U.PORTO

MASTER
ECONOMICS OF BUSINESS AND STRATEGY

The Tool Is Universal, The Capability Is Not

Individual practice and the limits of organisational AI capability in early enterprise adoption

Nuno Veiga Coutinho

M

2026



U.PORTO
FEP FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

The Tool Is Universal, The Capability Is Not

Individual practice and the limits of organisational AI capability in early enterprise adoption

Nuno Veiga Coutinho

Dissertation

Master in Economics of Business and Strategy

Supervised by

Nuno Alexandre Meneses Bastos Moutinho, PhD

2026

Acknowledgments

This dissertation marks the end of a master's programme that took me across three continents. Throughout that experience, and in bringing this work to completion, there were always people I could count on, wherever I was.

My family, especially my parents and my uncle Miguel, have always been there for me. They supported my choices without asking me to justify them, and their support and trust continue to inspire me to do better. My friends kept me close. They checked in when I was hard to reach, answered calls at strange hours, and made it easy to be honest about how I was doing. To my partner, thank you for sharing the weight of this international life, for your patience and unconditional support, and for making the busiest days feel lighter.

Professor Nuno Moutinho, both a supervisor and a friend, guided the work with precise feedback and immense patience, always making time for me, in person or online, whatever time zone I was writing from. When the research needed doors opened, he kept knocking until they did. He was as dedicated and stubborn about this dissertation as I was, which is the highest compliment I can pay him. I also thank the professors and administrative staff at FEP, the external specialists, and the students who gave their time and shared their experiences.

To all of you, in Portugal and beyond, thank you.

Abstract

Enterprise generative artificial intelligence is widely available to knowledge workers, yet evidence that individual task-level gains cumulatively generate greater value remains limited. This dissertation investigates how a knowledge-intensive organisation might begin to develop organisational capability through individual AI use, and which conditions enable or constrain this early development. It examines the first months of Microsoft 365 Copilot use at a Portuguese university faculty through a retrospective instrumental case study with embedded academic and administrative contexts. The framework method was applied deductively to 17 sessions: 7 interviews with academic staff, 1 administrative staff focus group and 9 interviews with AI specialists from other organisations. The analysis included institutional documents and, descriptively, exported institutional usage data and a student questionnaire. Evidence was recorded passage by passage, linking each judgement to its source and examining it against negative cases and a rival explanation.

Workflow redesign recurred without spreading among colleagues; competence circulated through isolated exchanges without producing reusable artefacts; verification was consistently reported as rigorous and individually adopted, but lacked shared procedures, escalation routes, or accountability; and personal curation substituted for organisational data arrangements. Every configured artefact was built, held and maintained by one person.

The study identifies an individual-organisational gap: the components of capability are present, but the relationships needed to make practices visible, transferable and maintainable remain underdeveloped. Complementarity, it argues, depends on both level and combination. As a single early-stage case analysed by one researcher, it establishes neither causality nor realised value, nor whether practices would survive their originators; the findings therefore indicate only how such capability may begin to form.

JEL Codes: O33, D23, D83, M15, I23.

Keywords: Enterprise Generative AI; Organisational capability; Complementarity; Technology Assimilation; Knowledge-intensive Organisations; AI Governance; Case Study.

Resumo

A inteligência artificial generativa empresarial está amplamente disponível para profissionais cujo trabalho assenta em conhecimento especializado. Contudo, é limitada a evidência de que ganhos individuais nas tarefas geram, cumulativamente, maior valor. Esta dissertação investiga como uma organização intensiva em conhecimento pode começar a desenvolver capacidade organizacional através da utilização individual de IA e que condições favorecem ou limitam esse processo. Examina os primeiros meses de utilização do Microsoft 365 Copilot numa faculdade portuguesa, através de um caso de estudo instrumental retrospetivo que integra os contextos académico e administrativo. O método de *framework* foi aplicado dedutivamente a 17 sessões: 7 entrevistas a docentes, 1 *focus group* com administrativos e 9 entrevistas a especialistas em IA de outras organizações. A análise incluiu documentos institucionais e, descritivamente, dados exportados de utilização institucional e um questionário a estudantes. A evidência foi registada por excerto, associando cada avaliação à fonte e confrontando-a com casos negativos e uma explicação concorrente.

A reformulação de fluxos de trabalho foi recorrente, sem disseminação entre colegas; as competências circularam em trocas pontuais, sem produzir artefactos reutilizáveis; a verificação foi consistentemente descrita como rigorosa e individual, mas faltaram procedimentos comuns, canais de encaminhamento de problemas e mecanismos de responsabilização; e a seleção e organização pessoal de dados substituíram a organização institucional. Cada artefacto configurado foi criado, detido e mantido por uma única pessoa.

O estudo identifica um desfasamento entre indivíduo e organização: existem os componentes da capacidade, mas as relações necessárias para tornar as práticas visíveis, transferíveis e passíveis de manutenção permanecem pouco desenvolvidas. Argumenta-se que a complementaridade depende do nível e da combinação desses componentes. Circunscrito a um caso de adoção inicial, analisado por um investigador, o estudo não demonstra causalidade nem valor gerado, nem determina se as práticas persistiriam sem os criadores; indica apenas como essa capacidade poderá começar a formar-se.

Códigos JEL: O33, D23, D83, M15, I23.

Palavras-chave: Inteligência Artificial Generativa Empresarial; Capacidade Organizacional; Complementaridade; Assimilação Tecnológica; Organizações Intensivas em Conhecimento; Governação da IA; Caso de Estudo.

Table of Contents

1. Introduction 1

2. Literature Review 3

2.1. From Artificial Intelligence to Enterprise Generative AI 3

2.2. Task-Level Effects of Generative AI on Knowledge Work 4

2.3. From Adoption to Assimilation 5

2.4. Organisational Complementarity and the Translation of Task-Level Effects 6

2.5. Complementarity Mechanisms 7

2.6. Adoption in Knowledge-Intensive Organisations (KIOs) 8

2.7. Defining AI-Enabled Organisational Capability 10

2.8. Research Gap and Research Questions 11

3. Methodology 13

3.1. Research design 13

3.2. The case, the participants and the evidence 14

3.3. Data collection 16

3.4. The analytical framework 18

3.5. Building the evidence base 19

3.6. From evidence to judgements 21

3.7. Testing the account 23

3.8. Credibility: auditability and consistency 24

3.9. Ethics and positionality 25

3.10. Limitations 26

4. Findings 27

4.1. The setting and task-level baseline 27

4.2. Workflow redesign: recurring individual redesign that does not travel laterally 28

4.3. Skills and absorptive capacity: competence without conversion 30

4.4. Governance: a consistent norm without an arrangement 31

4.5. Data readiness: personal curation in place of organisational arrangement 33

4.6. The knowledge-intensive conditions and the coordination of the mechanisms 34

4.7. Early indications of capability formation 36

4.8. The external comparison and synthesis 37
5. Discussion 39
5.1. The individual-organisational gap 39
5.2. Ownership and authority: why nothing travelled 40
5.3. What the rival account explains, and what it cannot 42
5.4. Professional knowledge and the boundary of codification 43
5.5. What the account establishes, and its reach 44
6. Conclusion 45
References 47
Annex A. Source notes for the quantitative figures 51
A1. The tenant usage export 51
A2. The student questionnaire 52
Annex B. Consistency controls and the composition of the evidence base 54
B1. Automated consistency checks 54
B2. Composition of the evidence base 55
Annex C. What each research instrument reached 56
C1. Professor interviews 57
C2. The administrative focus group 57
C3. Specialist interviews 58
C4. Fieldwork chronology 59
Annex D. Context-level judgements and the evidence behind them 60
D1. The judgement grid at informant level 60
D2. Basis for each judgement: the academic context 61
D3. Basis for each judgement: the administrative context 63
D4. Domain-level synthesis 66
Annex E. Knowledge-intensive conditions and the coordination of the mechanisms 67
E1. The 4 conditions against each mechanism 68
E2. Coordination among the mechanisms 71
Annex F. Evidence behind the capability indicators 72

F1. Enabling, limiting and contradictory evidence 73
F2. Evidence held outside the indicator judgements 73
Annex G. Testing the account 74
G1. The 12 propositions and their outcomes 74
G2. Rejection conditions and negative-case searches 76
G3. The relation each evidence stream bears to each proposition 78
G4. The rival-pattern comparison 78
Annex H. The external comparison 80
H1. Specialist accounts by position in the adoption trajectory 80
H2. Standpoints and source criticism 81

Index of Tables

Table 1. Context-level judgements by mechanism and precondition 34
Table 2. The four capability indicators at case level 36

1. Introduction

Knowledge-intensive organisations rely on analysis and professional judgement (Starbuck, 1992; Alvesson, 2004; Lönnqvist & Laihonen, 2017), making generative AI particularly relevant to their work. Studies show that generative assistants can improve performance on bounded knowledge-work tasks (Noy & Zhang, 2023; Peng et al., 2023; Brynjolfsson et al., 2023). Neither task-level gains nor usage figures show that the resulting practices are embedded in collective work. Technology may be widely available and actively used while effective practices remain fragmented, locally held, and dependent on their originators. This dissertation examines that transition during the first months of Microsoft 365 Copilot use at the Faculty of Economics and Management of the University of Porto.

Generative AI may improve performance where system and task fit but impair it where the system is unreliable (Dell'Acqua et al., 2026). Effective use therefore requires users to judge when assistance is appropriate and verify outputs independently. Existing research mainly examines individuals rather than the processes through which outputs are reviewed, coordinated, approved, and implemented. It establishes task-level effects but says little about how isolated improvements become sustained changes in collective work. Gallivan's (2001) distinction between adoption and assimilation clarifies this limitation. Assimilation comprises breadth, reflected in the number of adopters, and depth, reflected in changes to processes and structures. Licensing establishes availability and observed use indicates breadth, but neither demonstrates depth. Substantial uptake may therefore coexist with limited integration into how work is performed. Complementarity research suggests how such depth might develop. Information technology creates value through complementary investments in processes, work practices, skills, and structures (Brynjolfsson & Hitt, 2000). Sociotechnical accounts similarly explain how arrangements emerge and change through technology use (Trist & Bamforth, 1951; Orlikowski, 1992). In knowledge-intensive settings, however, professional autonomy, ambiguous outputs, tacit expertise, and decentralised coordination may enable local adaptation while binding effective practices to particular users and tasks (Alvesson, 2004; Lönnqvist & Laihonen, 2017). Mikalef and Gupta (2021) define AI capability as an organisation's ability to select, orchestrate, and leverage AI-specific resources, but do not explain how it develops from early task-level use. Existing research therefore demonstrates individual effects, distinguishes adoption from assimilation, and identifies the importance of complementary investments. What remains underexplored is the process

connecting these levels: how individually developed practices become coordinated, shared, reproducible, and sustainable beyond their originators, or why that transition stalls during early adoption. The study therefore asks how a knowledge-intensive organisation might develop AI-enabled organisational capability from individual, task-level use, and which conditions appear to shape that process during early adoption. Capability is defined as the ability to embed AI-assisted practices in recurring work so that they can be shared, reproduced, and sustained beyond their originating users and tasks. The analytical framework examines three complementary domains: workflow redesign, skills and absorptive capacity, and governance. Data readiness is treated as a cross-cutting condition for information-dependent embedding. The sub-questions consider how these domains develop and become coordinated at FEP, how knowledge-intensive conditions shape that process, and how the resulting pattern compares with specialists' accounts from other settings.

This study adopts a retrospective, single instrumental case study with an embedded design. Its unit of analysis is the process through which AI-assisted work becomes, or fails to become, embedded beyond its originators. Academic and administrative work provide the embedded contexts. Evidence comprises seventeen interview and focus-group sessions involving professors, administrative staff, and external specialists, together with institutional documents and two descriptive quantitative sources. Qualitative material was analysed using the framework method described by Gale et al. (2013), while quantitative evidence contextualised the scale and distribution of use. This thesis provides a process-level account of the transition from individual practice to organisational capability. It examines how workflow redesign, competence, governance, and data readiness may develop locally without becoming coordinated across the institution. It thereby explains why availability, substantial uptake, and sophisticated individual applications are not equivalent to capability, and identifies how local practices may begin to become visible, transferable, reproducible, and sustainable. This contribution connects evidence of task-level effects with the wider process through which enterprise generative AI begins to acquire a collective and enduring form.

The dissertation proceeds as follows. Chapter 2 reviews the literature and develops the analytical framework. Chapter 3 presents the case-study design and establishes the evidential boundaries. Chapter 4 reports the findings on the development and coordination of AI-assisted practices at FEP. Chapter 5 interprets the pattern and discusses its contribution and implications. Chapter 6 presents the overall conclusion.

2. Literature Review

2.1. From Artificial Intelligence to Enterprise Generative AI

Artificial intelligence is a broad and shifting category. Systems within it vary in autonomy, learning, and inscrutability and therefore raise different managerial questions (Berente et al., 2021; Haenlein & Kaplan, 2019). To maintain analytical precision, this dissertation focuses on generative AI: computational techniques that learn patterns from training data to produce seemingly new and meaningful content, such as text, images, audio, and code (Feuerriegel et al., 2024). Enterprise generative AI refers here to the systems and applications that an organisation provides, configures, integrates, or evaluates for use in its work. "Enterprise" thus identifies only the organisational setting.

Many contemporary generative AI systems are built on foundation models, which are trained on broad data, generally at scale, and can be adapted to varied downstream tasks (Bommasani et al., 2021; Schneider et al., 2024). However, a foundation model is only the technical core of what an organisation uses. A generative AI system combines the model with infrastructure, data processing, and user-facing components, while an application situates that system within a particular use case (Feuerriegel et al., 2024). For example, GPT-4 is a foundation model, while Microsoft 365 Copilot is the generative AI system through which users interact. Using Copilot to summarise emails, draft documents, or analyse spreadsheets represents an application of that system.

Properties of foundation models can carry over into systems and, consequently, into applications. Two are particularly relevant: emergence and homogenisation (Bommasani et al., 2021). Emergence refers to capabilities or behaviours that were neither explicitly designed nor anticipated. Homogenisation arises when multiple applications rely on the same model or model family, allowing not only improvements, but also inherited biases, vulnerabilities, and defects to spread across them. In addition, outcomes are further shaped by prompt sensitivity, as instructions from providers, developers, and users can materially shape the system's behaviour (Schneider et al., 2024).

These distinctions suggest that performance cannot be treated as a fixed property of the underlying model. It reflects the combined influence of the model's capabilities, the system's infrastructure and data processing, the application's configuration and available information, and the task, user, and organisational context in which it is used (Berente et al., 2021;

Feuerriegel et al., 2024; Schneider et al., 2024). Empirical findings must accordingly be interpreted in relation to the conditions under which they were observed. The next section examines these findings at the most immediate level of use: bounded knowledge-work tasks.

2.2. Task-Level Effects of Generative AI on Knowledge Work

Controlled experiments and field studies show that generative AI can improve performance on bounded knowledge-work tasks. Noy and Zhang (2023) found that college-educated professionals with access to a general-purpose assistant completed incentivised writing tasks approximately 40% faster and produced higher-rated outputs. Peng et al. (2023) found that developers using a coding assistant completed a standardised programming exercise 55.8% faster than those in the control group. Across three field experiments involving 4,867 developers, Cui et al. (2025) estimated an approximately 26% increase in completed tasks. Brynjolfsson, Li, and Raymond (2023) similarly found that access to a generative assistant increased issues resolved per hour by approximately 14% among 5,179 customer-support agents. While these measures are not directly comparable, they support that generative assistance can improve speed, quality, or throughput.

These effects vary with the task and the capabilities of the system being used. Dell'Acqua et al. (2026) found that consultants assisted by a generative AI system worked faster and produced higher-quality outputs on tasks that prior testing showed the system could perform effectively. However, on one complex task that it could not solve reliably, assisted consultants were less likely to reach the correct conclusion. Their recommendations were nevertheless rated as more coherent and persuasive, even when incorrect. The same system can therefore improve or impair performance across similar tasks. However, effective use requires users to judge where assistance is appropriate and to verify outputs independently.

Benefits also differ across users. Noy and Zhang (2023) found that generative assistance reduced performance differences across participants, and Brynjolfsson et al. (2023), Peng et al. (2023), and Cui et al. (2025) each document disproportionately large gains for novice and lower-skilled workers. While generative assistance can narrow some task-specific performance gaps, expertise has not become dispensable. When a system performs more of the initial work, expertise may shift partly from producing an output to specifying, evaluating, verifying, and correcting it.

Most studies examine bounded tasks – a written draft, a coding exercise, or a customer request – rather than the wider processes through which outputs are reviewed, approved, coordinated, and implemented. Their findings therefore only establish task-level performance effects, not organisational value. They do not show whether these gains persist through downstream activities, become embedded in recurring workflows, or contribute to organisation-level outcomes. The poorer result on a task beyond the system's capabilities (Dell'Acqua et al., 2026) further illustrates why task-level gains may not carry through to the wider process: verification and correction may consume resources that the reported measures do not capture.

Generative AI has conditional task-level potential, but evidence of that potential cannot establish whether localised improvements acquire an organisational form. The first step to answer that question is to distinguish the provision and use of technology from its assimilation into organisational work.

2.3. From Adoption to Assimilation

Adoption research distinguishes an organisational decision to adopt a technology from its subsequent use and assimilation. Gallivan (2001) describes the organisational decision as primary adoption and adoption by intended users as secondary adoption. Assimilation extends beyond either decision. It concerns how extensively the innovation is used and how deeply that use affects organisational processes, structures, and culture. It therefore has two dimensions: breadth, the number of adopters, and depth, the extent of use and its organisational impact.

Applied to enterprise generative AI, organisational provision indicates that a primary adoption decision has been made. Usage can indicate secondary adoption and, when considered across users, breadth. Neither establishes depth. Because breadth and depth are distinct, an application may be used widely without becoming deeply embedded in organisational work, while deeper use may remain confined to one setting. Provision and usage therefore do not, by themselves, show that AI-assisted practices have become assimilated.

Gallivan developed this framework for complex technologies selected by organisational authorities and often mandated for intended users. It cannot therefore be transferred wholesale to enterprise generative AI. This dissertation uses its distinctions as an analytical

guide without importing the complete explanatory framework or assuming that all practices follow a fixed sequence. These distinctions clarify the translation problem established in Section 2.2: task-level performance, organisational provision, and individual uptake do not show whether use has altered wider organisational work or produced organisational value. Explaining movement towards assimilation requires shifting attention to the organisational changes that develop around use.

2.4. Organisational Complementarity and the Translation of Task-Level Effects

Brynjolfsson and Hitt (2000) synthesise firm-level evidence that a significant component of IT's value arises through complementary organisational investments, including changes in business processes, work practices, skills, and organisational structures. A sociotechnical perspective adds a processual account of this relationship. Trist and Bamforth (1951) examine a work system as an interaction between its technical arrangements and the social organisation of work. Orlikowski (1992) provides a related account in which human action shapes technology, while technology, through its use, enables and constrains subsequent action. The contribution of this perspective is therefore to show that organisational practices and arrangements may be produced, reinforced, or altered as users incorporate a technology into their work.

The productivity J-curve adds a temporal qualification to the complementarity argument. Brynjolfsson, Rock, and Syverson (2021) argue that general-purpose technologies require complementary intangible investments, including new processes and human capital. Because conventional productivity measures do not capture these investments as output, measured productivity growth may initially understate true productivity growth. Drawing on this reasoning, Simons et al. (2024) suggest that limited firm-level evidence of AI-related productivity effects may partly reflect the cost and time required to develop complementary assets. This explanation offers one reason why organisational effects may lag AI adoption and task-level benefits, but it does not establish that the relevant complements will develop or that organisational value will ultimately result.

Together, these perspectives frame the movement from localised use towards assimilation as a process in which complementary organisational arrangements develop through use and over time.

2.5. Complementarity Mechanisms

To examine this process, the analytical framework centres on three mechanisms: workflow redesign, skills and absorptive capacity, and governance. Each respectively addresses a distinct organisational domain: how work is structured, how knowledge is developed and shared, and how AI use is directed and controlled. Their operation is supported by data readiness, which is treated as a cross-cutting condition rather than as a fourth mechanism. The mechanisms are not intended to be exhaustive. Leadership, change management, infrastructure, and performance measurement also feature prominently in the literature. Nonetheless, they are treated as influences on the domains rather than separate mechanisms.

Workflow redesign changes how recurring work is sequenced, allocated between people and AI, coordinated across roles, and verified (Demirer et al., 2026). Davenport (1993) distinguishes incremental improvement to an existing process from process innovation that redesigns how work is performed, arguing that the latter offers greater potential gains. Because general-purpose assistants can be incorporated readily into existing tasks, augmentation may occur without altering the surrounding process. Workflow redesign nonetheless goes further by redistributing tasks, handoffs, responsibilities, and verification across roles or units. Accordingly, generative AI use begins to acquire an organisational character when it changes how recurring work is coordinated, not only how an individual task is performed.

