The Cognitive Atrophy Project™

A Cognitive Equilibrium Framework for Human–AI Interaction in Professional Environments

Bruno Horta Soares

University Professor and Executive Advisor in Digital Governance, AI and Cybersecurity

Abstract

Generative AI has moved, in less than three years, from peripheral curiosity to structural participant in professional cognition. The dominant discourse celebrates productivity, augmentation and acceleration; the implicit assumption is that what is delegated is recovered, and that what is automated is improved. This paper questions the symmetry of that assumption. It argues that the equilibrium between human cognition and AI-assisted delegation is, itself, a legitimate object of governance, and that the absence of a vocabulary to recognise it is already a governance failure. Drawing on the literatures of distributed cognition (Hutchins, 1995; Hollan, Hutchins, & Kirsh, 2000), automation bias (Bainbridge, 1983; Parasuraman & Manzey, 2010), cognitive offloading (Risko & Gilbert, 2016; Sparrow, Liu, & Wegner, 2011), deskilling (Zuboff, 1988), and the emerging literature on meaningful human oversight (European Union, 2024; Fink, 2025), the paper introduces three interrelated constructs — Cognitive Atrophy, Cognitive Elasticity and Cognitive Equilibrium — and articulates them as the dimensions of a dynamic, behavioural and governance-oriented framework for understanding human–AI cognitive interaction in professional environments. The contribution is conceptual, not empirical. It is positioned within information systems, AI governance and socio-technical research, and it deliberately refuses the vocabulary of neuroscience and clinical pathology. The framework does not diagnose individuals; it offers organisations and professions a language to observe, govern and design the cognitive consequences of working alongside intelligent systems.

Keywords: Cognitive Atrophy; Cognitive Elasticity; Cognitive Equilibrium; Generative AI; Human–AI Interaction; AI Governance; Cognitive Delegation; Human Oversight; Professional Augmentation; Socio-Technical Systems.

1. Introduction

1.1 The Acceleration We Did Not Negotiate

Few transformations in the recent history of professional work have been adopted with as little deliberation as the integration of generative AI into cognitive workflows. Within months, professionals across law, medicine, finance, engineering, public administration and academia began outsourcing first drafts, then analytical synthesis, then judgment-adjacent tasks, to systems whose internal reasoning they could not inspect and whose error patterns they had not yet learned to recognise. The verbs we use to describe this shift — adopt, integrate, leverage, augment — share a common rhetorical move: they neutralise the act of delegation by presenting it as a tooling decision rather than a cognitive one.

It is not. The choice to let an external agent perform an intellectual operation that one previously performed oneself is, in any serious account of professional practice, a transfer of authority. It may be a wise transfer, a careless one, or a coerced one, but it is never simply a

productivity gain. The premise of this paper is that the field needs a vocabulary equal to that recognition.

1.2 The Productivity Story and Its Blind Spot

The dominant narrative around generative AI in professional environments is built around productivity, augmentation, and the redistribution of cognitive load from humans to machines. Industry research, consulting reports and a substantial portion of the academic literature converge on a single dominant question: how much faster, cheaper or more scalable does professional cognition become once AI is in the loop? This is a legitimate question. It is not, however, the only question, and it is not the most important one for governance.

The blind spot is structural. A productivity frame treats the human as a fixed asset whose output rises when assisted. It does not ask what happens to that human when the assistance is sustained, repeated and embedded into the daily texture of work. It does not ask what muscles are being unused while others are being amplified. It does not ask whether the conditions of professional judgment — the slow, friction-laden, sometimes uncomfortable practices of validation, challenge and reasoning — survive when they are no longer required for the immediate output.

1.3 What We Are Quietly Losing

There is a recurring observation among practitioners who have worked with generative AI for two or three years: it becomes harder, not easier, to start from a blank page. The acceptance threshold for a draft drops. The instinct to reread, to challenge, to suspect, weakens with use rather than strengthens. None of this is yet documented at the level of robust empirical evidence specific to generative AI, although the cognate literatures on automation complacency (Parasuraman & Manzey, 2010) and externalised memory (Sparrow et al., 2011) have established structurally similar patterns in adjacent domains. This paper does not claim more than the literatures support. What it does claim is that the patterns are coherent enough, and consequential enough, to deserve a conceptual frame before the empirical work catches up.

