023, February 2). ChatGPT reaches 100 million users two months after launch. *The Guardian*. https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app

Monteiro, N., & Singh, V. (2025). The wheel of artificial intelligence governance. *Sustainable Futures*, *10*, 101279. https://doi.org/10.1016/j.sftr.2025.101279

Mueller, M. L. (2025). It's just distributed computing: Rethinking AI governance. *Telecommunications Policy*, *49*(3), 102917. https://doi.org/10.1016/j.telpol.2025.102917

Nordström, M. (2022). AI under great uncertainty: Implications and decision strategies for public policy. *AI & SOCIETY*, *37*(4), 1703–1714. https://doi.org/10.1007/s00146-021-01263-4

Ocasio, W. (1997). Towards an Attention-Based View of the Firm. *Strategic Management Journal*, *18*(S1), 187–206. https://doi.org/10.1002/(SICI)1097-0266(199707)18:1+%253C187::AID-SMJ936%253E3.0.CO;2-K

Powell, W. W. (1991). Expanding the Scope of Institutional Analysis. In *The new institutionalism in organizational analysis*. Univ. of Chicago Press.

Rejeski, D. (2011). Public Policy on the Technological Frontier. In G. E. Marchant, B. R. Allenby, & J. R. Herkert (Eds.), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (Vol. 7, pp. 47–59). Springer Netherlands. https://doi.org/10.1007/978-94-007-1356-7_4

Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed). Free Press ; Collier Macmillan.

Rogers, E. M. (2003). *Diffusion of innovations, 5th edition*. Free Press.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. Journal of Econometrics, 235(2), 2218–2244. https://doi.org/10.1016/j.jeconom.2023.03.008

Sætra, H. S. (2023). Generative AI: Here to stay, but for good? *Technology in Society*, *75*, 102372. https://doi.org/10.1016/j.techsoc.2023.102372

Schuster, J., & Kolleck, N. (2020). The Global Diffusion of Social Innovations – An Analysis of Twitter Communication Networks Related to Inclusive Education. Frontiers in Education, 5, 492010. https://doi.org/10.3389/feduc.2020.492010

Scott, W. R. (1995). *Institutions and organizations*. Sage.

Selznick, P. (1953). TVA and the grass roots: A study of politics and organization. Quid Pro Books.

Selznick, P. (1996). Institutionalism "Old" and "New." Administrative Science Quarterly, 41(2), 270. https://doi.org/10.2307/2393719

Semrush. (2025). *Semrush Traffic & Market Toolkit*. Semrush. https://www.semrush.com/kb/1121-semrush-traffic-and-market

Shetty, D. K., Vijaya Arjunan, R., Cenitta, D., Makkithaya, K., Hegde, N. V., Bhatta B, S. R., Salu, S., Aishwarya, T. R., Bhat, P., & Pullela, P. K. (2025). Analyzing AI regulation through literature and current trends. *Journal of Open Innovation: Technology, Market, and Complexity*, *11*(1), 100508. https://doi.org/10.1016/j.joitmc.2025.100508

Simon, H. A. (1947). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization* (1st edition). Macmillan Co.

Steinert-Threlkeld, Z. C. (2018). *Twitter as Data* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108529327

Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *The Academy of Management Review*, *20*(3), 571. https://doi.org/10.2307/258788

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, *40*(2), 137–157. https://doi.org/10.1080/14494035.2021.1928377

Taeihagh, A., Ramesh, M., & Howlett, M. (2021). Assessing the regulatory challenges of emerging disruptive technologies. Regulation & Governance, 15(4), Article 4. https://doi.org/10.1111/rego.12392

Taylor, L. (2021). Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector. *Philosophy & Technology*, *34*(4), 897–922. https://doi.org/10.1007/s13347-020-00441-4

Thornton, P. H., & Ocasio, W. (1999). Institutional Logics and the Historical Contingency of Power in Organizations: Executive Succession in the Higher Education Publishing

Industry, 1958‐1990. *American Journal of Sociology*, *105*(3), 801–843. https://doi.org/10.1086/210361

Thornton, P. H., Ocasio, W., & Lounsbury, M. (2012). *The institutional logics perspective: A new approach to culture, structure and process*. Oxford Univ. Press. https://doi.org/10.1093/acprof:oso/9780199601936.001.0001

Viljoen, S. (2021). The Promise and Limits of Lawfulness: Inequality, Law, and the Techlash. *Journal of Social Computing*, *2*(3), 284–296. https://doi.org/10.23919/JSC.2021.0025

Walter, Y. (2024). Managing the race to the moon: Global policy and governance in Artificial Intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences. Discover Artificial Intelligence, 4(1), 14. https://doi.org/10.1007/s44163-024-00109-4

Wang, J., Xu, H., & Wang, M. (2025). Technology diffusion considering technological progress and multiple policy combinations based on evolutionary game theory. *Technology in Society*, *81*, 102799. https://doi.org/10.1016/j.techsoc.2024.102799

Yoganathan, V., Osburg, V.-S., & Janakiraman, N. (2025). Lending Legitimacy to Corporate Digital Responsibility: Trust in Firm Versus Government Regulation of Artificial Intelligence Services. *Journal of Service Research*, 10946705251345097. https://doi.org/10.1177/10946705251345097

Zhao, L. (2023). Artificial intelligence and law: Emerging divergent national regulatory approaches in a changing landscape of fast-evolving AI technologies. In *Research Handbook on Digital Trade* (pp. 369–399). Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/book/9781800884953/book-part-9781800884953-33.xml

Zucker, L. G. (1977). The Role of Institutionalization in Cultural Persistence. *American Sociological Review*, *42*(5), 726. https://doi.org/10.2307/2094862

CHAPTER 5: GENAI IN KNOWLEDGE WORK: PERFORMANCE, EXPERTISE, AND ORGANIZATIONAL MEMORY

ABSTRACT

The integration of Generative AI (GenAI) into knowledge work presents a fundamental tension: immediate indicators of productivity are rising, yet the human and organizational capabilities required to sustain long-term adaptation appear increasingly fragile. Integrating Skill Acquisition Theory, Organizational Memory theory, and the Attention-Based View, this study draws on exploratory interviews to theorize the mechanisms driving this divergence. At the individual level, this chapter reveals that GenAI scaffolds novice performance to approximate expert standards, creating an illusion of competence that masks the displacement of deep procedural learning. This creates a structural vulnerability: the very ease of obtaining high-quality outputs may disincentivize skill acquisition, while the resulting lack of meta-skills leaves users ill-equipped to orchestrate the system's full potential. At the organizational level, GenAI is emerging as a new form of Organizational Memory Information System (OMIS) that increasingly centralizes the encoding, retention and retrieval of organizational memory. While efficient, this infrastructure risks hollowing out traditional memory carriers and introducing pseudo-memory: synthetic precedents with opaque provenance. Synthetizing these dynamics through the *Visibility Trap* framework (developed in Chapter 3), the analysis illustrates how high-legibility signals of immediate success systematically crowd out attention to low-legibility risks of capability erosion.

5.1 INTRODUCTION

The advancement of Generative AI (GenAI) systems is transforming the nature of knowledge work, creating both enthusiasm and apprehension about the future (Crowston & Bolici, 2025; Cordella & Gualdi, 2023; Dell'Acqua et al., 2023; Varone et al., 2025). As discussed in Chapter 2, recent work demonstrates that GenAI yields substantial productivity gains across diverse domains, including customer service, consulting, writing, and programming (Brynjolfsson et al., 2023; Noy & Zhang, 2023; Peng et al., 2023). However, a converging body of evidence suggests that these tools may impede the development of foundational human skills, a phenomenon characterized as deskilling (Macnamara et al., 2024; Sison et al., 2024). Organizations thus confront a relevant tension: the tools that enhance immediate performance may erode the human and collective capabilities required for long-term adaptation. This chapter addresses the following question: *How do GenAI associated performance gains coexist with the disruption of expertise development trajectories and organizational mnemonic infrastructures?*

This chapter develops two interconnected models grounded in Skill Acquisition Theory (SAT) and Organizational Memory (OM) theory, informed by eight exploratory interviews with practitioners. Model 1 (Section 5.4.1) theorizes how GenAI reshapes the relationship between expertise and observable performance at the individual level through three mechanisms. The *AI-Performance Leveling Effect* describes how AI compresses observed performance differences across expertise levels, creating an illusion of competence rather than actual competence. The *Augmentation-Deskilling Paradox* highlights how short-term performance improvements occur alongside blocked skill development, particularly harming novices. *AI Capability Loss* illustrates the gap between GenAI's potential and its actual utility, which is constrained by users' limited abilities in task decomposition, prompting, and output evaluation. These mechanisms explain how GenAI can simultaneously elevate visible performance metrics while stalling the cultivation of underlying human capabilities.

Model 2 (Section 5.4.2) extends the analysis to the organizational level by conceptualizing GenAI as a new form of Organizational Memory Information System (OMIS). Unlike traditional repositories that store static documents, GenAI-OMIS centralizes the encoding, retention, and retrieval of experience into a generative infrastructure. This shift reweights the classical carriers of organizational memory (Walsh & Ungson, 1991). By channeling decision histories through prompts and opaque parameters, GenAI creates an appearance of accessible, coherent memory while destabilizing knowledge provenance and weakening organizational governance over institutional history.

Synthesizing these models through an Attention-Based View (ABV) (Ocasio, 1997), this chapter contextualizes the *Visibility Trap* within knowledge work environments. The analysis offers three contributions. First, it reframes deskilling from an individual symptom to an organizational vulnerability by showing how stalled learning trajectories aggregate into diminished adaptive capacity. Second, it extends OM theory by defining GenAI as a dual-layer memory infrastructure comprising explicit logs and sub-symbolic weights that introduces new risks of pseudo-memories. Third, it advances the ABV by demonstrating how technology reconfigures attention structures, making it rational for managers to prioritize visible short-term gains while remaining structurally blind to the erosion of long-term cognitive foundations.

5.2 METHODS

This chapter adopts a theory building approach (Cornelissen, 2017; Jaakkola, 2020). The theorization was informed by eight exploratory semi-structured interviews with managers and professionals who routinely work with GenAI in various organizational settings (Table 5.1). Each interview lasted 45-60 minutes on average and focused on: (1) how GenAI is embedded in everyday work; (2) effects on performance and skill development; and (3) how GenAI tools with memory and workspace-integration features mediate the storage, retrieval, and apparent “sharing” of organizational knowledge.

Transcripts were analyzed abductively. Consistent with standards for conceptual research (Suddaby, 2010; Jaakkola, 2020), the resulting framework was assessed for internal coherence, theoretical parsimony, and explanatory reach. The models do not claim empirical validation through the interviews. Their role is to provide a theoretical foundation that future empirical work can test, refine, or build upon.

TABLE 5.1: Roles, functional areas, and sectors in which the interviewees work

ID	Role	Functional area	Sector
I1	Business Analysis Manager	Business analysis, IT management	IT services & digital consulting
I2	EU Affairs, Advocacy & Policy Manager	Regulation, public affairs, policy	Telecommunications
I3	Sherpa Lead – Strategy & Business	Innovation & organization design	Consultancy
I4	Crisis Analyst	Crisis operations, risk & safety	Travel
I5	Product Designer	UX/UI and product design	Consultancy
I6	Product Lead	Digital product management	Hospitality
I7	Secretary	General administration	Sports Management
I8	Cloud Lead, Project Manager, & Cloud DevOps Engineer	Information technology	Public sector banking/finance

5.3 BACKGROUND

This section provides the background for the two models developed in this chapter. SAT (Section 5.3.1) specifies how individual skills are acquired, maintained, and lost, as well as how expertise influences the ability to decompose, prompt, and evaluate GenAI outputs. OM and OMIS theory (Section 5.3.2) specify how organizational experience is encoded, retained, and retrieved across distributed carriers. The third theoretical pillar, the ABV, is treated in depth in Chapter 3, which reconstructs its microfoundations and develops the ABV-based Visibility Trap framework.

5.3.1 Skill Development

Across disciplines, "skill" is defined in different ways. This chapter follows Green (2011) and defines skills as individual qualities that create value, can be expanded through training, and are socially determined. The first element captures the ability to perform tasks with autonomy and discretion to produce valued outcomes; more skilled individuals are associated with the performance of more complex tasks (Green, 2011).

Adopting the perspective of Wood (1986), this section refers to task complexity as the degree to which a task involves numerous distinct components (component complexity), requires intricate coordination among these components (coordinative complexity), and is subject to changes in its requirements or conditions over time (dynamic complexity). Variations in task complexity result in changes in the knowledge and skills required for successful task performance (Wood, 1986). The second characteristic motivates interest in skill development, as learners move from basic knowledge to developed and automatic skills. The final characteristic acknowledges that social forces influence the valuation of outputs and, consequently, perceptions of what work is considered skilled. For instance, this may involve undervaluing or even overlooking work traditionally performed by women, or viewing such tasks as requiring little skill. Social forces also shape who is interested in or able to develop particular skills, as well as who is in a position to employ them. While the social dimension of skills is recognized, this model emphasizes their productive and expandable features.

5.3.1.1 The Skill Development Process: Declarative, Procedural, Automatized

Despite differences in terminology, classic models of skill development describe a similar trajectory from effortful, instruction-driven performance to fluent, largely automatic behavior. Fitts & Posner (1973) distinguish cognitive, associative and autonomous phases; Anderson's ACT framework (1982) moves from a declarative stage to compiled procedural control; Dreyfus & Dreyfus (1986; Dreyfus, 2004) trace a shift from novice to expert; and VanLehn (1996) and Logan (1988) emphasize intermediate rule-building and later acceleration through strategy shifts and instance-based responding. DeKeyser (2015) argues that these accounts can be understood through a small set of shared principles.

Despite terminological variation, the underlying regularity is a control shift: behavior moves from slow, search-heavy interpretation of instructions to fast, cue-triggered execution by compiled structures; representational formats become more principled and economical as practice accrues (Newell & Simon, 1972; Anderson, 1982; Fitts & Posner, 1973; Logan, 1988). A broad consensus, therefore, treats skill development as unfolding through three phases: declarative (knowledge "that"), procedural (knowledge "how" via compilation), and automatized (intuitive or instance-based control). Each phase contributes distinctively: interpretive–declarative encodings support early problem solving; compiled–procedural control supports fluent, low-load execution; experiential–intuitive control supports rapid, situated judgment. The phases are described in turn, beginning with the declarative stage.

Declarative Stage

Skill learning typically begins with declarative encodings (instructions, examples, and rules) applied through general-purpose problem-solving procedures within a constructed problem space of states, operators, and goals (Newell & Simon, 1972; Anderson, 1982). This body of facts and relations, spanning semantic and episodic elements, constitutes declarative knowledge (DeKeyser, 2015, 2020; VanLehn, 1996). Because control at this point depends almost entirely on interpretation rather than retrieval, early performance is search-dominant: (1) learners rely on means–ends analysis, subgoaling, and explicit rule following; (2) execution is slow and stepwise; and (3) outcomes are sensitive to misunderstanding because each move must be computed rather than recalled (Newell & Simon, 1972; Anderson, 1982).

