Adopting Without Governing

Artificial Intelligence in Slovak Enterprises
Governance, Measurement, and the Configurations of Practice
A Two-Wave Interview Study

Peter Štetka
Nora Grisáková

Scientific Monograph



2026

AI IMPACT SK

Funded by the European Union
NextGenerationEU

RECOVERY AND RESILIENCE PLAN

Adopting Without Governing: Artificial Intelligence in Slovak Enterprises

Peter Štetka & Nora Grisáková

Governance, Measurement, and the Configurations of Practice: A Two-Wave Interview Study

2026

Title: Adopting Without Governing: Artificial Intelligence in Slovak Enterprises

Subtitle: Governance, Measurement, and the Configurations of Practice: A Two-Wave Interview Study

Authors: Peter Štetka, Nora Grisáková

Edition

1st edition, 2026

Published by

Vydavateľstvo EKONÓM, Bratislava University of Economics and Business, Dolnozemská cesta 1, 852 35 Bratislava, Slovak Republic

Publication type

Scientific monograph

Reviewed by

doc. Ing. Jindra Peterková, Ph.D.

Ing. et Ing. Benita Beláňová, PhD.

Copyright: © 2026 Peter Štetka and Nora Grisáková.

License: This work is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/.

ISBN: 978-80-225-5354-4

Bratislava, 2026

University of Economics in Bratislava
Faculty of Business Management
Dolnozemská cesta 1
852 35 Bratislava, Slovak Repubic

ai-impact.sk

Funding

This work was supported by the European Union – NextGenerationEU through the Recovery and Resilience Plan for Slovakia under project No. 09I05-03-V02-00003/2025/VA (AI-ImpactSK); 100% share.

Data availability

The AI-ImpactSK interviewes dataset (Štetka et al., 2026) is openly available via Zenodo under a CC BY 4.0 licence: https://doi.org/10.5281/zenodo.21891565

Ethics

The research protocol was reviewed and approved by the Ethics Committee of the Bratislava University of Economics and Business (Reference: EKEUBA 5/2025).

Use of AI tools

Data preparation and statistical analyses were performed using standard scientific software. The authors are responsible for all data, analyses, interpretations and conclusions, and have reviewed and approved the final manuscript; any use of digital assistance was confined to routine support and does not affect the integrity or originality of the scientific content.

Abstract

Artificial intelligence moved from novelty to normal in the world's enterprises between 2023 and 2026, yet the organizational practices that should accompany responsible adoption — governing the technology and measuring its value — have lagged conspicuously behind. This monograph examines that lag in a Central-European, small-and-medium-enterprise (SME) setting that the dominant Western, large-firm literature has largely overlooked. Drawing on the **AI-ImpactSK** project, it analyzes **58 semi-structured interviews** with Slovak organizations collected in two waves (autumn 2025 and spring 2026) and funded by the European Union's NextGenerationEU Recovery and Resilience Plan. The analysis combines framework-based reflexive thematic analysis (eight categories, twenty-four themes, 629 coded segments) with a computational triangulation of 58 English-translated transcripts and a suite of association tests with effect sizes, bootstrap confidence intervals, and multiplicity control.

Three findings define the adoption regime. First, a **governance gap**: explicit AI governance appears in only 13.8% of interviews and regulation in 6.9%, with governance constituting just 1.9% of all coded content; the gap is systemic rather than confined to any firm subgroup, and governance surfaces as diffuse, informal vigilance rather than codified practice. Second, a **measurement paradox**: 65.5% of firms articulate concrete benefits from AI, but only 20.7% measure them, and just 10.5% of benefit-claimers measure their gains (McNemar $p < .001$) — adoption proceeds largely without evaluation. Third, **configurations of practice**: adoption is informal and bottom-up by default, information-and-communication-technology firms show modestly higher implementation formality, and four indicative adopter configurations emerge from a coding-independent typology. A methodological contribution accompanies these results: the study demonstrates AI-assisted framework coding triangulated with computational text analysis, and transparently reports a wave-level coding-depth artifact while isolating a robust, coding-independent signal — the sharp rise of agentic-AI discourse (mentioned by 37.9% of firms in 2025 and 69.0% in 2026). The monograph argues that Slovak — and likely wider CEE/SME — firms are *adopting without governing or measuring*: a legitimacy-driven, capability-thin pattern that leaves value undefended and risk unmanaged on the eve of the EU AI Act.

Keywords: artificial intelligence adoption; generative AI; agentic AI; AI governance; EU AI Act; benefits measurement; productivity paradox; small and medium-sized enterprises; Slovakia; Central and Eastern Europe; thematic analysis; computational text analysis.

Acknowledgements

This monograph rests on the work of many hands, and it is a pleasure to acknowledge them. Our first and greatest debt is to the AI-ImpactSK project team, whose collaboration made this study possible: for conceiving and fielding the enterprise survey, for the careful collection, cleaning, and documentation of the data, and for the many conversations that sharpened how we read it. Without their groundwork this book would have had nothing to stand on.

We are equally grateful to the enterprises and their representatives who gave their time to respond. Their willingness to describe, candidly, how they actually use artificial intelligence in their day-to-day work is what gives this study its evidentiary base, and we hope the findings repay that openness.

We thank our colleagues and home institutions for the intellectual environment in which this research took shape, and the reviewers whose questions and corrections improved the manuscript in ways large and small. Any errors that remain are, of course, our own.

Finally, and most personally, we thank our families. Their patience, encouragement, and quiet understanding through the long hours that a project like this demands were a gift we did not take for granted. This book is, in no small part, theirs as well.

Foreword

When we began this study, artificial intelligence had already crossed the line from novelty to fixture in the working lives of enterprises around us. What had not crossed that line, we suspected, were the quieter disciplines that ought to accompany a powerful new technology: governing it deliberately, and measuring what it is actually worth. This book grew out of that suspicion, and out of a conviction that the place to test it was not the well-documented world of large Western firms but the small and medium-sized enterprises of Slovakia and Central Europe, where most of the region's economic life is conducted and where the evidence base has been thinnest.

The window in which we worked — the autumn of 2025 to the spring of 2026 — turned out to be unusually informative. Generative tools were maturing into agentic ones, the EU AI Act was moving from text to obligation, and firms were making consequential choices faster than they were writing them down. Listening to fifty-eight Slovak organizations describe, in their own words, how they adopt and live with these tools, we heard a consistent and slightly unsettling pattern: enthusiastic, capable, often bottom-up adoption, accompanied by very little codified governance and almost no measurement of value. Firms were adopting without governing, and adopting without evaluating.