Skills and absorptive capacity concern how the knowledge required for effective use develops and spreads. At the individual level, relevant skills include framing tasks, selecting suitable applications, formulating prompts, and evaluating outputs. Research on workplace AI associates the development of these skills with experimentation, role-specific training, and organisational support (Mayer et al., 2025; Microsoft & LinkedIn, 2024). Individual competence, however, does not show that learning has extended beyond the user. Cohen and Levinthal (1990) define absorptive capacity as the organisational ability to recognise the value of new external knowledge, assimilate it, and apply it. This capacity builds on related prior knowledge and communication across organisational units. In the present framework, it explains how local learning can inform shared organisational practice. Without such processes, generative AI use may remain ad hoc and isolated (Mayer et al., 2025).

Governance directs and controls AI use. Mäntymäki et al. (2022) define AI governance as the rules, practices, processes, and technological tools through which organisations align AI use with strategic, legal, and ethical requirements. These arrangements define acceptable uses, allocate decision rights, and connect AI oversight with data governance. Their influence, however, depends on how they are enacted. Mittelstadt (2019) argues that high-level AI principles have limited practical effect unless organisations translate them into operational mechanisms and establish accountability. Thus, governance supports embedding when standards, responsibilities, and accountability become part of recurring work.

Data readiness differs from the three mechanisms because it concerns the informational conditions under which they operate. Jöhnk et al. (2021) identify data availability, quality, accessibility, and flow, together with suitable infrastructure, as central dimensions of AI readiness. Mikalef and Gupta (2021) similarly emphasise data integration, sharing, cleansing, storage, and security. These conditions require continuing organisational work, and their importance varies across applications. Data limitations do not prevent all generative AI use. They do, however, constrain forms of embedding that depend on shared organisational information, retrieval, or system integration. Data readiness is therefore a cross-cutting precondition for information-dependent embedding, not a fourth mechanism.

Although distinct, the three mechanisms may reinforce one another. Milgrom and Roberts (1990, 1995) define activities as complements when strengthening one increases the return to strengthening another. Redesigned workflows can create recurring settings in which users apply and share what they learn. Skills can support evaluation and verification within those workflows, while governance can assign responsibilities and establish consistent standards across them. Where these arrangements draw on organisational information, their operation also depends on accessible, usable, and appropriately governed data (Mikalef & Gupta, 2021).

Potential complementarity does not ensure coherent development. Coordination still depends on how knowledge moves, how discretion is exercised, and where authority resides. Organisational context therefore shapes how these mechanisms come together.

2.6. Adoption in Knowledge-Intensive Organisations (KIOs)

As developed economies have shifted from industrial manufacturing towards knowledge- and service-based activity, KIOs have gained greater economic and managerial

importance (Lönnqvist & Laihonen, 2017). Accordingly, their structural properties provide a relevant context for examining how these mechanisms develop and become coordinated.

Starbuck (1992) defines an organisation as knowledge-intensive when knowledge contributes more to its work than either capital or labour. Its core activities therefore rely heavily on the intellectual and analytical skills of educated and experienced professionals (Alvesson, 2004; Starbuck, 1992). These activities commonly involve creating, applying, and disseminating knowledge or developing novel solutions to complex problems (Lönnqvist & Laihonen, 2017). Universities, consultancies, law firms, and R&D-intensive organisations are common examples. Nevertheless, knowledge intensity has no fixed boundary and may vary across organisations and among units within them (Alvesson, 2004).

Four properties of KIOs are particularly relevant to enterprise generative AI adoption. First, they often combine professional autonomy with adaptable organisational forms. Hierarchy is relatively subdued, and substantial discretion rests with the individuals and teams closest to the work (Alvesson, 2004). Second, their tasks and outcomes are frequently ambiguous and intangible. Because quality matters more than output volume, knowledge-work performance can be difficult to define and measure (Drucker, 1999; Lönnqvist & Laihonen, 2017). Third, professional expertise includes tacit and context-dependent judgement that may be difficult to codify or transfer independently of practice and interaction (Alvesson, 2004; Lönnqvist & Laihonen, 2017). Fourth, knowledge workers are often expected to plan and manage their own work. Coordination consequently relies heavily on communication and mutual adjustment, while formal management focuses on clarifying objectives and supporting the conditions under which professionals operate (Alvesson, 2004; Lönnqvist & Laihonen, 2017).

These properties do not introduce additional mechanisms. Rather, they shape how the mechanisms identified in Section 2.5 operate. Professional autonomy allows enterprise-AI use to be adapted to local tasks, but ambiguous outcomes make it difficult to determine which practices merit wider adoption. Tacit expertise constrains their articulation and transfer, while decentralised coordination complicates their embedding across roles and units. Workflow redesign must therefore connect locally developed uses to interdependent processes. Skills and absorptive capacity must support the recognition, adaptation, and diffusion of useful practices beyond their originators. Governance must provide consistent boundaries without displacing the professional judgement on which the work depends.

Data readiness is similarly conditioned by the nature of professional knowledge. Where enterprise generative AI applications depend on organisation-specific information, relevant tacit knowledge may first need to be articulated in an accessible form – a process that Nonaka and Takeuchi (1995) describe as externalisation. However, such articulation remains selective. It may be costly, contested, or incomplete and can separate knowledge from the context that gives it meaning (Alvesson, 2004; Lönnqvist & Laihonen, 2017).

KIOs therefore present a tension between local adaptation and organisational integration. Autonomy and context-specific expertise allow professionals to tailor AI use to their own work, but may also tie effective practices to particular users and tasks. The central question is whether such practices can move beyond their points of origin to become shared, reproducible, and sustained – the organisational outcome the next section defines.

2.7. Defining AI-Enabled Organisational Capability

Mikalef and Gupta (2021) define AI capability as a firm's ability to select, orchestrate, and leverage its AI-specific resources. This places capability in the organisation's ability to mobilise resources rather than in any resource alone. Applications, data, individual skills, and isolated practices may contribute to capability formation, but none demonstrates it by itself. Their resource-based definition identifies capability at the organisational level but does not specify their early formation process. This dissertation therefore adopts a narrower, study-specific definition: AI-enabled organisational capability, understood as the organisation-level ability to embed AI-assisted practices in organisational work so that they may be shared, reproduced, and sustained beyond the users and tasks in which they originated. The three mechanisms identified in Section 2.5 explain how this ability may develop. Data readiness conditions information-dependent forms of embedding, while KIO characteristics shape how the mechanisms develop and become coordinated.

To examine whether this ability is beginning to form, the definition is translated into four study-specific indicators. Evidence of emerging capability is sought where AI-assisted practices are (i) embedded in recurring processes rather than confined to individual experimentation; (ii) subject to shared verification arrangements; (iii) distributed beyond isolated users or units; and (iv) retained in routines or supporting arrangements that do not depend entirely on their originators. These indicators may develop unevenly; emerging capability therefore denotes partial formation of the defined capability.

AI-enabled organisational capability is a proximate organisational outcome, not evidence of realised organisational value or advantage. It may enable later value creation under conditions not observed here, but early adoption cannot establish whether such value is realised. The remaining gap concerns how the three mechanisms begin to develop and become connected around enterprise generative AI use, and whether their development indicates that this capability is beginning to form during early adoption in KIOs.

2.8. Research Gap and Research Questions

The reviewed literature supports a central argument. Enterprise generative AI can improve performance on bounded knowledge-work tasks, but such gains do not, by themselves, establish organisational value. An organisation may provide the technology and employees may use it widely, without it becoming assimilated.

Movement towards assimilation may depend on complementary developments in workflow redesign, skills and absorptive capacity, and governance. Data readiness enables or constrains forms of embedding that depend on organisational information, while knowledge-intensive characteristics shape how these mechanisms develop and become coordinated. When the resulting arrangements allow for embedding AI-assisted practices in organisational work so that they may be shared, reproduced, and sustained beyond the users and tasks in which they originated, an AI-enabled organisational capability may begin to form. This capability remains distinct from realised organisational value, although it may support its later creation.

While the available literature allows us to identify these mechanisms and distinguishes localised use from organisational capability, it offers limited explanation of how they begin to develop around task-level enterprise generative AI use, become coordinated, and interact with data readiness. The remaining gap is therefore processual: how individual use begins to acquire an organisational form, move towards assimilation, and contribute to capability formation.

The study examines this process within two connected boundaries. The knowledge-intensive context defines the organisational setting in which professional autonomy, task ambiguity, tacit expertise, and decentralised coordination may shape their development. Early adoption defines the temporal setting in which organisational arrangements are beginning to form.

The study investigates this process at FEP – Faculty of Economics and Management of the University of Porto – during the first months of its Microsoft 365 Copilot deployment. As a university faculty, FEP provides a knowledge-intensive setting that combines academic and administrative work. The case therefore allows for examining early assimilation and capability formation across different forms of work within one organisational setting.

Specialists working with enterprise generative AI outside FEP provide a limited external reference point. Their retrospective accounts help interpret the conditions associated with use that remains localised and use that becomes more widely shared and embedded. These accounts provide neither evidence about FEP nor a general path through which assimilation or capability develops.

Against this background, the study asks:

Primary research question: How might a knowledge-intensive organisation begin to develop AI-enabled organisational capability from individual, task-level use of enterprise generative AI, and what conditions appear to enable or constrain this process during early adoption?

Sub-question 1 – Mechanisms and data readiness: In what forms, if any, are workflow redesign, skills and absorptive capacity, and governance evident in FEP's early adoption, and how does data readiness appear to enable or constrain their development?

Sub-question 2 – Knowledge-intensive conditions: How do FEP's knowledge-intensive features – professional autonomy, task ambiguity, tacit expertise, and decentralised coordination – shape the development and coordination of these mechanisms?

Sub-question 3 – External comparison: How do specialists' retrospective accounts characterise the conditions under which enterprise generative AI use remains localised or becomes more widely shared and embedded, and where does FEP's early-stage pattern align with or diverge from those accounts?

These questions concern the early formation of an organisational ability. They do not address realised organisational value or appropriable advantage, neither of which can be inferred from the early-stage developments examined here. The following chapter presents the case design, evidence base, and analytical approach used to investigate this process.

3. Methodology

3.1. Research design

The research design had to capture changes in practice, distinguish topics examined without relevant evidence from topics not examined, and reconstruct the professional reasoning behind accepted, rejected, delegated, or limited uses. These requirements are addressed through a retrospective, single instrumental case study with an embedded design. FEP constitutes the bounded case, enabling early capability formation to be examined in depth (Yazan, 2015). The study is instrumental because the inquiry is guided by what FEP can reveal about a process that may also occur elsewhere, rather than by the case alone (Stake, 1995; Yazan, 2015). The unit of analysis is the organisational process through which AI-assisted work becomes, or fails to become, embedded beyond its originating users and tasks. Academic and administrative work serve as embedded contexts, allowing this process to be examined and compared across two settings, without treating them as separate cases (Yazan, 2015; Yin, 2018). Context-specific observations inform the comparison, but the resulting judgements concern FEP as a single case. The study reconstructs the process retrospectively from participants' accounts, institutional records, and usage traces collected after the relevant events.

The evidence is analysed through a primarily deductive application of the framework method (Gale et al., 2013), whose matrix structure organises summarised qualitative evidence by case and category. In this study, participant-by-construct displays connect each analytical judgement to the relevant passages in the evidence register. The framework's four top-level categories – the complementarity mechanisms, the precondition, the knowledge-intensive conditions, and the capability indicators – were derived from Chapter 2 and established before fieldwork. The method is therefore suited to examine whether these literature-derived concepts are evident in the case, the forms they may take, and the gaps that may remain.

The conceptual structure remained fixed, but its operationalisation was refined through engagement with the evidence. No category was added, removed, or renamed, and a workbook formula verifies that only the four coding categories – the three analytical domains and data readiness – remain in use. When passages exposed ambiguity at category boundaries, inclusion and exclusion rules were clarified. Limited inductive interpretation was permitted when evidence coded within the framework suggested a relationship or explanation that the framework had not anticipated. This process generated one proposition, which is explicitly

identified as inductive. The analysis was therefore deductive in its conceptual structure but iterative in its operationalisation. Section 3.5 documents the resulting refinements.

Three procedures addressed distinct risks: cross-source comparison prevented non-equivalent sources from being treated as interchangeable; negative-case analysis searched the within-case evidence for material that challenged the emerging propositions; and rival-pattern comparison assessed the framework-derived account against a plausible alternative. Section 3.7 explains how the latter two procedures were conducted and the limits of the conclusions they support.

The study is predominantly qualitative. Quantitative sources, namely the institution's Microsoft tenant usage export and a student questionnaire, were used descriptively. For each reported percentage or distribution, the denominator, treatment of missing responses, applicable time window, and data-quality checks are stated in Annex A. These figures provide contextual patterning and do not generate independent construct judgements. Neither is analysed through an autonomous procedure, used to generate standalone findings, or allowed to determine a qualitative judgement. The study is therefore not classified as mixed methods, which requires distinct qualitative and quantitative components to be systematically integrated (Fetters et al., 2013). Sections 3.2 and 3.5 specify the scope and evidentiary limits of each quantitative source.

3.2. The case, the participants and the evidence

FEP was selected because it met the two boundaries established in Chapter 2: a knowledge-intensive setting in Starbuck's sense and an early-adoption context defined by the pilot conducted during the first half of 2026. Professors and administrative staff participating in the pilot were invited without selection by discipline, service, or intensity of use. The guides accommodated regular, occasional, and non-users, including users of other tools. Each account is treated as evidence about the participant's own work rather than as representative of a discipline or service. Any extension beyond FEP therefore rests on analytical generalisation rather than statistical representativeness.

Professors were included because academic work strongly reflects knowledge-intensive conditions. Teaching, supervision, and research involve substantial professional discretion, are difficult to define in advance, and rely heavily on tacit expertise. Because every approached professor took part, the group's composition reflects who was approached and

accepted. All seven participants used Copilot to some extent, with the resulting limitation addressed in Section 3.10. The group comprised three professors from quantitative disciplines, two from applied and management disciplines, and two from normative disciplines, enabling comparison across different forms of academic work. Administrative staff offered a contrasting perspective of more standardised and coordinated functions common to many organisations, including information systems, marketing and communication, back-office and academic services, internationalisation, staff development, and executive support. Invitations were spread across services, and because the format required a single session, the group that met was formed by those who accepted and were available at the same time. Eight current Copilot users from six services took part, representing markedly different levels and forms of use. To support open discussion, the group did not include line managers alongside their direct reports and preserved room for disagreement. Sections 3.3 and 3.6 explain how the group format limits the claims that can be drawn from this evidence.

Nine specialists working with enterprise generative AI outside FEP, recruited individually through non-random selection, provided an external basis for comparison. Participants had to be professionally involved with enterprise generative AI in an organisational setting and able to describe a specific adoption trajectory. Although recruitment was non-random, the resulting group was heterogeneous, spanning technology providers, integrators, consultants, an SME partner, a regulated-product developer, a retail-group adopter and a software engineering leader. Their experience was relevant because each had encountered, at a more advanced stage, conditions that this study examines as they begin to emerge. However, only two participants described adoption from within their own organisations. The others spoke as providers, consultants, or product owners. Each participant's standpoint was recorded so that the nine accounts were not treated as equivalent observations of comparable situations. While their retrospective accounts concerned other organisations, providing no evidence about FEP, they supported Sub-question 3 and informed the interpretation of the primary research question.

The student questionnaire received 1,503 responses from across three faculties of the University of Porto. An opening question asked whether respondents currently used any AI tool. The 1,430 who answered yes completed the substantive questions, while the remaining 73 did not. Percentages reported from those questions are therefore based on 1,430

respondents (Annex A2). Three limitations follow. First, because the questionnaire did not record respondents' faculty, its findings cannot be attributed specifically to FEP students. Second, only current AI users answered the substantive questions, so the results describe their reported experiences but cannot explain why others do not use these tools. Third, the questionnaire covers a broader population under different conditions from those examined in the FEP case. It therefore provides university-level context rather than triangulation: its patterns may align with the case evidence, but they cannot corroborate it.

The usage export provided a narrow, time-bound view of Copilot activity within the institution's tenant. It showed that an interaction had occurred through prompt count, but not its purpose, outcome, or whether it reflected experimentation, an abandoned attempt, or completed work. Its coverage window ended on June 12, 2026 (Annex A1), three weeks before the first research session. As participants could describe practices developed outside this period, the data could not verify or challenge their accounts, and a lack of recorded activity did not establish non-use at the time of interview.

3.3. Data collection

Between July 3 and August 5, 2026, 17 sessions were conducted in Portuguese: seven professor interviews, one administrative focus group, and nine specialist interviews (Annex C4). All sessions were voluntary and followed thematic guides derived from the analytical framework, covering the same core themes while allowing their order and depth to vary. Prepared questions provided consistency, while optional probes pursued raised relevant points.

The administrative focus group was held on July 3 (Annex C2). A group format was chosen because participants worked within the same institutional system and could extend, qualify, or challenge one another's accounts. The moderator explicitly encouraged disagreement, periodically summarised contributions, and distinguished questions addressed to the group from individual responses. The transcript records agreement, dissent, correction, and elaboration. These interactional patterns inform the analysis, although no formal method of interaction analysis is claimed. Professors were interviewed individually between July 10 and 18 after scheduling constraints made a group session impossible. The seven semi-structured interviews varied in length and depth, but all followed themes fixed by the analytical framework before fieldwork. Questions answered during the conversation

were not repeated, and each session's coverage was recorded question by question so that omissions remained visible.

Eight specialists were interviewed between July 8 and 15, followed by a ninth on August 5 (Annex C3). The later interview was arranged when an opportunity arose to speak with someone working in a technology company where enterprise generative AI had become significantly integrated into work practices. Data collection developed iteratively across the interview sequence. Later interviews gave greater attention to themes that earlier accounts had covered less fully, so the ninth could focus on the areas where the evidence remained thinnest. Its purpose was to strengthen coverage rather than collect further examples of established points, and no conclusions were drawn from it about questions it did not address.

For both professors and specialists, recruitment ended when the evidence base covered the analytical categories relevant to each participant group, contained sufficient variation and negative cases to challenge the emerging account, and supported examination of the propositions used in the analysis. This establishes adequacy for the study's analytical purpose, not saturation or the absence of further possible insights. Adequacy was therefore assessed against the analytical requirements rather than a numerical target or a claim of saturation (Braun & Clarke, 2021). By contrast, the administrative context was examined through one focus group, so its adequacy depended on how fully that session covered the analysis' issues. The coverage grid described in Section 3.5 supports both assessments by showing what each session addressed and where the evidence remains limited. All sessions were recorded with consent and transcribed. Transcripts were cleaned of transcription noise – false starts, repetitions, unclear fragments – without altering meaning or speaker voice. Each carries a record of posed, unposed, rephrased, reordered or combined questions. Because the guides were used flexibly, an absence in an interview can have two meanings. If a topic was raised but the participant reported nothing relevant, the analysis concludes only that no relevant evidence emerged under the questioning undertaken. This does not establish that the arrangement or practice was absent. The participant may have been unaware of it, unable to recall it, unwilling to disclose it, had no relevant experience, or unable to discuss it within the available time. If the topic was not raised, it represents a limit on what the study can conclude.

The questionnaire was designed for this study and administered, with the supervisor's assistance, to University of Porto students from three faculties between June 20 and July 15, 2026. Before analysis, responses were checked for completion time, straight-lining across

item blocks, and harmonisation of freely typed tool names. Excluding flagged responses changed no reported figure by more than 0.8 percentage points, so all responses were retained. The limited documentary record available from FEP comprised four institutional governance documents, the presentation used in Copilot training, a national policy recommendation included for context, and the final report of Inov-Norte, an institutional evaluation of generative AI use at FEP conducted alongside this study. The faculty's information service supplied the usage export described in Section 3.2. Material not produced for this study situates the case without carrying the evidential weight of material generated to answer its research questions.

3.4. The analytical framework

The coding framework follows the structure established in Chapter 2 and comprises four components with distinct analytical roles. The first component consists of three analytical domains: workflow redesign, skills and absorptive capacity, and governance. Chapter 2 presents these as proposed complementarity mechanisms. Empirically, however, coding identifies what participants report within each domain without treating them as directly observed causal mechanisms. The second component is data readiness, which Chapter 2 treats as a cross-cutting precondition for information-dependent embedding rather than as a fourth mechanism. It is not a precondition for use in general: its informational conditions may enable or constrain particular forms of embedding, but limited data readiness does not prevent embedding altogether. It is therefore examined in relation to the three analytical domains. The third component comprises the four knowledge-intensive conditions: professional autonomy, task ambiguity, tacit expertise, and decentralised coordination. The analysis examines how these conditions may shape the development and coordination of the analytical domains. Finally, the fourth component consists of the four capability indicators. Early signs of capability formation are assessed by examining whether practices become part of recurring processes, undergo shared verification, spread beyond isolated users, and persist independently of their creators.