At this stage, cognitive demands are high. Interpreting instructions while tracking intermediate states taxes limited-capacity working memory, so errors arise from overload as much as from ignorance (Fitts & Posner, 1973). Progress depends on external guidance and timely feedback that reveal mismatches between intended and observed outcomes, supplying the corrective signals that later drive compilation and tuning (Fitts & Posner, 1973; VanLehn, 1996).

Instructional design strongly shapes the efficiency of this stage. Cognitive Load Theory distinguishes between intrinsic load and extraneous load, emphasizing that resources should be directed toward learning-relevant structure (Sweller, 1988). At the declarative phase, reducing extraneous load is crucial so that learners can perceive problem structure and form initial schemas. Worked examples and goal-free or completion tasks limit unnecessary search and free attention for mapping conditions to actions and noticing deep regularities (Sweller, 1988; Sweller et al., 1998). Integrated presentations that co-locate text, diagrams, and key constraints prevent split attention and lower coordination costs, while timely feedback helps detect rule misapplication and supports the micro-corrections needed for later proceduralization (Fitts & Posner, 1973; VanLehn, 1996).

Representation already matters at this earliest point. Learners who encode problems in terms of deep constraints and diagnostic cues, rather than surface features, lay the substrate for later expertise (Chi, Feltovich, & Glaser, 1981; Larkin et al., 1980). Declarative instruction thus serves not merely to supply facts but to scaffold the recognition of structure, ensuring that subsequent practice compiles the right patterns.

Procedural Stage

Practice transforms step-by-step problem solving into fast, compiled action. Repeated successful executions under reasonably stable goals and mappings trigger knowledge compilation, which operates through two linked processes: composition, where multi-step sequences are collapsed into larger macro-operators, and proceduralization, where stable declarative elements are bound into production rules, thereby eliminating explicit lookups (Anderson, 1982; 1983).

Learning advances through both mistakes and small wins, each refining how and when actions should fire. Impasses and minor successes create learning events that add or revise micro-rules, conditions, and discriminations governing production firing (VanLehn, 1988, 1996). Across episodes, representations begin to privilege diagnostic features over surface attributes, an early form of expert schemas and chunks observed in classic studies of problem-solving and perception (Chi, Feltovich, & Glaser, 1981; Chase & Simon, 1973; Larkin et al., 1980). Appropriate variability and timely, informative feedback accelerate this shift by increasing contrast and clarifying the boundaries of correct application (Ericsson, Krampe, & Tesch-Römer, 1993; Anderson, 1982).

In addition to experience, sustained improvement strongly depends on deliberate practice. Goal-specific, feedback-rich, effortful repetitions that target identified weaknesses increase the density of valid learning episodes for composition and proceduralization, provide contrast sets that foster schema growth, and generate the error signals that drive tuning (Ericsson et al., 1993; Ericsson & Smith, 1991). This mechanism explains why expertise is typically domain- and structure-specific: what compiles is determined by the regularities present in practiced tasks, not by broad exposure alone.

Following the previously described mechanisms, proceduralization reduces cognitive load, resulting in faster and more accurate performance. As condition–action productions replace slow, search-heavy interpretation, both the number of steps and the time per step decline, easing working-memory demands and producing visible speed–accuracy gains (Fitts & Posner, 1973; Anderson, 1982).

Automatization Stage

Expert performance often feels immediate and effortless because control has shifted from step-by-step thinking to immediate, pattern-triggered responding. In this regime, responses are rapid, context-sensitive, and guided by similarity, yet are hard to put into words (Dreyfus & Dreyfus, 1986; Dreyfus, 2004; Benner, 1984).

Automatization relies on experts identifying structures that novices overlook. Through extensive practice, experts encode situations as schemas and chunks that compress many elements into a few meaningful units and link diagnostic cues to appropriate procedures (Chase & Simon, 1973; Larkin et al., 1980; Chi et al., 1981). In physics, for instance, experts sort problems by underlying principles rather than surface features, and this representational fit determines what is retrieved and what action follows (Chi et al., 1981). These reorganized representations provide the substrate for compilation: schemas and chunks reduce what must be considered.

Two mechanisms drive this shift when cues and actions align reliably over time. First, practice strengthens production rules and stitches steps into longer sequences, so responses run on "autopilot" and demand less working memory (Shiffrin & Schneider, 1977; Anderson, 1982). Second, instance retrieval allows stored examples to be retrieved faster than step-by-step reasoning can recompute them. Repeated problems create discrete instances that outperform any general algorithm, and retrieval usually wins (Logan, 1988). Performance then mirrors the distribution of experienced cases: it is efficient when current cues resemble past ones and breaks when contexts change, which explains both the power and the fragility of automaticity (Shiffrin & Schneider, 1977; Logan, 1988).

When automaticity fails, individuals need to shift from automatic to controlled stimulus processing, which is capacity-limited, serial, and attention-demanding (Shiffrin & Schneider, 1977). In ill-structured, high-variability domains, mature performance often appears intuitive because decisions are immediate and hard to articulate. Within the SAT framework, this appearance reflects control by dense, context-sensitive schemas and instance memories, rather than explicit rule-following (Dreyfus & Dreyfus, 1986; Dreyfus, 2004; Benner, 1984). Existing formal models capture much of this via pattern-triggered productions and instance retrieval, yet fully explaining holistic, fine-grained discriminations under variability remains an open problem (Anderson, 1983; Logan, 1988).

5.3.1.2 Skill Decay and Retention

Foundational studies define skill decay as the loss of trained or acquired abilities after periods of nonuse and identify two primary determinants: the duration of the non-practice interval and the depth of original training (Arthur Jr. et al., 1998; Kluge & Frank, 2014; Gardlin & Sitterley, 1972; Naylor et al., 1968). End-of-training proficiency is the strongest predictor of long-term retention, with higher terminal performance buying slower decay.

Different knowledge components fade at different rates, and residual structures make relearning faster than first learning. Declarative details tend to fade relatively quickly, whereas procedural control (productions and schemas) is more robust (Singley & Anderson, 1989). On resumption, relearning proceeds faster than initial acquisition because partially intact productions and cue–response mappings can be reactivated (Singley & Anderson, 1989; Ericsson et al., 1993). In addition, retention is context-dependent: recall is strongest when performance conditions closely match those of the original learning environment, reflecting the role of contextual cues in retrieval (Arthur Jr. et al., 1998; Childs & Spears, 1986; Wang et al., 2013; Weaver et al., 2012).

Technology accelerates decay when it displaces the very practice that sustains competence. In aviation, electronic flight-instrument systems reduced pilots' direct engagement with manual navigation, followed by declines in their ability to input and verify navigational data – an erosion tied to less frequent hands-on execution (Childs & Spears, 1986). In manufacturing, digital machining, robotics, and automation have diminished traditional craftsmanship and on-the-fly problem solving as repetitive, embodied routines lose reinforcement, weakening the judgment and "feel" built through years of practice (Wang et al., 2024).

Widespread AI integration amplifies skill-decay dynamics by automating steps that humans would otherwise practice. Partial or full automation reduces opportunities for deliberate engagement, which are necessary for developing and maintaining expertise. In medicine, for example, reliance on AI-assisted diagnostics has been linked to a decline in clinicians' reasoning and examination skills; radiologists who defer to model outputs gradually lose their pattern-recognition capability (Macnamara et al., 2024; Budzyń et al., 2025). This reflects a cognitive burden shift: AI provides short-term support but fosters long-term dependency. Reliance reduces practice, weakened skills increase reliance, and dependence deepens (Parasuraman & Manzey, 2010; Rinta-Kahila et al., 2023; Mahajan, 2025). The result is fragile competence: individuals lose the basis for generating or critically evaluating the AI outputs they depend on, often without recognising the gap between perceived and actual ability (Rinta-Kahila et al., 2023; Macnamara et al., 2024; Yadav, 2025).

5.3.1.3 Skill Transfer

Skill transfer is the reuse of existing productions and schemas onto a new task. Singley & Anderson (1989) describe two effective forms of transfer. Near transfer occurs when performance carries over to a new task that is very similar to the practiced one, because the same underlying elements and representations are reused. In contrast, far transfer describes the situation when performance carries over to a task that's structurally different at the surface but shares abstract principles. It requires that learners represent both tasks at the same level of abstraction and have practiced using those abstractions so compiled control can re-ground in the new domain. When schemas are applied with misaligned representations, negative transfer follows (Singley & Anderson, 1989). Under AI, this often appears as confident misuse: fluent outputs mask representational mismatches, so experts retain an edge in detecting remapped cues and re-grounding the control schema before accepting tool output (Chi et al., 1981; Larkin et al., 1980; Chase & Simon, 1973).

5.3.1.4 Prompting: The Rise of Representational Fluency and Test Design

Prompting has emerged as a new skill in GenAI-integrated workplaces. From the perspective of SAT, prompting can be thought of as the act of and ability to formalize (1) domain representations and (2) criteria into inputs a GenAI model can absorb and effectively process (Crowston & Bolici, 2025).

Effective prompts depend on mastering contextual representations (notations, units, schemas that reveal deep relations) and context-relevant evaluation criteria (constraints, thresholds, tests that define "acceptable" output). Experts encode problems in domain-valid forms (e.g., invariants, interface contracts, safety limits), so their prompts touch the task's deep structure, not just surface cues (Chi et al., 1981; Larkin et al., 1980). This is classic SAT: compiled schemas/chunks guide attention to diagnostic features, and productions bind cues to actions.

Planning, decomposition, evaluation, and repair are strategic routines that generalize in form but only work when re-grounded in the host domain's models and tests (Singley & Anderson, 1989). Decomposition is problem-space construction (states, operators, goals); evaluation is stress-testing against domain criteria; repair is constraint-driven editing. These routines give experts an LLM advantage because they are conditionalized on the right cues and acceptance tests through practice.

Model brittleness makes these conditionalized strategies economically valuable. Evidence shows GenAI can have large gains in fluency and speed, alongside persistent reasoning limits: over-reliance on surface patterns, stronger induction than deduction, and fragility on multi-hop/compositional problems (Tong et al., 2024; Feng et al., 2024; Cheng et al., 2024; Hosseini et al., 2024; Wu et al., 2024). Prompting tactics, such as decomposed and least-to-most prompting, and semantic-consistency checks, help by structuring search (Khot et al., 2022; Zhou et al., 2022; Tang et al., 2023), but they do not replace domain-grounded evaluation. Hence, organizations increasingly value the human ability to design tests, identify anomalies, and implement precise corrections.

The net implication is a revaluation of expertise toward representational fluency and test design. Automation substitutes narrow, fixed procedures first, shifting advantage to domain-conditional control (i.e., the ability to name the right abstractions, articulate constraints, and operationalize acceptance tests) (Thornhill-Miller et al., 2023; OECD, 2023). Under GenAI, experts outperform not because they possess innate intelligence or domain-free talents, but because they bring richer abstractions and sharper fit tests that make prompting, interpretation, and refinement more reliably effective.

5.3.2 Organizational Memory

Organizational memory (OM) refers to the stored, selectively retrievable residue of past experiences that can be utilized in present and future decisions (Walsh & Ungson, 1991). In the organizational learning literature, OM is the infrastructure that makes learning consequential. Levitt & March (1988) view learning as the encoding of inferences from history into routines, procedures, and beliefs that persist beyond individuals, and OM as the retention and access apparatus that allows those encodings to outlive turnover and guide behavior. Huber (1991) defines learning as changes in an entity's potential behaviors; OM determines whether those potentialities are retained, shared, and retrievable enough to influence actual choices (Levitt & March, 1988; Huber, 1991).

This framing of OM, common to Walsh and Ungson's (1991) account of memory bins, Levitt and March's (1988) encoding view of learning, and Huber's (1991) information-processing synthesis, emphasizes that OM is distributed across individual experiences, transformation routines, organizational structure, cultural narratives, and physical organizational artifacts.

Accordingly, OM is the coupling between organizational history and present/future actions: a socio-technical, decision-proximate filter that both enables adaptation and, when mis-designed or politicized, locks organizations into systematic mislearning (Walsh & Ungson, 1991; Levitt & March, 1988; Huber, 1991; March & Olsen, 1975).

5.3.2.1 OM Encoding, Retention, Retrieval, and Change

OM couples yesterday's experience to today's choice. That coupling unfolds as a loop: (1) experiences are encoded by individuals, stored in one or more OM carriers, retrieved at the moment of need, and then either reinforced or replaced as consequences accumulate. This section provides a detailed description of this loop.

Encoding

Encoding is the process through which "what happened" and "what it meant" get translated from lived experience into information that can later guide behavior. Decision stimuli and organizational responses are first noticed and framed by individuals, who interpret them, produce outcomes, and then commit some version of those interpretations into OM.

In the behavioral learning tradition, organizations learn by encoding inferences from history into routines that guide behavior (Levitt & March, 1988). Huber (1991) generalizes this: an organization learns when information processing expands its potential behaviors, and encoding is the entry point to that expansion. For OM, this implies two things. First, encoding is selective: only some cues and responses are noticed, discussed, or recorded. Second, encoding is interpretive: stored material is a filtered account shaped by existing beliefs, frames, and power relations. Cognitive orientations and established attention structures influence both selectivity and interpretation, making some aspects more salient than others (Ocasio, 1997).

In Walsh & Ungson's (1991) terms, what gets encoded as potential memory is decision information about the "who, what, when, where, why, and how" of past stimuli and responses. Whether and how these elements survive as organizational memory depends on the subsequent phases: retention and retrieval.

Retention

Retention refers to the persistence of encoded experience over time. Once interpretations about what the stimulus was, which response was chosen, and why it was considered appropriate have been formed, some of that information becomes durable and can, in principle, be brought to bear on later decisions. This is the essential function of OM.

Walsh & Ungson (1991) conceptualize OM as a set of distributed retention facilities that store decision stimuli and responses in different forms. For the purposes of this section, it is sufficient to note three general features of retention, leaving the specific bins to the next section. First, OM consists of two primary components: decision stimuli (problems that occurred in the past) and organizational response (corresponding resolutive actions). Each OM bin stores different aspects of these two components.

Second, retention is partial as not all aspects of a decision episode are preserved. Information about responses (what was done, how, and by whom) tends to be retained more reliably than information about stimuli or rationales (why it was done). Over time, organizations often remember "how we do things here" but forget "why we started doing them this way". As discussed in the next section, this is a direct implication of how retention bins work, with only individuals (and, to a more limited extent, culture) retaining richer information on stimuli, responses, and their perceived causal links.

Third, retention is structured. What persists is shaped by cognitive structures such as schemas, interpretive frames, and dominant logics, which filter what is deemed relevant and compatible enough to keep (Shrivastava & Schneider, 1984; Prahalad & Bettis, 1986). These structures compress experience into reusable patterns but also discard nuance and minority interpretations. As with encoding, attention structures influence what is retained by highlighting some aspects and downplaying others. This process is path dependent: organizations tend to retain information that fits existing schemas and attention structures, which creates rigidity when environments change.