The frame proposed here is not anti-AI. It assumes, on the contrary, that AI integration is durable and that resistance to it is neither realistic nor desirable. The question is what professionals and organisations preserve, deliberately, while the integration deepens.

1.4 Argument and Contribution

This paper argues that the appropriate object of governance is not the AI system, nor the human user in isolation, but the equilibrium between them. It introduces the Cognitive Equilibrium Framework™ as a conceptual lens for that equilibrium, organised around three constructs: Cognitive Atrophy, Cognitive Elasticity and Cognitive Equilibrium itself. The contribution is foundational and definitional. It is intended to provide later empirical, methodological and pedagogical work with a stable conceptual ground.

The paper proceeds as follows. Section 2 reviews the literatures on which the framework draws and identifies the gap it addresses. Section 3 defines the three constructs and states the boundary conditions of the framework. Section 4

articulates the framework itself, including its visual representation and the delegation–verification erosion loop. Section 5 develops governance and professional implications. Section 6 introduces the broader Cognitive Atrophy Project™ as the operational extension of the framework. Section 7 sets a research agenda. Section 8 states the limitations of the contribution. Section 9 concludes.

2. Literature Review

The framework proposed in this paper does not emerge from a single discipline. It draws on four established literatures and one strategic counterweight, each of which is reviewed below not as exhaustive synthesis but as conceptual scaffolding for the constructs developed in Section 3.

2.1 Cognitive Offloading and Distributed Cognition

The idea that cognition is not contained within a single skull is older than computing. The tradition of distributed cognition, developed most influentially by Hutchins (1995) in his ethnographic study of naval navigation and extended by Hollan, Hutchins, and Kirsh (2000) into a foundation for human-computer interaction research, established that cognition routinely operates across people, artefacts and environments rather than within individual minds alone. Writing, lists, calendars, spreadsheets, databases and search engines are all instances of this externalisation. A more recent strand, formalised by Risko and Gilbert (2016), names the underlying behaviour cognitive offloading: the use of physical action, including the use of external tools, to alter the information-processing requirements of a task and reduce cognitive demand. Empirical work in this tradition has shown that offloading is mediated by metacognitive evaluation of one's own capabilities, that those evaluations are systematically prone to error, and that the consequences of offloading for unaided performance depend on what is offloaded and how habitually.

This literature is the first reason the framework rejects alarmism. Offloading is not pathological; it is what cognition does when it has tools. Sparrow, Liu, and Wegner (2011) documented this with particular clarity in their analysis of the so-called Google effect, showing that the availability of external information reorganises memory toward where information can be found rather than what it contains. The relevant question has never been whether to offload, but what to offload, to what, under what conditions, and with what residual competence retained in the human.

2.2 Automation Bias and the Discipline of Verification

A second, more cautionary literature has accumulated around automation bias and complacency in high-stakes domains. Bainbridge's (1983) seminal paper on the ironies of automation observed that automating most but not all of a task produces a paradoxical configuration in which the human operator's residual responsibility — typically for monitoring, intervention and recovery from abnormal conditions — requires precisely the skills that automation has rendered uncalled for in the routine course of work. Subsequent work in aviation (Mosier, Skitka, Heers, & Burdick,

1998), simulated flight environments (Skitka, Mosier, & Burdick, 1999), and decision-support systems more generally (Parasuraman & Manzey, 2010) has documented the predictable consequence: humans operating alongside reliable automation tend, over time, to verify less, to challenge less, and to adopt the system's output as default. The phenomenon is not a failure of intelligence. It is, as Parasuraman and Manzey (2010) argue, a predictable adaptation of attention to a tool that is usually right, and one that cannot be overcome by simple practice.

Generative AI presents a sharper version of the same problem. It produces fluent, plausible, contextually appropriate output across an enormous range of domains, and it does so without the explicit confidence calibrations that mature decision-support systems were forced to develop. The conditions for automation bias are, if anything, more favourable in the generative AI context than in the systems where the literature first identified it. Notably, the EU AI Act itself recognises this risk: Article 14(4)(b) requires that personnel assigned to oversight remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (European Union, 2024).