As argued in Section 2.5, the three analytical domains may complement and reinforce one another, but they need not develop in parallel. The analysis therefore considers each domain independently and examines relationships only where the evidence supports a connection. Development in one domain without corresponding change in the others is treated as a finding (Annex E2). Clear inclusion and exclusion rules keep the categories

distinct. Faster task completion, for example, is not workflow redesign unless the task steps also change; learning how to use Copilot for a task is coded under skills and absorptive capacity. By contrast, being unable to complete that task because Copilot lacks a required function is not a data-readiness issue. Data readiness concerns whether the organisation has the information needed to support the task, not what the software itself can do.

Each evidence record includes a claim-boundary field stating what cannot be inferred from it. Introduced during coding as an evidentiary safeguard, it is a note attached to a record and is never counted as a construct. Its purpose is to enforce the boundary established in Chapter 2: task-level effects do not demonstrate organisational value and that an early-adoption study can identify only initial indications of capability formation. A passage may concern an analytical domain without establishing the outcome it describes. Participants may report benefits or costs, such as time savings being absorbed by increased output, while specialists may explain that value would require measurement before and after adoption. Coding identifies what these statements concern, not whether organisational value was realised. Records referring to sustained use, mature capability, realised value, or financial returns therefore carry an explicit note limiting the conclusions they can support. That note restricts the use made of the record. While the passage can inform a judgement about the domain it concerns and about how far the practice it describes has developed, it is never used to establish that the persistence, value, or return it mentions actually occurred.

3.5. Building the evidence base

Evidence identification began during fieldwork, as potentially relevant passages were flagged with a provisional note on the construct they might inform. These flags oriented the researcher to the material and shaped later interviews but did not determine the final analysis. Formal selection occurred afterward, reviewing every transcript and document systematically against the predefined framework. Some flags were confirmed, others reassigned, and further passages added where consistent application revealed connections missed initially. A passage qualified when it spoke to a construct directly enough to be coded under a stated definition, and material outside the framework was excluded rather than accommodated by broadening one. Where a passage tested the very edge of a definition, the definition was sharpened and the decision recorded; no refinement added, replaced, or renamed a category.

Extraction rules differed by evidence stream and governed inclusion in the register rather than participant selection or analytical weight. Targets and ceilings were introduced to keep

the register analytically manageable while preserving construct coverage and source diversity. Exceptions were permitted where additional passages altered a construct judgement, supplied contrary evidence, or documented a claim boundary. The full decision rules and affected records are reported in the workbook. Professor interviews used an initial target of ten passages, extended where further construct-relevant material was present. The focus group was treated as one source and subject to an overall ceiling; three weaker or duplicative passages were set aside. External accounts were allocated differently. Each began with a baseline of five passages, which increased when the account was longer and more detailed, described adoption from within the specialist's own organisation, or provided the only perspective on a particular point in the adoption trajectory. The number of passages therefore varies deliberately across the nine specialists, from five to twelve, since an equal share would have underrepresented the most informative cases. The allocation was a practical rule for deciding how much of each transcript to register, and each figure was the point at which extraction stopped. These allocations determined only how much material entered the evidence register, not the analytical weight assigned to each specialist. Judgements did not depend on the number of passages contributed, and passage counts were never interpreted as measures of development. Documentary extraction followed a more open-ended approach. While it began with a smaller target, this was not treated as a ceiling, and further passages were registered whenever verification or negative-case searches identified relevant evidence.

Applying the defined rules to the corpus produced 198 registered passages, of which 193 are active (Annex B2). A record is active when it contributes to findings, and only active records enter a judgement, a display, or a count. The remaining five are retained as an audit trail: one reserved at the participant's request, one excluded as outside the framework, and the three dropped at the focus-group ceiling, each cross-referencing the passage retained in its place. While these records carry no analytical weight, keeping them makes the composition of the evidence base fully inspectable. Every registered passage carries the text in its original language, a source locator, an English translation, the relevant construct, its contribution, and the rationale for its coding. No passage was registered without a locator, providing the foundation of the study's traceability. Each empirical claim in the chapters that follow can be traced to one or more registered passages or documentary records. A separate coverage grid, built from the question mapping in every transcript, records what was asked rather than

how fully it was answered, so a topic marked as reached does not guarantee that the participant addressed it substantively.

Because a dissertation can accommodate only a limited amount of direct quotation, thirty passages were selected before drafting (Annex B1). This decision affected presentation only; it did not change which evidence informed the analysis. This shortlist was selected under the following rules: (i) a passage that reports an enacted practice is preferred to one that reports a perception of the same point; (ii) a point already carried by a stronger passage in the same stream is cited once and quoted once; (iii) evidence that qualifies or contradicts a claim is quoted in the informant's own words wherever such a passage exists; (iv) a claim about FEP is never quoted from an external account; (v) and a passage a reader could not have anticipated is preferred to one that merely confirms it. A further eighteen were shortlisted without being selected, each recorded with the reason why.

Quantitative evidence either contextualises the scale and distribution of use or compares a pattern in another population with the within-case interpretation. For each of the 600 licensed accounts, the usage export records the total number of prompts submitted across applications, the number of active days, and the date of most recent activity. The questionnaire provides a student-level view of broader generative AI use, covering perceived effects, task types, changes in how work is organised, verification practices, awareness of institutional rules, and the conditions respondents associate with organisational value. Neither set of fields is reproduced in full here.

3.6. From evidence to judgements

Context-level judgements were reasoned syntheses based on the evidential requirements, contrary evidence, and coverage gaps. A registered passage records what one person said about one thing. It does not, by itself, establish how far an analytical domain has developed. That assessment is a judgement, formed by bringing together everything registered for one participant on one construct: the passages that show development, those that limit it, those that contradict it, and the areas the guide did not reach. Once all participants have been assessed, the same process is repeated across the seven professors and across the administrative focus group. This allows the academic and administrative settings distinguished by the embedded design to be characterised as wholes and then compared. While participant-level cells exist for both groups, their evidential basis differs. A professor's cell rests on an individual interview whose question mapping is known. An administrative

participant's cell rests on what that person contributed within a single group discussion, where the guide was put to the room rather than to each participant in turn. Such a cell is therefore a participant-level indication within the focus-group context. Where a construct was not reached with that participant, the cell records that the group format did not bring the construct to that person, and, in three instances noted in Section 3.5, that the participant's only passage on the construct was dropped at the extraction ceiling. Contributions that the transcript records as collective, together with the group's ranking exercise, are registered as utterances of the group rather than attributed to individuals. For the administrative context, the operative judgement is accordingly the group-level one, which also carries the convergences, divergences, and qualifications noted in Section 3.3.

Judgements used six evidence-based categories: not examined; examined with no evidence found; nascent or experimental; localised but recurring; shared across users or units; and institutionalised (Annex D1). Each judgement records supporting and contrary evidence and the distinction on which it rests. The categories are ordered by evidential requirements but do not constitute a numerical maturity scale: recurrence, social reach, and independence from an originator may develop unevenly. They should therefore be read as qualitative judgements rather than measurements that can be averaged across constructs. The same body of evidence is then assessed against the four capability indicators. Each indicator is judged separately, allowing capability formation to appear as partial or uneven. The knowledge-intensive conditions are examined next: first by asking how each condition appears to shape each analytical domain, then by examining the relations among the domains directly – whether the development of one created the conditions for another, and whether any occasion exists at which more than one was addressed together. Two domains appearing separately in the same interview are not treated as coordinated. Coordination is only recorded where the evidence connects them.

The analysis established judgements about FEP solely from within-case evidence. External accounts were subsequently organised according to the analytical function they served, including ownership, diffusion, governance, maintenance, and downstream integration. These functions are interpretive descriptors rather than stages, do not form a validated sequence, and are not used to rank organisations or infer a trajectory for FEP. Because the accounts concern heterogeneous organisations that the researcher did not observe, their relevance depends on contextual comparability.

3.7. Testing the account

An analysis can organise supporting evidence into a coherent account and still reach an incorrect conclusion. Two procedures were therefore used to test it. Both procedures followed the cross-source comparison rule established in Section 3.1. Evidence from different streams was treated as convergent only when independent sources addressed the same claim about the same object. Each source was assessed within its evidential limits and according to whether it supported, qualified, contradicted, or merely contextualised the claim under examination. The first procedure was disconfirmation, which tested the limits of the emerging interpretations. Twelve case-based propositions were developed after fieldwork and examined through predefined rejection conditions. Because the same corpus informed both their development and assessment, the procedure provided no independent confirmation; its purpose was to challenge and delimit the emerging account. The corpus was searched for evidence that would reject, narrow, reformulate, or leave each proposition unresolved. Three propositions stood as formulated, eight were narrowed, and one remained inconclusive (Annex G1). All contrary evidence and revisions were retained in the workbook. Of the twelve, three stood as formulated, each carrying a stated limit on how far it reaches; eight were narrowed, in several cases substantially; and one was reported as inconclusive. Because the propositions and the corpus are not independent, this procedure had to guard against claims surviving through repeated qualification until they became impossible to contradict. To safeguard against this risk, every narrowing was recorded alongside the evidence that required it, allowing the reader to compare the original and revised claims; when a shortfall could be classified only by relying on the judgement whose adequacy was under examination, the proposition was not narrowed to accommodate it and reported as inconclusive. This rule accounts for the one inconclusive result. All evidence that ran against a proposition was retained as a negative case as it helped define a finding's limits.

The second procedure is rival-pattern comparison. The framework developed in Sections 2.4-2.7 offered one explanation for limited organisational embedding: in a knowledge-intensive organisation at an early stage of adoption, this outcome could reflect complementary organisational investments not yet being made, with knowledge-intensive conditions shaping how the pattern emerged. A simpler rival attributed the same outcome entirely to resource scarcity and uncertainty over the pilot's duration – constraints that could affect any small organisation, regardless of knowledge intensity. Because both readings

accounted for much of the same evidence, the comparison was confined to seven points at which they imply different observations. The framework-derived reading was preferred only where it accounted for evidence the rival could not; where both fitted equally well, there was no basis for preferring it. Six of the seven comparisons separated the two readings, and all six favoured the framework-derived one; the seventh could not separate them. That result does not establish that the framework-derived reading is correct and the rival false, and the two are not mutually exclusive: resource scarcity and an uncertain licence horizon remain consistent with the severity and the timing of what was observed. Section 5.3 sets out what each reading accounts for and what neither can attribute.

Three qualifications shape how this comparison is reported. First, because both interpretations and the seven points of divergence were formulated during the analysis, the procedure assesses their relative fit rather than testing independently specified alternatives. Second, one of the seven comparisons cannot distinguish between the two accounts (Annex G4). The instrument on which it relies offered participants four options but omitted the condition that the rival account would identify. The result is therefore reported only as consistent with the framework-derived reading, not as evidence that discriminates between them. Third, two of the remaining six comparisons rely principally on external accounts. Because these accounts concern other organisations, they establish nothing about FEP. Overall, the comparison indicates which interpretation fits the case more closely; it is not a confirmatory test in the experimental sense. The study observes coexistences and associated patterns, but it does not identify the causal effects of KIOs' characteristics.

3.8. Credibility: auditability and consistency

Credibility in Lincoln and Guba's (1985) sense turns on whether a reader who did not conduct the analysis can examine how it was performed, and here it rests on an auditable analytical record and on consistency controls. The reasoning is held in a single workbook, whose principal components are the codebook, evidence register, coverage grid, judgements, capability assessment, conditions and their coordination, external comparison, proposition tests, disconfirmation record, and rival-pattern comparison. Each states what it contains, how it should be read, and what conclusions it cannot support. Because every judgement is recorded with the evidence for and against it and the distinction on which it turns, a reader can check what was concluded and why. This audit trail does not validate the reasoning itself, and another analyst might reach different interpretations. Traceability runs in both directions,

from each judgement to the passages it rests on and from each passage to every display that uses it, so passages registered but never used remain visible. Identifiers carry this into the dissertation. Every registered passage has an identifier such as P1.1, where the letter marks the evidence stream, the first number the source, and the second the passage. P covers professors, S administrative participants, E specialists, D documents, SG the focus group's collective statements, and SC its ranking exercise. The workbook maps each identifier to its internal record code and to the passage's exact location in its source. All evidence was coded and analysed in Portuguese, and English translations, prepared for traceability and quotation, were each checked against the original.

Automated checks tested identifier integrity, source traceability, permitted coding values, separation of active and set-aside records, resolution of display references, completeness of proposition tests, and quotation fidelity. These checks establish internal consistency and compliance with the stated inclusion rules, not the substantive validity of coding decisions or analytical judgements. The full list of checks is reported in Annex B1. Contrary and unresolved evidence remains visible throughout. The disconfirmation record retains the evidence found against each proposition, and the rival-pattern comparison preserves the one test that did not separate the competing explanations. Limitations were recorded as they arose, including ambiguous attribution, incomplete coverage, and reliance on a single source. Each is carried as a qualification on the claim it affects and is stated wherever that claim is reported; Section 3.10 draws them together.

3.9. Ethics and positionality

Participation was voluntary and informed, and all recordings were made with consent. One participant requested that part of his remarks not be attributed. The passage remains in the register to document that the request was honoured; it is neither quoted nor does it contribute to any judgement. When a quoted passage names someone other than the speaker, the name is replaced with the person's role in square brackets in both the original language and the English translation. Both the researcher and dissertation supervisor are connected to FEP. As a master's student at the faculty, the researcher was able to identify and approach relevant professors and administrative staff, which may also have encouraged participation. However, some participants might have been less open when discussing their work with a student from the same institution. The supervisor served on FEP's Pedagogical Innovation Committee, led the Copilot pilot, and delivered training to professors and administrative

staff, so he produced some of the institutional documents later analysed in the dissertation, although those documents were treated within the evidential limits stated in Chapter 3.

3.10. Limitations

The study's limitations concern the design's reach, the evidence's strength, and the conclusions that can be drawn at this early stage. All interviewed professors had used Copilot to some extent. The study can therefore examine how academic use develops, but not why some academics choose not to use it. Evidence on the administrative context comes from a single focus group with eight participants and consequently supports group-level indications rather than claims equal to those from eight individual interviews. Evidence about technical provisioning is also less direct because no separate interview was conducted with FEP's information service. Instead, it draws on one service member's contribution to the focus group, institutional documents, and participants' accounts of the information the tools could access. Informal conversations with the service and with the supervisor who led the pilot helped clarify its organisation, but were not recorded, transcribed, or coded and therefore support no analytical claim.

Further limitations apply to particular analytical questions. The focus group and professors were not asked whether a practice would survive its originator's departure. Retention is therefore carefully recorded as not established, rather than interpreted as absent. The tacit-knowledge probe was likewise not put to the focus group (Annex C2), so the relevant proposition draws on only two evidence streams, a qualification stated wherever the result is reported. The usage export covers an earlier period and only one institutional tool, subject to the limits detailed in Section 3.2. Claims based on a single source or a contested attribution are flagged where they appear. More fundamentally, the timing of the study limits what can be concluded. By examining only the first months of adoption, it cannot establish that AI caused the observed changes, increased productivity, matured into an organisational capability, or created organisational value. Such conclusions would require longer-term observation, comparative evidence, or measures not included in this study. The analysis therefore makes a narrower inquiry: whether an organisational capability may be beginning to form at FEP and which conditions enable or constrain that process. It asks whether embedding remains localised, whether verification arrangements exist, whether diffusion is episodic, and whether practices persist independently of those who created them.

4. Findings

4.1. The setting and task-level baseline

This chapter presents the findings on FEP's early adoption of Microsoft 365 Copilot. Sections 4.2 to 4.5 and 4.7 each close with a brief external comparison, placed after the judgement it follows and never used as evidence about FEP. Section 4.6 carries none because the knowledge-intensive conditions were coded only against within-case evidence. Each claim is linked to a record in the analysis workbook, and a judgement of "not examined" identifies a limit in the research instrument without implying that the construct was absent.

Two descriptive observations provide context within the boundaries established in Chapter 3. The first is that uptake was broad but generally light and highly concentrated. Out of 600 licensed accounts, 349 submitted at least one prompt, representing 58.2% (Annex A1). The median active account submitted five prompts across two active days, while the most active 10% generated 90.3% of the 28,045 prompts recorded. The 65 accounts whose identifiers indicate staff or faculty generated 54.7% of all prompts. By the end of the coverage window (12 June 2026), 22.3% of previously active accounts had been inactive for more than six weeks. Because the export identifies neither the purpose nor the outcome of prompts and cannot show whether work processes changed, it provides context for the qualitative findings without establishing effective use or capability formation. The second observation concerns individual task-level benefits that largely fail to reappear as freed time. An applied/management-band professor captures this pattern: "To be honest, I think that in lesson preparation it does not save me time either, because I end up doing more things than I did before. In reality it improves the type, the quantity and the quality of the materials I create for classes" [P1.1]. Across the professor interviews, gains were generally reinvested in greater or better output or absorbed by verification [P1.2, P7.3]. One sustained user also reported continuing the same work without structural change, as discussed in Section 4.2. Two exceptions are retained. The same professor reported time released through translation [P1.11], and an administrative participant saved time on a bounded summarisation task [S3.2]. A further professor reported a net positive time impact without describing any change to the task steps [P2.12]. Overall, released time was reported mainly where verification was inexpensive; elsewhere, gains supported greater or better output or were absorbed by review.

The parallel student cohort of 1,430 students who use AI reported a similar task-level pattern (Annex A2). Although 92.9% agreed that tasks became faster, 68.3% also spent more

time reviewing and correcting outputs. Their reported changes were mainly individual, as 49.1% had removed a manual step and 38.9% had reorganised their own tasks, but only 16.2% had changed how work was divided with others. A majority, 60.3%, nonetheless expected individual speed gains to raise the productivity of the whole organisation. These figures provide context from a different population and no evidence about organisational work or outcomes at FEP.

4.2. Workflow redesign: recurring individual redesign that does not travel laterally

Recurring AI-assisted redesign is evident at FEP, but no registered practice was adopted laterally by a colleague or service. Changed sequences recur in the work of four of the seven professors and are judged "localised but recurring" or higher. Two administrative participants also use configured agents within their regular routines. The two strongest academic cases extend beyond the builder, reaching students. First, a quantitative-band professor reorganised supervision around systematic recording: "In fact, I did not record meetings. I think I did one or two last year, and this year every meeting is recorded, systematically" [P3.1]. The recordings are transcribed and summarised, and students receive the summaries [P3.2-4]. Second, an applied/management-band professor uses a configured agent continuously: "I use a Socratic agent to support students. In some curricular units it answers basic questions continuously and helps resolve doubts, like the old doubt classes, but with knowledge controlled from the materials I gave it" [P7.1]. Both cases meet the requirement for "shared across users or units" because they reach a second party. Nonetheless, their reach remains student-facing and follows a relationship in which the professor holds authority. Each artefact is built and maintained by one person, with no assigned owner, maintenance arrangement, or handover to a colleague.

Academic redesign also varies with the character of the work. Recurring cases are concentrated in quantitative and applied/management bands. A further quantitative-band professor maintains a revised article-review sequence and a retrieval-constrained benchmarking system for his own use [P2.1-3]. In the normative band, one professor uses the technology for consultation and validation but considers his practice impossible to reproduce at scale. His account was therefore judged "examined with no evidence found" for redesign [P6.5]. Another professor attempted to configure an agent but abandoned it after it failed [P5.3], while a third planned an agent for one of her curricular units and went no further than outlining it [P1.3]. Standards also remained implicit where student-facing

redesign fell short. One professor considered her tutor agent to be insufficiently configured [P3.12], while another developed an artefact that his students later abandoned [P2.15].

Administrative redesign follows the same pattern of individual ownership without the student channel. One executive-support participant has embedded configured agents in the institutional process for authorising expenses: "I use two or three agents for expense-authorisation requests. […] You have to create a title and justify the expense in administrative language. I created a prompt for that and it formats the text" [S3.1]. These expense-authorisation agents sit inside an institutional process without altering how that process is performed. Another participant reported the clearest change to daily work: "I have an agent that runs several queries and organises my data into tables. Whenever anyone asks me for information, it goes through that agent, and I did not do that before. So it has drastically changed the way I work day to day. Today, everything I am asked for goes through AI" [S7.1]. His account of this query agent names AI generally and does not identify the application it runs on, and his one other reported attempt at information-dependent use, connecting a folder of his own material, failed and was abandoned [S7.2]. A third participant organises his exchanges by theme so that context does not have to be re-established at each stage [S5.2].

Three records further define the limits of the findings. A normative-band professor who uses the assistant regularly reported that "there was no structural change" [P5.7]. The focus group's most intensive user similarly reported no change to his processes or data organisation, demonstrating that intensive use can occur without redesign [S2.4]. Workflow redesign was therefore real and recurring, but its diffusion beyond the originating professional was not observed.