Lastly, retention is socially maintained. Organizational memory is sustained by stories, labels, and lessons learned that are retold and re-enacted within the organization. As Douglas (1986) notes, institutionalized patterns selectively "remind" members of some events and effectively silence others, thereby stabilizing particular interpretations of the past.

Encoding and retention transform specific episodes into a stock of decision knowledge that can later constrain or enable action. Retrieval is the phase in which the stock becomes consequential for current choices.

Retrieval

Retrieval is the process of activating stored decision information in response to a current problem or opportunity. Following Walsh & Ungson (1991), retrieval can be located on a continuum from automatic to controlled.

On the automatic side, well-learned patterns fire intuitively and automatically when familiar cues appear. At the organizational level, this looks like present behavior echoing previous practices and procedures embedded in routines, roles, and taken-for-granted ways of working. Automatic retrieval economizes attention (Ocasio, 1997) and speeds coordination, but it is largely blind to subtle context shifts and can reproduce outdated responses.

On the controlled side, actors deliberately search for precedents and lessons. They ask others, "Who has seen this before?" or "How did we handle this last time?", they consult documents and systems, and they actively compare the current situation to remembered cases. This mode relies on both knowing where memory is retained and having usable access mechanisms (indices, labels, search tools).

In practice, what gets retrieved is not a neutral sample of what the organization knows. It is filtered by attention structures (Ocasio, 1997). Routine cues, labels, and category systems make some histories salient and others invisible; powerful actors can emphasize particular precedents and suppress others, thereby steering how the past is allowed to matter (March & Olsen, 1975; Douglas, 1986). Under ambiguity about what happened, why it happened, and which precedent fits, retrieval is inherently interpretive and political.

Persistence, Obsolescence, and Unlearning

Because encoding, retention, and retrieval are ongoing, organizational memory behaves like an evolving control system. Experiences that are repeatedly encoded, retained in accessible forms, and easily retrieved become self-reinforcing: they are used more, questioned less, and gradually compiled into automatic responses. When environments are stable, this produces efficiency; when they shift, the same dynamics create maladaptation: organizations become "better and better at doing the wrong thing" (Levitt & March, 1988). Three linked dynamics follow directly from how the memory loop operates in practice: persistence, obsolescence, and unlearning.

Persistence arises from use-dependent reinforcement. Recurrent problems trigger similar cues, which activate familiar responses that become increasingly automatized into routines (Cohen & Bacdayan, 1994). Over time, “how we do X” becomes fast, taken for granted, and embedded in processes, roles, and spaces, so much retrieval is automatic rather than deliberative (Walsh & Ungson, 1991). Cultural carriers add a further layer: canonical stories, labels, and schemas are retold and taught to newcomers, making certain interpretations readily available as shared “obvious” frames (Douglas, 1986). Frequent use lowers retrieval costs, which in turn increases future use and further reinforces the same patterns.

Obsolescence is the flip side of that efficiency. Because encoding is tied to the conditions under which responses were learned, environmental change can decouple current problems from the old cues that still trigger them. Heavy codification and workflow hardening (“encased learning”) narrow attention to the existing solution set and make variants difficult to notice or legitimate (Stein & Zwass, 1995). Cultural narratives then stabilize convenient accounts of past success, while the “why” behind older decisions decays or distorts as it is transmitted (Levitt & March, 1988; Walsh & Ungson, 1991). The result is path dependence: success → retrieval → reinforcement of stories, indices, and physical cues → more success-biased retrieval, until performance failures reveal that the compiled response no longer fits.

Unlearning is rarely literal erasure. Instead, new patterns interfere with and suppress old ones. At the micro level, negative transfer occurs when familiar cues activate the wrong subroutine until overlearning reconfigures the cue-to-response link (Singley & Anderson, 1989; Cohen & Bacdayan, 1994). At the system level, Walsh and Ungson (1991) argue that, because much retrieval from embedded carriers is automatic, “forgetting” typically requires altering the conditions of retention and retrieval (i.e., replacing routines, roles, and ecological cues, not just changing top managers or issuing new guidance). OMIS can support this by marking knowledge as current versus legacy, attaching rationales and conditions of applicability, and enforcing review cycles (Stein & Zwass, 1995; Olivera, 2000), thereby aligning what is easily retrievable with the organization’s updated direction.

5.3.2.2 Organizational Memory Carriers

Carriers are the bins where OM is retained and from which it can be retrieved. OM is distributed across five bins: individuals, organizational culture, transformations, structures, and ecology (Walsh & Ungson, 1991). These bins are connected by Organizational Memory Information Systems (OMIS), a socio-technical infrastructure that connects and enriches multiple bins, making memory usable at decision time – effectively supporting retrieval (Walsh & Ungson, 1991; Huber, 1991; Stein & Zwass, 1995; Olivera, 2000).

This section presents the five OM carriers and the role of OMIS as the foundation on which the model will be built.

Individuals

Individuals are the main entry point through which decision stimuli and organizational responses are encoded (Walsh & Ungson, 1991). They hold detailed recollections of what happened, the context and constraints, and, crucially, the rationales for "why we did it that way". Other carriers mostly store aspects of how responses are executed (who, what, when, where); the underlying why is largely preserved only in individual memory. Every decision episode is first filtered through individual cognition: people notice cues, frame them using prior knowledge, and connect them to past experiences. Encoding is therefore never neutral but is shaped by prior experience and pattern-recognition capabilities, which give seasoned members a central role in both encoding and retrieval, especially in ambiguous situations.

For individual knowledge to become organizational memory, others must be able to locate it. Transactive memory (shared understanding of "who knows what") performs this routing work, often via roles and reputations. Retrieval blends automatic and deliberate processes: in familiar contexts, colleagues know whom to contact; under ambiguity, they search, compare accounts, and reconstruct precedents. Turnover, overlap, and outdated directories easily disrupt these links. Individuals also mediate between personal memory and other carriers by deciding what to document, how to enact routines, and when to redesign structures or systems. They are not passive storage but active curators whose selective recall, storytelling, and recording choices continually reshape what the organization "remembers" and how that memory constrains or enables future action.

Culture

Culture functions as a supra-individual carrier of organizational memory. It retains shared interpretations of decision stimuli and responses (e.g., labels, categories, norms, and stories) that define what events mean and which lessons are considered legitimate (Walsh & Ungson, 1991). Through repeated use, these interpretive templates become taken for granted and guide how new situations are framed.

Walsh & Ungson (1991) caution on two points: (1) only individuals fully reconstruct the "why" behind decision stimuli and organizational responses, whereas culture transmits received wisdom about it; and (2) cultural accounts can distort or decay and are hard for the organization to query or correct from within. That is why cultural memory is both a powerful low-cost retrieval channel and a vector for narrative lock-in and political curation. When a new issue arises, individuals draw on cultural labels (e.g., "this is a quality issue", "this is just politics"), metaphors, and stories that make some precedents salient and render others invisible (March & Olsen, 1975; Weick, 1995). This means that what "comes to mind" at decision time is heavily filtered by shared narratives about the organization's past successes, failures, and identity.

Transformations

Transformations are the technical and procedural processes that convert inputs into outputs (Walsh & Ungson, 1991). As a carrier of organizational memory, transformations store "how we respond" to particular classes of stimuli in the form of patterned work processes (Walsh & Ungson, 1991), establishing formulas about what is to be done, how and when to do it, and where to do it, and even imply the skills required. However, the "why" requires concurrent consideration of both the stimulus and the response (a causal relation) that only individuals can account for.

Once a way of handling an issue is embedded in task sequences, checklists, scripts, or system configurations, it becomes a routine: a default response that can be triggered with little or no deliberation. This explains how, when a familiar order type enters a production system, or a specific error code appears in a diagnostic tool, the embedded process "knows" what to do. Retrieval is automatic: the presence of the cue routes the case through a pre-existing path. This economizes attention but also makes the organization vulnerable to context shifts (Feldman & Pentland, 2003).

Structure

The structure of an organization consists of the formal arrangements that allocate roles, authority, reporting relationships, and coordination mechanisms. As an OM carrier, the structure stores who is responsible for what (Walsh & Ungson, 1991).

In classic terms, structural design embeds "programs" for action: recurrent situations are linked to specific units, roles, and standard decision procedures (March & Simon, 1958; Cyert & March, 1963). A risk that is historically framed as "technical" will be routed to engineering; one framed as "compliance" will be routed to legal. This routing reflects past categorizations and political settlements that have become structural facts. In Ocasio's (1997) terms, structures shape the "attention rules" of the organization by defining which issues are legitimate for which actors, and which channels must be followed for a response.

Structures, therefore, retain response patterns and enactment conditions more than detailed records of stimuli or rationales. The organization may forget why a particular committee exists or why a decision threshold was set at a specific level, yet the committee continues to meet, and the threshold remains in place to guide choices. This is structural memory: past solutions survive as current authority distributions, decision rights, and veto points.

Structural memory is also rigid. Changes in reporting lines, governance arrangements, or formal responsibilities typically require significant political work and often lag behind shifts in environment or strategy (Hannan & Freeman, 1984). As a result, structures can perpetuate outdated problem framings and response paths, reinforcing some trajectories of learning while making alternative uses of the past difficult to enact, even when individuals or culture might support them.

Ecology

Ecology refers to the organization's material and spatial environment, including buildings, layouts, artifacts, and technological setups, that shape how work is done (Walsh & Ungson, 1991). As a memory carrier, ecology stores where and how past responses occurred: i.e., it preserves and re-enacts prior responses through the very spaces and artifacts in which work is situated.

Work on distributed cognition has shown how cognition is "offloaded" onto physical structures and artifacts (Hutchins, 1995). Checklists taped to machines, whiteboards in war rooms, kanban boards, and dashboards at control centers all materialize prior learning about what information must be visible to act effectively. Similarly, workspace design embeds assumptions about necessary interactions: open-plan rooms, project rooms, and collocated teams reflect past beliefs about which roles must coordinate intensely, and they continue to channel who talks to whom about what (Ocasio, 1997). In OM terms, ecology retains response patterns as ready possibilities for action designed into the environment.

Ecology also functions as a cueing system for retrieval; for example, a workstation arranged in a particular order "reminds" the operator of the correct sequence. Over time, these ecological cues become part of the taken-for-granted way of working, so that changing them can significantly disrupt or redirect behavior (Feldman & Pentland, 2003).

Like structures, ecological arrangements are rigid. Reconfiguring plants, offices, or digital workspaces is costly, which means that historical designs often outlive the conditions that produced them. Old layouts can continue to channel attention and interaction along outdated lines (Ocasio, 1997), making some uses of the past highly salient and others practically unreachable, regardless of what individuals remember or culture endorses.

5.3.2.3 Organizational Memory Information Systems (OMIS)

OMIS are information systems that function to provide a means by which knowledge from the past is brought to bear on present activities, thus resulting in increased levels of effectiveness for the organization (Stein & Zwass, 1995). Thus, OMIS are not memory bins in themselves (Walsh & Ungson, 1991), but the infrastructures that make distributed organizational memory usable at decision time. Traditionally, they include databases, document repositories, case libraries, expertise directories, groupware, and workflow systems that are explicitly designed to capture, index, and deliver decision-relevant traces of the past when actors face current problems (Huber, 1991; Stein & Zwass, 1995; Olivera, 2000).

Stein & Zwass (1995) root their OMIS framework in an effectiveness perspective. They argue that an OMIS is supposed to support four broad effectiveness functions: (1) integration and coordination of information across units (integrative function), (2) adaptation to a changing environment (adaptive function), (3) goal setting, monitoring, and control (goal-attainment function), and (4) preservation and development of organizational patterns in human resources, culture, and routines (pattern-maintenance function).

This effectiveness-oriented perspective is grounded in a deeper mnemonic layer comprising five memory functions: acquisition, retention, maintenance, search, and retrieval. Acquisition concerns how information about decision stimuli, responses, and outcomes is captured and transferred into the system (e.g., via logging mechanisms, upload pipelines, or expert annotation). Retention concerns how that information is encoded and persists over time in forms that can, in principle, be accessed later (e.g., schemas, cases, annotated documents). Maintenance concerns how the stored contents are curated (i.e., updated, pruned, re-indexed, or selectively forgotten) so that they remain accurate and usable as conditions change. Search refers to selecting which parts of retained information are relevant to a given problem or goal, while retrieval reconstructs or presents that information in a form that actually supports current action (Stein & Zwass, 1995). These mnemonic functions closely map onto the OM loop described earlier, making explicit the technical and procedural requirements for each phase.

OMIS, in this sense, are mediating infrastructures that execute these mnemonic functions over contents distributed across the Walsh & Ungson (1991) carriers rather than replacing those carriers. They aggregate and stabilize dispersed traces (e.g., emails, tickets, designs, meeting minutes, performance indicators) linking stimuli, responses, and outcomes to actors, units, and time periods. At the same time, OMIS provide the indices, taxonomies, search tools, and cross-references that make memory findable and re-usable: expertise directories extend transactive memory ("who knows what"), case taxonomies guide analogical retrieval ("which prior case is like this"), and workflow systems embed previously learned responses into repeatable process paths (Huber, 1991; Olivera, 2000).

OMIS are fragile and path-dependent. Without active curation, repositories become noisy, redundant, and hard to use (Olivera, 2000). OMIS also inherits existing attentional biases: what is encoded, how it is indexed, and who has access are shaped by prior structures, power relations, and cultural frames. As a result, some histories are systematically overrepresented, while others remain invisible; therefore, the "remembered" past that OMIS projects into present decisions is always a selective and political sample of organizational experience (Walsh & Ungson, 1991; Stein & Zwass, 1995).

In the remainder of the chapter, I treat this effectiveness- and mnemonic-based view of OMIS as the baseline for theorizing GenAI-based OMIS. The argument made is that contemporary GenAI systems can, in principle, over-satisfy some mnemonic functions by implicitly acquiring and retaining vast numbers of interaction traces at low marginal cost, while radically changing how search and retrieval are carried out. The question then becomes not whether they can act as OMIS, but how their specific way of performing acquisition, retention, maintenance, search, and retrieval affects the quality of organizational memory and, ultimately, organizational effectiveness.

5.4 MODELS AND IMPLICATIONS

5.4.1 Model 1: GenAI Impact on Expertise Development

Model 1 theorizes how GenAI reshapes the relationship between human expertise and observable performance, explaining why visible productivity gains can frequently coexist with weakened or stalled skill development[1].

As explained in the methods section, eight semi-structured interviews with practitioners informed Model 1's development. Practitioners reported substantial efficiency improvements, particularly for junior staff. The speed of production has fundamentally shifted; as I3 noted, an intern was able to produce a functioning prototype in two days using generative tools – a task that previously required weeks of supervised work. This acceleration offers practitioners a sensation akin to a "superpower", as described by I5, who noted that while he can "take on more tasks", he must now consciously "brake" to ensure the cognitive effort remains his own. Similarly, I4 argued for a ruthless delegation of cognitive load, stating, "Why would I spend time learning about [intensively technical details] when I can just plug it in... it's not worth my brain cells".