2.3 Deskilling and the Erosion of Tacit Knowledge

A third literature, drawn from labour studies, manufacturing, aviation and medicine, examines what happens to professional skill when procedures that were once learned through repetition are absorbed by machines. Zuboff's (1988) study of computer-mediated work in factories, offices and professional settings made a foundational distinction that remains directly relevant: information technology can either automate, substituting for human action while preserving its underlying logic, or informate, generating new layers of information about the work that demand new intellectual skills from the people who interact with it. Her central argument was that the cognitive consequences of the technology are determined not by the technology itself but by the managerial choices that surround its deployment. Bainbridge's (1983) earlier observation about the ironies of automation runs in parallel: explicit procedural skill survives reasonably well in such transitions; tacit knowledge — the unwritten judgment that distinguishes the experienced professional from the trained one — does not. It is precisely the layer most exposed to erosion when intermediate cognitive steps are no longer performed.

This is the literature most directly relevant to the construct of cognitive atrophy. It establishes that erosion is real, observable and consequential, and that it has been studied with care in domains where the stakes were already high before generative AI existed.

2.4 Human Oversight and AI Governance

A fourth literature, more recent and still consolidating, addresses the governance of AI systems in operational and regulatory contexts. The vocabulary of human-in-the-loop, meaningful human oversight, AI assurance and operational resilience has moved rapidly from research into regulation, most visibly in Article 14 of the EU AI Act (European Union, 2024) but also in sectoral guidance from financial supervisors, health regulators and standards bodies. Article 14 requires that high-risk AI

systems be designed in a way that allows them to be effectively overseen by natural persons, that those persons remain aware of automation bias, and that they retain the ability to interpret outputs, decide not to use the system, and intervene to halt it. Recent legal scholarship has, however, identified a structural tension in the regime. Fink (2025) argues that empirical evidence on human cognitive constraints and automation bias places significant limitations on the effectiveness of oversight as a standalone safeguard, and that overreliance on it as a governance mechanism risks producing oversight that is formally compliant but substantively hollow.

The implicit assumption of much of this literature, however, is that the human in the loop is cognitively available. The framework proposed here argues that this availability is not given. It is produced or destroyed by the way professionals interact with AI systems over time. Oversight that is mandated by regulation but undermined by atrophy is a governance fiction, and the field does not yet have a fully developed vocabulary to name that fiction.

2.5 Augmentation and the Counterweight

A fifth literature, partly overlapping with the productivity discourse but conceptually distinct, addresses human augmentation: the genuine expansion of cognitive capability through interaction with intelligent systems. This literature, traceable through Engelbart's early framing of the augmentation of human intellect and developed across decades of research on cognitive tools, documents cases in which AI assistance does not erode but extends professional capacity — accelerating learning curves, broadening synthetic reasoning, exposing professionals to perspectives and data they would not otherwise encounter. Zuboff's (1988) own analysis of the informating capacity of information technology, properly distinguished from its automating use, belongs in this lineage.

This counterweight is essential to the framework. Without it, atrophy becomes a one-sided narrative and the framework collapses into nostalgia. The construct of cognitive elasticity, introduced in Section 3, is the conceptual instrument by which this counterweight is incorporated.

2.6 The Gap

Each of these literatures addresses a piece of the problem. None addresses the equilibrium between them. Offloading research describes the act; automation bias research describes the trust failure; deskilling research describes the erosion; governance research describes the mandate; augmentation research describes the upside. What is missing is an integrated, dynamic frame that treats human–AI cognitive interaction as a continuous negotiation between forces that pull in opposite directions, and that situates that negotiation as a legitimate object of organisational and professional governance. The Cognitive Equilibrium Framework™ is proposed as a candidate for that frame.

3. Conceptual Foundations

Three constructs anchor the framework. Each is defined functionally rather than mechanistically, behaviourally rather than neurologically. The reason is methodological discipline: the

framework operates at the level at which professional governance can act, which is the level of observable interaction patterns in a working environment.

3.1 Cognitive Atrophy

> *Cognitive Atrophy is a progressive reduction in the active exercise of critical professional cognitive practices, resulting from sustained overreliance on AI-assisted cognitive delegation.*

The construct is deliberately narrow. It refers neither to intelligence nor to neurological function. It does not assert that AI causes cognitive decline in any clinical sense, and the framework explicitly refuses such assertions. What it names is a behavioural and professional pattern, conceptually adjacent to the deskilling effects identified by Bainbridge (1983) and Zuboff (1988), and to the complacency dynamics documented by Parasuraman and Manzey (2010): the gradual disappearance, from the daily texture of work, of the practices through which professional judgment is exercised and renewed.