We have tried to report that pattern honestly, including where our own methods have limits. The reader will find, alongside the findings, a frank account of what our data can and cannot support — most pointedly a wave-level coding-depth artifact that we chose to surface rather than smooth over, and the robust, coding-independent signals we could still trust. We combined a structured, AI-assisted reading of the interviews with a computational analysis of the transcripts, not because either alone was sufficient, but because triangulating the two let us hold our interpretations to a higher standard. If this book has a methodological argument, it is that such transparency is not a confession of weakness but a condition of trust.

We wrote for several readers at once. For managers and practitioners, we hope the configurations of practice described here offer a recognizable mirror and a few proportionate next steps. For policymakers in Bratislava and Brussels, we hope the governance gap and the measurement paradox read as a call for support that is proportionate to the firms it is meant to reach. And for our fellow researchers, we offer both a Central-European, SME-centered complement to the prevailing literature and a replicable template for studying adoption as it unfolds.

No single study settles a question this large, and we do not pretend otherwise. What we offer is a careful, situated reading of a particular moment in a particular place — a moment we believe is widely shared. We invite the reader to take it in that spirit, and to carry the questions it raises further than we have been able to.

Peter Štetka & Nora Grisáková
Bratislava, 2026

Table of Contents

Abstract 5
Acknowledgements 6
Foreword 7
1 Introduction 11
1.1 An adoption without a ledger 11
1.2 The 2025–2026 inflection 12
1.3 Why Slovakia, and why a CEE, SME-heavy site 13
1.4 The AI-ImpactSK study 14
1.5 Aim and research questions 14
1.6 Contributions 15
1.7 Structure of the monograph 16
2 Context: Artificial Intelligence in the Slovak and Central-European Enterprise Landscape 17
2.1 The Slovak economy as an SME-heavy, manufacturing-and-ICT system 17
2.2 Digital maturity in the V4 and CEE region: a qualitative framing 18
2.3 The policy backdrop: the EU AI Act and its SME proportionality problem 19
2.4 Policy, funding, and the 2025–2026 observation window 20
2.5 A Slovak SME lens: empirical value and boundaries 20
3 Literature Review and Theoretical Framework 22
3.1 Theoretical foundations of technology adoption 23
3.2 AI governance, responsible AI, shadow AI, and regulation 26
3.3 AI and IT business value and the measurement problem 28
3.4 Generative and agentic AI and configurational approaches 30
3.5 Research gaps 32
3.6 Conceptual framework 34
4 Research Design and Methods 38
4.1 Research Philosophy and Overall Design 38
4.2 Sampling and Recruitment 39
4.3 The Interview Instrument 40
4.4 Data Preparation: Transcription, Translation, and Anonymization 41
4.5 Analytic Approach: Framework-Based Reflexive Thematic Analysis 41
4.6 AI-Assisted Coding and the Human-in-the-Loop 42
4.7 Computational Triangulation 43
4.8 Quantitative Procedures 43
4.9 Reflexivity and Methodological Transparency: The Coding-Depth Artifact 45

4.10 Rigor, Trustworthiness, and Ethics 46
5 The Landscape of AI Adoption: Descriptive Findings 47
5.1 Sample composition 47
5.2 AI technology types and use areas 49
5.3 The thematic structure: twenty-four themes 50
5.4 The category structure 52
5.5 Theme co-occurrence: the adoption-narrative core 53
5.6 Coded-segment totals and the coding-depth difference 55
5.7 The landscape in summary 56
6 The Governance Gap 57
6.1 Research question, hypotheses, and the shape of the argument 57
6.2 The low salience of codified governance (H3a) 58
6.3 Risk talk without a spine: Ethics and security against governance 59
6.4 No detected subgroup difference (H3b) 60
6.5 Governance as diffuse vigilance, not codified practice (H3c) 61
6.6 The texture of informal vigilance 62
6.7 Interpretation: Principles without practice, shadow without sanction 63
6.8 Robustness and the governance-salience thesis 65
6.9 Governance-related theme coverage at a glance 66
7 The Measurement Paradox 68
7.1 The question and the core results 68
7.2 The Breadth and Correlates of the Gap 71
7.3 Non-Measurement in the Respondents' Own Words 73
7.4 Interpretation: The Productivity Paradox and Measured GenAI Gains 74
7.5 Adoption Without Evaluation: Consequences and Synthesis 75
8 Configurations of Adoption 77
8.1 Research question and the informal default 77
8.2 Formality, technological breadth, and the ICT effect 78
8.3 Four indicative adopter configurations and sourcing choices 80
8.4 Interpretation: TOE, dynamic capabilities, and a configurational reading 83
9 The Temporal Shift, 2025 to 2026 85
9.1 Research question and the burden of methodological caution 85
9.2 The coding-depth artifact 85
9.3 Coding-independent signals that are robust 87
9.4 What is artifact, what is real, and what it means 92
9.5 The agentic turn and the governance tension 93

10 Discussion 96
10.1 The adoption–governance–evaluation triad 97
10.2 Theoretical contributions 98
10.3 Dialogue with the literature and the maturation trajectory 102
10.4 The early-adoption-stage rival explanation 103
10.5 Synthesis 105
11 Implications for Practice, Policy, and Method 106
11.1 Implications for managers and practitioners 106
11.2 Implications for policymakers (European and Slovak) 108
11.3 Implications for method and researchers 110
11.4 Synthesis 111
12 Limitations and Future Research 112
12.1 Limitations 112
12.2 Future research 118
Conclusion 121
12.3 Three findings, in brief 121
12.4 What the method contributes 122
12.5 The pattern, and its stakes 122
12.6 From use to capability 123
References 124
Appendix A. Thematic Codebook (8 categories, 24 themes) 132
Appendix B. Confirmatory Tests, Effect Sizes, and FDR 134
Appendix C. List of Figures 135

1 Introduction

Artificial intelligence is no longer entering firms as a distant technological promise, but as a routine element of organizational work. Yet the organizational disciplines that make a powerful technology accountable — governance, measurement, and deliberate integration — have not diffused at the same pace. This chapter introduces the empirical and conceptual puzzle at the heart of the monograph: in a two-wave study of Slovak enterprises, AI adoption appears energetic, bottom-up, and increasingly agentic, while codified governance and systematic value measurement remain thin. It situates this puzzle in the 2025–2026 inflection, explains why Slovakia offers a revealing Central-and-Eastern-European, SME-heavy setting, introduces the AI-ImpactSK study, and states the research questions, contributions, and structure of the book.