The external accounts describe comparable artefacts embedded in organisational processes. One specialist reports a configured sales agent that holds solution positioning, target audiences, and product features, while another describes end-to-end invoice and project monitoring feeding a dashboard used by the rector [E1.1; E5.3]. In both accounts, the artefact is described as embedded in an organisational process. Comparable artefacts at FEP remained individually held.

4.3. Skills and absorptive capacity: competence without conversion

Competence is substantial among some individuals. However, it is also unevenly distributed, with no functioning arrangement for converting it into shared practice. One professor completed more than a dozen courses on prompting, partly through her own initiative before FEP offered training [P1.4]. Others developed their skills through trial and error or conversation with the system [P2.4, P7.9]. Yet this learning remained largely individual: "Something is said, but very superficially. There is no sharing culture as such" [P1.5]. The same professor reported no facilitators or mechanisms for sharing and suggested that a platform containing prompts, agent skills, and practical examples would be more useful [P1.6]. The pool's most prominent generative-AI practitioner identified the same institutional gap: "One failing I feel at FEP is the training of the teaching staff" [P2.5].

Three episodes provide evidence of limited diffusion. Peer-led sessions were organised on one colleague's initiative, but produced no reusable material, designated reference person, or further session [P2.13, P2.14]. A second professor adopted the technology after hearing colleagues describe their experiences, although those accounts conveyed motivation without practical guidance [P3.11]. A more practical exchange occurred during the focus group, when one participant described working with Copilot directly in applications [S2.2]. What drew the room was his account of connecting Copilot to SharePoint, which several colleagues had not explored and asked him to explain. One of them later proposed repeated sessions with practice between them and a collective review of what each person had tried [S3.4]. The interest was therefore real, but the use was not new, since the same in-application use had already been introduced through the pilot training materials [D6.1-3]. An earlier initiative had already created a sharing arrangement of this kind. After a training session, a space was created for the purpose: "We tried. We had a first training session, last year I think, with [the colleague who led the training] and two other colleagues, and we even had an area in Teams that was supposedly for us to share in" [P3.7]. Although she had helped establish the space, the participant could not recall posting any content or seeing colleagues do so. Informal discussion among professors was frequent and centred on the ethics of AI use, leaving practical methods largely undiscussed [P7.11].

The episodes have not, however, accumulated into a functioning recurring arrangement for organisational learning. Each episode moved competence, not a practice. The sessions circulated no reusable material, testimony conveyed motivation without method, and the

demonstration produced only an intention to consult. In none of the three episodes a colleague adopted another person's redesigned sequence or configured artefact. The distinction matters because competence can move between peers while the practice built on it stays with the person who built it. The reach of the finding is set by what the instrument reached: skills and absorptive capacity were examined with all seven professors and with four of the eight administrative participants; one of those four was examined and yielded no evidence, and the group format did not bring the construct to the other four. External accounts describe such arrangements. One specialist uses telemetry to uncover practices that users develop without reporting, allowing those practices to be identified and extended [E3.2]. Another describes an organisation-wide assessment of adoption maturity, with results returned to each participant individually [E4.9]. Each makes locally developed practices visible to the organisation, addressing the point at which the three episodes at FEP stalled.

4.4. Governance: a consistent norm without an arrangement

Wherever governance was examined, rules remained specific to the individuals or services that established them. Across the examined academic and administrative accounts, verification was the most consistently reported norm. When asked who was accountable for the system's output, the focus group answered together: "The person themselves. We always have to validate; we really do have to validate everything" [SG.1]. Professors applied the same principle independently. One audits every output [P7.4], another refuses to expose student scripts [P3.9], a third avoids submitting other authors' work out of ethical and security concerns even though it would save her time [P4.4], and a fourth treats the system as a junior legal assistant and checks its legal claims against their sources [P6.1, P6.2]. These practices are grounded mostly in professional obligation: "That [student] deserves to be assessed properly, and I do not want [him] to be the victim of an artificial intelligence" [P5.5]. However, for the faculty's own work, a search of the professor and administrative transcripts and every available institutional document returned no agreed method, documented procedure, escalation route, or formal allocation of accountability. The institutional evaluation conducted alongside this study records the same verification burden among teaching staff, with every substantive response acknowledging the need for careful review before use [D4.3].

The available institutional guidance covers academic outputs, including assignments, reports, dissertations, and disclosure of AI use in research [D5.1]. Its clearest ethical

requirement is directed at students through a declaration that generative AI was used in a "responsible, ethical, critical, and non-abusive" manner [D2.1]. These principles are, however, broad and provide limited guidance on how they should be applied [P1.7; P7.5]. On the other hand, guidance directed at the faculty's professionals was sent by email and went largely unread. One professor received about eight University documents and did not know what they contained [P1.8], another had left two guidance emails unopened [P5.9], and an administrative participant had not read the director's message on AI in higher education [S6.3]. As most of this material was unavailable to the study, the analysis is limited to its reported scope and reach. The one participant who reports having read the guidance describes it as being "not so much about administrative use or use by staff. It is more about pedagogical purposes and the use of AI in dissertations and similar questions" [S2.3].

One documented operational decision shows how provisioning proceeded without corresponding rules for use. The responsible service chose not to extend Copilot to locally installed applications: "To guarantee that everyone had access to Copilot in the applications installed on their computer, we would have had to reinstall or update Office on every machine, which would have meant a great deal of time and effort" [S2.1]. While the application was available online, the decision shaped how staff accessed the tool without establishing responsibilities, rules, or expectations concerning its use. The only written requirement shown to influence professional conduct came from outside FEP. One participant followed an external accreditation body's requirement to disclose AI use through a disclaimer [S4.1]. This isolated case has no internal equivalent, but it shows that a binding written requirement can guide professional practice. FEP's own requirements govern student submissions and research publications without extending to the faculty's everyday work.

AI use at FEP is mostly governed through individual professional judgement. The practice therefore depends on the individuals who apply it, and the available evidence cannot establish whether it would endure under pressure or in their absence. External accounts from more advanced settings help clarify what this missing organisational layer involves. Speaking from an internal-audit standpoint, one specialist identifies the formal allocation of responsibility as the first requirement for governance [E4.10]. Another specialist links the need for governance to the technology's non-determinism [E4.12], echoing the reasoning several FEP participants used when defining their own boundaries.

4.5. Data readiness: personal curation in place of organisational arrangement

Within the interviewed FEP groups, data-readiness constraints were most visible in information-dependent embedding, emerging when practices depended on access to the University's files and systems. A normative-band professor found that his institutional file store carried no clear time boundary, so that "It went and fetched the work that a student of mine did thirteen years ago and gave me that as the basis for the answer" [P5.1]. Because older and current files were equally accessible, the system could not distinguish between them. He concluded that provisional student work could not remain in OneDrive without clearer scoping.

Other constraints arise at the boundaries of institutional systems and permissions. A quantitative-band professor must download work submitted through Moodle before using it [P3.10]. In the administrative context, one participant abandoned an intended use after failing to connect a folder containing his own information following the training session [S7.2]. Another knows that information remains within the University of Porto but is sceptical of how it may circulate internally, so he restricts what he provides [S5.1]. No arrangement above the individual addresses any of these constraints. Accordingly, information-dependent practice currently rests on personal curation. Professors check sources at page level [P1.9], create topic folders that restrict where the system searches [P2.8], and reorganise scattered materials into cloud projects [P7.7]. These arrangements support their creators but remain unavailable to others. One professor reorganised her materials into cloud projects while retaining numerous scattered folders and creating no master repository [P7.8], while another uploaded lessons and case law directly into a chat session without creating a persistent resource [P6.4]. Such fragmented and temporary character helps explain why the configured artefacts discussed in Section 4.2 remain individual.

Three cases trace the range of knowledge externalisation. At one end, a professor converted photographs of her notebooks into structured manuals organised by knowledge area for future reuse [P3.6]. Since the system can learn from written material [P2.9], this case locates the constraint earlier, when knowledge must be made explicit. The same professor illustrates a more limited form of externalisation. She curated which sites her tutor agent could draw on and which to avoid, then conveyed that filtering to her students orally and, in her own words, did not systematise it [P3.13]. The agent's information base was curated, while the rule used to curate it remained oral. At the furthest end of the range, knowledge

remained unwritten because a written position becomes binding where opinions are commercially contestable, so the professional incentive runs against codifying it [P6.3], suggesting a possible barrier to knowledge transfer in that field.

Evidence on externalisation comes mainly from the professor interviews. The focus group was not asked about tacit knowledge, so the administrative evidence offers no basis for assessing this issue. The data-readiness records and the professor and administrative transcripts were also searched for organisational information curated above the individual, but none was found. The written provision that does exist addresses what individuals may upload and how sources must be handled [D5.2, D5.3], without allocating responsibility for organisational information, and one administrative participant ties his own restraint directly to that absence [S7.3]. Externally, the information-dependent embedding the case lacks was reached where organisational information was already accessible. One account describes internal search across the intranet and information repositories turning an organisation's dispersed knowledge into something its people could actually use [E8.7]. Reachability carries its own exposure, and one specialist raises the converse risk that an indexer will answer questions about individual salaries that were technically accessible and practically unfindable [E4.3], which is the condition the professor above met from the other direction when the system returned work from thirteen years earlier. Two further accounts describe arrangements built for codification itself, one placing the act of recording inside the work so that it happens as a by-product [E9.4], and one creating the conditions for an individual's knowledge base to be held in a governed repository [E4.11]. The permission uncertainty is also not FEP's alone, since a specialist speaking from an audit standpoint reports that his organisation cannot establish what happens to the data it supplies [E4.7].

4.6. The knowledge-intensive conditions and the coordination of the mechanisms

Sections 4.2 to 4.5 reported each mechanism on its own. Table 1 condenses those judgements so that the pattern across them can be seen at once (Annex D).

Table 1. Context-level judgements by mechanism and precondition

Context	Workflow redesign	Skills & absorptive capacity	Governance	Data Readiness
Academic (7 professors)	Localised but recurring. Two instances reached students; one yielded no further evidence	Localised but recurring. diffusion is episodic and does not accumulate	Localised but recurring, always as a personal rule	Nascent to localised; no arrangement above the individual

Administrative (8 participants, 6 services)	Localised but recurring in two participants' routines; the most intensive user showed no further evidence on examination	Nascent, with a single demonstration inside the facilitated session	Localised but recurring as a consistent norm, with no formal arrangement	Nascent

The four knowledge-intensive conditions are consistent with, and help interpret, the pattern summarised in Table 1 (Annex E1). Professional autonomy enabled participants to adapt the technology, determine what to learn, and set their own boundaries, while redesigned sequences and governing rules remained locally held. Task ambiguity coincided with the absence of agreed effective-use standards; tacit expertise informed professional boundaries but was difficult to articulate for transfer; and decentralised coordination left guidance and redesigned practices weakly connected across colleagues. The evidence supports this interpretation of the observed pattern but does not identify the causal effect of any condition.

The proposed mechanisms developed with few observed connections. Redesigned practices generated no peer learning, and the two practices that reached another role were not laterally adopted by colleagues. Competence and governance were also disconnected: participants reported no training in the relevant rules, no rule specified the competence required, and the training material addressed capabilities without clarifying authorisation [D6.1]. No instance was found in which workflow change, required competence, and governing rules were decided together. A limited connection did emerge when professionals redesigned a sequence or configured an agent and simultaneously set personal rules for its use and information access [P7.1, P7.4, SG.2]. However, it remained within individual practice and provided no observed organisational basis for joint development of the mechanisms. Because the instruments did not examine how task ambiguity or information conditions shaped coordination among the mechanisms, both relationships were recorded as not examined. A further condition emerged from the same accounts, bearing not on how the mechanisms developed but on whether professionals invested in them. An administrative participant linked her willingness to invest to how long she expected access to continue [S6.4]. Such uncertainty was reported to discourage investment among participants who had not yet committed to the technology. Participants who invested the most had already decided that the technology mattered [P1.4, P3.5], so their investment is not attributable to the pilot.

When asked what would most support organisational value, participants generally prioritised competence. The focus group ranked skills first and workflow redesign last [SC.1]. A separate population of 1,430 students produced the same ordering, with 58.1% selecting training and 5.0% selecting workflow redesign. Professors divided: two placed training first or second, and one placed the regulation of use first [P1.10, P5.10, P6.10]. The professor operating a configured artefact in continuous use was the only one to place reliable data first and workflow redesign second [P7.10]. Teaching staff in the parallel institutional evaluation returned the same competence-first ordering, tying continuity most often to further training and then to clearer guidance [D4.6]. Most external specialists placed greater emphasis on process and vision [E3.5, E7.3, E9.3], while one product developer ranks training first [E6.5].

4.7. Early indications of capability formation

Chapter 2 identifies four indicators through which AI-enabled organisational capability can be observed. Table 2 assesses each indicator at the case level and presents its evidential basis and limits, allowing partial and uneven development to remain visible (Annex F1). The six categories used above apply to participant-level judgements and do not extend to the case level, so each judgement below is expressed in the form its evidence supports.

Table 2. The four capability indicators at case level

Indicator (from 2.7)	Judgement	Principal basis	Principal limit
(i) Embedding in recurring processes	Nascent to localised	Recurring in most contexts examined; genuinely embedded in only a few	Every instance remains with its originator; no artefact has an assigned owner, maintenance plan, or handover
(ii) Shared verification	Behaviour at ceiling; formal arrangements absent	Consistent, rigorous, unprompted personal verification in both contexts	No agreed check, procedure, escalation route, or accountability exists anywhere in the case
(iii) Distribution beyond originating users or units	Nascent – diffusion occurs but does not accumulate	Three unrecorded, one-off episodes, leaving no reusable artefact	The dedicated sharing space went unused; the one notable capability had already been delivered through training
(iv) Retention beyond originating individuals	Not established	Every artefact is built, held, and maintained by a single person	The counterfactual was not put within the case, so single-person dependency is observed and retention is untested.

The indicators develop unevenly, and Table 2 records where each stands. Only verification reaches a ceiling as behaviour: the norm is applied by every participant with whom it was examined, and it can develop no further without an arrangement to carry it.

Indicator (iv) needs one further distinction. Every configured artefact in the corpus, including those that failed or were abandoned, was traced to its holder, and each is built, held, and maintained by one person, with no assigned owner, maintenance arrangement, or handover. That is single-person dependency: an observed property of the artefacts. Retention is counterfactual about what would survive an originator's departure, and it was not raised with either the focus group or any of the seven professors. Accordingly, dependency does not demonstrate that retention failed. The evidence therefore supports partial indications consistent with early capability formation, not an organisational capability already in place. The external accounts did not establish institutional retention even in more embedded arrangements [E2.12, E6.9]. Because the four indicators developed unevenly, as Section 2.7 anticipated, they are reported as separate probes and not combined into a composite scale.

These indicators rest on evidence streams that Chapter 3 keeps distinct: the professor interviews, the administrative focus group and its collective statements, the specialist accounts, the institutional documents, and, descriptively, the usage export and the student questionnaire. The streams relate in three ways. They agree on verification, which the professors and the focus group report independently and which the parallel institutional evaluation records in the same terms [P7.4, SG.1, D4.3]; under the rule set in Section 3.1, agreement of this kind counts only because independent sources address the same claim about the same object. They are complementary on governance, where the documents establish what was written and for whom [D5.1] and only the interviews establish whether it was read [P1.8, P5.9, S6.3]. They differ on priorities: within the case competence is ranked first, while the specialists who have observed adoption elsewhere name process and vision [SC.1, E3.5, E7.3, E9.3], a difference of standpoint, since those accounts concern other organisations. The questionnaire and the export stand outside all three relations. Neither addresses the same population or the same period as the participant evidence, so they situate the case against broader patterns of use without bearing on the indicators themselves.

4.8. The external comparison and synthesis

The external accounts set beside earlier findings are specialists' retrospective descriptions of their own organisations and professional standpoints, providing no direct evidence about FEP. Read together, they show how arrangements missing at FEP operate elsewhere and how adoption stalls when those arrangements are absent. One specialist describes a pattern closely resembling the case: "If each person uses their own AI it does not change the process.

The company does not work with AI; it is the employees who have AI for themselves" [E9.1].

Across the accounts, movement beyond localised use was associated with three arrangements (Annex H1): named ownership and sponsorship; recurring structures for learning and diffusion; and formal review, auditability, and maintenance responsibilities. Stalling was associated with training without conversion, licences without time or enablement, and individual gains blocked by unchanged downstream processes. Engagement was also reported to depend on how specific the artefact was, as one specialist observed that "Specific agents had far more engagement than generic agents, because it was easier to see the added value compared with tools such as ChatGPT" [E6.2]. In a resource-rich software organisation, protected time granted without ownership was abandoned [E2.1], while an SME operating under tighter constraints built a configured agent and, in about three months, an employee portal [E1.2]. The same specialist attached a condition to such protected space: "A protected space has to be created for someone to experiment, test and improve, and that person needs the authority to change the workflows" [E2.4]. Resources alone therefore did not determine how far adoption travelled. FEP's pattern resembles the conditions these accounts associate with stalling, without occupying a position on any adoption sequence.

The strongest enforcement structures were already in place before AI adoption [E2.11], so the technology was incorporated into existing governance, and embedding came from settings where tasks could be readily verified and management could specify how work should be performed [E2.10]. In one such setting the requirement was attached to the work itself, as team leads were required to hold certification [E2.6]. These conditions apply less readily to the academic work examined at FEP. The reported partial integration was itself recent, since the practice had crossed a departmental boundary only weeks before the interview [E2.7, E2.8]. Even where practices were described as embedded in organisational processes, the accounts do not establish that they persisted beyond their originators.

Within these limits, the comparison clarifies the distinction between recurring individual practices and organisational capability. At FEP, each mechanism recurs at the individual level, and practices reach a second party only through an authority relation the professional already holds. Task-level gains therefore coexisted with practice that was recurring but individually held, and the evidence supports only partial and uneven indications consistent with early capability formation.

5. Discussion

5.1. The individual-organisational gap

The study asks how individual, task-level use of enterprise generative AI might begin to acquire an organisational form, and what enables or constrains the process during early adoption. At FEP the process began within individual practice. The raw material for each mechanism identified in the framework was present, but no arrangement carried it beyond the person who produced it. The central gap therefore lies not between acquiring the technology and using it, but between using it well as an individual and using it as an organisation. Sections 5.2 to 5.4 examine why the gap persists: ownership and diffusion, from Sections 4.2 to 4.4; the rival explanation, from Sections 4.1, 4.4 and 4.6; and codification, from Section 4.5.

The gap does not indicate a failure of adoption. In Gallivan's (2001) terms, FEP achieved primary adoption and a limited degree of secondary adoption: as reported in Section 4.1, 58.2% of licensed accounts submitted at least one prompt. What remained underdeveloped was depth. The findings therefore show that breadth and depth can develop independently. Some individual uses were highly developed, such as those reported in Section 4.2 – the recorded-supervision routine, the Socratic agent, and the query agent – but none of them altered how the faculty performs its work.

Task characteristics help explain this independence. The experiments reviewed in Section 2.2 examine bounded tasks with fixed outputs, for which generative assistance reliably increases speed or throughput (Brynjolfsson et al., 2023; Noy & Zhang, 2023; Peng et al., 2023). FEP's professionals, by contrast, produce open-ended outputs. Drawing on Dell'Acqua et al. (2026), Section 2.2 shows that effective use requires users to judge where AI assistance is appropriate and verify its outputs. At FEP, where output was elastic, time savings were reinvested in better work or absorbed by verification, and released only where verification was inexpensive, as Section 4.1 reports. The gains were therefore real for individuals but largely invisible to the organisation. This challenges a widely held expectation about aggregation: Section 4.1 reports that 60.3% of the 1,430 surveyed students expected individual speed gains to raise organisational productivity, which would require deliberate coordination of a kind not observed at FEP.

Artefact quality does not explain the absence of diffusion [E6.2]. FEP's artefacts were specific in the relevant sense, yet, as reported in Section 4.2, none reached a colleague. Davenport's (1993) distinction between augmentation and process innovation fits the case more closely: at FEP, augmentation spread while process innovation did not. The case nevertheless supports Chapter 2's complementarity argument, although at a level of aggregation that the framework did not anticipate. Drawing on Brynjolfsson and Hitt (2000) and Milgrom and Roberts (1990, 1995), value was proposed to emerge when workflow redesign, skills, and governance reinforce one another. At FEP, these complements cohered within individual practices. The professor who built the Socratic agent also learned how to configure it and established the rules governing the information it could use [P7.1, P7.4, P7.9]. The focus group's account of agent configuration describes the same coupling [SG.2].