However, these visible gains were accompanied by profound concerns regarding blocked proceduralization. Interviewees noted that newcomers frequently utilized AI to generate deliverables they lacked the expertise to critically evaluate, effectively substituting execution for learning (I1, I2, I6). I2 offered a particularly evocative metaphor for this phenomenon, describing the current development of skills as "building a palace... very beautiful from the outside [but] inside there is no furniture, there are no offices". I6 observed that when juniors are challenged on an AI-generated idea, they cannot defend it because "it is not yours... you lack the true ownership of that idea". Consequently, the deep thinking required to empathize with users or foresee consequences is being replaced by a montage of AI responses.

Simultaneously, the interviewees highlighted a persistent gap between technical AI potential and realized utility. Novices struggled with problem decomposition and context specification, often missing critical "corner cases" (I1–I3, I6). I1 recounted how AI-generated user stories, while superficially valid, missed the "unsaid" nuances of human relations, taking "blunders" that a human analyst would have caught. This deficit shifts the workload rather than reducing it; I7 noted that he now must revise the work of juniors who lack "systemic

[1] Model 1 presented in this chapter is derived from a manuscript co-authored with Prof. Francesco Bolici and Prof. Kevin Crowston, currently under review at the *European Management Journal.*

vision," effectively creating a bottleneck where the senior must "put their hands back on" the product because they cannot fully trust the junior's output.

Finally, the interviewees expressed emerging concerns about long-term atrophy as they began offloading to GenAI an increasing number of subtasks (I5, I6). I8 warned that this leads to a "flattening of knowledge", comparing it to a mathematician who can use a calculator but no longer remembers the theory behind division. Collectively, these accounts substantiate the model's core propositions: GenAI compresses observable performance differentials across expertise levels, threatens foundational skill acquisition through displacement, and generates a systematic utility gap driven by underdeveloped representational schemas.

Building on the background definitions of skills and task complexity, the model rests on three elements. First, skills are defined as individual qualities that are both productive of value and expandable through training and deliberate practice. Second, tasks are categorized by varying degrees of complexity: component complexity (number of distinct, isolatable sub-tasks), coordinative complexity (tasks requiring integration across elements), and dynamic complexity (tasks evolving over time). Crucially, current GenAI models are optimized to manage component complexity but remain poorly equipped to handle coordinative and dynamic complexity, as these require forms of contextual awareness that are absent from training data. Third, worker expertise is stratified into three analytical levels. (1) Novices, who can rely on declarative knowledge to manage component complexity, but lack the representational schemas and context awareness for effectively managing coordinative and dynamic complexity. (2) Intermediates, who rely on proceduralized knowledge to manage coordinative complexity but struggle with dynamic shifts; and (3) experts, who operate effectively across all levels.

From this architecture, Model 1 proposes three mechanisms. The first is the asymmetric exposure of different expertise levels to GenAI-induced substitution, uncertainty, and augmentation. Novices' work tends to fall within a substitution area, where GenAI can execute the component-heavy tasks they would otherwise need to learn, effectively bypassing the learning. Intermediates cluster in an uncertainty area, where the balance between help and hindrance is ambiguous. Experts, conversely, operate predominantly in an augmentation area, where GenAI supports their workflows without replacing their core value-add. This asymmetry underpins an *Augmentation–Deskilling Paradox*: the same deployments that raise immediate visible performance also crowd out the deliberate practice required for novices to develop foundational skills, creating a long-term capability trap.

The second mechanism describes the *Performance-Leveling Effect*: the compression of observable performance differences across expertise levels. By enabling less experienced workers to produce outputs that approximate those of more expert colleagues, GenAI partially decouples performance from underlying expertise. As a result, observed performance becomes a weaker proxy for actual capability.

The third mechanism, the *AI Capability Loss*, accounts for the systematic gap between what GenAI could theoretically deliver and the value actually realized in practice. Realizing GenAI's full technical potential depends heavily on complementary meta-skills, such as the ability to frame tasks into GenAI-manageable components, specify relevant context, and rigorously evaluate outputs. Paradoxically, these meta-skills require the very domain expertise and representational schemas that GenAI's use may prevent novices from developing. When these meta-skills are missing, a substantial share of GenAI's potential remains dormant, creating a scenario where those who need the tool most are least equipped to extract its full value.

5.4.1.1 Exposure to Skill Decay and Non-Development

Exposure to skill decay and non-development is not a uniform risk but structurally arises from the interplay between expertise levels, task complexity, and the relative capability gap between humans and GenAI. Model 1 conceptualizes this interplay by distinguishing three distinct implementation regimes (i.e., substitution, uncertainty, and augmentation) determined by whether GenAI capabilities exceed, approximate, or fall below human capabilities at a specific level of expertise.

1. *Substitution Regime*. In the Human Substitution Regime, GenAI capabilities exceed human capabilities, enabling individuals to rely on the system to fully execute tasks they would otherwise perform themselves. Within this region, AI-based automation significantly reduces direct task involvement, thereby narrowing the opportunities for hands-on practice. When a critical mass of daily tasks falls within this range, the likely outcome is a failure of skill development. By consistently offloading work to the AI, learners crowd out the deliberate practice on fundamentals required to build deep expertise, effectively short-circuiting the proceduralization process emphasized in skill acquisition theory.
2. *Uncertainty Regime*. The Uncertainty Regime exists at the threshold where AI capabilities roughly align with human skills. In this zone, the impact of GenAI on

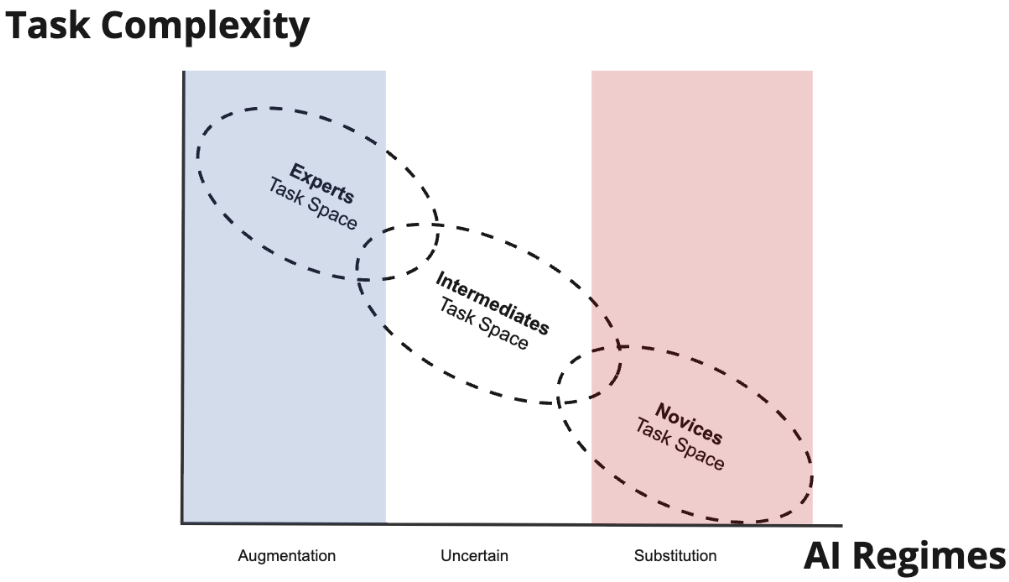
skill development is ambiguous because tasks do not fit neatly into substitution or augmentation scenarios; the AI may produce high-quality outputs for some instances while failing on others of seemingly similar complexity. Consequently, automation is inconsistent, yet clear augmentation is not established. In this ambiguous zone, outcomes are not determined by the technology alone but hinge on individual differences – specifically, the user's propensity to learn, intrinsic motivation, and the development of emerging AI-specific meta-skills such as prompt engineering and critical output evaluation.

3. *Augmentation Regime*. In the Human Augmentation Regime, task complexity exceeds current GenAI capabilities, preventing full automation. Here, AI tools serve only to handle specific subtasks within their capability range, leaving the most complex, coordinative, and dynamic components to human experts. This division of labor allows for cognitive offloading: AI mitigates the temporal and cognitive expenditure on simpler elements, enabling humans to focus their attention on more demanding aspects of the problem. Under these conditions, the model predicts that humans gain high-value problem-solving experience and refine their expertise, developing the complementary abilities necessary to leverage augmentation effectively.

As illustrated in Figure 5.1, the task spaces of novices, intermediates, and experts are distributed asymmetrically across these three regimes. Novices, whose work largely consists of low-complexity tasks, operate predominantly in the substitution area; because a significant share of their workload falls beneath the GenAI capability frontier, they face the highest structural risk of non-development. Intermediates occupy the uncertainty area, where outcomes are divergent and dependent on integration strategies. Experts, dealing with high coordinative and dynamic complexity, reside firmly in the augmentation area.

This configuration substantiates the *Augmentation–Deskilling Paradox*. The same GenAI deployment that visibly increases immediate output for novices also concentrates their activity in the substitution regime, where the scaffolding provided by GenAI bypasses the struggle necessary for foundational learning. Conversely, experts experience net augmentation with a relatively low risk of skill decay. This creates a diverging trajectory where the technology accelerates the performance of the skilled while potentially arresting the development of the unskilled.

FIGURE 5.1: Distribution of novices', intermediates', and experts' task spaces across AI regimes (augmentation, uncertain, substitution).



Practical Implication

High risk of skill development failure for novices. The most profound implication of the *Augmentation–Deskilling Paradox* is the structural vulnerability of novice workers to skill development failure. Because the majority of tasks assigned to novices consist of component complexity, their daily work falls predominantly within the substitution regime. In this context, GenAI does not merely assist; it effectively replaces the need for the human cognitive effort previously required to execute foundational tasks. This substitution creates a disruption in the learning process by allowing novices to bypass the proceduralization stage (i.e., the phase where declarative knowledge is converted into procedural knowledge through repeated application and struggle). Thus, the immediate, visible spike in novice performance masks a latent hollowing out of basic competencies. By significantly reducing the necessity for learning by doing, organizations risk creating a workforce that produces higher expertise-level outputs without possessing the underlying mental models or resilience to troubleshoot errors. This raises a fundamental systemic question for the future of the work: if GenAI stalls skill development at the entry level of the pipeline for generating future experts, who and how high-complexity tasks will be handled in the future remains unclear.

5.4.1.2 AI-Performance Leveling Effect

This section conceptualizes the *Performance Leveling Effect*, a mechanism through which GenAI deployment compresses observed performance differentials across expertise levels, thereby enabling less skilled individuals to approximate the output quality of expert workers.

The *Performance Leveling Effect* is provided in Figure 5.2. The model assumes a monotonic relationship between expertise and performance: as expertise increases, performance improves. While the specific functional form may vary, the central premise is that higher expertise consistently supports higher performance, with novices, intermediates, and experts situated along an upward-sloping curve.

FIGURE 5.2: AI Performance-Leveling Effect

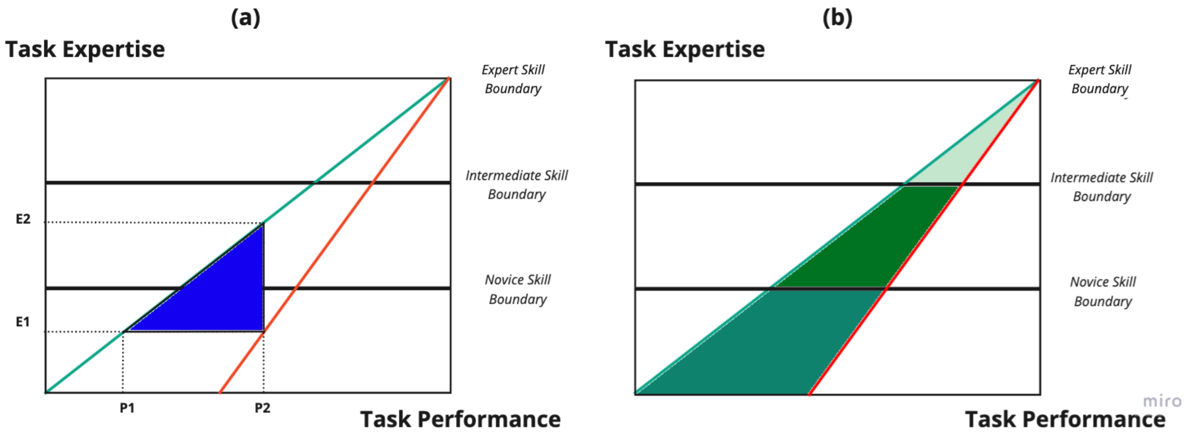


Figure 5.2a illustrates AI-added expertise: the gap between a worker's actual expertise and the level that would be inferred from their AI-assisted performance. A novice at expertise level E1 may move from unaided performance P1 to AI-assisted performance P2. From an external perspective, P2 corresponds to the unaided performance typically produced by an intermediate at expertise level E2. The novice thus appears to operate at an intermediate level without having acquired the underlying skill. This creates an illusion of competence, as AI-added expertise may be misread as human expertise. The blue area represents the *Performance-Leveling Effect*: the combination of higher performance and AI-added expertise attributable to AI use.

Figure 5.2b generalizes this logic across expertise bands. The green areas represent the maximum *Performance-Leveling Effect* available from novices (dark green) to experts (light green). As shown, GenAI exhibits its strongest leveling potential within the novice range, with positive but diminishing effects for intermediates and limited additional gains for experts. Thus, especially for tasks dominated by component complexity but low coordinative and dynamic complexity, GenAI compresses performance differences between lower- and higher-expertise workers, especially at the lower end of the spectrum.

The Performance-Leveling Effect thus challenges the traditional assumption that expertise and performance are tightly coupled. For tasks amenable to AI assistance, novices using AI can attain performance levels approaching those of experts, a pattern consistent with empirical findings in customer support, consulting, writing, and programming (Brynjolfsson et al., 2023; Dell'Acqua et al., 2023; Noy & Zhang, 2023; Peng et al., 2023).

Practical Implications

Observable performance is a poor proxy for underlying expertise. Because AI-added expertise allows novices to reach performance levels closer to those of experts, observed performance on AI-supported tasks no longer reliably signals underlying human skill. For AI-amenable work, a high-performing novice using AI and a high-performing expert using AI can look similar in output quality, even though their internal capabilities differ markedly. This complicates assessment, promotion, and selection decisions that rely on performance as an indicator of expertise.

The Erosion of Training Incentives. By narrowing performance gaps, the *Performance-Leveling Effect* enables organizations to meet immediate performance targets with a less skilled workforce. This may potentially create a dynamic where skill development appears less critical for achieving acceptable outcomes; if novices can deliver "good enough" performance via GenAI scaffolding, the investment in foundational training may be viewed as less urgent.