1.1 An adoption without a ledger

Two-thirds of the firms in this study could tell us, in concrete terms, what artificial intelligence had done for them. One in five could show us a number. That mismatch is where the book begins.

Across fifty-eight Slovak organizations interviewed in two waves between 2025 and 2026, 65.5% [52.7–76.4] articulated concrete benefits from their use of AI: time saved, output accelerated, drudgery offloaded, decisions better informed. Only 20.7% [12.3–32.8] could point to any measurement of those benefits. Among the firms that claimed a benefit, the conditional probability of having measured it was just 10.5% (Štetka et al., 2026). The distance between asserting value and demonstrating it is not a rounding error or a quirk of a few outliers. It is a structural feature of the data, robust under an exact McNemar test ($p = 6.9 \times 10^{-5}$), and it holds across firm sizes and sectors. These firms were, in the plainest terms, far more willing to believe in artificial intelligence than to count what it did for them.

The same firms had even less to say about governing the technology than about measuring it. Formal AI governance — a policy, an oversight arrangement, a codified rule for how the tools may be used — surfaced in only 13.8% [7.2–24.9] of organizations, and explicit awareness of regulation in just 6.9% [2.7–16.4]. In the thematic analysis underlying this study, the whole governance-and-regulation category accounted for 1.9% of all coded material, the smallest share of any analytical category by a wide margin. Two points of caution belong here from the outset. The interview protocol carried no standalone governance question; governance and regulation entered the record only when a respondent raised them unprompted, under the legal-and-ethical-challenges prompt or the forward-looking question about expected regulation. So these figures measure how often governance was foregrounded in firms' own accounts of adoption, not whether governance practices exist somewhere in the organization. And when governance did come up, it seldom took the form of a written standard or a named owner. It appeared instead as a diffuse, informal vigilance: a habit of checking outputs, a wariness about feeding sensitive data to a model. We characterize the resulting pattern as adoption running ahead of both governing and measuring, and that characterization, rather than a documented organizational fact, is what the working title compresses.

This is the puzzle the book sets out to document and explain. It is a puzzle because the two halves resist the standard narrative of cautious, deliberate technology adoption. The conventional account assumes that firms invest, then justify the investment with metrics, then build the rules and roles that keep it safe. What the interviews show is looser and stranger: a fast, bottom-up, enthusiastic embrace of capable tools, with little of the evaluative and governing

apparatus that management and information-systems scholarship has long treated as the connective tissue between adoption and value. This introduction situates that puzzle in its moment, explains why a small, SME-heavy Central and Eastern European economy is a revealing place to study it, introduces the study that gathered the evidence, and lays out the questions, contributions, and structure that organize the chapters to come.

1.2 The 2025–2026 inflection

The window in which these interviews were conducted is not incidental. The years 2025 and 2026 mark a genuine inflection in the diffusion of artificial intelligence into ordinary organizational life, and the two-wave design brackets that inflection with unusual precision. Two shifts deserve emphasis.

The first is the normalization of generative AI. What had been, only a few years earlier, a research curiosity or a consumer novelty became by 2025 a routine fixture of knowledge work. In the present sample, generative AI was the dominant modality by an overwhelming margin: 89.7% of the firms reported using it, far ahead of any other category of AI technology (Štetka et al., 2026). This tracks a broader scholarly recognition that generative systems form a distinct, general-purpose class of technology whose organizational reach is unusually wide and whose governance challenges existing IT practices do not fully anticipate (Feuerriegel et al., 2024; Banh & Strobel, 2023). Controlled studies across this period made the stakes concrete. They documented large, measurable productivity effects from generative tools in real and experimental workplaces: a roughly 15% increase in customer-support throughput (Brynjolfsson, Li, & Raymond, 2025), a 40% reduction in writing time alongside quality gains (Noy & Zhang, 2023), and a 55.8% acceleration of a defined programming task (Peng et al., 2023). Task-exposure analyses suggested the technology could touch a substantial fraction of the work of most of the workforce (Eloundou et al., 2023). Normalization is thus not simply more people using chatbots. It is the arrival of a general-purpose technology whose effects, where they have been rigorously measured, are real, sizable, and unevenly distributed across tasks and workers.

The second shift is the agentic turn. Over 2025 and into 2026, the frontier of practical interest moved from systems that generate content on request toward systems that take actions: tool-augmented language models that plan, retrieve, write code, and chain operations with growing autonomy (Wang et al., 2024; Liu et al., 2025). This monograph captures that turn in motion. Coding-independent text analysis shows that respondent mentions of agentic concepts — agents, autonomy, retrieval-augmented generation — rose sharply between the waves, from a mean of 0.72 in the first wave to 3.14 in the second (Mann–Whitney $p = .0022$, rank-biserial 0.44), and the share of interviews touching on agentic AI roughly doubled, from 37.9% to 69.0% (Štetka et al., 2026). This is one of the few cross-wave movements in the data that can be defended as a genuine shift rather than an artifact of how the two waves were coded, a distinction taken up below and handled with care throughout the book. The agentic turn matters because it widens what AI can do inside a firm and because it raises the governance stakes. As autonomy increases, so do the risks that come with ceding control to a system (Mitchell et al., 2025), and so does the relevance of the very governance and measurement practices these firms have largely not built. The turn also reopens long-standing questions of enterprise sourcing, whether to build, buy, or adapt, in newly unsettled terms (Klotz, 2026), a theme the firms here navigated through subscriptions, retrieval-augmented adaptation, and selective in-house construction (Lewis et al., 2020).

The core problem of the moment is one of pace and asymmetry. Adoption is racing ahead of the institutional machinery, governance and measurement, that is supposed to make adoption legible, accountable, and improvable. This is not a uniquely Slovak phenomenon. The principles-to-practice gap in AI ethics and governance recurs wherever it has been studied (Mittelstadt, 2019; Morley et al., 2020), and the difficulty of measuring the value of information technology has a thirty-year pedigree (Brynjolfsson, 1993; Brynjolfsson & Hitt, 2000). What 2025–2026 does is sharpen the asymmetry. The capability frontier is advancing faster, and into more hands, than at any prior moment, while the governing and measuring practices lag as they always have.

1.3 Why Slovakia, and why a CEE, SME-heavy site

If the problem is general, why study it in Slovakia? A small, open, SME-heavy economy in Central and Eastern Europe is an unusually revealing site for watching how AI adoption unfolds away from the technological frontier, and such sites are badly underrepresented in the empirical record.