Professional autonomy is consistent with that coupling remaining locally contained, since each of the three developed without extending into a colleague's work. Mikalef and Gupta (2021) locate AI capability in the ability to select, orchestrate, and leverage AI-specific resources. FEP possessed these resources but lacked the organisational orchestration that individual autonomy alone cannot provide. The more revealing indicator of early capability formation is therefore the level at which the complements cohere, not their presence. Section 2.4 supplies the frame: Orlikowski (1992) holds that use can itself produce and reinforce organisational arrangements, which was evident at FEP, but the arrangements were personal. The technical side of the work system changed, while the social organisation of the work did not, and Trist and Bamforth (1951) examine the two as one.

5.2. Ownership and authority: why nothing travelled

The two academic cases that extended beyond their builders – the recorded-supervision routine and the Socratic agent – crossed relationships in which the professor already holds authority and can change the work unilaterally. By contrast, as reported in Section 4.2, neither recurring administrative redesign crossed such a boundary, and no lateral movement of redesigned practices between peers was observed. The distinction matters because reaching students shows that an individual can extend AI-assisted work along a line of authority, but does not demonstrate that the organisation can move practices across roles over which no single actor controls.

The absence of diffusion cannot be attributed simply to a lack of goodwill, infrastructure, or exposure, all of which were present in the episodes reported in Section 4.3. The peer-led

sessions left nothing behind, the Teams sharing space was never populated, and the capability whose demonstration surprised the focus group had already been delivered in the pilot training materials. What each lacked was conversion, without which generative AI use may remain ad hoc and isolated (Mayer et al., 2025). In Cohen and Levinthal's (1990) terms, the participants recognised the value of the new knowledge, but the case showed no recurring arrangement through which it was assimilated or applied beyond its originators. Accordingly, the capacity they locate in communication across units had no channel here. Chapter 2's account of decentralised coordination may explain why those processes did not arise, as diffusion has no default carrier where hierarchy is subdued and coordination rests on mutual adjustment (Alvesson, 2004; Lönnqvist & Laihonen, 2017).

Governance shows the same divide more sharply, coinciding with a stronger underlying norm. Even a rule upheld across every examined account does not constitute an organisational arrangement if each user holds it independently and responsibility remains unallocated. Measured against Mäntymäki et al.'s (2022) definition, FEP determines acceptable uses individually, allocates no decision rights, and connects AI oversight to data governance only through a rule addressed to individuals. Task ambiguity may partly explain this absence, as a rule binding several actors is hard to formulate where appropriate use cannot be specified in advance. Yet, an administrative participant complied with an external accreditation body's requirement to disclose AI use [S4.1], demonstrating that rules can bind professionals. What the case lacked was an internal body that issued an equivalent rule.

External accounts associated shared organisational practice with a dedicated internal actor who had responsibility for adoption and authority to influence work. This authority became consequential when it was attached to the work itself [E2.6]. Section 4.3 reports two further arrangements that perform the same function by making locally developed practice visible, which is what FEP's episodes left to individual initiative.

These comparisons support prospective implications, bounded by two limitations. The first is sequence. The strongest external arrangement relied on an enforcement architecture that predated generative AI [E2.11]. It therefore demonstrates how AI adoption can be absorbed into an established governance system, not how its adoption might generate such a system. The second concerns the nature of the work. The same arrangement governed verifiable tasks over which managers already possessed legitimate authority to prescribe methods and standards [E2.10]. Academic work offers no equivalent concentration of

authority. Together, these limitations show that the challenge extends beyond a failure to assign ownership, because authority is dispersed where professional autonomy is found. The comparison identifies the functional gap – concentrated ownership with enough authority to make emerging practices visible – without providing an arrangement a faculty could adopt.

5.3. What the rival account explains, and what it cannot

Chapter 3 compared the framework-derived account with a plausible rival at seven points where the two implied different observations (Annex G4). The rival account starts from FEP's circumstances: it is a small organisation conducting a pilot of uncertain duration. On this view, scarcity and uncertainty should explain the findings without reference to knowledge-intensive conditions. The comparison supports part of this account. Resource scarcity and the uncertain licence horizon are consistent with both the extent and the timing of the observed pattern. As reported in Section 4.1, use remained limited and concentrated among a few individuals. Practices lapsed quickly, and, as Section 4.6 reports, participants who were not already convinced of the technology's value were reluctant to invest in learning it. This horizon effect is an inductive finding because the framework does not predict it. Its practical implication is direct: uncertainty about a pilot's continuation may discourage the investment that the pilot is intended to elicit. This findinding also supplies a microfoundation for an aggregate argument. Brynjolfsson, Rock, and Syverson (2021) attribute the lag between technology adoption and measured value to costly and largely unmeasured investments in intangible assets. When the expected period of access is short, withholding such investment may be a rational response, so the J-curve's lag may arise partly from many individual decisions of that kind.

Scarcity alone fits the shape of the observed pattern less well. The rival account loses explanatory power at the boundary of appropriate use. If cost were the binding constraint, professionals would be expected to prioritise the applications that offered the greatest time savings. Several participants instead rejected precisely such applications on professional grounds. A boundary imposed by cost falls where a use becomes too expensive; one imposed by professional judgement falls where the work would cease to be the professional's own, and can therefore exclude a use that is inexpensive, available, and valuable. The same distinction applies to the codification discussed in Section 5.4, which halted at its least expensive step. External accounts likewise described organisations in which authority and ownership carried adoption further under comparable constraints, whereas spare capacity

did not [E1.2, E2.1]. Because both readings were formulated during the analysis, the comparison establishes their relative fit, supporting no causal attribution.

5.4. Professional knowledge and the boundary of codification

Drawing on Nonaka and Takeuchi (1995), Chapter 2 defined externalisation as the process through which organisation-specific knowledge becomes available to AI-assisted work. It was expected that the process would be selective, because codification can be costly, contested, or incomplete (Alvesson, 2004; Lönnqvist & Laihonen, 2017). The findings are consistent with selective externalisation and suggest a further source of that selectivity. In the clearest cases examined at FEP, the boundary appeared to depend partly on who stood to benefit from codification, not only on how difficult codification was.

The case's clearest instance of externalisation demonstrates both how far the process could proceed and where it halted. The structured manuals one professor built from photographs of her notebooks, reported in Section 4.5, were substantial and produced explicit material in a form that could support AI-assisted work. Yet the same professor curated which websites her tutor agent could use, conveyed that filtering to students orally, and did not systematise it [P3.13]. Externalisation proceeded where the author could capture the return through future reuse and stopped before the practice became transferable. The contrast is informative because the omitted step was the less demanding of the two.

Where the incentive ran against the author, the boundary was expressed as professional standing. The professor in the normative band, who articulated the corpus's most developed rule set, described it as an asymmetry: "From me to it, it is like a trainee; from it to me, it is not like a trainee" [P6.1]. The asymmetry is a governing rule grounded in the purpose of his expertise and inseparable from the judgement that produced it. The same participant's account of declining written doctrine in his field shows the liability codification can create: "A professor gives his opinion in his textbook; then he issues an opinion, and the other side says: but in your book you said this. That is, I became bound to that position, so I stopped writing" [P6.3]. Articulation may therefore be obstructed not only by cost, contestation, or loss of context, but also by the liability created by the written record itself. The implication is structural, as the information-dependent embedding that Jöhnk et al. (2021) and Mikalef and Gupta (2021) treat as a matter of infrastructure and data readiness may depend on an incentive that infrastructure alone cannot create. It also helps explain the form that data readiness assumed at FEP, as reported in Section 4.5: material assembled for its author's own

use can support individual embedding without becoming an organisational asset. However, as Section 3.10 notes, the tacit-knowledge probe was not put to the focus group, so the claim rests on two streams: the professor interviews (within one participant's account of his own field), and the specialist accounts of arrangements built for codification. Where effort or habit is the obstacle, an organisational arrangement may lower it by making documentation a by-product of the work, as one external account in Section 4.5 mentions [E9.4]. No arrangement in the corpus addresses liability and lowering the cost of writing answers neither why one professor stopped writing nor why the professor who does write should write for anyone else.

5.5. What the account establishes, and its reach

At the level supported by the evidence, the three sub-questions converge on one finding: the components of AI-enabled capability were present at FEP, but workflow redesign, skills and absorptive capacity, and governance cohered only within individual or locally held practices. Data readiness constrained information-dependent embedding rather than with use altogether; individuals substituted their own curation for the organisational arrangements that FEP did not provide. The knowledge-intensive conditions were consistently associated with this configuration and provided an interpretive account of its form, without establishing causal effects. Professional autonomy enabled individuals to adapt to the technology but provided no route through which one person's redesign could become another's practice. Task ambiguity left no clear criterion for deciding which practices should be generalised. Tacit expertise located the boundary of appropriate use within professional judgement and weakened the incentive to codify that judgement. Decentralised coordination coincided with workflow redesign, skill development, and governance proceeding in parallel without converging. The external accounts identified a similar point of stalling and described organisational arrangements associated with wider embedding built on ownership, authority, and enforcement, which arose in settings where work was verifiable and authority concentrated – conditions that academic work does not reproduce. Two limits bound that account. Because the professor interviews included only Copilot users, Sections 5.2 to 5.4 describe constraints encountered after adoption, so the account captures a minimum set, rather than an exhaustive set of constraints. And because the administrative evidence rests on a single focus group, it supports only group-level indications about ownership and governance rather than claims equal to those from eight individual interviews.

6. Conclusion

This study examined how individual, task-level use of enterprise generative AI may develop into AI-enabled organisational capability in a knowledge-intensive setting, and which conditions appear to enable or constrain this process during early adoption. During the first months of Microsoft 365 Copilot use at FEP, recurring and sophisticated practices developed while their wider embedding remained limited. Although 58.2% of licensed accounts submitted at least one prompt, usage alone did not establish assimilation in Gallivan's (2001) sense. No redesigned practice was observed spreading laterally to a colleague or across an administrative service. Competence circulated through isolated exchanges that produced no recurring arrangement or reusable practice. Verification was consistently reported as rigorous and independently adopted, but was not supported by shared procedures, formally allocated responsibility, and escalation routes. Personal curation enabled information-dependent use without creating shared data arrangements.

The principal theoretical contribution is that complementarity depends on both level and combination. Workflow redesign, skills and absorptive capacity, and governance may reinforce one another within individual practices while remaining weakly coordinated across roles and units. What matters is whether they cohere in arrangements that make practices visible, transferable, reproducible, maintainable, and less dependent on their originators. The individual-organisational gap names a distinct condition in which components associated with capability are present while the relationships required to connect and sustain them remain underdeveloped. It draws attention to a transition from individual use to wider orchestration that existing accounts of adoption, assimilation, complementarity, and AI capability do not fully explain.

Knowledge-intensive conditions were consistently associated with this pattern, without being established as causal determinants. Professional autonomy accompanied local adaptation but provided no observed route for lateral diffusion. Task ambiguity coincided with difficulty in determining which practices should be generalised. Tacit, context-dependent expertise was associated with limited articulation and transfer, while decentralised coordination coincided with the proposed mechanisms developing without converging in shared arrangements. External accounts provided an interpretive comparison, linking wider embedding to explicit ownership, authority, learning structures, review mechanisms, and maintenance responsibilities; they offered neither a maturity model nor a transferable

blueprint. The practical implication is to assess early adoption by whether useful practices become visible, shared, maintained, and connected to responsibilities. Thus, training, licence allocation, and increased use remain insufficient indicators of capability formation. The findings extend Mittelstadt's (2019) argument that principles have limited practical effect unless translated into operational mechanisms and accountability. At FEP, professionals reported verifying outputs and setting boundaries for appropriate use, while shared procedures and allocated responsibilities remained absent from the available evidence. Effective AI governance in knowledge-intensive work may therefore require arrangements that coordinate professional judgement without replacing or standardising it.

The study's scope is bounded by the early months of a single case. It establishes no causal effects, realised organisational value, mature capability, or appropriable advantage. The professor sample included only Copilot users, and the administrative evidence came from one focus group. Retention beyond originators was not directly examined, so dependence on one person does not demonstrate that practices would fail to endure. The analysis was also conducted by one researcher and, although supported by an inspectable audit trail, was not independently validated. The early timing nevertheless captures a phase that may be difficult to reconstruct once shared arrangements become established.

The study connects task-level effects with organisational capability and provides a framework for distinguishing recurring use from shared practice, personal verification from coordinated governance, and individual curation from shared data arrangements. It argues that availability and sophisticated use alone do not establish this capability. By examining how practices might spread beyond their creators, become connected across the faculty, and endure, it advances the literature beyond adoption and task-level performance, and provides a basis for understanding how enterprise generative AI may become embedded in collective work.

References

Alvesson, M. (2004). *Knowledge work and knowledge-intensive firms*. Oxford University Press.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly, 45(3), 1433–1450*. https://doi.org/10.25300/MISQ/2021/16274

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., et al. (2021). *On the opportunities and risks of foundation models* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2108.07258

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications.

Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives, 14(4), 23–48*. https://doi.org/10.1257/jep.14.4.23

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work* (NBER Working Paper No. 31161). National Bureau of Economic Research. https://doi.org/10.3386/w31161

Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics, 13(1), 333–372*. https://doi.org/10.1257/mac.20180386

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly, 35(1), 128–152*. https://doi.org/10.2307/2393553

Cui, K. Z., 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* [Working paper]. SSRN. https://doi.org/10.2139/ssrn.4945566

Davenport, T. H. (1993). *Process innovation: Reengineering work through information technology*. Harvard Business School Press.

Dell'Acqua, F., McFowland, E., III, Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge

worker productivity and quality. *Organization Science, 37(2), 403–423*. https://doi.org/10.1287/orsc.2025.21838

Demirer, M., Horton, J. J., Immorlica, N., Lucier, B., & Shahidi, P. (2026). *Chaining tasks, redefining work: A theory of AI automation* (NBER Working Paper No. 34859). National Bureau of Economic Research. https://doi.org/10.3386/w34859

Drucker, P. F. (1999). Knowledge-worker productivity: The biggest challenge. *California Management Review, 41(2), 79–94*. https://doi.org/10.2307/41165987

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—Principles and practices. *Health Services Research, 48*(6, Pt. 2), 2134–2156. https://doi.org/10.1111/1475-6773.12117

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

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology, 13*, Article 117. https://doi.org/10.1186/1471-2288-13-117

Gallivan, M. J. (2001). Organizational adoption and assimilation of complex technological innovations: Development and application of a new framework. *The DATA BASE for Advances in Information Systems, 32(3), 51–85*. https://doi.org/10.1145/506724.506729

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review, 61(4), 5–14*. https://doi.org/10.1177/0008125619864925

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes – An interview study of organizational AI readiness factors. *Business & Information Systems Engineering, 63(1), 5–20*. https://doi.org/10.1007/s12599-020-00676-7

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Lönnqvist, A., & Laihonen, H. (2017). Management of knowledge-intensive organizations: What do we know after twenty years of research? *International Journal of Knowledge-Based Development, 8(2), 154–167*. https://doi.org/10.1504/IJKBD.2017.10006153

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics, 2(4), 603–609*. https://doi.org/10.1007/s43681-022-00143-x

Mayer, H., Yee, L., Chui, M., & Roberts, R. (2025). *Superagency in the workplace: Empowering people to unlock AI's full potential*. McKinsey & Company. https://www.mckinsey.com/…/superagency-…-full-potential-v4.pdf

Microsoft & LinkedIn. (2024). *2024 Work Trend Index Annual Report: AI at work is here. Now comes the hard part*. https://assets-c4akfrf5b4d3f4b7.z01.azurefd.net/…/2024_Work_Trend_Index_Annual_Report.pdf

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management, 58*(3), 103434. https://doi.org/10.1016/j.im.2021.103434

Milgrom, P., & Roberts, J. (1990). The economics of modern manufacturing: Technology, strategy, and organization. *American Economic Review, 80(3), 511–528*. https://www.jstor.org/stable/2006681

Milgrom, P., & Roberts, J. (1995). Complementarities and fit: Strategy, structure, and organizational change in manufacturing. *Journal of Accounting and Economics, 19(2–3), 179–208*. https://doi.org/10.1016/0165-4101(94)00382-F

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1(11), 501–507*. https://doi.org/10.1038/s42256-019-0114-4

Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research, 25(9), 1212–1222*. https://doi.org/10.1177/1049732315588501

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press.

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

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science, 3(3), 398–427.* https://doi.org/10.1287/orsc.3.3.398

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). *The impact of AI on developer productivity: Evidence from GitHub Copilot* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2302.06590

Ridder, H.-G. (2017). The theory contribution of case study research designs. *Business Research, 10, 281–305.* https://doi.org/10.1007/s40685-017-0045-z

Schneider, J., Meske, C., & Kuss, P. (2024). Foundation models: A new paradigm for artificial intelligence. *Business & Information Systems Engineering, 66, 221–231.* https://doi.org/10.1007/s12599-024-00851-0

Simons, W., Turrini, A., & Vivian, L. (2024). *Artificial intelligence: Economic impact, opportunities, challenges, implications for policy* (European Economy Discussion Paper No. 210). European Commission, Directorate-General for Economic and Financial Affairs. https://op.europa.eu/…/0fb59a42-55f6-11ef-acbc-01aa75ed71a1

Stake, R. E. (1995). *The art of case study research*. SAGE Publications.

Starbuck, W. H. (1992). Learning by knowledge-intensive firms. *Journal of Management Studies, 29(6), 713–740.* https://doi.org/10.1111/j.1467-6486.1992.tb00686.x

Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations, 4(1), 3–38.* https://doi.org/10.1177/001872675100400101

Yazan, B. (2015). Three approaches to case study methods in education: Yin, Merriam, and Stake. *The Qualitative Report, 20(2), 134–152.* https://doi.org/10.46743/2160-3715/2015.2102

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.

Annexes

These annexes hold the analytical apparatus behind the empirical chapters. Annex A documents the basis of every quantitative figure the study reports. Annex B reports the consistency controls applied to the analytical record and the composition of the evidence base. Annex C reports what each research instrument reached. Annex D opens the context-level judgements summarised in Table 1, giving the evidence for and against each one. Annex E reports how the 4 knowledge-intensive conditions bear on each mechanism and whether the mechanisms developed together. Annex F gives the evidence behind the 4 capability indicators summarised in Table 2. Annex G reports the 2 procedures used to test the account. Annex H reports the external comparison and the standpoint from which each specialist account was given.

Evidence identifiers follow the convention defined in Chapter 3: P denotes professors, S administrative participants, SG the focus group's collective statements, SC its ranking exercise, E specialists and D documents. The analytical workbook itself is not reproduced: it holds participant names, service attributions and verbatim passages in the original Portuguese, and it is not anonymised.

Annex A. Source notes for the quantitative figures

Both quantitative sources are descriptive. Neither arose from a design that would support an inferential test, and none is reported. The notes below give, for each figure the dissertation reports, the denominator it rests on, the treatment of missing values, the time window it covers and the data-quality checks that bear on it. Figures held in the workbook but not reported in any chapter are not documented here.

A1. The tenant usage export

The export is a Microsoft 365 Copilot usage report covering the 600 accounts licensed under the pilot. For each account it records the number of prompts submitted across all Copilot surfaces, the number of distinct days on which any Copilot activity occurred, and the date of most recent activity in each application. It reports a trailing 180-day window and was refreshed on 12 June 2026; both values are constant across all 600 rows.

No account is absent from the export. Where an account registered no activity in the period, its last-activity date is blank. Blank dates are read as an absence of recorded activity

inside the institution's own tenant, never as an absence of use, because the export cannot observe other services.

Two non-interchangeable denominators are used. A total of 352 accounts carry a last-activity date, but only 349 submitted at least one prompt, 3 accounts having registered activity without a recorded prompt. Each figure below states the denominator it rests on. Recency is computed strictly against the refresh date, so an account counts as active in a closing window only where its last activity falls after that boundary; the convention is applied uniformly.

Table A1. Descriptive figures from the Microsoft 365 Copilot tenant usage export, with the denominator on which each rests.

Figure as reported in the dissertation	Value	Denominator	Reported in
Accounts submitting at least one prompt	349 of 600 (58.2%)	All 600 licensed accounts	Chapters 1, 4.1, 5.1, 6
Total prompts recorded	28,045	All surfaces, all accounts	Section 4.1
Median active account	5 prompts across 2 active days	The 349 accounts submitting at least one prompt	Section 4.1
Share of all prompts held by the most active 10% of accounts	90.3%	All 600 accounts, ranked by prompt volume	Section 4.1
Accounts whose identifier format indicates staff or faculty, and their share of prompts	65 accounts, holding 54.7% of prompts	All 600 accounts	Section 4.1
Accounts inactive for more than 6 weeks at the close of the window	78 (22.3%)	The 349 accounts submitting at least one prompt	Section 4.1

Note. The staff-or-faculty split is inferred from the format of the account identifier, which distinguishes an 8-character local part from an 11-character one. The split is exhaustive across the 600 accounts but is not a verified personnel roll. The concentration figure describes a distribution of contact with the tool: the export records that a prompt was submitted, not the task it served, the quality of what returned, or whether any process changed.