The relevant indicators are visible in interaction. Validation discipline weakens; the professional who once read a draft suspiciously now scans it for fluency. Challenge behaviour declines; the impulse to ask whether the AI's framing is the right framing fades into the impulse to refine the framing the AI offered. Reasoning persistence shortens; the willingness to sit with a problem before reaching for assistance gives way to the reflex to consult. Understanding becomes shallower; the professional knows what the answer is without being able to reconstruct the path to it. The pattern is consistent with what Skitka, Mosier, and Burdick (1999) documented as errors of omission and commission in operators working alongside automated decision aids — errors not of capacity but of attention reallocated by the presence of a usually-reliable system.

None of these symptoms is dramatic in any single instance. Their significance is cumulative, and their cost becomes visible only when the professional is required to operate without the assistance — to defend a position, to recognise an error in the system's output, to teach a junior colleague how the work is actually done.

3.2 Cognitive Elasticity

> *Cognitive Elasticity is the adaptive reinforcement and expansion of professional cognitive capability through deliberate and balanced interaction with AI systems.*

Elasticity is not the absence of atrophy. It is a distinct, positive phenomenon. Where atrophy describes capability that fades from disuse, elasticity describes capability that grows from a particular kind of use — one in which the AI is treated as an interlocutor rather than as a substitute. Its indicators are the mirror image of atrophy's, but with a generative rather than restorative quality: synthesis becomes broader because the professional has been exposed to more frames; reasoning becomes more flexible because it has been challenged by an interlocutor that does not share the professional's training; learning curves shorten because feedback is faster and cheaper than it has ever been. The construct connects directly to what Zuboff (1988) called the informating capacity of information technology — the capacity not merely to automate work but to render it more visible,

more analysable and, when properly governed, more demanding of intellectual skill.

Elasticity is the reason this framework is not a lament. It is the empirical and conceptual basis for asserting that human–AI interaction can produce professionals who are, by any reasonable measure, more capable than their unassisted predecessors. The discipline of the framework is to insist that this outcome is not automatic.

3.3 Cognitive Equilibrium

> *Cognitive Equilibrium is a dynamic state of balanced human–AI cognitive interaction in which augmentation benefits are achieved without excessive degradation of critical professional cognitive practices.*

Equilibrium is the central construct of the framework, and the most easily misread. It is not a midpoint, an average, or a compromise between atrophy and elasticity. It is a separate condition characterised by the simultaneous presence of augmentation benefits and preserved professional integrity. A professional in equilibrium does not use AI less than one in atrophy; in many cases the volume of use is identical. What differs is the configuration of the relationship: validation persists, challenge persists, reasoning persists, and the professional retains the capability to operate without the system when circumstances require it.

Equilibrium is dynamic. It is not achieved once and held; it is sustained or lost in the daily decisions of how, when and why the AI is consulted. The governance implication, developed in Section 5, is that equilibrium cannot be regulated into existence. It can only be made more or less probable by the design of professional practice and organisational incentives.

3.4 Boundary Conditions of the Framework

Before extending the constructs into a framework, it is useful to state, in compact form, what the framework is not. These boundary conditions are not afterthoughts; they are constitutive of the contribution. The framework's defensibility depends on them being held with discipline throughout.

Boundary Conditions

- The framework does not evaluate intelligence, talent or general capability.
- The framework does not diagnose cognitive impairment, neurological decline or any clinical condition.
- The framework operates at the level of observable professional interaction patterns, not at the level of internal cognitive states.
- The framework is governance-oriented rather than therapeutic, educational or psychological.
- The framework studies interaction configurations between professionals and AI systems, not the systems alone and not the humans alone.

These boundaries are restrictive by design. They preserve the framework's scientific prudence, insulate it from misappropriation by adjacent discourses, and clarify that its proper jurisdiction

is the professional and organisational level at which governance can act.

4. The Cognitive Equilibrium Framework™

4.1 The Logic of the Framework

The framework rejects two seductive simplifications. The first is the linear maturity model, in which organisations and professionals progress through stages of AI adoption, with later stages being uniformly superior to earlier ones. Maturity models are useful for procurement and for political theatre; they are poor descriptions of cognitive dynamics, which can deteriorate as easily as they can develop.