The first reason is structural. Much of what is known about AI adoption derives from large firms, frontier economies, and survey instruments calibrated to organizations with the resources to formalize. Yet the modal European firm is small, and SMEs adopt technology differently: more informally, more bottom-up, more through accessible platforms than through strategic programs (Schwaeke et al., 2024; Wei & Pardo, 2022). Maturity models built for large enterprises map poorly onto them (Mittal et al., 2018). A study designed around SMEs observes the phenomenon where most economic actors live, and it observes the informality that the large-firm literature tends to design out. The dominant pattern in this study, adoption that is real but unformalized and enthusiastic but unevaluated, is what recent SME research has begun to identify as the new normal, and what these data document in unusual depth.

The second reason is regional. Central and Eastern Europe, and Slovakia specifically, remain thinly covered in the AI-adoption literature, especially in qualitative, theory-anchored work spanning more than a single moment in time. The handful of directly Slovak studies tend to be single-function or single-firm in scope (Dvorský, 2025; Marcineková, Sujová, & Ďurica, 2025), and the broader V4 evidence base, while growing (Jaciow et al., 2026; Horváth & Szabó, 2019), leaves a clear gap for in-depth, cross-sector, two-wave evidence from a transition economy. Slovakia is a useful case on its own terms: an EU member state with a manufacturing-and-services economy, operating under the same emerging European AI governance regime as every other member, but without the dense AI ecosystem of the largest economies. It offers a view of an ordinary economy meeting a powerful new technology, which is the encounter most of the world is now having.

The third reason is regulatory. The firms in this study operate under the European Union's Artificial Intelligence Act (Regulation (EU) 2024/1689), a horizontal, risk-tiered regime phased in from 2025 onward, with provisions meant in part to ease the burden on smaller actors (European Parliament & Council, 2024). The study is itself funded through the EU's NextGenerationEU instrument. That an EU-funded study of EU firms operating under a new EU AI law should find regulation foregrounded by fewer than one firm in fourteen is not a minor descriptive fact. It points to a readiness gap at the precise interface where European AI policy meets the small firms it most affects, though, given how regulation entered the interviews, it is a gap in salience as much as in practice. The Slovak case thus opens onto a question of continental significance: how far, and how fast, does an emerging governance regime actually diffuse to the shop floor of the ordinary economy?

Honesty about the site requires one prominent caveat, repeated wherever it matters. The sample is skewed toward information and communication technology and software firms, which make up 44.8% of the fifty-eight organizations. This is an early-adopter bias, a genuine limitation, and it inflates, relative to the wider economy, the apparent sophistication and breadth of AI use. The ICT skew is not hidden here; it is analyzed. Several findings turn on the contrast between ICT and non-ICT firms, and the more sophisticated, multi-modal, technically deep configurations of adoption are disproportionately populated by the ICT segment. Where the analysis compares groups or generalizes, it does so cautiously, foregrounding the skew rather than papering over it.

1.4 The AI-ImpactSK study

The evidence in this monograph comes from the AI-ImpactSK study, a two-wave qualitative investigation of AI adoption in Slovak organizations conducted by the authors at the Faculty of Business Management, Bratislava University of Economics and Business, with funding from the European Union–NextGenerationEU through the Recovery and Resilience Plan for Slovakia (project No. 09I05-03-V02-00003/2025/VA) and ethics approval EKEUBA 5/2025. The dataset comprises fifty-eight semi-structured computer-assisted telephone interviews, gathered in two waves: a first wave of twenty-nine interviews between August and October 2025, and a second wave of twenty-nine between April and June 2026. Interviews were conducted in Slovak; transcription, translation, and coding were AI-assisted and author-reviewed, a workflow whose methodological standing is discussed below. The transcripts were analyzed through framework thematic analysis in the tradition of Braun and Clarke (2006), yielding eight categories, twenty-four themes, and 629 coded segments (Štetka et al., 2026).

Two features of the design merit emphasis at the outset. First, the sample is purposive, not representative. With fifty-eight participants recruited through telephone fieldwork, and with a candidly documented collapse in response rate between the waves (from 22.2% to 6.4% on numbers dialed, the second wave drawing on a much larger dialing frame), the study makes no claim to statistically represent the population of Slovak firms. Its strength lies in the depth, contextual richness, and cross-case comparability of qualitative evidence gathered at a pivotal moment, comfortably exceeding conventional thresholds for thematic saturation (Guest, Bunce, & Johnson, 2006). Every inferential claim in this monograph is hedged against the N = 58 purposive caveat, and the reader should carry that caveat into each chapter. Second, the two-wave structure, spanning roughly eight months of an exceptionally fast-moving period, lets the study observe change, most notably the agentic turn, in something close to real time, a vantage point that single-snapshot studies cannot offer.

1.5 Aim and research questions

The aim of this monograph is to characterize and explain a distinctive regime of AI adoption, thin in governance, thin in measurement, and dependent on the configuration of the adopting firm, as it took shape in Slovak enterprises during the 2025–2026 inflection, and to draw out its theoretical, practical, and policy implications. The study is organized around four research questions, each anchoring a substantive strand of the analysis.

RQ1 (Configurations). How does AI adoption vary in form and structure across Slovak firms, and what shapes it? This question treats adoption not as a binary event but as a set of qualitatively distinct configurations, and asks what firm characteristics — sector, capability, technological breadth — pattern them.

RQ2 (The measurement paradox). Do firms that claim benefits from AI actually measure those benefits? This question targets the dissociation between articulated and measured value that opens this chapter, and asks how wide, how robust, and how conditioned it is.

RQ3 (The governance gap). How prevalent is explicit, codified AI governance in firms' accounts of their own adoption, and is its salience predicted by firm traits? This question examines whether governance is uniformly low-profile across firms or foregrounded more in some than in others.

RQ4 (The temporal shift). How did AI discourse and practice change from 2025 to 2026, net of how the data were coded? This question asks what genuinely shifted across the waves, and insists, as a matter of method, on separating real change from artifacts of analysis.

These four questions are not independent. They describe one phenomenon from four angles: a population of firms adopting a powerful general-purpose technology in varied configurations (RQ1), seldom measuring what it does (RQ2), seldom foregrounding how its use is governed (RQ3), and visibly maturing toward agentic uses over a short window (RQ4). The thesis the chapters build toward is that these are facets of a single adoption regime, one in which capability has run ahead of both evaluation and control.

1.6 Contributions

This monograph makes three substantive contributions and one methodological contribution.