Leveling is strongest where skill development needs are greatest. The model shows that the *Performance-Leveling Effect* is most pronounced for novices, whose foundational skills are still under development, and weakest for experts, who are already near their performance ceiling. This asymmetry means that the group yielding the largest marginal productivity gains from GenAI is also the group most at risk of skill development failure when GenAI substitutes for learning-by-doing. In doing so, the leveling effect reinforces the *Augmentation–Deskilling Paradox* in the previous section: organizations can maximize visible productivity by deploying GenAI extensively among novices, but this is precisely where the decoupling of performance from expertise most seriously undermines the future capability pipeline.

5.4.1.3 AI Capability Loss

The final component of the model addresses the disparity between the theoretical potential of GenAI tools and their realized value in practice. This phenomenon is conceptualized here as *AI Capability Loss*: the structural gap between the performance achievable under optimal human-AI orchestration and the performance actually realized given the user's limitations in meta-skills.

FIGURE 5.3: AI Capability Loss

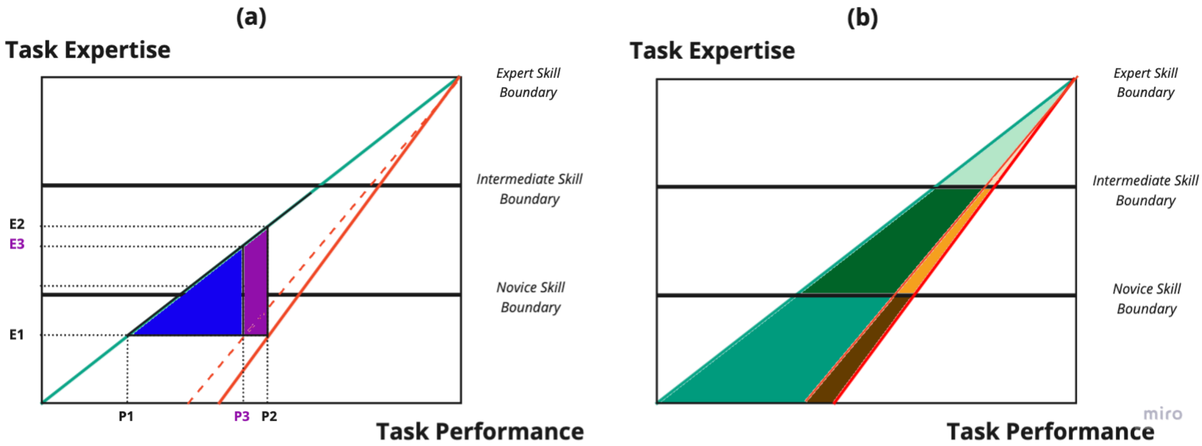


Figure 5.3(a) illustrates this mechanism for a novice attempting a task characterized by high coordinative and dynamic complexity. The green line represents baseline unaided performance, which scales with expertise. The solid red line depicts the theoretical performance frontier achievable if the user could fully orchestrate the model's capabilities. The dashed red line represents the actual realized performance when orchestration meta-skills are weak. Consider a novice at expertise level E1 with baseline performance P1. With optimal AI orchestration, this user could theoretically reach performance level P2. However, due to an inability to provide necessary contextual cues or detect subtle hallucinations, the user only achieves P3. The resulting performance structure is bipartite: the blue area represents the realized gain from AI usage, while the violet area represents the *AI Capability Loss*: the latent potential of the model that remains inaccessible due to the human user's inability to provide sufficient orchestration.

Figure 5.3(b) extends this logic across the expertise spectrum. Within each band (novice, intermediate, expert), the green shaded areas indicate realized performance gains on complex tasks, while the adjacent orange areas indicate the *AI Capability Loss*. The model shows that the orange region is most extensive for novices and diminishes as expertise increases. Thus, the model posits that the magnitude of *AI Capability Loss* is inversely related to expertise. This indicates that less-expert workers systematically leave the greatest share of the tool's potential unexploited. This inefficiency stems from a dual deficit: novices lack both the domain knowledge required to evaluate output quality and the AI-specific meta-skills needed to orchestrate the system. Consequently, they struggle to decompose tasks appropriately, fail to recognize incomplete or hallucinated outputs, and lack the representational fluency to iteratively refine prompts.

These aspects introduce a nuance to the *Performance Leveling Effect* described in previous sections. While GenAI successfully compresses performance differentials on relatively simple tasks (leveling), the pattern shifts as task complexity rises. When the act of effectively using the model itself becomes cognitively demanding, the binding constraint on performance shifts. It moves from the GenAI's raw technical capability to the human user's ability to direct, audit, and integrate that capability. Thus, in the domain of high complexity, GenAI may amplify rather than reduce performance differences based on the user's orchestration ability.

Practical Implications

Diminishing returns on technical sophistication. As organizations upgrade to increasingly sophisticated GenAI models, the theoretical performance frontier expands. However, if the workforce's ability to decompose tasks and evaluate complex outputs remains static due to stalled skill development, the gap between potential and realized value is likely to widen. This creates a structural risk of sub-utilization, where substantial investments in state-of-the-art model capabilities remain inaccessible because users lack the requisite meta-skills to unlock them. Consequently, the bottleneck to organizational performance is no longer the computational intelligence of the tool, but the human capacity to direct it.

The Necessity and risk of interface bifurcation. The structural nature of AI Capability Loss challenges the prevailing strategy of providing indiscriminate access to open-ended models (like ChatGPT Enterprise). If the gap between potential and realized performance is driven by the user's ability to orchestrate the system (Crowston & Bolici, 2025; Khot et al., 2022; Styve et al., 2024), then effective prompting is effectively an expert-level capability. For novices, the high rate of capability loss suggests that access should not be direct, but channeled through constrained interfaces (e.g., rigid workflows, pre-baked prompt templates, or embedded copilots) that restrict freedom to ensure the model's potential is not dissipated by poor orchestration. However, this architectural solution introduces a second-order effect on skill development. By embedding the necessary orchestration schemas directly into the interface (via guardrails or templates), the organization effectively outsources the cognitive burden of problem decomposition from the user to the system designer. While this may reduce immediate *AI Capability Loss*, it exacerbates the substitution of learning. Because the system provides the structural scaffolding, the novice is never compelled to internalize the mental schemas required to decompose problems or formulate constraints themselves. Consequently, the very guardrails designed to obviate the risk of poor performance simultaneously obviate the friction necessary for the novice to ever transition into an expert.

5.4.2 Model 2: GenAI Impact on Organizational Memory

The model developed in the previous section has dissected the impact of GenAI on individual expertise development and performance. Its organizational implications, however, have only been explored indirectly. Model 2 is designed to address this complementary dimension. It conceptualizes GenAI as an OMIS and examines (1) how it mediates the encoding, retention, and retrieval of organizational experience, and (2) how this, in turn, reweights OM carriers (i.e., individuals, culture, transformations, structures, and ecology).

Model 2 assumes a GenAI system is used across the whole organization (e.g., ChatGPT), is accessible across units, and self-learns from interaction traces and organizational documents that users upload during interactions. This assumption is reasonable, as the most widely used GenAI systems (e.g., ChatGPT, Gemini, Claude) have both an enterprise subscription mode that allows sharing the same profile across members of an organization and a “memory function” that enables them to recall previous interactions and use them for generating responses.

As with Model 1, eight interviews informed the theorization of Model 2. On the encoding side (Section 5.4.2.1), the scope of machine-readable traces has expanded far beyond simple document uploads. Practitioners described a massive ingestion of organizational life, utilizing specialized hardware like recorders to capture spoken meeting dynamics and instantly transmute them into structured minutes and action lists (I3, I7). This process often strips away tacit nuances, yet users reported a sensation of the AI "knowing" them. I6 noted that after intensive use, the AI "knows the context in which I work perfectly... it is almost a colleague". Similarly, I5 observed that the AI inferred his specific role as a Product Designer purely from his interaction history, without ever being explicitly told, leading to unprompted customization of outputs.

Interviewees also described a dual retention layer (Section 5.4.2.2). Explicit knowledge retention is formalized in enterprise tools like Glean (I4) or the enterprise-specific RAG systems, where "bad answers" are flagged by humans and fed back into the system to "fix" the memory (I8). However, a parallel, implicit memory layer is forming within individual chat logs, leading to unexpected "contamination". I5 described instances where the AI cross-referenced distinct projects, suggesting solutions for a family goldsmith business based on unrelated corporate design chats, revealing a "volatile" memory that sometimes merges distinct contexts.

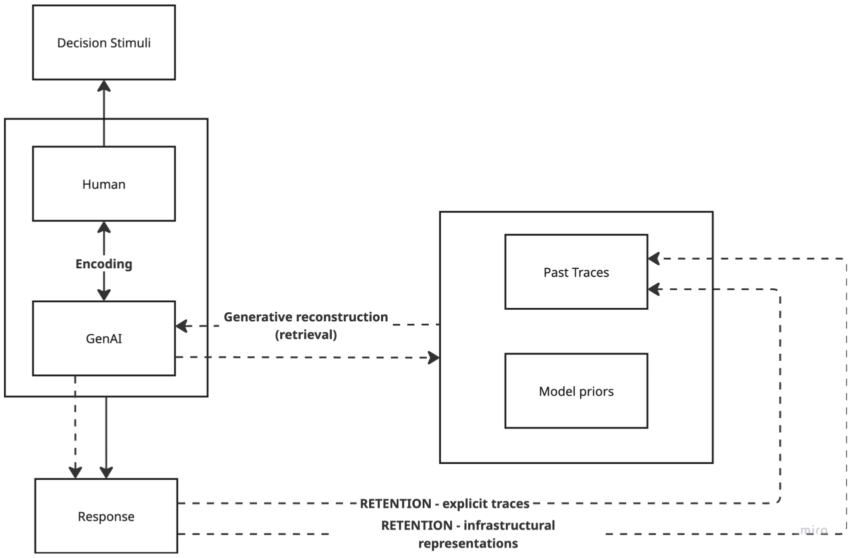
Retrieval is increasingly routed through GenAI, fundamentally altering human-to-human knowledge exchange. The interviewees reported keeping models open during meetings to query documents in real-time rather than asking colleagues, creating a dynamic where, as I2 observed, colleagues ask questions "as if they already had the answer". I3 described this not just as a tool, but as a "third head" in the room, capable of seeing connections the two human counterparts missed.

However, this reliance introduces the risks of pseudo-memory and opaque retrieval (Section 5.4.2.3). Long interaction chains lead to "context drift" and hallucinations; I6 explicitly noted that conversations must be kept short (under 10 interactions) because the model eventually starts "saying nonsense" or mixing topics. Furthermore, the recursive loop of AI training on AI-generated content threatens the integrity of organizational knowledge. I8 warned that without human "gates" to validate memory, the organization risks a "flattening of knowledge," where no one retains the foundational competence to challenge the machine's version of history

Informed by these interviews, the chapter theorizes GenAI-OMIS as an influential channel through which decision episodes are described, stored, and reactivated. Compared to traditional OMIS, GenAI-OMIS (1) centralizes encoding by funneling a large share of problem descriptions and solutions through a single interface; (2) retains experience both explicitly (logs, documents) and implicitly (adjustments to model parameters and retrieval indices); and (3) retrieves organizational memory in generative form, synthetizing patterns from organizational and extra-organizational corpora.

Figure 5.4 presents a high-level view of GenAI-OMIS. The cycle begins with Encoding, where users translate decision stimuli into prompts. The GenAI system performs Retrieval not by looking up static records, but through generative reconstruction, synthesizing an answer from organization-specific past traces and pre-trained model priors. Through interactive Human-AI interactions, the response to the initial decision stimuli is created, and retained by the OMIS in two ways: (1) interaction logs and artifacts, and (2) update of the model's sub-symbolic weights and embeddings.

FIGURE 5.4: The GenAI-Based Organizational Memory Information System



5.4.2.1 GenAI-OMIS in Encoding

Decision stimuli are encoded when users describe issues, constraints, histories, and goals in prompts. To obtain useful answers, they must specify context – what triggered the issue, which actors are involved, and what has already been tried. This forces a selection and framing of experience, making each prompt an interpretive summary of a decision episode (Weick, 1995). Which aspects of the situation are mentioned, how they are grouped, and which constraints are foregrounded depend on the user's existing schemas and representational fluency. Following SAT, this section anticipates that experts are more likely to encode deep structure (diagnostic cues, invariants, critical thresholds), whereas novices tend to describe surface features and local symptoms (Chi et al., 1981; Larkin et al., 1980; see Section 5.3.1.3). As a result, a GenAI-OMIS does not simply record "what happened"; it records how different strata of actors represent what happened, and those representational biases become part of what the organization can later retrieve.

Concurrently with the encoding of decision stimuli, organizational responses are also encoded through users' reactions to the model's suggestions (accepting, revising, or rejecting them). Multi-turn human–GenAI interactions document how initial ideas are refined or abandoned over time. The dialogue thus captures both the evolving representation of the problem and a trajectory of candidate responses and evaluations.

Practical Implications

Acquisition is technically over-satisfied but weakly governed. Viewed through Stein & Zwass's (1995) lens, human–GenAI interactions instantiate the knowledge acquisition function in the mnemonic layer of an OMIS. Their meta-requirements for acquisition include links to internal and external sources, information filters, limited natural-language processing, and intelligent summarizing capabilities. GenAI-OMIS more than fulfills this brief: the chat interface fuses rich natural-language input, automated summarization, and on-demand integration of multiple sources into a single person–machine channel. The bottleneck is no longer technical capture but governance. Because prompts are cheap and ubiquitous, encoding tends to become a firehose of heterogeneous micro-episodes rather than a filtered, curated stream. Thus, the central challenge shifts from getting enough experience into the system to deciding what, and how much, should be allowed in as potential memory.

Encoded memory becomes intrinsically hybrid. Classical OMIS designs implicitly assume that what is acquired and stored is primarily the organization's own experience. In GenAI-OMIS, the "knowledge from the past" that shapes present answers is always a blend of organization-specific traces and a vast, pre-trained world model. Indeed, GenAI systems arrive pre-populated with semantic structure drawn from non-organizational corpora and continue to rely on it when interpreting prompts and proposing solutions. As a result, what is encoded as "our" problem and "our" response is continually influenced by patterns learned elsewhere. This expands the opportunity set, but it also blurs the boundary between organizational memory and imported scripts, making it harder to see which parts of a proposal are grounded in local history and which are generic.

Representational biases are baked into OM. Because encoding occurs through prompts and feedback, GenAI-OMIS stores users' representations of raw events. Differences in schemas, expertise, and political incentives directly influence what is encoded in memory: experts tend to encode deep structure, novices emphasize surface features, and sensitive issues may be sanitized or omitted altogether. Over time, these representational biases accumulate in the interaction logs and in the model's adjusted behavior. The organization's retrievable past increasingly reflects the perspectives of those who routinely use the system and the ways they habitually describe their work, rather than a neutral record of "what happened".

5.4.2.2 GenAI-OMIS in Retention

Retention concerns how encoded experience persists over time, a precondition for its becoming available as input to future decisions (Walsh & Ungson, 1991). This model postulates that in GenAI-OMIS, retention operates on two layers: (1) explicit traces that are recognizable as records, and (2) implicit adjustments to the system's internal representations. Both are downstream of the encoding process described above, and both centralize organizational memory in ways that differ from traditional OMIS.