The second simplification is technological determinism, in which the introduction of a sufficiently capable AI system produces predictable cognitive outcomes in the humans who use it. The empirical record across four decades of automation research suggests the opposite (Bainbridge, 1983; Parasuraman & Manzey, 2010; Zuboff, 1988): outcomes are produced by the interaction between system capability, organisational context, professional practice and individual disposition. The framework is built on the assumption that cognitive outcomes are negotiated, not delivered.

4.2 The Continuum

The framework organises human–AI cognitive interaction along a continuum bounded by two opposing pulls and centred on a sustained condition. Figure 1 represents this structure. It is deliberately not a maturity scale: there is no progression from left to right, and the centre is not the average of the two ends. The centre is a distinct configuration, sustained by deliberate practice and lost without it.

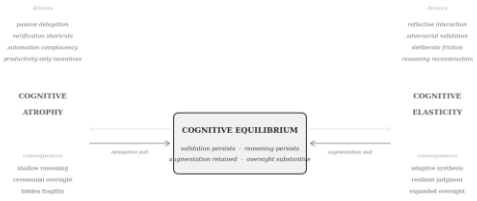

Figure 1. *The Cognitive Equilibrium Framework™. A dynamic tension between two opposing pulls; equilibrium is a sustained condition, not a midpoint.*

The Atrophy Zone

In the atrophy zone, delegation is dominant and validation is residual. AI outputs are accepted with low scrutiny. Challenge behaviour is rare and, when it occurs, is directed at surface form rather than substance. Oversight, where it exists, is procedural — a signature on a document whose reasoning the professional could not reconstruct. Productivity is high; resilience is low. The professional appears competent until the system fails, at which point the absence of underlying capability becomes visible — a configuration that maps directly onto the irony Bainbridge (1983) identified: that the operator most needed in abnormal conditions is precisely the one whose skills have been least exercised in routine ones.

The Equilibrium Zone

In the equilibrium zone, AI is heavily used but never trusted by default. Validation is internalised as professional habit. Challenge is treated as a contribution rather than as friction. Oversight is substantive: the professional retains the capability to detect, explain and correct system errors, and exercises that capability routinely. Productivity is high and durable; resilience is preserved.

The Elasticity Zone

In the elasticity zone, AI interaction actively expands professional capability. The professional uses AI not primarily to accelerate known tasks but to access frames, perspectives and synthetic operations that were previously out of reach. Learning is faster, cross-domain reasoning is more accessible, and judgment becomes more rather than less rigorous over time. The elasticity zone is rare, requires deliberate practice, and is unstable without organisational reinforcement.

4.3 Drivers of Movement Along the Continuum

Movement between zones is produced by drivers that operate simultaneously at the level of the individual, the team and the organisation. The framework identifies two clusters, summarised in Table 1 and discussed in turn.

Atrophy Drivers	Elasticity Drivers
Passive AI consumption	Reflective AI interaction
Delegation normalisation	Independent framing
Verification shortcuts	Adversarial validation
Productivity-only incentives	Deliberate reasoning persistence
Reduced ambiguity tolerance	Structured cognitive friction

Table 1. *The two clusters of drivers that shape movement along the continuum.*

Atrophy drivers operate, for the most part, by default. They do not need to be installed; they emerge from the natural pressures of professional environments under productivity expectations. Elasticity drivers do not. They require explicit cultivation, organisational patience, and a theory of professional development that treats reasoning as something to be exercised rather than merely produced. The asymmetry is consequential: in the absence of deliberate counter-practice, the gravitational pull of professional environments is toward atrophy, not toward equilibrium. This asymmetry is consistent with the broader finding from accountability research in the automation bias tradition, that internalised perceptions of accountability — rather than externally imposed ones — are the most reliable predictor of sustained verification behaviour (Skitka, Mosier, & Burdick, 1999).

4.4 The Delegation–Verification Erosion Loop

The asymmetry between atrophy drivers and elasticity drivers is not merely additive. It is reinforced by a self-perpetuating dynamic that the framework names the delegation–verification erosion loop. The loop runs as follows.