The first substantive contribution concerns the governance gap. The study provides rare firm-level evidence on how little codified AI governance surfaces in the accounts of SMEs operating under an emerging European regulatory regime. It also characterizes what tends to stand in its place: not a complete absence of concern, but a diffuse, informal vigilance centred on verifying outputs and protecting data. The contribution is to move the long-noted principles-to-practice gap from the level of published guidelines to the level of ordinary firms' own accounts of what they actually do, and to name the informal vigilance that occupies the space where codified governance is not foregrounded.

The second substantive contribution is the measurement paradox, the flagship finding of the study. Firms articulate benefits from AI far more often than they measure them, even though measurement was probed directly through questions about goals, success, and tracked indicators. The contribution is to supply quantified field evidence of adoption without evaluation: firms perceive and claim AI-related benefits, but often do not instrument them. Read against the information-technology productivity paradox and benefits-realization literature, the finding suggests that the problem is not necessarily the absence of value, but the absence of disciplined evaluation.

The third substantive contribution concerns the configurations of adoption. Against any tidy story of uniform diffusion, the study finds informality to be the default mode of adoption and identifies four indicative adopter configurations that differ in technological breadth, implementation formality, measurement, governance, and sourcing. ICT firms show more formal and broader AI use than non-ICT firms, while the configurations themselves are treated as indicative rather than definitive. The contribution is to bring configurational and capability-theoretic reasoning to bear on AI adoption in a non-frontier economy, and to show that governance and measurement gaps are part of how adopter types differ from one another.

The methodological contribution runs alongside these three. The study demonstrates a workflow of AI-assisted framework coding triangulated with computational text analysis, using large language models to support deductive coding against an expert codebook, with human review throughout. This triangulation allows the study to defend the agentic turn as a real textual signal while declining to over-read other cross-wave movements. The monograph also reports a coding-depth artifact transparently and treats it as a reflexivity point rather than a problem to be hidden. Temporal claims are advanced only where coding-independent evidence supports them, and the governance and measurement findings are explicitly assessed against this artifact. In doing so, the monograph offers a practical contribution to standards for AI-assisted qualitative research.

1.7 Structure of the monograph

The book proceeds in four movements. The opening chapters establish context and foundations. Chapter 2 situates AI adoption in the Slovak and wider Central and Eastern European enterprise setting, sketching the economy, its digital maturity, and the EU AI Act and NextGenerationEU backdrop against which the study unfolds. Chapter 3 reviews the relevant literatures and builds the theoretical framework, drawing together adoption theory, AI governance and responsible-AI scholarship, the information-technology value and productivity-paradox tradition, and configurational and capability perspectives, and ending with the research gaps that motivate each strand. Chapter 4 sets out the research design and methods in full, including the two-wave sampling and its limitations, the framework thematic analysis, the AI-assisted coding workflow and its triangulation, the quantitative procedures, and an explicit treatment of the coding-depth artifact.

The second movement establishes the empirical landscape. Chapter 5 reports the descriptive results: sample composition, AI types and use areas, the theme structure, and the co-occurrence of themes, establishing the overall shape of adoption before the deeper analyses begin. The third movement comprises the three analytical deep dives that correspond to the substantive contributions: Chapter 6 on the governance gap (RQ3), Chapter 7 on the measurement paradox (RQ2), and Chapter 8 on the configurations of adoption (RQ1). Chapter 9 then takes up the temporal shift from 2025 to 2026 (RQ4), placing the coding-depth artifact front and center and separating what is artifact from what is real, including the agentic rise.

The final movement synthesizes and extends. Chapter 10 draws the three angles into a single thesis, a thin-governance, thin-measurement, configuration-dependent adoption regime, and develops its theoretical implications. Chapter 11 turns to implications for managers, for policy, and for method. Chapter 12 states the study's limitations and maps an agenda for future research. Chapter 13 concludes by returning to the problem, the findings, and the central message that gives the book its title: that across these fifty-eight Slovak enterprises, at a decisive moment in the diffusion of artificial intelligence, adoption outpaced both the governing and the measuring of it, and that the distance between what these firms were doing and what they could account for is the most important thing the evidence has to say.

2 Context: Artificial Intelligence in the Slovak and Central-European Enterprise Landscape

Every claim about how firms adopt artificial intelligence is, implicitly, a claim about a particular kind of firm in a particular place at a particular moment. The international evidence base has been assembled largely from large organizations, often Western-European, North-American, or East-Asian, and disproportionately from manufacturing or from firms that build AI rather than merely consume it (Oldemeyer et al., 2024; Schwaeke et al., 2024). The empirical site of this monograph differs on every one of those axes. It is a small, open, manufacturing-and-services economy on the eastern edge of the European single market. Its enterprise population is overwhelmingly small and medium-sized. And the firms observed here are, in the main, AI-*adopting* organizations that reach for general-purpose tools through subscriptions and platforms — not AI-*developing* laboratories with internal audit functions and model-governance committees. This chapter sets out that context. It characterizes the Slovak economy and its position within the Visegrád Four (V4) and the wider Central-and-Eastern-European (CEE) region; it frames the region's digital maturity qualitatively rather than through invented league-table figures; it describes the policy backdrop in which the study unfolds, including the European Union's Artificial Intelligence Act and the NextGenerationEU recovery instrument that funds the project; and it explains why the autumn-2025-to-spring-2026 observation window is well suited to watching the generative-to-agentic transition in ordinary firms. Throughout, the dataset's own composition is used to describe the empirical site honestly, including the biases that make it a partial, purposive, hypothesis-generating window rather than a representative census.

2.1 The Slovak economy as an SME-heavy, manufacturing-and-ICT system

Slovakia is a small, highly open economy whose enterprise base, like that of the European Union as a whole, is dominated numerically by small and medium-sized enterprises (SMEs). The overwhelming majority of Slovak firms are micro, small, or medium-sized; large enterprises, while economically weighty, are a small minority of the firm population. This structural fact matters for the whole argument of the book, because the frameworks most often used to study technology adoption were not designed for such firms. Maturity models, governance scaffolds, and benefit-realization instruments have been calibrated predominantly on large enterprises with dedicated information-technology functions, formal project-management offices, and the organizational slack to document, pilot, and evaluate (Mittal et al., 2018). Mittal et al. (2018) make the point directly in the Industry 4.0 context: most existing smart-manufacturing and digital-maturity models are ill-suited to SMEs, which differ from large firms not only in resources but in culture and in the very way technology enters and diffuses through the organization. The same caution applies with redoubled force to general-purpose AI, which arrives in the SME not through a procurement committee but through an individual employee's browser. A central interpretive commitment of this monograph, then, is to read "formality," "governance," and "measurement" in ways sensitive to firm size, rather than penalizing small firms for failing to behave like large ones.