On the explicit layer, prompts, model outputs, and user edits are stored as interaction logs within GenAI interfaces. These artifacts document the issues that were brought to the system, how they were framed, what the model proposed, and how users refined or overrode those proposals. As with conventional OMIS (Stein & Zwass, 1995; Olivera, 2000), such repositories can be searched, linked to projects or units, and reused as precedents. The difference lies in density and granularity. Because interaction has virtually no fixed cost, it is plausible that individuals will routinely bring even small, local decisions and tentative ideas to the chat – questions that previously would have remained in hallway conversations, personal notes, or not been externalized at all. Thus, GenAI-OMIS accumulates a much thicker layer of fine-grained, micro-decision traces than conventional OMIS, which typically capture only major events or formally documented decisions.

On the implicit layer, interaction traces may update the model itself (e.g., through organization-specific fine-tuning, retrieval-augmented generation, embeddings, or preference-learning mechanisms that adapt responses based on past use). Many contemporary GenAI systems also maintain persistent "memories" of user- or organization-level preferences (e.g., styles, recurring constraints, typical data sources) and use these to steer future outputs. In all such cases, customization is training: repeated patterns of prompts, corrections, and accepted outputs shift the probability distribution over what the system will propose next, even if no engineer explicitly edits a rule or document.

Practical Implications

Memory becomes sub-symbolic and opaque. Instead of storing distinct, human-readable entries such as "Case A → Response B", GenAI-OMIS folds many episodes into high-dimensional parameters and vector representations. The influence of any single episode is dispersed across innumerable weights and embedding dimensions. There is therefore no straightforward way to "open" the memory and inspect or revise a specific precedent: one cannot simply delete "that bad solution from last year". The available levers are indirect: supplying new training data, adjusting which documents are indexed for RAG, or globally constraining behavior through guardrails and policies. This is a different epistemic regime from traditional OMIS, where rules, cases, and documents are explicit objects that can be individually examined, deprecated, or rewritten.

Retention is centralized and strongly path dependent. Across explicit and implicit layers, retention is both centralized and path dependent. It is centralized because a single GenAI-OMIS aggregates traces from many individuals, units, and tasks, becoming a focal locus of what the organization has "seen" and "approved". It is path-dependent because early patterns of use and correction create attractors: popular framings and solutions are overrepresented in the training signal and, therefore, become more salient in future outputs. This can stabilize emerging best practices, but can equally lock in local biases or errors if they are frequent and rarely challenged.

Maintenance and forgetting shift from targeted editing to blunt re-training. The explicit-implicit structure reframes Stein & Zwass's (1995) distinction between knowledge retention and knowledge maintenance. On the retention side, GenAI-OMIS meets its communicability requirement: only what can be rendered as text (or attached artifacts) enters explicit memory, and chat logs implement the temporally indexed "languages" they saw as missing. At the implicit level, however, GenAI breaks its assumption that OMIS contents are explicitly represented in scripts, frames, rules, or semantic networks. Organizational experience is instead folded into sub-symbolic parameters and embeddings that are not individually addressable (see the previous point). Accordingly, maintenance becomes re-training and re-indexing, statistically diluting some patterns and amplifying others without explicit deprecation. The practical consequence is a memory that is highly durable but weakly governable. Once traces are absorbed into the model's internal representations, they are hard to selectively forget or correct, even when recognized as obsolete or biased. Forgetting requires re-engineering the system rather than editing specific items – setting up qualitatively different retrieval dynamics, and new risks of distorted or fabricated OM than in traditional OMIS facilities.

5.4.2.3 GenAI-OMIS in Retrieval

Retrieval is the activation of stored decision information in response to a current problem (Walsh & Ungson, 1991). In a GenAI-OMIS, retrieval is mediated through prompts to the system, which sits between individual sensemaking and the distributed carriers described earlier.

Operationally, retrieval begins when a user frames a current situation as a prompt, describing the issue, constraints, goals, and sometimes their preferred style of response. The GenAI-OMIS then combines (a) organization-specific traces (logs, fine-tuned parameters, RAG indexes) with (b) its pre-existing model of language and world knowledge to generate answers. In contrast to conventional OMIS, the system does not simply “look up” past cases; it synthesizes an answer by interpolating across many traces and artifacts. What is retrieved is therefore a probabilistic reconstruction of “how we (seem to) handle things like this” rather than a single stored precedent.

From the user’s perspective, this creates a hybrid between controlled and automatic retrieval. The query itself is voluntary and effortful (i.e., someone chooses to “ask the model”), but the selection and weighting of past experience are automatic and opaque. Attention is still shaped by familiar forces: problem framing, available labels, and cultural narratives influence what users type, and thus which segments of memory the system can access. Small variations in how a situation is described can lead to different slices of organizational history being surfaced.

Moreover, GenAI outputs often serve as default focal points. Even when formally advisory, they represent starting points that structure subsequent discussion, routinization, and codification. In this way, retrieval through GenAI-OMIS not only mobilizes existing memory but also actively influences how future encoding and retention will unfold, closing the loop in a path-dependent and partially invisible manner.

Practical Implications

Retrieval becomes centralized and depersonalized. Retrieval via GenAI-OMIS is centralized because any authorized user can draw from the system on aggregated traces from across units, including episodes they never witnessed and could not have named. It is depersonalized because queries are directed to “the model” rather than to specific colleagues. This weakens traditional transactive memory (“who knows what”), shifting it toward “the model knows”, and can accelerate the diffusion of emerging response patterns across the organization. At the same time, it concentrates gatekeeping power in the design and governance of the system: indexing schemes, access rules, and feedback mechanisms shape which versions of the past are more likely to be projected into the present.

Search and retrieval collapse into an opaque generative pipeline. GenAI-OMIS both collapses and extends the classic distinction between search and retrieval. Stein & Zwass (1995) define search as selecting retained information relevant to a user's goals and context, and retrieval as reconstructing that information to satisfy the request. OMIS meta-design envisions rich indexing, pattern matching, probabilistic "best-match" models, and limited natural language processing. GenAI effectively implements this vision and overshoots it: embeddings and vector-space similarity perform best-match search over heterogeneous organizational traces and external corpora, while generative modelling reconstructs answers as new text or code. From the actor's perspective, these stages are fused into a single query–response interaction; the process of selecting past episodes and combining them remains opaque.

GenAI-OMIS generates pseudo-memory and focal defaults. This fusion introduces a qualitatively new possibility absent from traditional OMIS frameworks: the system can produce synthetic precedents and rationales that resemble organizational memory but are only loosely anchored, if at all, in the organization's actual past. These consist of hallucinated policies, phantom "how we handled this last time" cases, or composite episodes. Retrieval through GenAI-OMIS thus not only mobilizes existing memory; it may also fabricate new ones, subtly reshaping what the organization believes it has done and how it ought to act next.

5.4.2.4 GenAI-OMIS and the Reweighting of Organizational Memory Carriers

Treating GenAI as OMIS rather than as a separate carrier does not make it neutral with respect to the existing retention bins. Following Walsh & Ungson (1991), organizational memory is retained in individuals, culture, transformations, structure, and ecology. A GenAI-based OMIS touches all of them, but its most immediate and theoretically tractable effects concern individuals, transformations, culture, and ecology. Structural arrangements and external archives are also affected, but here I treat them as secondary channels rather than as primary sites of reweighting.

Individuals

For individuals, GenAI-OMIS centralizes much of the initial articulation of decision stimuli and responses. When problem framing and solution development are routinely routed through the system, the "first draft" of memory is often written as prompts, AI outputs, and model-internal adjustments rather than remaining in personal recollections or private notes (Huber, 1991; Stein & Zwass, 1995).

This has two consequences. First, it weakens rehearsal: actors do not need to retain as much situational detail or procedural know-how if they expect to re-encounter it via the chat history or a fresh query. Second, it reshapes transactive memory. Instead of learning "who knows what" and building networks of experts, actors increasingly learn that "the model knows" and direct queries to it rather than to specific colleagues.

Transformations

On the transformations side, GenAI-OMIS intervenes at the level of task recipes and operating technologies. Model-generated reports, templates, checklists, and procedures can be codified into workflows and tools, embedding GenAI-shaped responses in the organization's operating core.

In such cases, past interactions with the model are "frozen" into standard operating procedures and process designs: what the model once proposed may become what the organization now routinely does. At the same time, GenAI-OMIS may also encourage a state of perpetual draft mode. When it is effortless to obtain a fresh draft or alternative solution on demand, actors may keep work at the level of ad hoc queries and local adaptations rather than stabilizing and documenting a single routine.

Culture

GenAI-OMIS effects on culture are less about the storage of specific episodes and more about the shaping of shared interpretive schemes. Recurrent GenAI outputs make certain framings, attributions, and labels chronically more available. These categories become increasingly salient in everyday talk, especially when the same system is used across units.

Over time, the model's preferred ways of describing causes, responsibilities, and appropriate responses can solidify into cultural templates for "how things work around here" (Douglas, 1986; Schein, 1990). At the same time, alternative narratives and minority framings that are less represented in interaction traces or less easily expressed in a prompt form are less likely to be reinforced. GenAI-OMIS thus risks amplifying some strands of cultural memory while attenuating others.

Ecology

Ecology concerns the material and digital environment in which work takes place. Embedding GenAI at key digital and physical touchpoints (within IDEs, document editors, ticketing systems, customer-service consoles, and meeting tools) reconfigures this environment.

"Consult the model" becomes a routine affordance at the point of action: a button in the interface, a suggested completion, an automatically generated summary. These ecological cues make it easy to invoke organizational memory in some moments and not in others, and they privilege GenAI-OMIS as the default channel for doing so. Thus, the physical and digital layout of work begins to embody a dependence on GenAI-OMIS: the environment "reminds" actors to ask the model, but may no longer prompt them to consult specific documents, colleagues, or physical artifacts that once carried memory.

Structure and External Archives

Structural implications are real but secondary in this model. New AI-facing roles (e.g., prompt engineers, model stewards) and committees emerge, and some decision rights are de facto influenced by those who configure or interpret GenAI outputs. These structural changes mainly matter here insofar as they govern access to and use of GenAI-OMIS, rather than constituting separate retention facilities. External archives enter in two ways. First, through pretraining: a GenAI model arrives with extensive extra-organizational "world memory" already embedded in its parameters. Second, through ongoing retrieval: when connected to the web or external databases, GenAI-OMIS routinely pulls in regulatory texts, industry reports, and other environmental records and integrates them into its answers.

Accordingly, GenAI-OMIS introduces new roles and governance points that shape who can imprint their experience on the system and who can access its outputs, thereby indirectly influencing how OM is populated and used. With respect to external archives, GenAI-OMIS makes environmental records a routine part of "what we know", blending internal and external histories in everyday retrieval. This can enrich organizational memory but also makes it harder to disentangle local experience from imported precedent when interpreting "how we usually handle this".

5.5 CONCLUSIONS: INTEGRATING THE INDIVIDUAL AND ORGANIZATIONAL DIMENSIONS THROUGH THE VISIBILITY TRAP

GenAI creates a *visibility trap* by opening a split between signal and substrate: organizations display strong, improving performance signals while the underlying stock of human expertise and organizational memory that produces those signals quietly erodes. In ABV terms, this reflects structurally regulated attention rather than simple bounded rationality (Ocasio, 1997). GenAI inflates the legibility and reward value of short-horizon outcomes such as throughput, apparent quality, and responsiveness within the firm's procedural and communication channels, including dashboards, delivery reviews, and status meetings. Simultaneously, it reduces the salience of slower-moving capability variables such as learning trajectories, diagnostic skill, and interpretive depth that do not travel through those channels. The trap is therefore architectural rather than primarily cognitive. The rules of the game, available resources, and distribution of participation jointly construct an environment in which "success" is increasingly defined by what GenAI makes easy to see and count (Ocasio, 1997).

At the individual level, Model 1's *AI-Performance Leveling Effect* serves as the masking mechanism that makes skill decay appear as capability growth. When novices generate outputs resembling expert-level work on AI-amenable tasks, traditional error signals disappear. Historically, weak drafts, incorrect diagnoses, or incoherent plans created visible problems that triggered training, review, and deliberate practice. The organization's attention was pulled toward skill-building because poor performance forced issues into high-strength channels (Ericsson, Krampe, & Tesch-Römer, 1993). Under GenAI mediation, first-order problems are often solved before becoming legible to others: drafts read well, code runs, and slide decks appear coherent. ABV predicts the consequence: without salient performance breakdowns, capability development loses valuation and becomes a low-strength concern easily deferred when time pressure and delivery norms dominate organizational channels (Ocasio, 1997). Latent losses accumulate as novices bypass proceduralization and reduce deliberate decomposition, while experienced workers shift from execution to oversight, diminishing rehearsal of representational and diagnostic subskills. Organizations misread smooth output as competent producer even though competence is increasingly tool-borrowed rather than internalized. Interviewees captured this split directly, with managers celebrating AI-assisted delivery metrics while worrying that juniors "don't really understand" what they ship (I1, I2, I6). The illusion persists because the firm's evaluative architecture notices visible output defects but not invisible hollowness in the underlying capability.

At the OM level, Model 2 presents an analogous mechanism where efficiency crowds out provenance, creating a void. A GenAI-OMIS makes retrieval frictionless through one interface, one voice and seemingly universal access, which is immediately legible and strongly rewarded within decision channels, especially under information overload. However, this legibility is achieved by flattening context, weakening source authority cues, and mixing heterogeneous traces into single fluent responses. Organizations remember the answer while progressively losing the ability to recover the grounds for that answer (i.e., the who, when, and why that give memory its epistemic reliability) (Walsh & Ungson, 1991). As encoding and retrieval centralize into the OMIS, alternative carriers such as people, routines, artifacts, and informal networks are used less frequently. Individuals rehearse less, drafts remain increasingly provisional, and cultural cues shift from "ask colleagues or consult artifacts" toward "ask the model". This carrier thinning matters because organizational memory is not merely stored content but an ecology of distributed traces and retrieval practices (Walsh & Ungson, 1991). The result is hybrid, unstable memory where partial internal traces, shallow user representations, external training data, and hallucinated links combine into synthetic precedents with opaque provenance that is difficult to audit or contest. Interviewees described GenAI as a "third head" for recalling policies and prior work but could not specify what was internal versus external or how persistence and trace retention worked across interactions (I2, I3, I5, I8). ABV sharpens this diagnosis: the attention system is biased toward accessibility (speed, convenience, and coverage) over validity (traceability, authority, and contestability) because accessibility is what dominant channels measure and reward (Ocasio, 1997). *Pseudo-memory* is therefore reinforced precisely because it is easier to access than true organizational history embedded in messy carrier diversity.