Increasing AI reliability reduces the perceived cost of delegation. As delegation expands, the marginal effort required for verification rises relative to the apparent benefit of verifying. Verification, as a result, declines. Reduced verification weakens the habit of challenge, because there is no occasion on which challenge is exercised. Weakened challenge, in turn, shortens reasoning persistence: the professional no longer rehearses the cognitive operations that challenge would have required. Reasoning persistence, once shortened, normalises dependency, because the alternative — sustained

independent reasoning — has become unfamiliar. Normalised dependency degrades oversight quality, because oversight that is not grounded in active reasoning becomes ceremonial. And ceremonial oversight produces hidden operational fragility: the appearance of governance without its substance.

The loop is probabilistic, not deterministic. It is moderated by individual disposition, professional training, organisational incentives and the design of the AI systems themselves. None of these moderators is automatic, and the absence of any of them allows the loop to run. The framework's claim is not that the loop is inevitable, but that it is the default trajectory in environments that have not deliberately chosen otherwise.

The implication for governance is direct. Interventions aimed at any single point in the loop — better verification tools, mandated review, training in challenge behaviour — are necessary but rarely sufficient, because the loop has multiple reinforcing mechanisms. Equilibrium is sustained by interventions distributed across the loop, not concentrated at any one point.

4.5 The Object of Assessment

A clarification is necessary, because it is the point at which most frameworks of this kind fail. The Cognitive Equilibrium Framework™ does not assess organisations. It assesses professionals interacting with AI systems within organisational contexts. The distinction matters because it determines what can be done with the framework's outputs.

Organisations cannot be in atrophy or in elasticity; only people can. What organisations can do is shape the probability that the people within them move toward one zone or the other. This is, properly understood, the substance of human capability governance in the AI era, and it is the bridge between the conceptual framework and the operational implications developed in the next section.

5. Governance and Professional Implications

5.1 Why Governance, Not Wellness

A reasonable reader might ask why the framework insists on governance vocabulary when its constructs sound, at points, like territory better served by educational psychology, organisational development or even cognitive coaching. The reason is precise. The phenomena described here have direct consequences for accountability, oversight, regulatory compliance and operational resilience. Treating them as wellness concerns subordinates them to discretionary budgets and personal initiative. Treating them as governance concerns places them where they belong: in the formal architecture by which organisations decide who is responsible for what, and how that responsibility is sustained.

Table 2 makes the reframing explicit. Each row identifies a way in which the phenomena described by this framework would conventionally be categorised, and the parallel category through which the framework proposes they should be understood instead. The shift is not cosmetic; it relocates the phenomena from a register in which they are addressed by individual and discretionary means to one in which they are addressed by structural and accountable ones.

Traditional Framing	Cognitive Equilibrium Framing
Wellness issue	Governance issue
Learning challenge	Oversight challenge
Productivity optimisation	Human capability preservation
Tool adoption	Professional judgment sustainability

Table 2. *The reframing performed by the Cognitive Equilibrium Framework™.*

5.2 Implications for Professional Development

Professional development practices designed before generative AI assumed that the principal risk to capability was insufficient learning. The risk in the AI-integrated environment is different: it is the gradual disappearance of the practices that capability requires. This shifts the design of professional development from accumulation toward maintenance. Curricula need to include not only what professionals should learn but what they should continue to do, even when the AI could do it for them, in order to preserve the capability that makes them professionals in any meaningful sense.

5.3 Implications for AI Governance and Oversight

The current generation of AI governance frameworks treats human oversight as a control mechanism. The framework proposed here treats it as a capability that decays. The implication is that oversight obligations cannot be discharged by structural design alone — by appointing reviewers, defining checkpoints, or installing approval workflows. They require the active maintenance of the cognitive practices that make oversight substantive rather than ceremonial. An oversight regime that produces signatures without scrutiny is not oversight; it is liability transfer dressed as accountability. The concern is not hypothetical: Article 14 of the EU AI Act explicitly requires that personnel assigned to oversight remain aware of automation bias (European Union, 2024), and recent legal scholarship has noted that the effectiveness of the entire oversight regime depends on cognitive conditions that the regime itself does not produce (Fink, 2025).

5.4 Implications for Operational Resilience

Resilience, in operational and regulatory terms, is the capability to continue functioning under stress, including the stress of system failure. In an AI-saturated professional environment, this capability is increasingly co-located with the human's preserved ability to perform tasks the AI normally performs. An organisation in which professionals have moved deep into the atrophy zone is, by definition, less resilient to AI failure than one in which equilibrium has been preserved, regardless of how sophisticated its technical redundancy may be. This is the contemporary form of the irony Bainbridge (1983) identified four decades ago: that the human capability most needed in failure is the one most likely to have eroded in success.