Sectorally, Slovakia is frequently described as one of the most industrialized economies in the European Union. Its manufacturing core, anchored historically in automotive and supplier industries, gives the country a pronounced exposure to the technologies of advanced

manufacturing and, increasingly, to the analytics and automation layered on top of them. At the same time, a substantial and growing information-and-communication-technology (ICT) and software-services sector has emerged, concentrated in the larger urban centers and oriented toward both domestic and export markets. This manufacturing-plus-ICT duality is consequential for AI adoption, because it implies two rather different adoption logics operating in the same national economy. In manufacturing and the broader "ordinary economy," AI tends to enter as a discrete, function-specific augmentation: a chatbot in customer service, a drafting aid in administration, an analytics layer in risk or quality work. This echoes the way Industry 4.0 technologies were found to diffuse in identifiable, layered, sequenced patterns rather than uniformly (Frank et al., 2019). In the ICT and software sector, by contrast, AI sits closer to the core of the product and the production process itself, and firms there can be expected to integrate it more deeply and more technically. The empirical chapters of this book report exactly this fault line within the sample: ICT firms reach a degree of implementation formality and a breadth of AI use that the non-ICT firms do not — a difference that holds up under within-wave robustness checks (Štetka et al., 2026). The Slovak case thus offers, within a single small economy, a natural contrast between a technical-sector adoption pattern and a broad-economy one.

2.2 Digital maturity in the V4 and CEE region: a qualitative framing

It is tempting, when situating a country's digital position, to reach for a single composite index and a rank. This monograph deliberately refrains. The project's evidentiary discipline restricts quantitative claims to those grounded in its own dataset, and composite digital-maturity scores are a poor instrument for the qualitative, mechanism-level questions pursued here. What the regional literature supports is not a precise coordinate but a robust qualitative pattern. Across the V4 and CEE, SME digital and Industry 4.0 uptake is real but uneven, with awareness and interest consistently outrunning structured implementation, and with the binding constraints being organizational, managerial, and human rather than the sheer availability of technology (Horváth & Szabó, 2019; Ghobakhloo et al., 2022; Pirola et al., 2019).

Several strands of regional scholarship converge on this picture. Horváth and Szabó (2019), writing from a V4 (Hungarian) vantage point, show that multinationals and SMEs do not enjoy equal opportunities in technology adoption: even when the driving forces are present, smaller firms face stronger barriers in resources, competences, and management attention. Cross-V4 and neighbouring-country evidence reinforces this asymmetry. Snieška et al. (2020), in one of the few directly Slovak (and Slovak–Polish comparative) studies, find that SMEs perceive Industry 4.0 chiefly through technical, information, and innovation *risks*, and that these risk perceptions materially dampen adoption. That cautious, risk-aware posture recurs in the present interviews when respondents speak of verifying outputs and protecting data. Türkeş et al. (2019) document, for Romanian SMEs, the same wide gap between interest and implementation, propelled by expected efficiency gains but held back by cost, limited skills, and weak awareness. Tick (2023) shows, across CEE more broadly, that Industry 4.0 familiarity among SMEs is heterogeneous and segmentable by firm characteristics rather than uniformly diffused. That kind of internal heterogeneity is precisely what motivates a configurational reading of adoption. And maturity-assessment work from the CEE neighbourhood, such as Kljajić Borštnar and Pucihar's (2021) Slovenian multi-attribute study, demonstrates that SMEs cluster at distinctly different maturity levels driven by organizational readiness and managerial perception rather than by technology access alone. The cumulative message is clear. The region is not technologically excluded from the AI wave; rather, its firms, especially its smaller ones, adopt in a patchy, lightly formalized, risk-conscious manner. That is the setting the AI-ImpactSK study sets out to examine in depth.

The AI-specific Slovak and V4 evidence, though thin, points the same way and is worth foregrounding as the closest national comparator. Dvorský (2025), studying the Slovak SME segment directly, finds that AI is beginning to reshape enterprise risk-management practice but that its use remains partial, function-specific, and not yet embedded in formal organizational processes. That is a near-exact precis of the broader pattern this monograph documents across functions and firms. Marcineková et al. (2025) provide a Slovak micro-enterprise instance from the opposite, formal-implementation end of the spectrum, showing that measurable customer-service gains from AI chatbots are attainable but depend on deliberate, structured integration rather than ad-hoc tool use; situated within a population of firms, such structured cases turn out to be the exception rather than the rule. And the most recent V4 work, Jaciow et al.'s (2026) study of Polish SMEs, identifies perceived benefits and leadership engagement under competitive pressure, not technology availability, as the binding factors in AI maturity, a socio-technical framing that anticipates the governance and benefit-perception levers the present study finds decisive. Read together, these studies establish both the regional fit of the Slovak case and the gap it fills: each is single-function, single-firm, or cross-sectional, whereas AI-ImpactSK offers a systematic, multi-firm, two-wave account.

2.3 The policy backdrop: the EU AI Act and its SME proportionality problem

The firms observed in this study do not adopt AI in a regulatory vacuum. As member-state enterprises, they operate under an emerging European AI-governance regime whose centrepiece is Regulation (EU) 2024/1689, the Artificial Intelligence Act (European Parliament & Council, 2024). The Act establishes a horizontal, risk-tiered legal architecture. It prohibits a narrow set of unacceptable-risk practices, imposes extensive obligations on systems classified as high-risk, and sets lighter transparency duties for certain systems, including generative ones. The timing matters for this study. The Regulation applies in phases from 2025 onward rather than all at once, so the obligations are arriving on a known but staggered schedule across precisely the period the interviews span. The Act does not stand alone. It interacts with the pre-existing General Data Protection Regulation, and systems that process personal data face simultaneous, partly distinct obligations whose interaction creates real compliance complexity for organizations (Schwartmann et al., 2024). For many Slovak SMEs, which already internalize the GDPR only thinly, this combined data-and-AI regime is a substantial conceptual and administrative load.