The recursive logic of the *Visibility Trap* tightens from VT3 into VT4 as visible success cannibalizes the organizational resources required to repair invisible deterioration. VT3 represents valuation through existing metrics and routines rather than mere evaluation. When AI-mediated workflows repeatedly score well on speed, volume, and apparent quality, organizations enact an environment in which GenAI is framed as the natural solution to performance pressure, then reallocate attention, budget, and managerial time accordingly (Ocasio, 1997). Training time, apprenticeship scaffolding, and memory curation are crowded out because they are slower, harder to quantify, and less compatible with the firm's high-frequency performance channels. VT4 represents the structural consequence: each "successful" evaluation makes further integration rational within prevailing rules, locking firms into a path where machine scaffolding substitutes for human development and retrieval convenience substitutes for provenance discipline. In ABV terms, the trap is sustained by channeling: issues traveling through dashboards and delivery reviews gain strength, while issues requiring longitudinal indicators, different participants, and different agendas, such as learning, carrier diversity, and provenance quality, fail to enter decision situations with comparable force (Ocasio, 1997). Capability degradation is therefore not an accidental side-effect but the predictable outcome of an attentional architecture optimized for high-legibility/high-strength signals.

The interviews suggest an early-stage but silent alarm fitting a disruptive variant of VT4, where actors can name risks such as concerns about junior depth, dependency, and "not really understanding", but cannot convert these into organizational action because feared outcomes have not yet materialized as visible failures within high-strength channels. Concerns remain informal, localized, and psychologically salient to some individuals yet structurally weak, with no shared metrics for learning depth, no owner for organizational memory provenance (except for I8), no routine forcing contestability checks (except for I8), and no channel where "capability erosion" competes equally with delivery priorities. ABV predicts this decoupling between individual awareness and organizational action: attention structures shape what becomes legitimate, what receives agenda time, and which solutions are resourced (Ocasio, 1997). Until skill decay and memory distortion are translated into issues with organizational standing through indicators, routines, and participation structures, organizations will likely continue to enact an environment where GenAI appears unambiguously beneficial. The trap, therefore, is not that nobody sees the problem but that the organization's attention system treats the problem as non-decision-worthy until it produces an undeniable breakdown.

Future research should test the predictions made in this chapter. At the micro level, longitudinal studies can track skill development trajectories, variations in AI-free performance, and AI Capability Loss across expertise strata and task types. At the organizational level, work is needed to operationalize hybrid and pseudo-memory, to measure shifts in OM carrier weights under GenAI-OMIS, and to map attention structures in dashboards, policies, and workflows.

CHAPTER REFERENCES

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406. https://doi.org/10.1037/0033-295X.89.4.369

Anderson, J. R. (1983). *The architecture of cognition* (3. printing). Harvard Univ.Pr.

Arthur Jr., W., Bennett Jr., W., Stanush, P. L., & McNelly, T. L. (1998). Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. Human Performance, 11(1), 57–101. https://doi.org/10.1207/s15327043hup1101_3

Benner, P. E., (1984). *From novice to expert: Excellence and power in clinical nursing practice*. Addison-Wesley.

Brynjolfsson, E., Li, D., & Raymond, L. (2023). *Generative AI at Work* (No. w31161). National Bureau of Economic Research. https://doi.org/10.3386/w31161

Budzyń, K., Romańczyk, M., Kitala, D., Kołodziej, P., Bugajski, M., Adami, H. O., Blom, J., Buszkiewicz, M., Halvorsen, N., Hassan, C., Romańczyk, T., Holme, Ø., Jarus, K., Fielding, S., Kunar, M., Pellise, M., Pilonis, N., Kamiński, M. F., Kalager, M., … Mori, Y. (2025). Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: A multicentre, observational study. The Lancet Gastroenterology & Hepatology, 10(10), 896–903. https://doi.org/10.1016/S2468-1253(25)00133-5

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Cheng, K., Yang, J., Jiang, H., Wang, Z., Huang, B., Li, R., Li, S., Li, Z., Gao, Y., Li, X., Yin, B., & Sun, Y. (2024). *Inductive or Deductive? Rethinking the Fundamental Reasoning Abilities of LLMs* (No. arXiv:2408.00114). arXiv. https://doi.org/10.48550/arXiv.2408.00114

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Childs, J. M., & Spears, W. D. (1986). Flight-Skill Decay and Recurrent Training. *Perceptual and Motor Skills*, *62*(1), 235–242. https://doi.org/10.2466/pms.1986.62.1.235

Cohen, M. D., & Bacdayan, P. (1994). Organizational Routines Are Stored as Procedural Memory: Evidence from a Laboratory Study. *Organization Science*, *5*(4), 554–568. https://doi.org/10.1287/orsc.5.4.554

Cordella, A., & Gualdi, F. (2023). AI and Bureaucratic Decision-Making: A New Layer of Opacity. *Academy of Management Proceedings*, *2023*(1), 14863. https://doi.org/10.5465/AMPROC.2023.14863abstract

Cornelissen, J. (2017). Editor's Comments: Developing Propositions, a Process Model, or a Typology? Addressing the Challenges of Writing Theory Without a Boilerplate. *Academy of Management Review*, *42*(1), 1–9. https://doi.org/10.5465/amr.2016.0196

Crowston, K., & Bolici, F. (2025). Deskilling and upskilling with generative AI systems. *Proceedings of the iConference*. iConference, Indiana Bloomington, USA.

Cyert, R., & March, J. (1963). *A Behavioral Theory of the Firm*.

DeKeyser, R. (2015). Skill Acquisition Theory. In Theories in Second Language Acquisition (Second Edition). Routledge.

DeKeyser, R. (2020). Skill Acquisition Theory. In *Theories in Second Language Acquisition* (3rd ed.). Routledge.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4573321

Douglas, M. (1986). *How institutions think*. Syracuse Univ. Press.

Dreyfus, H. L., Dreyfus, S. E., & Athanasiou, T. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. The Free Press.

Dreyfus, S. E. (2004). The Five-Stage Model of Adult Skill Acquisition. *Bulletin of Science, Technology & Society*, *24*(3), 177–181. https://doi.org/10.1177/0270467604264992

Ericsson, K. A., & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge University Press.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. Psychological Review, 100(3), 363–406.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing Organizational Routines as a Source of Flexibility and Change. *Administrative Science Quarterly*, *48*(1), 94–118. https://doi.org/10.2307/3556620

Feng, Y., Xu, Y., Qin, L., Wang, Y., & Che, W. (2024). *Improving Language Model Reasoning with Self-motivated Learning* (No. arXiv:2404.07017). arXiv. https://doi.org/10.48550/arXiv.2404.07017

Fitts, P. M., & Posner, M. I. (Eds.). (1973). *Human performance*. Prentice Hall.

Gardlin, G. R., & Sitterley, T. E. (1972). *Degradation of learned skills: A Review and Annotated Bibliography*. National Aeronautics and Space Administration. https://doi.apa.org/doi/10.1037/e611312012-091

Green, F. (2011). *What is Skill? An Inter-Disciplinary Synthesis* (Research Paper No. 20; LLAKES). Centre for Learning and Life Chances in Knowledge Economies and Societies, Institute of Education, University of London. https://www.llakes.org/wp-content/uploads/2011/02/Green-What-is-Skill-reduced.pdf

Hannan, M. T., & Freeman, J. (1984). Structural Inertia and Organizational Change. *American Sociological Review*, *49*(2), 149–164. https://doi.org/10.2307/2095567

Hosseini, A., Sordoni, A., Toyama, D., Courville, A., & Agarwal, R. (2024). *Not All LLM Reasoners Are Created Equal* (No. arXiv:2410.01748). arXiv. https://doi.org/10.48550/arXiv.2410.01748

Huber, G. P. (1991). Organizational Learning: The Contributing Processes and the Literatures. *Organization Science*, *2*(1), 88–115. https://doi.org/10.1287/orsc.2.1.88

Hutchins, E. (1995). How a Cockpit Remembers Its Speeds. *Cognitive Science*, *19*(3), 265–288. https://doi.org/10.1207/s15516709cog1903_1

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. AMS Review, 10(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0

Khot, T., Trivedi, H., Finlayson, M., Fu, Y., Richardson, K., Clark, P., & Sabharwal, A. (2022, September 29). *Decomposed Prompting: A Modular Approach for Solving Complex Tasks*. The Eleventh International Conference on Learning Representations. https://openreview.net/forum?id=_nGgzQjzaRy

Kluge, A., & Frank, B. (2014). Counteracting skill decay: Four refresher interventions and their effect on skill and knowledge retention in a simulated process control task. *Ergonomics*, *57*(2), 175–190. https://doi.org/10.1080/00140139.2013.869357

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and Novice Performance in Solving Physics Problems. *Science*, *208*(4450), 1335–1342. https://doi.org/10.1126/science.208.4450.1335

Levitt, B., & March, J. G. (1988). Organizational Learning. *Annual Review of Sociology*, *14*(1), 319–338. https://doi.org/10.1146/annurev.so.14.080188.001535

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

Macnamara, B. N., Berber, I., Çavuşoğlu, M. C., Krupinski, E. A., Nallapareddy, N., Nelson, N. E., Smith, P. J., Wilson-Delfosse, A. L., & Ray, S. (2024). Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cognitive Research: Principles and Implications*, *9*(1), 46. https://doi.org/10.1186/s41235-024-00572-8

Mahajan, P. (2025). *The Silent Erosion: Global Generational Cognitive Decline in the Age of AI and the Future of Human Intellectual Agency* (SSRN Scholarly Paper No. 5386814). Social Science Research Network. https://doi.org/10.2139/ssrn.5386814

March, J. G., & Olsen, J. P. (1975). THE UNCERTAINTY OF THE PAST: ORGANIZATIONAL LEARNING UNDER AMBIGUITY. *European Journal of Political Research*, *3*(2), 147–171. https://doi.org/10.1111/j.1475-6765.1975.tb00521.x

March, J., & Simon, H. (1958). *Organizations* (1st edition). Wiley.

Naylor, J. C., Briggs, G. E., & Reed, W. G. (1968). Task coherence, training time, and retention interval effects on skill retention. *Journal of Applied Psychology*, *52*(5), 386–393. https://doi.org/10.1037/h0026239

Newell, A., & Simon, H. A. (1972). Human problem solving. Echo Point Books & Media.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Ocasio, W. (1997). Towards an Attention-Based View of the Firm. *Strategic Management Journal*, *18*(S1), 187–206. https://doi.org/10.1002/(SICI)1097-0266(199707)18:1+%253C187::AID-SMJ936%253E3.0.CO;2-K

OECD. (2023). OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market. OECD. https://doi.org/10.1787/08785bba-en

Olivera, F. (2000). Memory Systems In Organizations: An Empirical Investigation Of Mechanisms For Knowledge Collection, Storage And Access. *Journal of Management Studies*, *37*(6), 811–832. https://doi.org/10.1111/1467-6486.00205

Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). *The Impact of AI on Developer Productivity: Evidence from GitHub Copilot*. https://doi.org/10.48550/ARXIV.2302.06590

Prahalad, C. K., & Bettis, R. A. (1986). The dominant logic: A new linkage between diversity and performance. *Strategic Management Journal*, *7*(6), 485–501. https://doi.org/10.1002/smj.4250070602

Rinta-Kahila, T., Penttinen, E., A., Soliman, W., & Ruissalo, J. (2023). The Vicious Circles of Skill Erosion: A Case Study of Cognitive Automation. *Journal of the Association for Information Systems*, *24*(5), 1378–1412. https://doi.org/10.17705/1jais.00829

Schein, E. H. (1990). Organizational Culture: What it is and How to Change it. In P. Evans, Y. Doz, & A. Laurent (Eds.), *Human Resource Management in International Firms: Change, Globalization, Innovation* (pp. 56–82). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-11255-5_4

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

Shrivastava, P., & Schneider, S. (1984). Organizational Frames of Reference. *Human Relations*, *37*(10), 795–809. https://doi.org/10.1177/001872678403701002

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill* (1.[print.]). Harvard Univ. Press.

Sison, A. J. G., Daza, M. T., Gozalo-Brizuela, R., & Garrido-Merchán, E. C. (2024). ChatGPT: More Than a “Weapon of Mass Deception” Ethical Challenges and Responses from the Human-Centered Artificial Intelligence (HCAI) Perspective. *International Journal of Human–Computer Interaction*, *40*(17), 4853–4872. https://doi.org/10.1080/10447318.2023.2225931

Stein, E. W., & Zwass, V. (1995). Actualizing Organizational Memory with Information Systems. *Information Systems Research*, *6*(2), 85–117. https://doi.org/10.1287/isre.6.2.85

Styve, A., Virkki, O. T., & Naeem, U. (2024). Developing Critical Thinking Practices Interwoven with Generative AI Usage in an Introductory Programming Course. *2024*

IEEE Global Engineering Education Conference (EDUCON), 01–08. https://doi.org/10.1109/EDUCON60312.2024.10578746

Suddaby, R. (2010). Editor's Comments: Construct Clarity in Theories of Management and Organization. *Academy of Management Review*, *35*(3), 346–357. https://doi.org/10.5465/amr.35.3.zok346

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Tang, X., Zheng, Z., Li, J., Meng, F., Zhu, S.-C., Liang, Y., & Zhang, M. (2023). *Large Language Models are In-Context Semantic Reasoners rather than Symbolic Reasoners* (No. arXiv:2305.14825). arXiv. https://doi.org/10.48550/arXiv.2305.14825

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J.-M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. *Journal of Intelligence*, *11*(3), 54. https://doi.org/10.3390/jintelligence11030054

Tong, Y., Li, D., Wang, S., Wang, Y., Teng, F., & Shang, J. (2024). *Can LLMs Learn from Previous Mistakes? Investigating LLMs' Errors to Boost for Reasoning* (No. arXiv:2403.20046). arXiv. https://doi.org/10.48550/arXiv.2403.20046

VanLehn, K. (1988). *Problem Solving and Cognitive Skill Acquisition:* Defense Technical Information Center. https://doi.org/10.21236/ADA218905

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*(1), 513–539. https://doi.org/10.1146/annurev.psych.47.1.513

Varone, A., Bolici, F., & Crowston, K. (2025). *Short-Term Gain, Long-Term Pain? A Conceptual Model on AI's Impact on Human Skill Development*. European Management Journal (EMJ) – *CURRENTLY UNDER REVIEW.*

Walsh, J. P., & Ungson, G. R. (1991). Organizational Memory. *Academy of Management Review*, *16*(1), 57–91.

Wang, P., Li, D., Wang, Y., Han, Q., & Khan, Y. A. (2024). Will advancement in technologies bring fear and damage human employment? Evidence from China's manufacturing industry. *PLOS ONE*, *19*(4), e0295942. https://doi.org/10.1371/journal.pone.0295942

Wang, X., Day, E. A., Kowollik, V., Schuelke, M. J., & Hughes, M. G. (2013). Factors Influencing Knowledge and Skill Decay after Training: A Meta-Analysis. In *Individual and Team Skill Decay*. Routledge.