A2. The student questionnaire

The questionnaire was designed for this study and distributed by e-mail, through the supervisor, to students of 3 faculties of the University of Porto. A total of 1,503 responses were filed between 20 June and 15 July 2026. An opening item asked whether the respondent currently used any AI tool; the 1,430 who answered yes were routed to the substantive items, and the remaining 73 saw none of them. Every percentage reported from the substantive items rests on the 1,430 base.

The substantive items were complete for all 1,430 routed respondents, so no figure required imputation and no case was removed. Where an item permitted multiple selections, shares are computed on the 1,430 base and do not sum to 100. The instrument asked about

practice current at the time of completion; it is not aligned to the tenant export's window, and the 2 are never compared.

Three data-quality checks were run on the analytical base. First, completion duration: the median was 5.1 minutes, with a 10th percentile of 3.2 minutes and a 90th of 9.4 minutes, and 1 case fell below 60 seconds. Computed on all 1,503 filed responses the median is 5.0 minutes, and 71 cases fall below 60 seconds, all of them non-users routed past the substantive block. Second, straight-lining: 17 respondents, or 1.2%, answered flat across 4 of the 5 item blocks, with flat response rates by block of 10.8% for effects, 4.2% for tasks, 12.9% for skills, 2.8% for verification and 39.4% for information conditions. The last figure is expected rather than anomalous, because the 3 items in that block are near-consensus and highly correlated. Third, tool-name harmonisation: 76 raw strings in the open tool-name field were reduced to 53 labels, an adjustment material for 1 tool, which rises from 2.3% to 6.2% once 15 distinct spellings are merged.

No case was removed on the basis of any check, and the flags were retained rather than acted upon. Under a sensitivity test excluding flagged cases, no figure reported in the dissertation moves by more than 0.8 percentage points.

Table A2. Figures from the student questionnaire, with the item type from which each derives.

Figure as reported in the dissertation	Value	Item type	Reported in
Tasks became faster	92.9% agree or strongly agree	5-point agreement scale	Section 4.1
More time spent reviewing and correcting outputs	68.3% agree or strongly agree	5-point agreement scale	Section 4.1
Dropped a manual step, such as writing a first draft	49.1%	Multiple selection	Section 4.1
Began performing tasks in a different order	38.9%	Multiple selection	Section 4.1
Changed how work is divided with others	16.2%	Multiple selection	Section 4.1
Individual speed gains will raise organisational productivity	60.3% agree or strongly agree	5-point agreement scale	Sections 4.1, 5.1
Training ranked the most important condition for organisational value	58.1%	Single choice among 5 options	Section 4.4
Redesigning work and its division around AI ranked most important	5.0%	Single choice among 5 options	Section 4.4

Note. All percentages rest on the 1,430 respondents routed to the substantive items. The instrument records no faculty or unit variable, so no subset of responses can be identified as FEP students. It reached a broader population under different conditions from the case and establishes nothing about FEP's organisational mechanisms, assimilation, capability or value.

Annex B. Consistency controls and the composition of the evidence base

The analytical record was subjected to 19 automated checks, each computed by formula from the evidence register or from the workbook's traceability record, and each carrying an expected value stated independently of the result. A check with no expected value cannot fail, and a check that cannot fail is not a check. These checks establish that the evidence base is internally consistent and correctly gated. They establish nothing about whether a coding decision was right, whether a judgement is sound, or whether the corpus is sufficient; those questions are answered by the material in Annexes D to G.

B1. Automated consistency checks

Table B1. The 19 automated consistency checks applied to the analytical record, with the expected and returned value of each.

Family of check	No.	What the check tests	Expected	Returned
Identifier integrity	1	Evidence identifiers occurring more than once in the register	0	0
Source traceability	2	Records carrying no source locator, so that the passage could not be found again in its source file	0	0
Source traceability	3	Records missing the passage itself where one is required. Tenant-analytics records and the elicited ranking legitimately carry none	0	0
Source traceability	4	Records carrying no recorded verbatim status, that status being the result of testing the excerpt as an exact substring of its source	0	0
Permitted coding values	6	Distinct analytical categories in use, which must be the 3 mechanisms and the 1 cross-cutting precondition and no more	4	4
Permitted coding values	11	Pathway positions applied to records that are not external comparison, which would silently convert a position into an embedding judgement	0	0
Permitted coding values	12	External-comparison records carrying no standpoint, without which 9 unlike observers would read as equivalent	0	0
Separation of active from set-aside records	7	Records not marked active that nevertheless carry an analytical category, and so could be returned by a filter on a mechanism	0	0
Separation of active from set-aside records	8	Records not marked active that carry a knowledge-intensive condition or capability-indicator flag	0	0
Separation of active from set-aside records	13	Records flagged for quotation that are not active, one participant's remarks being reserved at his own request	0	0
Resolution of display references	9	Identifiers cited in an analytical display that do not resolve to a register record, enforcing the rule that nothing enters a chapter that is not a record	0	0
Resolution of display references	10	Identifiers cited in a display that resolve to a record the analysis had set aside	0	0

Family of check	No.	What the check tests	Expected	Returned
Completeness of proposition tests	18	Propositions carrying a disconfirmation test but no definition, which would let a numbered claim reach a chapter with no statement behind it	0	0
Completeness of proposition tests	19	Propositions defined but never tested against any evidence stream	0	0
Quotation fidelity	5	Portuguese-language records carrying no English rendering	0	0
Quotation fidelity	14	Quotations on the citation shortlist, fixed in advance against the argument rather than against the prose	30	30
Quotation fidelity	16	Shortlisted records whose excerpt was not verified as exact at its recorded locator, and which are therefore not eligible to be quoted	0	0
Quotation fidelity	17	Shortlisted quotations carrying no English rendering	0	0
Claim control	15	Records flagged as unable to support a claim about persistence, value or return that carry no memo stating that limit	0	0

Note. Checks are numbered as they stand in the workbook. Three carry more weight than the rest. Check 10 is what prevents a record the analysis set aside from carrying an argument in a display; check 9 enforces the rule that nothing enters a chapter that is not a record; and check 16 enforces the condition that a passage becomes eligible for quotation only once its wording has been verified at source.

B2. Composition of the evidence base

The counts below describe the corpus on which the checks were computed. They are not checks, and no arithmetic is performed on them. A record is a passage, so a record count measures how much was said about a domain and never how far that domain has developed.

Table B2. Composition of the evidence register, computed on active records only.

Count	Value	Basis and reading
Records in the register	198	Every registered passage, including the 5 that do not contribute to findings
Records active	193	The only records that contribute to judgements, displays and claims
Records set aside	5	3 dropped at the extraction ceiling, 1 reserved at the participant's request, 1 excluded as falling outside the framework
Within-case records	101	The 7 professor interviews and the administrative focus group
External-comparison records	70	The 9 specialist interviews. Never evidence about FEP
Document records	16	Drawn from 6 institutional and pilot documents
Tenant-analytics records	6	Distribution and intensity of use only
Workflow redesign	47	Mechanism
Skills and absorptive capacity	53	Mechanism
Governance	48	Mechanism
Data readiness	31	Cross-cutting precondition, not a fourth mechanism

Count	Value	Basis and reading
Records carrying no analytical category	14	Tenant analytics, the elicited ranking, national policy context, and records carrying only an inferential boundary
Records of configured-artefact use	18	The filter on which the configured-artefact argument runs
Records of open use of the assistant	75	Assigned per passage rather than per informant
Records carrying a value-inference boundary	17	Claim-control metadata. These contribute to no category count and to no judgement as independent evidence
Excerpts verified exact at their locator	191	Tested programmatically against the named source file at the recorded line or page
Renderings audited verbatim against the quoted span	29	Of the 30 shortlisted quotations, 29 render their Portuguese span word for word. The 30th, P5.5, was checked against source but is not verbatim: the speaker names a student twice and the rendering substitutes [student] and [him]. The substitution is marked in the quotation wherever it appears.
Records used in no analytical display	7	Not an error. It records how much verified evidence the analysis does not draw on
Professors interviewed	7	Individual semi-structured interviews
Administrative focus-group participants	8	Speakers drawn from 6 services, each stating their own service in the opening round
Specialist informants	9	All 9 transcripts used in the register
Institutional and pilot documents	6	4 governance documents, 1 national policy recommendation, 1 parallel-pilot report
Licensed accounts in the tenant export	600	Trailing 180-day window, refreshed 12 June 2026
Questionnaire responses filed	1,503	Filed between 20 June and 15 July 2026
Questionnaire analytical base	1,430	Respondents routed to the substantive items

Annex C. What each research instrument reached

Six judgement categories are used in this study, and 1 of them records that a construct was not examined. The distinction between a construct examined and found absent and a construct the instrument never reached is the one the analysis depends on most heavily, and it can be assessed only against a record of what was actually asked. This annex is that record. It was transcribed from each transcript's own question-mapping table and skipped-question list.

Coverage records what was asked. A question recorded as reached does not guarantee that the informant addressed it substantively, and a question not reached is never read as evidence that the thing itself is absent.

Note. The 3 instruments are not symmetrical, and no equivalent presentation is possible. The professor guide was administered 7 times individually, so coverage is reported as the number of interviews in which each question was reached. The focus group was a single session, so coverage is reported per question for that one session and cannot be expressed as

a count across participants. The specialist guide was adaptive, so coverage is reported as how directly each theme was reached across the 9 interviews. This asymmetry is itself a finding, and Annex D1 shows its consequence: every one of the 15 cells recorded as not examined falls in the administrative context.

C1. Professor interviews

Table C1. Coverage of the professor interview guide across the 7 interviews, with the analytical purpose and stated priority of each question.

Question	Analytical purpose	Priority in the guide	Interviews in which it was reached
Q1	Profile of use, non-use and initial heterogeneity	Essential	6 of 7
Q2	Work redesign as against acceleration of the same task, and the reabsorption of time saved	Essential	7 of 7
Q3	Configured agents as explicit statements of sources, limits, data and responsibility	Conditional	7 of 7
Q4	Competences and the development of learning	Essential	7 of 7
Q5	Absorptive capacity: diffusion as against individual retention of knowledge	Essential	7 of 7
Q6	The boundary of the task and professional judgement	Essential	6 of 7
Q7	Governance as an enabling or limiting condition, compatible with autonomy	Essential	7 of 7
Q8	Conditions for acceptable delegation in sensitive contexts	Conditional	2 of 7
Q9	Data readiness, the organisation of materials and tacit knowledge	Essential	7 of 7
Q10	Institutional value as against individual gain, and the principal brake	Essential	7 of 7
Q11	Empirical prioritisation of the conditions for organisational value	Essential	7 of 7
Q12	Heterogeneity of adoption between professors and the contextual conditions for it	Conditional	4 of 7
Q13	Closing open question on an unanticipated theme	Essential	2 of 7

Note. Questions were frequently rephrased, reordered, combined with another or extended by a follow-up rather than asked as written, and the workbook records which of these occurred question by question. A further question outside the guide was raised in 4 of the 7 interviews. Of the 3 questions the guide marks conditional rather than essential, 2 are the least often reached; the third, on configured agents, was reached in every interview. The closing open question was the item most often displaced by time.

C2. The administrative focus group

Table C2. Coverage of the administrative focus-group guide, question by question, for the single session held.

Block	Question	How it was reached	Topic and note on how it was put
B1	Q1	Asked as planned	Round-table: service, frequency, most-used task
B1	Q2	Not reached	Why frequency differs – never asked comparatively
B2	Q3	Rephrased	Speed across the whole process including rework

Block	Question	How it was reached	Topic and note on how it was put
B2	Q4	Rephrased and reordered	Process change – displaced to after learning and governance
B2	Q5	Reordered	Expectations unmet – asked only after the priority card
B3	Q6	Rephrased	Professional judgement versus delegable task
B4	Q7	Rephrased	Learning beyond the training session
B4	Q8	Rephrased	Sharing within services – answered very briefly
B4	Q9	Not reached	Prior experience that helps or hinders – not explored
B4	Q10	Rephrased, then followed up	Agent configuration decisions
B5	Q11	Reordered	Responsibility for a serious error – brought forward
B5	Q12	Rephrased	How results are judged and errors detected
B5	Q13	Rephrased and reordered	Tasks deliberately withheld from the tool
B5	Q14	Raised as a follow-up	What they do when unsure – asked only as a hypothetical about rules
B5	Q15	Combined with another question	Reach into FEP systems – combined with Q16
B5	Q16	Combined with another question	Reorganisation of documents – combined with Q15
B5	Q17	Not reached	Tacit institutional knowledge requiring externalisation – not explored
B6	Q18	Asked as planned	Priority card, 2 ranked choices, written in silence
B6	Q19	Rephrased	Institutional value – reframed as one more year of licences
B6	Q20	Not reached	Silent-failure probe – not asked
B6	Q21	Not reached	Value embedded in FEP versus retained by individuals – not asked
B7	Q22	Not reached	Open closing question replaced by a question about other tools

Note. Six of the 22 questions were not reached. Two of these bear directly on findings the dissertation reports. The item on tacit institutional knowledge requiring externalisation was not put, which is why the proposition on selective externalisation rests on 2 evidence streams rather than 3, a qualification that travels with that claim wherever it is reported. The item on whether value is embedded in FEP or retained by individuals was not put. The same counterfactual was not put to any of the 7 professors either, which is why the retention indicator is recorded as not established rather than as absent.

C3. Specialist interviews

Table C3. Coverage of the 14 specialist interview themes across the 9 specialist interviews.

Theme	Subject of the theme	Reached directly	Reached partially or indirectly	Not pursued
Theme ID	Theme	0	0	9
T1	Experience and standpoint	8	1	0
T2	Adoption trajectory	4	5	0
T3	Process before and after	9	0	0

Theme	Subject of the theme	Reached directly	Reached partially or indirectly	Not pursued
T4	Generic versus configured	6	3	0
T5	Skills	8	0	1
T6	Collective learning	6	1	2
T7	Human–AI boundary	4	5	0
T8	Governance	9	0	0
T9	Information and knowledge	8	0	1
T10	Value and evidence	8	1	0
T11	Priority conditions	7	0	2
T12	Missing piece / uneven adoption	0	8	1
T13	Knowledge-intensive organisations	3	6	0
T14	Faculty diagnosis and advice	4	3	2

Note. The specialist themes are numbered separately from the propositions tested in Annex G and are not related to them. The interview sequence was adaptive: the last of the 9 interviews was arranged when the opportunity arose, later than the rest, and was deliberately directed at the themes on which the earlier accounts had been weakest. A theme recorded as not pursued was therefore a deliberate omission in that interview and not a failure to ask.

C4. Fieldwork chronology

Table C4. Chronology of the 17 fieldwork sessions, in the order in which they were held.

Date	Type of session	Informant	Note
3 July 2026	Administrative focus group	S1 to S8, SG, SC	8 participants, 6 services
8 July 2026	Specialist interview	E8	–
9 July 2026	Specialist interview	E3	–
10 July 2026	Professor interview	P4	Quantitative band
10 July 2026	Professor interview	P1	Applied / management band
10 July 2026	Specialist interview	E6	–
11 July 2026	Professor interview	P7	Applied / management band
13 July 2026	Specialist interview	E1	–
13 July 2026	Specialist interview	E9	–
13 July 2026	Professor interview	P3	Quantitative band
14 July 2026	Professor interview	P6	Normative band
14 July 2026	Specialist interview	E4	–
14 July 2026	Professor interview	P2	Quantitative band
15 July 2026	Specialist interview	E5	A second session is recorded at 17 July
15 July 2026	Specialist interview	E7	–
18 July 2026	Professor interview	P5	Normative band

Date	Type of session	Informant	Note
5 August 2026	Specialist interview	E2	–

Note. Participant names are replaced throughout by the identifiers defined in Chapter 3. All 17 sessions are dated: 9 dates are printed in the transcript itself and 8 were supplied from the researcher's own records; where both exist they agree, in all 9 cases. Two features of the sequence bear on the analysis. The administrative focus group was the first session, 5 days before any individual interview, so it could not be informed by the professor accounts. The 9th specialist interview stands 3 weeks after the 8th, which is why the specialist coverage in Table C3 is uneven by design.

Annex D. Context-level judgements and the evidence behind them

Table 1 in Chapter 4 reports these judgements synthesised to the level of the 2 domains. This annex opens that synthesis. D1 gives the judgement reached for every informant against every construct, including the data-readiness column that Table 1 does not carry. D2 and D3 give the evidence for and against each of those judgements, for the academic and the administrative contexts respectively, together with the distinction on which each judgement turns. D4 gives the domain-level synthesis that Table 1 condenses.

The cells are qualitative categories. They are not scores, they are not ordered on a scale that permits arithmetic, and they are never counted, averaged or compared across columns. A judgement is never the sum of its supporting column: it synthesises supporting, limiting, contradictory and missing evidence together.

D1. The judgement grid at informant level

Table D1. Context-level judgement for every informant against the 3 mechanisms and the data-readiness precondition.

Informant	Context	Workflow redesign	Skills and absorptive capacity	Governance	Data readiness
P4	quantitative band	Nascent or experimental	Localised but recurring	Localised but recurring	Nascent or experimental
P1	applied / management band	Localised but recurring	Localised but recurring	Nascent or experimental	Localised but recurring
P7	applied / management band	Shared, reaching students	Localised but recurring	Localised but recurring	Localised but recurring
P2	quantitative band	Localised but recurring	Localised but recurring	Localised but recurring	Localised but recurring
P3	quantitative band	Shared, reaching students	Localised but recurring	Localised but recurring	Localised but recurring
P6	normative band	Examined, no evidence found	Localised but recurring	Localised but recurring	Nascent or experimental
P5	normative band	Nascent or experimental	Nascent or experimental	Localised but recurring	Nascent or experimental

Informant	Context	Workflow redesign	Skills and absorptive capacity	Governance	Data readiness
S6	administrative context	Nascent or experimental	Nascent or experimental	Nascent or experimental	Not examined
S4	administrative context	Not examined	Not examined	Nascent or experimental	Not examined
S8	administrative context	Nascent or experimental	Not examined	Not examined	Not examined
S7	administrative context	Localised but recurring	Not examined	Nascent or experimental	Nascent or experimental
S5	administrative context	Localised but recurring	Not examined	Not examined	Nascent or experimental
S2	administrative context	Examined, no evidence found	Localised but recurring	Localised but recurring	Examined, no evidence found
S1	administrative context	Not examined	Nascent or experimental	Not examined	Not examined
S3	administrative context	Localised but recurring	Examined, no evidence found	Not examined	Not examined
SG	collective utterance	–	–	Localised but recurring	–

Note. The 6 categories are, in ascending order: not examined; examined with no evidence found; nascent or experimental; localised but recurring; shared across users or units; and institutionalised. No cell in the case reaches the highest category, so the grid records localisation rather than institutionalisation. All 15 cells recorded as not examined fall in the administrative context and none among the professors. That distribution is a property of the instrument rather than of the setting: a single group session with 8 participants drawn from 6 services cannot reach every construct with every participant, whereas 7 individual interviews can. It is the principal reason the administrative evidence is treated as supporting group-level indications rather than claims equal to those from 8 individual interviews.

D2. Basis for each judgement: the academic context

Table D2. Evidence for and against every context-level judgement in the academic context, with the distinction on which each turns.

Informant	Construct	Judgement	Supporting evidence	Limiting or contrary evidence	Distinction on which the judgement turns
P4	Workflow redesign	Nascent or experimental	P4.1	P4.2	Isolated use only: no repetition of a changed sequence, no second party, no formalisation.
P4	Skills and absorptive capacity	Localised but recurring	P4.6	P4.7	Repetition established at the individual level; no diffusion beyond immediate contacts.
P4	Governance	Localised but recurring	P4.3, P4.4, P4.5	P4.10	Consistency across instances gives repetition; nothing above the personal rule is evidenced.
P4	Data readiness	Nascent or experimental	–	P4.8	Problem recognised, no action evidenced.