The scholarly analysis of the Act flagged, from the outset, a proportionality concern that bears directly on the present setting. Veale and Zuiderveen Borgesius (2021), in the most-cited legal analysis of the then-draft Regulation, argue that its risk-based architecture mixes genuine protections with vague, self-assessment-heavy obligations and dependencies on technical standards that risk uneven, burdensome, and under-enforced compliance. Those concerns fall most heavily on smaller actors with the least capacity to interpret and operationalize them. The Regulation itself anticipates this and includes measures intended to soften the burden on SMEs, such as regulatory sandboxes and proportionate documentation expectations (European Parliament & Council, 2024). Yet there is a gap between an obligation existing on paper, on a known timeline, and that obligation being understood and acted upon inside a resource-constrained firm — exactly the gap this study is positioned to observe. The point is not that the firms in this sample are non-compliant; most operate tools whose obligations are light. Rather, awareness and readiness stay minimal even as adoption scales, so formal requirements and informal, bottom-up practice drift apart. The governance regime supplies the external pressure

whose weakness helps explain why these firms steer AI through diffuse vigilance rather than documented control: the corporate-governance literature's internal/external taxonomy (Cihon et al., 2021) predicts that where external pressure is faint and internal structures thin, oversight, if it happens at all, runs through informal channels. Chapter 6 develops this governance theme in full.

2.4 Policy, funding, and the 2025–2026 observation window

The policy setting of this study is double-sided. The same European environment that now regulates artificial intelligence also, through a separate instrument, funds the study of its diffusion. AI-ImpactSK is supported under NextGenerationEU via the Recovery and Resilience Plan for Slovakia (project No. 09I05-03-V02-00003/2025/VA), the EU's post-pandemic recovery facility through which member states finance reforms and investments, including in digital transformation and research capacity (Štetka et al., 2026). This funding context sharpens the book's argument. A study financed by the recovery-and-resilience machinery meant to accelerate digital and AI transition is well placed to ask whether that transition, as it unfolds in firms, is accompanied by the governance and evaluation that responsible-adoption policy presumes. In this sense, policy ambition and firm-level reality can be held in the same frame.

The timing of the study reinforces the point. The years 2025 and 2026 sit at an inflection in enterprise AI. By 2024, generative AI had been characterized in the information-systems literature as a distinct and broadly general-purpose class of technology whose organizational adoption raises governance and risk-management challenges that prior IT-governance practice does not fully address (Feuerriegel et al., 2024). Exposure analyses had already suggested that a large share of work tasks could be affected by large language models, especially as those models are embedded in additional software (Eloundou et al., 2023). That embedding — the wrapping of language models in tools, memory, and orchestration to produce semi-autonomous "agents" — is the qualitative shift the study's two closely spaced waves were positioned to catch in motion.

Between autumn 2025 and spring 2026, agentic AI moved from a frontier curiosity toward a live enterprise concern. The present data record this movement directly: coding-independent measures show respondents' agentic discourse rising substantially across waves, with agentic AI mentioned by a minority of interviews in the first wave and by a clear majority in the second (Štetka et al., 2026). Because this signal is built from respondents' own words rather than from coding intensity, it is robust to the coding-depth differences between waves that Chapter 9 treats with care. For firms that already adopt informally and govern thinly, the arrival of more autonomous tools raises the study's central question with unusual immediacy: does the move toward agents arrive with any corresponding move toward oversight and evaluation, or does adoption simply race further ahead?

2.5 A Slovak SME lens: empirical value and boundaries

The value of the Slovak empirical site is ultimately comparative. The dominant adoption frameworks, including the technology–organization–environment perspective and its acceptance-model and diffusion extensions, are most often operationalized in firm-level survey designs that treat adoption as a managed, top-down decision (Tornatzky & Fleischer, 1990; Chatterjee et al., 2021; Sánchez et al., 2025). They struggle to account for the informal, bottom-up, employee-initiated adoption of platform-delivered AI that recent SME syntheses report as increasingly common (Schwaeke et al., 2024; Wei & Pardo, 2022). A Slovak SME vantage point therefore complements the existing evidence in three ways. It shifts the unit of observation from

AI-developing technology firms to AI-adopting ordinary firms. It supplies CEE-specific, two-wave evidence in a regional record that remains sparse and often confined to single-function or cross-sectional studies. And it brings a resource-constrained setting to phenomena — informal adoption, unmeasured benefits, and uncodified governance — that the international literature has named but mostly observed in better-resourced firms.

A faithful account of this value requires equal candor about the dataset's boundaries. The fifty-eight interviews are a purposive, convenience-based sample, not a representative draw from the Slovak enterprise population, and they carry biases that shape every inference in the chapters that follow (Štetka et al., 2026). The most consequential is a pronounced sectoral skew toward ICT and software, which accounts for 26 of the 58 cases (44.8%). Because ICT firms are early and sophisticated adopters, this composition tilts the sample toward the leading edge of practice. The analysis treats that tilt as a standing limitation rather than waving it away: the results describe an ICT-skewed set of adopters, not Slovak firms at large, and still less the CEE region.

The size distribution also requires caution. SMEs predominate, with 28 firms classifiable as small or medium-sized against 9 large firms, while company size could not be reliably determined for 21 cases. That unknown category is large enough that size effects must be read carefully and, where they are confounded with wave, not interpreted at all. Respondents are weighted toward management, founders, and owners, with strong first-hand involvement in the AI initiatives they describe; 43 of the 58 had directly implemented the technology. This gives the accounts unusual closeness to practice, even as it concentrates the perspective among decision-makers and implementers. The AI footprint of the sample is, as one would expect for this period, dominated by generative AI, present in 52 of 58 firms (89.7%), with AI agents the second most common modality at 26 firms (44.8%) and a mean of roughly two-and-a-half AI types per firm.

These features make the dataset a particular kind of instrument: an ICT-skewed, SME-heavy, practitioner-close, generative-AI-saturated window onto Slovak enterprise practice at a moment of technological transition. It is well suited to generating and stress-testing hypotheses about how adoption, governance, and measurement relate in resource-constrained firms, and poorly suited to producing national prevalence estimates. The economic, regional, regulatory, and temporal context is now in place: an SME-heavy manufacturing-and-ICT economy, a region whose digital maturity is real but unevenly formalized, an EU AI-governance regime phasing in alongside the recovery funding that enabled the study, and a generative-to-agentic transition unfolding across the observation window. The book turns next to the theory and methods through which the themes of governance and measurement are developed.

3 Literature Review and Theoretical Framework

Three claims organize the literature this chapter assembles. AI adoption in firms now outpaces the structures meant to govern it and the instruments meant to measure its value, and what adoption looks like in practice is best understood as a small number of internally coherent configurations, not one managed event. The point is not to catalog every study that touches artificial intelligence in organizations. It is to build, from clustered bodies of work, the argument the AI-ImpactSK evidence is designed to advance — that adoption has outrun its own scaffolding.