5.5 Implications for Executive Education

Executive education has, in many institutions, treated AI as a content area: what it is, what it

does, what it costs, what it threatens. The framework suggests a different orientation. The most consequential question for senior leaders is not what AI can do but what their organisations are doing to the people who work alongside it, and whether the resulting cognitive configuration is the one they would consciously choose. This reframes AI literacy at the executive level from a technological subject to a governance subject.

5.6 Human Capability Governance

Across the implications above runs a single thread: the emergence of human capability as an explicit object of governance, distinct from talent management, learning and development, or HR strategy. Human capability governance asks who, in the organisation, is accountable for the cognitive condition of the professionals working with AI systems; what evidence that accountability is exercised against; and what the organisation's position is when the answer to either question is unsatisfactory. The framework does not provide that answer. It provides the language in which the question becomes formulable.

6. The Cognitive Atrophy Project™

The Cognitive Atrophy Project™ is the operational and research extension of the framework articulated in this paper. It is a deliberate naming choice. The Project is not a Cognitive Equilibrium Project, although equilibrium is its conceptual centre, because the work begins from the recognition that the deviation from equilibrium most likely to occur, and most likely to be ignored, is atrophy. The name is a discipline against forgetting why the framework exists.

The Project comprises five interconnected components. The Cognitive Equilibrium Framework™, set out in this paper, is its conceptual foundation. The Cognitive Equilibrium Check™ is an early reflective operationalisation layer of the framework, designed not as diagnostic instrument or psychometric evaluation but as structured observation of the interaction patterns associated with cognitive delegation, reasoning persistence and validation behaviour in AI-assisted environments. Cognitive Elasticity Practices™ is a curated set of professional and organisational practices intended to make movement toward the elasticity zone more probable. Executive Education Programmes translate the framework into the formats through which senior leaders most reliably encounter it. A longitudinal research vision, finally, anchors the Project's commitment to revising the framework as empirical evidence accumulates.

The relationship between the framework and the Cognitive Equilibrium Check™ deserves to be made explicit. The Check is reflective, not evaluative. It is intended to surface, for the professional and the organisation using it, the configuration of their interaction with AI systems against the constructs articulated here. It does not produce a score that ranks individuals against each other. It produces a structured observation that supports deliberation. This positioning is intentional and reflects the Project's commitment to the boundary conditions stated in Section 3.4.

The Project is not, and should not be read as, a product. It is a research-informed platform

whose components mature at different rates and whose authority depends on the rigour with which the conceptual foundation is held. This paper is the foundation.

7. Research Agenda

The framework is conceptual; it requires empirical examination it does not yet have. The most pressing research directions are the following.

Longitudinal equilibrium analysis. The atrophy and elasticity hypotheses are, at root, claims about change over time. They cannot be tested in cross-sectional designs. The most informative work in this area will be longitudinal, ideally tracking cohorts of professionals through the early years of intensive AI integration and after, in the spirit of the longitudinal turn that Risko and Gilbert (2016) identified as a priority for cognitive offloading research.

Profession-specific dependency patterns. The constructs are general; their manifestations are not. The validation behaviours of a radiologist, an investment analyst, a tax lawyer and a software engineer differ in form even when they are similar in function. A productive research programme will produce profession-specific elaborations of the framework, sensitive to the cognitive practices that define each domain.

Cognitive elasticity intervention studies. If elasticity drivers can be identified, they can be tested. Designed interventions — at the level of practice, team or organisation — that aim to move interaction patterns toward elasticity, and that are evaluated against pre-specified indicators, would constitute the most direct empirical engagement with the framework.

Validation behaviour measurement. Validation is the single most diagnostic behaviour in the framework. Methods for observing it without distorting it — through interaction analysis, structured think-aloud protocols, or unobtrusive workflow telemetry — are an instrumentation priority.

Oversight robustness indicators. The link between cognitive condition and the substantive quality of human oversight is asserted in this paper but not yet quantified. The development of indicators that connect interaction patterns to oversight outcomes, in dialogue with the legal and regulatory literature on Article 14 (Fink, 2025), would strengthen both the framework and the regulatory regimes that depend on the assumption that oversight is real.