Informant	Construct	Judgement	Supporting evidence	Limiting or contrary evidence	Distinction on which the judgement turns
P1	Workflow redesign	Localised but recurring	P1.1, P1.2, P1.11	P1.3	Repetition established; every instance is held by the informant; the artefact that would have reached students was not implemented.
P1	Skills and absorptive capacity	Localised but recurring	P1.4	P1.5, P1.6	Individual competence is not organisational absorptive capacity; no diffusion or retention is evidenced.
P1	Governance	Nascent or experimental	–	P1.7, P1.8	No supporting passage exists in the register; the judgement rests on the 2 limiting records.
P1	Data readiness	Localised but recurring	P1.9	–	Repetition established at the individual level; no unit-level curation evidenced.
P7	Workflow redesign	Shared, reaching students	P7.1, P7.2	P7.3	Second-party use evidenced, which distinguishes this from repetition; formalisation and persistence are not evidenced.
P7	Skills and absorptive capacity	Localised but recurring	P7.9	P7.11	Communication exists and transmits normative concern rather than practice.
P7	Governance	Localised but recurring	P7.4	P7.5	Consistent personal rule; no shared arrangement evidenced.
P7	Data readiness	Localised but recurring	P7.7	P7.6, P7.8	Personal reorganisation evidenced; nothing at unit level.
P2	Workflow redesign	Localised but recurring	P2.1, P2.2, P2.3	P2.12, P2.15	Repetition established; no second party uses either practice.
P2	Skills and absorptive capacity	Localised but recurring	P2.4, P2.13	P2.5, P2.14, P2.16, P2.10	One diffusion event is evidenced and does not accumulate: no materials, no reference person, no repetition.
P2	Governance	Localised but recurring	P2.7	P2.6	Personal rule evidenced; no institutional rule evidenced.
P2	Data readiness	Localised but recurring	P2.8	P2.9	Personal curation evidenced; no organisational information arrangement.
P3	Workflow redesign	Shared, reaching students	P3.1, P3.2, P3.3, P3.4	P3.12	Second-party use evidenced; nothing is documented for reuse by a colleague and no owner is assigned.
P3	Skills and absorptive capacity	Localised but recurring	P3.11	P3.7	A diffusion event without a reusable object, against an unused mechanism.
P3	Governance	Localised but recurring	P3.9	–	Consistency gives repetition; no shared arrangement.
P3	Data readiness	Localised but recurring	P3.5, P3.6	P3.13, P3.10	Substantial personal externalisation; the one act that could have made it transferable was declined.

Informant	Construct	Judgement	Supporting evidence	Limiting or contrary evidence	Distinction on which the judgement turns
P6	Workflow redesign	Examined, no evidence found	–	P6.5	The construct was reached and nothing was found; distinct from 'not examined'.
P6	Skills and absorptive capacity	Localised but recurring	P6.6	–	Ad hoc sharing evidenced; no mechanism, no reference role.
P6	Governance	Localised but recurring	P6.1, P6.2, P6.8, P6.9	–	Repetition and consistency evidenced; nothing above the individual. One further governance record from this informant is reserved at his request and contributes to no judgement.
P6	Data readiness	Nascent or experimental	P6.4	P6.3	No organisational information arrangement; the constraint identified is an incentive, not a cost.
P5	Workflow redesign	Nascent or experimental	–	P5.3, P5.7	Isolated experiment; no repetition of any changed sequence.
P5	Skills and absorptive capacity	Nascent or experimental	–	P5.8	No learning beyond his own use is evidenced.
P5	Governance	Localised but recurring	P5.5, P5.6	P5.9	Consistent personal rule; no institutional arrangement.
P5	Data readiness	Nascent or experimental	–	P5.1, P5.2, P5.4	Conditions recognised, nothing enacted.

Note. A dash records that no passage of that kind exists in the register. Where the supporting column is empty the judgement rests on limiting evidence alone, which is stated rather than concealed.

D3. Basis for each judgement: the administrative context

Table D3. Evidence for and against every context-level judgement in the administrative context, with the distinction on which each turns.

Informant	Construct	Judgement	Supporting evidence	Limiting or contrary evidence	Distinction on which the judgement turns
S6	Workflow redesign	Nascent or experimental	S6.2	S6.1	One changed practice described without evidence of repetition, against one failure.
S6	Skills and absorptive capacity	Nascent or experimental	–	S6.4	No competence development beyond the individual is evidenced.
S6	Governance	Nascent or experimental	–	S6.3	Awareness without application.
S6	Data readiness	Not examined	–	–	Substantiated by Annex C2.
S4	Workflow redesign	Not examined	–	–	Substantiated by Annex C2.

Informant	Construct	Judgement	Supporting evidence	Limiting or contrary evidence	Distinction on which the judgement turns
S4	Skills and absorptive capacity	Not examined	–	–	Substantiated by Annex C2.
S4	Governance	Nascent or experimental	S4.1	–	One instance; no evidence of repetition or of an internal arrangement.
S4	Data readiness	Not examined	–	–	Substantiated by Annex C2.
S8	Workflow redesign	Nascent or experimental	S8.1	–	One reallocated step; the passage does not establish that it recurs.
S8	Skills and absorptive capacity	Not examined	–	–	Substantiated by Annex C2.
S8	Governance	Not examined	–	–	Substantiated by Annex C2.
S8	Data readiness	Not examined	–	–	Substantiated by Annex C2.
S7	Workflow redesign	Localised but recurring	S7.1	–	Repetition evidenced – 'sempre que me pedem informação' – and held by one person.
S7	Skills and absorptive capacity	Not examined	–	–	Substantiated by Annex C2.
S7	Governance	Nascent or experimental	–	S7.3	Personal caution without a reference rule.
S7	Data readiness	Nascent or experimental	–	S7.2	Condition recognised, unresolved, no further action.
S5	Workflow redesign	Localised but recurring	S5.2	–	Repetition evidenced at the individual level.
S5	Skills and absorptive capacity	Not examined	–	–	Substantiated by Annex C2.
S5	Governance	Not examined	–	–	Recorded so the absence is visible.
S5	Data readiness	Nascent or experimental	–	S5.1	A permission condition recognised and unresolved.
S2	Workflow redesign	Examined, no evidence found	–	S2.4	The construct was reached and nothing was found: the participant states that neither his processes nor his organisation of data changed. This is breadth of use without depth in 1 person. The tenant export establishes that he was already the heaviest user among the 8 participants in the period before the session, and cannot corroborate the account given in it.

Informant	Construct	Judgement	Supporting evidence	Limiting or contrary evidence	Distinction on which the judgement turns
S2	Skills and absorptive capacity	Localised but recurring	S2.2	–	Competence evidenced; the consultation is an intention formed in the session.
S2	Governance	Localised but recurring	S2.1, S2.3	–	Formal provision and an applied reading of a document; both are decisions and rules, not enacted practice elsewhere.
S2	Data readiness	Examined, no evidence found	–	S2.4	Same passage as workflow redesign; the construct was reached and nothing was found.
S1	Workflow redesign	Not examined	–	–	Recorded so the absence is visible.
S1	Skills and absorptive capacity	Nascent or experimental	–	S1.1	A competence gap in relation to a capability that had been delivered in training [D6.1, D6.2].
S1	Governance	Not examined	–	–	Substantiated by Annex C2.
S1	Data readiness	Not examined	–	–	Substantiated by Annex C2.
S3	Workflow redesign	Localised but recurring	S3.1, S3.2	–	Repetition evidenced – 'praticamente a todas as horas passo algo pelo Copilot' – and retained by one person.
S3	Skills and absorptive capacity	Examined, no evidence found	–	–	The construct was reached; nothing about current learning beyond the individual was found.
S3	Governance	Not examined	–	–	Recorded so the absence is visible.
S3	Data readiness	Not examined	–	–	Substantiated by Annex C2.
SG	Governance	Localised but recurring	SG.1, SG.2	–	A norm held in common is evidenced. It is not an arrangement: no agreed check, documented procedure, escalation route or allocated accountability is described anywhere in the session.

Note. The table is structured identically to Table D2 so that the 2 contexts may be read against each other. The difference in evidential density between them is real and is the consequence of the instrument: judgements recorded as not examined arise where the single group session did not reach that construct with that participant, as Annex C2 records question by question.

D4. Domain-level synthesis

Table D4. Synthesis of the context-level judgements to the level of each domain, with the analytical reasoning behind each.

Domain	Construct	Domain-level judgement	Analytical reasoning
Academic	Workflow redesign	Localised but recurring, with 2 instances reaching a second role	Recurring assisted practice is present in 4 of the 7 academic contexts and is held by its originator in every registered instance. The 2 instances in which a practice reaches a second role both travel toward students, where the professor holds authority, and not toward colleagues, where the professor does not. Nothing is documented for reuse, and no artefact has an assigned owner.
Academic	Skills and absorptive capacity	Localised but recurring; diffusion episodic and non-accumulating	Individual competence is high and unevenly distributed; organisational absorptive capacity is not evidenced. Two diffusion events are registered – peer-led sessions on one colleague's own initiative, and adoption motivated by colleagues' testimony – and neither left a reusable object or recurred. A sharing space created after an earlier training was never populated.
Academic	Governance	Localised but recurring, uniformly across the domain	Every academic context evidences a consistent personal rule and none evidences anything above it. The institutional artefacts state a standard without operationalising it, and the guidance in force is scoped to academic work rather than to the use of the tool in the running of the faculty.
Academic	Data readiness	Nascent in 3 contexts, localised in 4; no arrangement above the individual	Where information-dependent use occurs, it runs on personal curation substituted for organisational readiness: controlled source sets, topic folders, photographed notebooks, cloud projects. The substitution is precisely why the resulting artefacts remain individual. Against it stand an unscoped store surfacing a 13-year-old student file, a second institutional system the assistant cannot reach, and one account of codification suppressed by professional incentive rather than by cost.
Administrative	Workflow redesign	Localised but recurring in 2 participants' own routines	Two participants describe configured agents embedded in recurring administrative routines – expenditure authorisation and data queries – each built and held by one person. The group's most intensive user reports no change to processes or to the organisation of data at all. No practice crosses a service boundary anywhere in the session.
Administrative	Skills and absorptive capacity	Nascent; one demonstration event inside a facilitated session	The competence that surprised the room had been delivered in the training [D6.1, D6.2], so what the session evidences is not an information gap but the absence of an arrangement that converts exposure into practice. The stated intention to consult a colleague was formed inside the session and is not an established consultation.

Domain	Construct	Domain-level judgement	Analytical reasoning
Administrative	Governance	Localised but recurring as a universally held norm; no arrangement	Validation is unanimous as a norm and absent as an arrangement. The one written provision anyone in the room can describe addresses academic work rather than administrative work. One provisioning decision is on record, taken under uncertainty about how long the pilot would last.
Administrative	Data readiness	Nascent	One attempt to connect organisational information failed and was abandoned; one participant limits what he supplies, citing his inability to establish how far it may circulate; the most intensive user reports no change to how data is organised.

Note. This synthesis is what Table 1 in Chapter 4 condenses. It summarises across participants without becoming a count: an account from a participant is evidence about that participant's work, never about a service or a department.

Annex E. Knowledge-intensive conditions and the coordination of the mechanisms

This annex supports Section 4.6. E1 sets out how each of the 4 knowledge-intensive conditions bears on each mechanism, on data readiness, and on the coordination among the mechanisms. E2 examines the mechanisms against each other directly, which is the complementarity question the discussion turns on.

One rule governs both tables. Two mechanisms appearing separately in the same interview are not coordinated: coordination is recorded only where the evidence connects them. The conditions add no mechanism of their own; they condition how the 3 mechanisms develop and whether they meet. Nothing in either table supports a causal claim that a condition produced an outcome.

E1. The 4 conditions against each mechanism

Table E1. How each knowledge-intensive condition bears on each mechanism, on data readiness, and on the coordination among the mechanisms.

Condition	Bearing on	Interpretation	Supporting evidence	Limiting or contrary evidence	Context in which it is evidenced
Professional autonomy	Workflow redesign	Autonomy was associated with each professional adapting the technology to their own work: every changed sequence in the corpus is designed by the person who performs it. The same autonomy coincided with the absence of any observed route by which a redesign becomes anyone else's.	P7.1, P3.1, S3.1, S7.1	P5.7, P6.5, S2.4	Academic and administrative contexts alike; the 2 artefacts that reach a second role travel toward students, where the professor holds authority.
Professional autonomy	Skills and absorptive capacity	Learning is self-directed and its direction is chosen individually; competence accumulates unevenly and no one is authorised to decide what others should learn.	P1.4, P7.9, P2.4	P5.8, P1.5, P2.16	Academic context; the administrative session shows the same pattern with a shorter history.
Professional autonomy	Governance	Rules are set by each professional for their own practice, and are stated as obligations rather than as preferences. Institutional guidance is treated as something that may be read rather than something that binds.	P6.1, P5.5, P4.4, SG.1	P1.8, P5.9, S6.3	Uniform across both contexts.
Professional autonomy	Coordination among the 3 mechanisms	Autonomy is consistent with the 3 mechanisms developing separately: a person may redesign their own sequence, learn what they choose to learn, and set their own rule, without any of these being connected to the others or to anyone else's.	P7.1 with P7.4 and P7.11	P1.4 with P1.5 and P1.6	The clearest instance is the informant with the most developed artefact, the strongest verification rule and no diffusion of either.
Professional autonomy	Data readiness	Each professional curates their own information rather than drawing on an organisational arrangement, which is workable individually and unusable by anyone else.	P1.9, P2.8, P7.7	P7.8, S7.2	Academic context; the administrative session evidences no curation at all.
Task ambiguity	Workflow redesign	Where outcomes are hard to specify, it is hard to say which redesign is worth generalising. Two configured artefacts were judged by their own builders to have fallen short, in both cases against an expectation the builder had not written down.	–	P5.3, P3.12, P1.3	Academic context; all 3 instances concern student-facing artefacts.

Condition	Bearing on	Interpretation	Supporting evidence	Limiting or contrary evidence	Context in which it is evidenced
Task ambiguity	Skills and absorptive capacity	Ambiguity about what good use looks like coincided with colleagues being unable to say what should be shared. Informal discussion is intense and its subject is the ethics of use rather than method.	P3.11	P7.11, P3.7	Academic context; a sharing space provided at no cost was never populated.
Task ambiguity	Governance	Rules cannot be written for cases nobody can specify in advance; one informant argues that guidance cannot yet be clear, and the institutional artefact stops at a principle.	–	P7.5, P1.7, D2.1	Academic context and institutional documents.
Task ambiguity	Coordination among the 3 mechanisms	Not examined.	–	–	No passage in the corpus addresses how ambiguity affects the relation between the mechanisms; the instruments did not ask.
Task ambiguity	Data readiness	Where it is unclear what a body of information is for, it is unclear what should be codified; the one informant who curates most declined to write the curation down.	–	P3.13, P4.8	Academic context.
Tacit expertise	Workflow redesign	Work that rests on judgement was not reallocated: the tasks informants refuse are refused on grounds of what the judgement is for, not on grounds of cost or risk of error.	P3.3, P3.4	P6.5, P5.5, P5.6	Academic context; the boundary is drawn inside teaching and assessment.
Tacit expertise	Skills and absorptive capacity	What advanced users know is difficult to state in a transferable form; demonstration was observed to transmit motivation rather than method.	P3.11	P7.11, P2.14	Academic context.
Tacit expertise	Governance	The most articulated rule set in the corpus is a professional standing rule, held individually and never written: the system is used as a junior assistant and never as a peer, and every legal claim is checked at source.	P6.1, P6.2, P3.9, P4.3	–	Academic context, normative discipline band.
Tacit expertise	Coordination among the 3 mechanisms	Tacit expertise is consistent with a personal rule and a personal practice reinforcing each other while neither becomes available to anyone else: the rule is inseparable from the judgement that produced it.	P6.1 with P6.2 and P6.6	–	Academic context; the reinforcement is real and entirely internal to one person.

Condition	Bearing on	Interpretation	Supporting evidence	Limiting or contrary evidence	Context in which it is evidenced
Tacit expertise	Data readiness	Externalisation happens, selectively and at personal cost, and was not observed where professional incentives oppose codification – a limit distinct from cost, contestation or loss of context.	P3.5, P3.6	P6.3, P3.13, P2.9	Academic context. This relation rests on 2 streams and must carry that qualification.
Decentralised coordination	Workflow redesign	Nothing connects one person's changed sequence to anyone else's work; redesign was observed to stop at the boundary of the person who performed it, in both the academic and the administrative context.	–	S2.4, P5.7, P1.3	Both contexts.
Decentralised coordination	Skills and absorptive capacity	Diffusion here rests on mutual adjustment, and coincided with episodes rather than accumulation: sessions arranged on one colleague's own initiative, a demonstration inside a facilitated meeting, a space nobody filled.	P2.13, S2.2	P2.14, P3.7, P1.6	Both contexts.
Decentralised coordination	Governance	Guidance travels by message and is not received; the norm that does hold is the one each person applies to themselves.	SG.1	P1.8, P5.9, S6.3, D4.5	Both contexts, plus documentary evidence from a parallel pilot.
Decentralised coordination	Coordination among the 3 mechanisms	The 3 mechanisms develop in parallel and do not meet: there is no occasion at which a redesign, a rule and a learning arrangement are decided together, and the one body holding a formal mandate has not executed its own project.	–	P2.10, P3.7, D3.1	Academic context; D3.1 evidences the mandate and its composition, not any activity. The status of the committee's project rests on a single informant.
Decentralised coordination	Data readiness	No one is positioned to decide what organisational information should be made available; each person solves the problem for themselves or abandons it.	–	S7.2, S5.1, P5.2	Both contexts.

Note. The context column is load-bearing. Several relations are evidenced in the academic context alone, and 2 of the 20 are recorded as not examined because no passage in the corpus addresses them and the instruments did not ask. Both are recorded so that the absence is visible rather than inferred. The relation between tacit expertise and data readiness rests on 2 evidence streams and carries that qualification wherever it is reported.

E2. Coordination among the mechanisms

Table E2. Each mechanism against each other mechanism, and data readiness against each, with the coordination status of every relation.

Relation	Coordination status	Interpretation	Supporting evidence	Limiting or contrary evidence	Analytical memo
Workflow redesign with skills and absorptive capacity	Absent	No instance in the corpus shows a changed process creating a setting in which others learn. The 2 practices that reach a second role reach students, who consume an output rather than acquire a practice.	P7.1, P3.1	P3.7, P2.16, P1.6	Two mechanisms appearing separately in the same interview are not coordinated. Raquel's agent and Raquel's account of learning are recorded together and are not connected by any evidence.
Workflow redesign with governance	Informal and localised	Where a person redesigns a sequence they also set the rule that governs it, and the rule stays with them. Configuring an agent coincided with explicit decisions about what it may reach – the one place in the corpus where a redesign was accompanied by a governance decision.	SG.2, P7.1 with P7.4	D2.1, D5.1	The coupling is real and internal to the individual. No institutional rule attaches to any redesigned process.
Skills and absorptive capacity with governance	Absent	Competence and rules develop separately. Nobody is trained in the rules and no rule specifies what competence is required; the guidance in force addresses academic work rather than the use of the tool in running the faculty.	–	S2.3, D5.1, P2.5	The training material covers capability and not authorisation [D6.1].
All 3 mechanisms together	Not evidenced in any context	No occasion is described at which a change to how work is done, the competence to do it and the rule governing it were decided together.	–	P2.10, P3.7	The body with the formal mandate to convene such an occasion has a project that has been designed, funded and unexecuted for about 2 years. Single-source for the project's status.
Data readiness, bearing on workflow redesign	Enabling where personal, constraining where organisational	Configured, information-dependent practice exists only where the individual has substituted personal curation for an organisational arrangement. That substitution is consistent with the resulting artefact remaining individual.	P7.1 with P7.7, P2.2 with P2.8	S7.2, P3.10, P5.4	Data readiness constrains the scope of embedding rather than the existence of use: every informant uses the technology.

Relation	Coordination status	Interpretation	Supporting evidence	Limiting or contrary evidence	Analytical memo
Data readiness, bearing on skills and absorptive capacity	Constraining	What one person has curated was not learned from by others; the curation is not written down.	–	P3.13, P6.3	Rests on 2 streams; carries that qualification.
Data readiness, bearing on governance	Formally intended, not enacted	The written provision that exists addresses what may be uploaded and how sources must be handled; no allocation of responsibility for organisational information accompanies it.	D5.2, D5.3	S5.1, S7.3, P5.2	Two participants limit what they supply because they cannot establish how far it may circulate – a permission question nobody in the corpus is able to answer.
Data readiness, bearing on coordination among the mechanisms	Not examined	The instruments did not ask how information conditions affect the relation between the mechanisms, and no passage addresses it.	–	–	Recorded so the absence is visible rather than inferred.

Note. This is the material behind the claim that the mechanisms cohered within individuals rather than across them. The single coupling that is real, between workflow redesign and governance, is retained rather than explained away: where a person redesigns a sequence they also set the rule that governs it. That coupling is internal to the individual, and no institutional rule attaches to any redesigned process.

Annex F. Evidence behind the capability indicators

Table 2 in Chapter 4 reports the judgement, the principal basis and the principal limit for each of the 4 indicators. This annex gives the evidence identifiers behind each, and separates 2 things Table 2 cannot show: evidence that bears on the indicator, and evidence that is deliberately not entered as indicator evidence at all. The 4 indicators are independent and are judged separately, so that partial and uneven formation can be reported as partial and uneven rather than averaged into a single verdict.

F1. Enabling, limiting and contradictory evidence

Table F1. Enabling, limiting and contradictory evidence for each of the 4 capability indicators.