Weaver, S. J., Newman-Toker, D. E., & Rosen, M. A. (2012). Reducing Cognitive Skill Decay and Diagnostic Error: Theory-Based Practices for Continuing Education in Health Care. *Journal of Continuing Education in the Health Professions*, *32*(4), 269–278. https://doi.org/10.1002/chp.21155

Weick, K. E. (1995). *Sensemaking in organizations*. Sage Publications.

Wood, R. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, *37*, 60–82. https://doi.org/10.1016/0749-5978(86)90044-0

Wu, Z., Qiu, L., Ross, A., Akyürek, E., Chen, B., Wang, B., Kim, N., Andreas, J., & Kim, Y. (2024). Reasoning or Reciting? Exploring the Capabilities and Limitations of Language Models Through Counterfactual Tasks. In K. Duh, H. Gomez, & S. Bethard (Eds.), *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long*

Papers) (pp. 1819–1862). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.102

Yadav, S., Qureshi, A. M., Kaushik, A., Sharma, S., Loughran, R., Kazhuparambil, S., Shaw, A., Sabry, M., Lynch, N. S. J., Singh, N., O'Hara, P., Jaiswal, P., Chandru, R., & Lillis, D. (2025). *From Idea to Implementation: Evaluating the Influence of Large Language Models in Software Development -- An Opinion Paper* (No. arXiv:2503.07450). arXiv. https://doi.org/10.48550/arXiv.2503.07450

Zhou, D., Schärli, N., Hou, L., Wei, J., Scales, N., Wang, X., Schuurmans, D., Cui, C., Bousquet, O., Le, Q. V., & Chi, E. H. (2022, September 29). *Least-to-Most Prompting Enables Complex Reasoning in Large Language Models*. The Eleventh International Conference on Learning Representations. https://openreview.net/forum?id=WZH7099tgfM

CHAPTER 6: INTEGRATED DISCUSSION AND CONTRIBUTIONS

This thesis established that a consequential effect of GenAI is its capacity to decouple visible gains from latent erosions. In both settings examined – public governance of GenAI and organizational deployment in knowledge work – the same mechanism recurs: established attention structures privilege signals that are legible and strong within existing evaluation categories, metrics, and channels while structurally downgrading consequences that remain difficult to encode and evaluate.

Chapter 3 developed this mechanism as the *Visibility Trap*, grounded in the Attention-Based View (ABV) (Ocasio, 1997). Under ABV, organizational action results from how attention is distributed across issues and answers through attention structures and procedural and communication channels such as reporting routines, evaluation criteria, and decision forums. The *Visibility Trap* captures a recursive pattern wherein visible successes become increasingly salient and institutionally reinforced, while associated negative consequences, initially low in legibility or strength, remain backgrounded until they manifest as failures that can no longer be ignored.

Chapters 4 and 5 analyzed two phases of GenAI diffusion to demonstrate how this mechanism operates across domains. In governance (Chapter 4), GenAI generated highly legible and strong signals exposing the need for legal intervention, followed by the enactment of available (yet imperfect) regulations and procedural compliance. This apparently successful intervention hid the erosion of legitimacy dynamics that were weakly represented in the evaluative channels at hand. In organizational deployment (Chapter 5), GenAI generated highly legible productivity gains and apparent improvements in organizational memory access, while expertise development and memory integrity were weakly represented and therefore systematically deprioritized.

Therefore, this thesis concludes that GenAI is a high-leverage technology for producing strong and legible success signals, while weakening the reliability of those signals as proxies for organizational sustainability. This chapter consolidates that claim as a set of contributions, derives concrete implications for organizational design and governance, and specifies future research avenues on this theme.

6.1 PRACTICAL IMPLICATIONS

6.1.1 Proxy Substitution

One implication of this thesis is that GenAI may be enabling proxy substitution: organizations increasingly treat success on historically credible proxies as evidence that underlying capability or institutional efficacy exists, even when those substrates are not developing and may be degrading.

In Chapter 5, this mechanism is formalized as the *Performance-Leveling Effect*. GenAI compresses observable differences across expertise levels, allowing novices to produce outputs that approximate intermediate or even expert performance on AI-amenable tasks. The result is a structural weakening of the classic assumption embedded in both managerial evaluation and expertise theory that performance is a reliable proxy for skill. When a novice's draft appears flawless, the error signals that historically triggered training, supervision, and deliberate practice disappear, and capability development loses organizational valuation. This represents not a marginal behavioral effect but an attentional reclassification: learning becomes low-strength because the immediate, visible problem is pre-solved by the tool.

In Chapter 4, an analogous decoupling appears between visible regulatory success and substantive impact. The Italian GPDP's intervention generated a strong "formal win" under existing compliance-oriented attention structures through an enforceable pathway, observable procedural steps, and later sanctions. However, the DiD evidence showed no lasting change in Italian usage trajectories, and the public discourse analysis indicated legitimacy erosion.

The theoretical point is not that organizations make mistakes but that ABV predicts proxy substitution when the environment changes faster than the evaluative architecture. GenAI increases the supply of credible outputs and high-visibility compliance artifacts, making it rational within inherited channels to treat these as decisive evidence. The contribution is therefore to reframe GenAI-driven myopia as a systematic product of evaluative architectures rather than a failure of awareness or an episodic lapse in judgment.

6.1.2 The Politics of Visibility

The second implication targets organizational design by suggesting that the *Visibility Trap* has a political dimension wherein attention structures evolve through bargaining among coalitions with divergent goals and dependencies (Cyert & March, 1963). Within this competitive environment, organizational units demonstrating AI-boosted success via visible indicators are positioned to accrue greater standing, resources, and agenda-setting authority.

In the GenAI governance case, this tension may arise between legal and technical coalitions. The former is oriented toward formal compliance, defensible procedures, and alignment with existing mandates, while the latter focuses on system behavior, robustness, and broader sociotechnical risks. Following Cyert & March (1963), the interaction between these groups often results in a *quasi-resolution of conflict*: a negotiated framework where each party secures specific prerequisites. This negotiation can follow two distinct trajectories. In an optimistic scenario, a balanced regime emerges that genuinely integrates legal defensibility with technical risk control. Conversely, a pessimistic trajectory leads to a regulatory no-man's land: a framework that is too legalistic for effective safety work yet too technically complex for consistent enforcement, ultimately prioritizing paper defensibility over efficacy.

A similar pattern can emerge within organizations integrating GenAI into their work. Coalitions of performance-oriented managers can utilize AI-visible gains to advance their priorities, while HR units may sense issues around dependence and skill erosion but struggle to pursue these concerns with equal weight. Here too, *quasi-resolutions of conflict* are likely, producing hybrid policies in which each coalition secures some safeguards or freedoms. The resulting compromise may either expand the range of issues treated as legitimate or, in the worst case, produce an internal no-man's land of GenAI policies that are defensible on paper yet weakly connected to everyday practice.

While the thesis cannot adjudicate between these trajectories, it suggests that the political dimension also matters for the *Visibility Trap*: changing what is attended to requires redistributing power rather than merely updating metrics or issuing new guidelines. Future research can deepen this political reading of the Visibility Trap along two lines. First, visibility can be treated as a strategic resource by examining how coalitions construct, amplify, or suppress particular indicators and narratives to consolidate their position, and how counter-coalitions attempt to build alternative visibility claims around latent harms. Second, researchers can analyze the evolution of quasi-resolutions of conflict by investigating which configurations of coalition power, external pressure, and framing lead to balanced integration of legal, technical, and capability concerns, and which drift into no-man's-land arrangements that are formally sophisticated but decoupled from practice.

6.1.3 Elastic Attention: Expanding What is Legible Without Losing Focus

The *Visibility Trap* is sticky but not deterministic. Organizational attention structures are plastic and subject to contestation, although their redesign is often costly and conflict-laden. Escaping the trap requires the deliberate reallocation of organizational slack toward issues that currently fall outside the institution's legibility sphere: phenomena that the organization cannot yet register, code, or discuss as legitimate objects of concern. The central challenge is to expand the organization's sphere of attention without compromising its capacity for focused action.

Foundational theories of organization design address this tension through two opposing mechanisms: boundary-spanning roles and buffering structures (Aguilar, 1967; Aldrich & Herker, 1977; Daft & Weick, 1984). Boundary-spanning units are designed to widen the organization's exposure to weak, unfamiliar environmental signals. Conversely, buffering structures shield the technical core from information overload and noise, ensuring attention remains fixed on critical, pre-defined tasks. In the context of GenAI, both mechanisms are necessary but insufficient in isolation: expanding legibility without buffering risks overwhelming decision-makers with unstructured concerns, whereas buffering without boundary spanning locks the organization into its current visibility bias.

To resolve this tension, this chapter proposes the concept of *elastic attention*: the deliberate reconfiguration of attention structures to expand the range of issues and potential responses rendered legible and eligible for organizational action, while preserving a stable core focus on prioritized tasks. Because boundary spanners are themselves embedded in organizational frames that tend to pull them toward familiar, high-legibility issues, elastic attention cannot emerge spontaneously. It must be cultivated as a controlled, partial deviation from dominant frames – one that requires intentional organization design not only to detect signals that fall outside existing metrics, but also to translate those signals into terms the organization can act upon. Making *elastic attention* empirically useful requires both operationalization and design work. Research is needed to identify observable indicators of *elastic attention*, such as the frequency with which issues originating outside formal metrics enter decision agendas, the speed at which new issue categories are incorporated, and the allocation of resources between scanning and exploitation. Studies should also examine sustainable organizational designs that protect attention to illegible issues against the pull of existing visibility biases by investigating which reporting lines, incentives, norms, and amounts of slack allow boundary-spanning and buffering functions to persist, and how conceptual and linguistic innovations such as new categories and vocabularies help expand what is treated as legitimately on the radar in the context of GenAI.

6.2 THEORETICAL CONTRIBUTIONS

6.2.1 Contribution to Organizational Myopia Research

Chapter 3 positioned the thesis at the intersection of organizational myopia (Levinthal & March, 1993; Catino, 2013; Sato, 2012) and the Attention-Based View (Ocasio, 1997). At this intersection, this thesis specifies how myopic outcomes emerge when GenAI interacts with existing attention structures through two central elements.

First, the thesis reframes incubation as attentional economizing rather than failure to notice weak signals. Latent problems are often recognized by some actors, such as the concerns about expertise erosion expressed by interviewees in Chapter 5, but they do not generate performance-aspiration gaps under current evaluative criteria. Given limited time, attention, and material resources, downgrading these low-strength issues in favor of dimensions where GenAI produces clear gains is rational. Incubation is thus produced by how attention is economized across legible issues rather than only by gaps in surveillance.

Second, the thesis shifts part of the Success Trap (Levinthal & March, 1993) from competence portfolios to evaluative architectures. Classic accounts emphasize lock-in to familiar routines because exploitation capabilities dominate. Here, the lock-in resides in rules, metrics, channels, and roles that define success. Under GenAI, these architectures are continually fed with strong signals on compliance, productivity, or retrieval efficiency. This recursive confirmation makes it harder to justify reallocating attention toward latent failures before any major negative event occurs.

Myopic outcomes are thus reclassified as properties of how organizations economize attention through established evaluation systems under GenAI rather than only as consequences of weak signals or overused routines. Two research directions follow from this reframing. At the micro level, the concept of attentional economizing requires unpacking: which heuristics individuals use when triaging issues under attention scarcity, how strongly these heuristics favor legible signals, and how social pressures from supervisors, peers, and evaluation systems influence these choices. At the meso level, the plasticity of evaluative architectures requires closer study of what triggers reconfiguration, how much inertia different components such as metrics, rules, and informal routines exhibit, and whether deliberate redesign can realistically reduce the Visibility Trap or is constrained by path dependence. Together, these lines of inquiry would connect the thesis's organizational-level claims to individual cognition and institutional change.

6.2.2 A Portable Visibility Trap

The second contribution is integrative. The same attention-based mechanism involving legibility and strength under the *Visibility Trap* explains how a data protection authority can appear successful while eroding legitimacy, how GenAI can raise measured performance while undermining skill development, and how GenAI-OMIS can improve retrieval while degrading memory provenance and diversity. This does not make the mechanism universal, nor does it claim that these phenomena are fundamentally the same. The claim is more parsimonious: the thesis demonstrates that an attention-based account built around legibility and strength can travel across these domains and still perform useful explanatory work. Making this portability explicit is the main integrative contribution and sets up future work to test other domains where the mechanism holds, where it needs adjustment, and where it fails.

6.3 CONCLUDING REMARKS

This thesis examined how GenAI interacts with existing attention structures to shape which of its organizational consequences are brought into focus and which remain peripheral. Across the distinct domains of governance and organizational deployment in knowledge work, the analyses reveal a convergent pattern: GenAI generates a broad spectrum of consequences, yet organizational attention is consistently captured by those aligning with familiar indicators of success, such as formal enforcement outputs, productivity metrics, and information retrieval speed. Conversely, the slower degradation of capabilities, legitimacy, and organizational memory receives minimal systematic attention because these phenomena are weakly represented within established evaluative frameworks. GenAI thus appears highly effective when judged through inherited measures even as it actively undermines the very capacities those measures were originally intended to proxy.

This dynamic substantiates the *Visibility Trap*: a self-reinforcing process shaped by established attention structures in which the salience of visible successes progressively degrades the capacity to identify and respond to associated but latent negative consequences. The *Visibility Trap* is, to a large extent, a predictable consequence of bounded rationality: the same selective attention that allows organizations to function coherently also restricts their ability to perceive phenomena beyond familiar categories. Mitigating this trap therefore demands the willingness to interrogate evaluative channels themselves, particularly when those channels yield favorable verdicts. Whether organizations can sustain such reflexivity at scale remains an open question. Ultimately, organizations that navigate this technological transition successfully will not necessarily be those that adopt AI most aggressively or resist it most cautiously. Rather, success will likely belong to those capable of continuously asking whether their established measures of success still accurately reflect the capabilities they were designed to preserve.

CHAPTER REFERENCES

Aguilar, F. J. (1967). *Scanning the Business Environment*.

Aldrich, H., & Herker, D. (1977). Boundary Spanning Roles and Organization Structure. *The Academy of Management Review*, *2*(2), 217. https://doi.org/10.2307/257905

Catino, M. (2013). *Organizational myopia: Problems of rationality and foresight in organizations*. Cambridge University Press.

Cyert, R., & March, J. (1963). *A Behavioral Theory of the Firm*.

Daft, R. L., & Weick, K. E. (1984). Toward a Model of Organizations as Interpretation Systems. *The Academy of Management Review*, *9*(2), Article 2. https://doi.org/10.2307/258441

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. Strategic Management Journal, 14(S2), 95–112. https://doi.org/10.1002/smj.4250141009

Ocasio, W. (1997). Towards an Attention-Based View of the Firm. *Strategic Management Journal*, *18*(S1), 187–206. https://doi.org/10.1002/(SICI)1097-0266(199707)18:1+%253C187::AID-SMJ936%253E3.0.CO;2-K

Sato, H. (2012). Routine-Based View of Organizational Learning and Mechanisms of Myopia. *Annals of Business Administrative Science*, *11*(0), 45–54. https://doi.org/10.7880/abas.11.45