Decision-quality correlations. Productivity is easy to measure; decision quality is not. Yet the framework's ultimate claim is that decision quality is what equilibrium preserves. Methodological work on decision-quality measurement, in environments where AI is in the loop, is among the most consequential research directions implied by the framework.

8. Limitations

The framework is deliberately bounded, and several boundaries deserve to be made explicit, beyond those stated as constitutive boundary conditions in Section 3.4.

First, the framework does not diagnose medical or neurological conditions, and it does not authorise such diagnoses. Its constructs are behavioural and professional. The use of biological metaphor — atrophy, elasticity — is rhetorical and conceptual, not clinical.

Misreading the framework as a theory of cognitive decline is a category error that the paper has tried, throughout, to forestall.

Second, the framework does not measure intelligence, capability or expertise in any general sense. It describes the configuration of a particular kind of professional interaction. A professional who is in the atrophy zone with respect to AI-mediated work is not, by virtue of that fact, less intelligent or less capable than one in equilibrium. The framework is silent on questions it is not designed to address.

Third, the framework does not claim that AI causes cognitive decline. It claims that sustained patterns of interaction with AI systems, in the absence of deliberate counter-practices, produce observable changes in professional cognitive behaviour, and that these changes have governance consequences. The causal language is interaction-level, not biological.

Fourth, the framework is conceptual. It has not been validated empirically, and it does not pretend to have been. Its function is to provide a stable conceptual ground for the empirical work that should follow. Premature operationalisation — the conversion of the framework into assessments, scores or interventions in the absence of empirical grounding — would betray the discipline the framework is trying to introduce.

Fifth, the framework is situated. It speaks most directly to professional and organisational environments in which generative AI is integrated into knowledge work. Its applicability to other settings — education, public administration in low-resource contexts, creative practice outside professional structures — is plausible but undemonstrated.

Sixth, the framework may manifest differently across professions, organisational cultures and regulatory environments. The pressures that drive interaction patterns toward atrophy or elasticity are not universal: they vary with the structure of professional accountability, the nature of AI adoption incentives, the regulatory regime in which the work is performed and the cultural conventions that govern challenge, deference and authority within the workplace. The framework's constructs are intended to be portable; their empirical manifestations are not, and any application of the framework outside the contexts in which it has been refined should proceed with that variance explicitly in view.

9. Conclusion

The most consequential question facing professional environments in the next decade is not whether artificial intelligence should augment human cognition. That question is settled by the facts on the ground. The question is whether organisations and professions are willing to take seriously what augmentation does to the humans inside it, over time, and in the absence of deliberate counter-practices.

Productivity is the most easily measured benefit of generative AI integration, and for that reason it has crowded out most other measures. Equilibrium — the simultaneous preservation of augmentation benefits and the cognitive practices on which professional judgment depends — is harder to measure, harder to govern, and harder to defend in environments under quarterly pressure. But equilibrium is what makes the

productivity sustainable, and it is what makes the human oversight that regulatory regimes increasingly demand into something more than a fiction.

The framework offered here is a beginning, not a conclusion. It names three constructs, articulates them as the dimensions of a dynamic interaction, and proposes that this interaction is a legitimate object of professional and organisational governance. It does so in conceptual terms, conscious that the empirical work has not yet been done and that the framework will require revision as it is.

The closing proposition of the paper is the one to be defended in everything that follows it. The future of AI maturity will not be decided by the sophistication of the systems alone. It will be decided, also, by whether the professions that work alongside those systems retain the cognitive practices through which judgment, oversight and adaptive reasoning are renewed. That retention is not given. It is governed, or it is lost.

References

Bainbridge, L. (1983). Ironies of automation. Automatica.

European Union. (2024). Artificial Intelligence Act (EU Regulation 2024/1689). Official Journal of the European Union.

Fink, M. (2025). Human oversight under Article 14 of the EU AI Act. SSRN.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. ACM Transactions on Computer-Human Interaction.

Hutchins, E. (1995). Cognition in the wild. MIT Press.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. The International Journal of Aviation Psychology.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? International Journal of Human-Computer Studies.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. Science.

Zuboff, S. (1988). In the age of the smart machine: The future of work and power. Basic Books.