Indicator	Enabling evidence	Limiting evidence	Contradictory evidence
(i) Recurring-process embedding	P7.1, P7.2, P3.1, P3.2, P3.3, P3.4, P1.1, P1.2, P2.1, P2.2, S3.1, S3.2, S7.1, S5.2	P5.7 (sustained use, no structural change), P6.5 (practice not mass-producible), S2.4 (the group's most intensive user reports no change to processes), P1.3 (designed, never built), P3.12, P2.12	None
(ii) Shared verification	SG.1 (unanimous, collective), P6.1, P6.2, P5.5, P7.4, P4.5, P1.9, SG.2	No arrangement anywhere in the case. D2.1 states a standard without operationalising it; D5.1 scopes the written provision to academic work.	None
(iii) Distribution beyond originating users or units	P2.13, P2.14 (peer-led sessions, one colleague's own initiative), P3.11 (adoption motivated by colleagues' testimony), S2.2 (demonstration and a stated intention to consult)	P3.7 (sharing space created, never populated), P1.5, P1.6 (no culture, no facilitators), P5.8, P4.7, P6.6, P2.16 (formal mechanism absent), P7.11 (informal traffic is about ethics, not method)	The 3 diffusion episodes run against the proposition that diffusion is absent and are recorded as partial falsifiers in Annex G2.
(iv) Retention beyond originating individuals	P3.6 (manuals written explicitly for future reuse, including by their author), P3.13 (allowed and excluded sources curated into an agent – with its own limitation: the curation was passed on orally and not written down)	P7.8 (no reusable base), P2.10 (committee project designed, funded and unexecuted for about 2 years – single-source), P6.3 (codification structurally disincentivised), S6.4 (uncertain horizon suppresses investment)	None

Note. Contradictory evidence is recorded separately from limiting evidence. For indicator (iii) the 3 diffusion episodes run against the proposition that diffusion is absent and are retained as partial falsifiers rather than absorbed, as Annex G2 records.

F2. Evidence held outside the indicator judgements

A fourth body of evidence is held deliberately outside these columns. The tenant analytics show contact with the tool to be broad and highly concentrated, but a distribution of contact is not embedding and is not entered as indicator evidence. The student questionnaire reports verification behaviour at high rates, but student behaviour is not an organisational arrangement. The external accounts show what an arrangement looks like where one exists, and are never used as evidence about FEP. Each of these bounds the case's evidence rather than supplying a standard the case fails to meet.

Two boundaries govern the reading of the whole table. The strongest permissible conclusion is that aspects of capability formation may be nascent, localised, uneven or formally intended; capability here is a proximate organisational ability, not realised value or appropriable advantage. Indicator (iv) must not be upgraded on the basis of absence: the counterfactual on which retention turns was not put in the focus group or to any of the 7 professors, so what the case observes is single-person dependency, a property of the artefacts, and not a demonstration that retention fails.

Annex G. Testing the account

Two procedures were used to test the account rather than to illustrate it, and this annex reports both. G1 states the 12 propositions with the outcome of each and the qualification that travels with it. G2 gives the condition that would have rejected each proposition, set before the result was considered, and the negative-case search actually performed. G3 shows which evidence streams bear on each proposition and in what relation. G4 reports the comparison with a rival account at the 7 points where the 2 imply different observations.

Neither procedure is a prospective test. All 12 propositions were formulated during analysis, after fieldwork was complete, and both accounts in the rival comparison were constructed by the researcher in the same period as the tests. Because the same corpus informed both the development and the assessment of the propositions, the procedure provides no independent confirmation; its purpose is to challenge and delimit the emerging account. What it establishes is that the account survived a deliberate attempt to break it. It establishes neither causation nor generality.

G1. The 12 propositions and their outcomes

Table G1. The 12 propositions, the question each bears on, the outcome of testing it, and the qualification that travels with the claim.

No.	Proposition as stated	Question it bears on	Outcome	Qualification that travels with the claim
P1	Task-level gains are real and are largely reabsorbed or converted into output rather than released as time.	Primary question – conditions that enable or constrain	Narrowed	Two contrary records stand, both concerning bounded tasks in which verification is cheap. Gains are released where verification is cheap and reabsorbed where it is not.

No.	Proposition as stated	Question it bears on	Outcome	Qualification that travels with the claim
P2	Verification is universal and rigorous and entirely personal – a norm rather than an organisational arrangement.	Sub-question 1 – governance	Stood as formulated	The absence of an arrangement is evidenced positively within the case. The external accounts show what such an arrangement looks like where one exists and are never used as evidence about FEP.
P3	Individual competence is high and no arrangement converts it into shared practice.	Sub-question 1 – skills and absorptive capacity	Narrowed	Three diffusion episodes stand as partial falsifiers and are retained. Each is episodic, leaves no reusable object and did not recur.
P4	Data readiness binds the information-dependent forms of embedding rather than use in general.	Sub-question 1 – data readiness as a precondition	Narrowed	Readiness was associated with the scope of embedding rather than its existence, since every informant uses the technology. Two informants achieve configured use by substituting personal curation for an organisational arrangement, which may help interpret why the resulting artefacts remain individual.
P5	Externalisation occurs selectively and individually and was not observed where professional incentives oppose codification.	Sub-question 2 – tacit expertise	Stood, with a stated evidential limit	Two evidence streams only. The administrative context is silent because the tacit-knowledge question was not put in the session.
P6	Configured use reaches further than open use where the configuration meets a specification set in advance and the sources are scoped.	Sub-question 1 – workflow redesign	Inconclusive	Reported as inconclusive rather than narrowed after the fact. The specification criteria were set independently of the outcome, and 2 of the 4 instances can be classified only through the builder's or the users' judgement of the output, which is the judgement under examination. An unsuccessful attempt is not evidence of incompetent configuration.
P7	Governance exists as personal norm and as documents that are not read.	Sub-question 1 – governance	Narrowed	Holds for institutional governance. One participant has read the guidance and reports that it addresses academic work rather than the running of the faculty, which the document itself corroborates; a second instance applies an external accreditation body's rule rather than an institutional one.
P8	Time and the licence horizon are conditions of investment rather than complaints.	Primary question – conditions that enable or constrain	Narrowed	Inductive, and the one proposition with no anchor in the framework. Two informants invested substantially during the pilot, but both had already decided that the technology mattered, so the effect is restated as operating on those who had not.
P9	Participants rank competence first; specialists observing many organisations rank process, leadership and problem definition first.	Sub-question 2 – task ambiguity and decentralised coordination	Narrowed, and entirely prospective	Entirely prospective: every source evidences what participants think matters and none evidences a current state. Both counter-instances stand. The ranking instrument offered 4 options and time was not among them.

No.	Proposition as stated	Question it bears on	Outcome	Qualification that travels with the claim
P10	Diffusion requires a dedicated, internal, authorised driver rather than a committee or volunteer champions.	Sub-question 3 – external comparison	Stood, as a prospective claim only	Single-source for the status of the committee's project. Where enforcement carried diffusion in the external comparison, that architecture predates the technology, so it is not an arrangement AI adoption produced.
P11	The 3 mechanisms develop in parallel and do not meet: no occasion is described at which a change to how work is done, the competence to do it, and the rule governing it were decided together.	Sub-question 2 – coordination among the mechanisms; Section 2.5 complementarity	Narrowed	One coupling is real and is retained rather than explained away: where a person redesigns a sequence they also set the rule that governs it. That coupling is internal to the individual, so the claim is narrowed to coordination at the organisational level.
P12	Where a redesigned practice reaches a second party, it travels toward those over whom the professional holds authority, and not laterally toward peers.	Sub-question 2 – professional autonomy and decentralised coordination; Sub-question 1 – workflow redesign	Narrowed	Four contrary instances stand and are retained. Each transmits competence rather than a redesigned practice.

Note. Three propositions stood as formulated, each carrying a stated limit on how far it reaches; 8 were narrowed, in several cases substantially; and 1 was reported as inconclusive. Propositions 11 and 12 were formulated later than the rest, on 10 August 2026, from findings the analysis had already recorded but had never tested. None of the 12 is a hypothesis registered in advance, and a proposition that survives is supported for this case, at this stage, on this corpus.

G2. Rejection conditions and negative-case searches

Table G2. The condition that would have rejected each proposition, stated before the result was considered, and the negative-case search performed against it.

No.	What would have rejected the proposition	Negative-case search performed	Result
P1	An informant reports a net time saving that was released to other work.	Register searched on workflow redesign and on every passage reporting a time effect; the 7 professor transcripts and the focus-group transcript searched for statements of time released.	Qualified
P2	An agreed, documented or allocated verification procedure exists that binds more than one person.	Register searched on governance and on the shared-verification indicator; both transcripts and all institutional documents searched for an escalation route, a documented check or an allocated accountability.	Supported
P3	Someone adopted a practice they learned from another person, or a sharing mechanism functioned.	Register searched on skills and absorptive capacity and on the distribution indicator; both transcripts searched for adoption from a colleague, for reference persons, and for any functioning sharing space.	Qualified

No.	What would have rejected the proposition	Negative-case search performed	Result
P4	Someone is prevented from using the technology at all by information conditions, or someone achieves information-dependent embedding despite poor information conditions.	Register searched on data readiness; both transcripts searched for non-use attributed to information conditions.	Qualified
P5	Systematic institutional externalisation exists, or the blocked instance is explained by cost rather than by incentive.	Register searched on data readiness and on tacit expertise; the focus-group transcript searched for the tacit-knowledge probe, which was not put.	Supported, with a stated evidential limit
P6	A configured artefact that met the specification and had scoped sources nevertheless failed, or open use produced comparable embedding.	Register searched on the configured-use mode and on every configured artefact in the corpus; each instance assessed against the specification criteria defined below before its outcome was considered.	Inconclusive
P7	An informant applies written institutional guidance to a decision.	Register searched on governance; both transcripts searched for any application of a written institutional rule.	Qualified
P8	Investment behaviour is unrelated to the horizon, or someone invested heavily despite it.	Register searched on skills and absorptive capacity and on every passage mentioning the pilot's duration; both transcripts searched.	Qualified
P9	A within-case informant ranks process or leadership first, or a specialist ranks training first.	All Q11-equivalent responses and the priority-card result reviewed.	Qualified
P10	A committee or a champion arrangement produced diffusion in the case or in the external comparison.	Register searched on skills and absorptive capacity and on the distribution indicator; the external accounts searched for any champion arrangement that worked.	Supported as a prospective claim only
P11	An occasion at which a change to how work is done, the competence to do it and the rule governing it were decided together; or an instance in which the development of one mechanism created the conditions for another's.	Every relation on Annex E1 rebuilt one by one; the register searched on workflow redesign, skills and absorptive capacity and governance for any passage connecting 2 mechanisms; both transcripts searched for any occasion – a meeting, a committee, a project, a training session – at which more than one was decided.	Qualified
P12	A redesigned practice adopted by a colleague, or a configured artefact used by a peer rather than by students.	Register searched on workflow redesign and on the distribution indicator; every configured artefact in the corpus traced to the party that uses it; both transcripts searched for any instance of a colleague taking up another's practice or artefact, as distinct from learning from them.	Qualified

Note. A framework analysis that collects only confirming instances proves nothing, and this is the study's principal credibility procedure. Where the evidence could not settle a proposition it is recorded as inconclusive rather than narrowed until it could no longer be contradicted. That rule accounts for the single inconclusive result: for proposition 6, 2 of the 4 relevant instances could be classified only by relying on the judgement whose adequacy was itself under examination.

G3. The relation each evidence stream bears to each proposition

Table G3. The relation that each evidence stream bears to each proposition.

No.	Professor interviews	Administrative focus group	Institutional documents	Tenant analytics	Student questionnaire	External comparison
P1	Supports and contradicts	Contradicts	–	Not applicable	Contextually consistent	External comparison
P2	Supports	Supports	Supports	–	Contextually consistent	External comparison
P3	Supports and contradicts	Contradicts	Supports	Contextually consistent	Not applicable	External comparison
P4	Supports	Supports	Contextually consistent	–	Contextually consistent	External comparison
P5	Supports	Silent	Contextually consistent	–	–	External comparison
P6	Supports and contradicts	Supports	–	–	Not applicable	External comparison
P7	Supports	Supports	Supports	–	Contextually consistent	External comparison
P8	Contradicts	Supports	Contextually consistent	–	–	External comparison
P9	Supports	Supports	Contextually consistent	–	Contextually consistent	External comparison
P10	Supports	Silent	Contextually consistent	–	–	External comparison
P11	Supports and contradicts	Contradicts	Supports	Not applicable	Not applicable	External comparison
P12	Supports and contradicts	Supports and contradicts	Silent	–	–	External comparison

Note. The columns record a relation. Streams are not interchangeable and are never counted: a proposition supported by 1 stream and silent in 3 is not the same object as one supported by 4. Specialists cannot corroborate conditions at FEP, and the student questionnaire cannot establish anything organisational, so both are recorded as external comparison or contextual consistency and never as within-case support. Contextual consistency means a source is compatible with a proposition without being able to establish it. Silence is recorded where an instrument did not reach the topic and is never read as agreement; a dash records that the stream carries no relation to that proposition; and 'not applicable' records a stream that cannot bear on the proposition by its nature, as the tenant analytics cannot speak to how mechanisms relate. Where a stream both supports and contradicts, the contrary evidence is retained as a negative case and is reported in Table G2.

G4. The rival-pattern comparison

The framework-derived account holds that a knowledge-intensive organisation in early adoption, without complementary investment, will show recurring practice at individual or local level, verification universal as a norm and absent as an arrangement, diffusion resting on individuals, no artefact surviving its originator, and workflow redesign nascent and least

salient to participants. The rival holds that resource scarcity and an uncertain pilot horizon alone would produce the same picture in any small organisation: low aggregate use, high concentration, no shared infrastructure, no redesign, weak governance, and diffusion resting on volunteers.

The 2 accounts converge on the surface evidence, so agreement with the framework proves nothing on its own. Only the points at which they imply different observations discriminate, and 7 were identified.

Table G4. The 7 points at which the framework-derived account and the rival imply different observations, with the result of each comparison.

No.	Point of divergence	What the framework-derived account expects	What the rival expects	Result
D1	The grounds on which the boundary of appropriate use is drawn	The boundary tracks the verifiability of professional judgement and disciplinary norms; professionals refuse uses they could afford.	The boundary tracks the cost and the risk of error.	Separates; favours the framework-derived account
D2	Which condition participants rank first	Competence, because complementary arrangements are not visible from inside a task.	Time, money, continuity.	Does not separate the accounts
D3	Externalisation behaviour	Pursued at personal cost where the professional sees a return, and blocked where professional incentives oppose codification.	Does not occur, for want of time.	Separates; favours the framework-derived account
D4	Response to a sharing mechanism provided at no cost	Unused: no criterion exists for what merits sharing, and no one owns it.	No sharing occurs because no mechanism exists.	Separates; favours the framework-derived account
D5	Whether the most capable practitioners embedded most (see P12)	No: autonomy provides no route from individual excellence to shared practice.	Yes: the most capable and best resourced embed most.	Separates; favours the framework-derived account
D6	Settings of comparable scale without knowledge-intensive properties	Embed further under comparable resource constraints, because management can specify the process.	The same constraints produce the same outcome.	Separates; favours the framework-derived account
D7	What the embedded endpoint required where one was reached	Coordination and authority.	Resources.	Separates; favours the framework-derived account

Note. Six of the 7 comparisons separate the 2 accounts and all 6 favour the framework-derived expectation. Two of those 6 rest principally on external comparison and therefore describe other organisations rather than FEP. The 7th is reported as failing rather than resolved in the framework's favour: the ranking exercise that generated it offered 4 options and time was not among them, so the instrument could not register the rival's own candidate condition and the result cannot be read

as evidence against it. Time is named as the principal obstacle elsewhere by several informants and by the parallel pilot's report, which is consistent with the rival.

Taken together the comparisons support a 2-part reading. Resource scarcity and an uncertain licence horizon are consistent with the severity and the timing of what was observed: how little use there is, how sharply it concentrates, how quickly it lapses, and why those who had not already decided that the technology mattered declined to invest in it. The knowledge-intensive conditions are consistent with its shape: which tasks are refused and on what grounds, why a sharing mechanism provided at no cost went unused, and why individual excellence did not propagate laterally.

Annex H. The external comparison

Nine specialists working with enterprise generative AI outside the faculty contributed 70 registered passages. This annex arranges those accounts by the position each occupies in an adoption trajectory and records the standpoint from which each was given. Its purpose is to show what an arrangement looks like where it exists, sharpening the reading of what is absent at FEP.

Nothing in this annex is evidence about FEP. These organisations were described retrospectively by one informant each and were not observed. Pathway positions are not embedding judgements and are not comparable with the within-case categories used in Annex D1.

H1. Specialist accounts by position in the adoption trajectory

Table H1. Specialist accounts arranged by the position each occupies in an adoption trajectory, by construct.

Position	What the position means	Workflow redesign	Skills and absorptive capacity	Governance	Data readiness
PC · enabling condition	Named as a precondition for the technology to reach organisational work	E9.3, E9.6	E1.5, E2.4, E3.1, E3.2, E3.4, E5.1, E5.2, E6.5, E8.1, E8.4, E4.6	E3.5, E4.2, E7.4, E4.10, E1.7	E5.4, E4.11
ST · stall or failure mode	Named as the point at which adoption stops	E9.1, E9.2	E1.4, E2.1, E3.3, E7.3, E2.12, E6.9	E4.12	E1.3, E4.1, E4.3, E6.4, E9.5, E4.7
EI · emerging integration	Partial integration observed elsewhere	E6.2, E6.10, E2.7, E2.8	E6.3, E1.6	–	–

Position	What the position means	Workflow redesign	Skills and absorptive capacity	Governance	Data readiness
EE · embedded endpoint	Arrangements described as reaching organisational form	E1.1, E1.2, E2.2, E5.3, E7.1, E8.2, E8.3	E2.6, E8.6, E4.8, E4.9	E2.3, E4.4, E6.1, E2.9, E2.10, E2.11, E6.6, E6.7	E7.2, E9.4, E8.7, E9.7, E6.8

Note. Governance is the best-evidenced construct in the comparison, and the accounts separate into 3 kinds of arrangement the case has none of: verification with named reviewers and an approval threshold; auditability designed in, so that checking does not depend on anyone's recall; and evaluation attached to a role as a standing obligation. One qualification governs how far any of it travels: the enforced architecture in the strongest case predates the technology. What the comparison shows is AI inserted into governance that already existed, not governance that AI adoption produced.

H2. Standpoints and source criticism

Table H2. The standpoint from which each group of specialist accounts was given, and the source-criticism tags applied to them.

Position in the trajectory	Standpoints represented	Source criticism in play
Enabling condition	All 9 specialists contribute records at this position. Provider-enablers, mostly large organisations of mixed verifiability, predominate.	Vendor positionality, consultant incentive and initiative-owner advocacy.
Stall or failure mode	8 of the 9 specialists contribute records here, including all 3 internal adopters, whose accounts of where adoption stops carry the most weight.	Consultant incentive, retrospective reconstruction and initiative-owner advocacy. The organisation described in one retrospective account could not be attributed and is not named.
Emerging integration	3 informants evidence this position: a regulated-product developer, a small technology partner and a large software organisation.	Initiative-owner advocacy. This position is measured from platform data rather than from organisational observation, and its weight is qualified accordingly.
Embedded endpoint	8 of the 9 contribute records here. The clearest accounts come from a large distributed software organisation of high verifiability and from a small technology partner: settings in which management can specify the process, which academic work does not permit.	Initiative-owner advocacy, proprietary constraint, vendor positionality and consultant incentive. Every tag qualifies the weight an account receives and none excludes it.

Note. A standpoint records where an informant speaks from, and has 4 parts in the workbook: the organisation, whether the informant is an adopter or a provider-enabler, the scale of that organisation, and the verifiability of the work described. Verifiability governs transferability: the more verifiable the work, the less directly the arrangement travels toward academic work, where outcomes are ambiguous and authority is dispersed. 3 of the 9 specialists speak as internal adopters; the other 6 speak as providers, consultants or product owners, and a standpoint is recorded for every record. Source criticism records the interest or limitation attaching to an account. It qualifies the weight a record receives and never removes it from the evidence base. The 5 categories are vendor positionality, consultant incentive, initiative-owner advocacy, retrospective reconstruction and proprietary constraint, abbreviated in the workbook as VEN, CNS, OWN, RET and PRP-C.

The comparison also addresses the one indicator the within-case evidence cannot reach. The retention counterfactual was put to 2 external informants and neither reports it solved: in the most embedded case in the corpus, undocumented decision history and know-how remain

held in individuals; and in an organisation that designed for auditability and assigned maintenance to a named role, the informant still reports no consolidated method for institutional retention. The comparison therefore bounds the thin retention evidence in the case rather than supplying a standard the case fails to meet.

FACULDADE DE **ECONOMIA**