Four streams of literature carry that argument, taken in sequence. The first is the theory of technology adoption, which supplies the deductive vocabulary for *why* firms and the people within them take up AI: the technology–organization–environment (TOE) framework, the diffusion of innovation, individual acceptance models, organizational capability theories, and institutional accounts of legitimacy-driven imitation. The second covers governance, responsible AI, and regulation, defining what it would mean to steer AI well and documenting why principles so often fail to become practice. The third concerns a long-running problem the information-systems field has wrestled with for decades and the generative moment has revived: measuring the value of technology investment, from the productivity paradox through benefits-realization management to the recent field experiments that show generative AI delivering measurable gains under the right instrumentation. The fourth addresses the newest objects of study, namely enterprise generative and agentic AI, the build-or-buy sourcing decision, and the configurational theory the monograph uses to read adopter heterogeneity. Each stream closes by sharpening the specific gap the Slovak two-wave study fills. Table 3.1 summarizes how the four streams contribute to the conceptual framework developed in this chapter. The chapter then states the research gaps per analytic angle and develops a conceptual framework that ties the four streams together and maps the study's four research questions onto one model of adoption, governance, measurement, and configuration.

Table 3.1 Literature streams informing the conceptual framework[1]

Literature stream	Main role in the monograph	Key concepts used	Link to the empirical argument
Technology adoption	Explains why firms and individuals take up AI	TOE, diffusion of innovation, TAM/UTAUT, dynamic capabilities, institutional pressure	Adoption is treated as more than a binary decision; it varies in form, depth, and configuration
AI governance and responsible AI	Defines what it means to steer AI responsibly	Codified governance, principles-to-practice gap, regulation, shadow AI	Helps interpret why governance appears as informal vigilance rather than formal policy
IT business value and measurement	Explains why technology value is difficult to evidence	Productivity paradox, benefits realization, measurement endogeneity	Frames the gap between perceived AI benefits and measured AI benefits
Generative/agentic AI and configurations	Captures the newest technological shift and adopter heterogeneity	Generative AI, agentic AI, build/buy/adapt, configurational theory	Supports the analysis of adopter profiles and the temporal shift toward agentic discourse

[1] **Source:** Authors' synthesis based on the literature reviewed in this chapter.

A word on how this chapter relates to the empirical ones that follow. The monograph is built on AI-ImpactSK, a corpus of 58 semi-structured computer-assisted telephone interviews with Slovak organizations gathered in two waves across 2025 and 2026 and analyzed through framework-based thematic analysis triangulated with computational text analysis (Štetka et al., 2026). Three orienting patterns recur in the synthesis below: explicit governance surfaces in roughly one interview in seven, benefits are articulated far more often than they are measured, and adoption sorts into a handful of distinct profiles. Chapters 5 through 9 document and qualify those patterns. Here they appear only as reference points, used to show where the existing literature predicts, brackets, or fails to anticipate what the firms in the sample report. The theoretical frameworks introduced here are not re-explained later; this chapter carries that load once, for the whole work.

3.1 Theoretical foundations of technology adoption

The dominant framework for reasoning about firm-level technology adoption remains the technology–organization–environment model articulated by Tornatzky and Fleischer (1990). TOE holds that a firm's decision to adopt and implement a technological innovation is shaped by three interacting contexts: the technological context (the characteristics of the technology itself and the firm's existing technology stock), the organizational context (size, structure, slack resources, managerial attitudes, and communication processes), and the environmental context (industry structure, competitive pressure, regulation, and the availability of skilled labor). Its durability owes much to its parsimony and its agnosticism about which factors dominate, qualities that have let it be operationalized across technologies and countries. Zhu and Kraemer (2005) supplied one of the most consequential extensions of the tradition. In cross-country firm data on e-business, they showed that the adoption decision is a poor predictor of realized value: what drives business outcomes is the *depth of post-adoption usage*, not the binary fact of adoption. That adoption-versus-usage distinction is foundational here, because it anticipates the disjuncture the Slovak interviews surface, where firms have unambiguously adopted AI yet realize, govern, and measure it shallowly. Frank, Dalenogare, and Ayala (2019) reinforce the lesson from the manufacturing Industry 4.0 setting. Firms adopt advanced technologies in layered, sequenced patterns, with base capabilities preceding and enabling more advanced applications, so that adoption follows a configurational logic and does not arrive uniformly or all at once. Read together, these works warrant treating "adoption" as a configuration of contexts and intensities rather than a yes/no event, which is the stance the configurations angle of this study makes central.

Two structural limitations of the classical framework matter here. TOE was articulated for pre-internet industrial innovation, in which adoption was a deliberate, capital-intensive, top-down decision, and it is largely silent on the bottom-up, employee-initiated uptake of general-purpose tools that defines the generative-AI era. It is also a variance framework: it specifies *which* contexts matter without modeling *how* adoption unfolds as a process, or how individual behavior aggregates into firm-level outcomes. Recent work has begun to relax both constraints. Sánchez, Calderón, and Herrera (2025) fuse TOE with diffusion-of-innovation logic to structure SME AI adoption, identifying organizational readiness, perceived relative advantage, and environmental and regulatory pressure as the principal determinants, a pairing that closely mirrors the apparatus the monograph adopts. Tiago and Almeida (2026) go further with a multilevel account in which environmental, organizational, and individual determinants operate at once, so that firm-level factors alone cannot explain adoption outcomes; individual-level drivers materially shape whether AI is taken up. The multilevel turn matters here because the interviews document adoption that frequently begins with individual employees experimenting with

consumer-grade tools, a dynamic firm-level TOE cannot represent and a multilevel extension can.

The diffusion-of-innovation tradition complements TOE by foregrounding the temporal and social process through which an innovation spreads across a population of potential adopters. It also draws attention to perceived innovation attributes, among them relative advantage, compatibility, complexity, trialability, and observability, that condition the rate of uptake. In the SME and CEE setting that frames this study, diffusion logic helps explain why platform-delivered, low-friction AI tools spread so quickly. High trialability and low up-front cost lower the threshold for experimentation, and visibility among peers fuels imitation. Wei and Pardo (2022) document the mechanism directly: B2B SMEs overcome resource and capability constraints by accessing AI through external platforms that lower technical and financial barriers, which explains the speed and informality of the adoption the firms describe. Diffusion thus links the macro-pattern of rapid spread to the micro-mechanism of platform-mediated trialability, and it anticipates the study's finding that breadth of adoption need not coincide with depth of implementation.

Move to the individual and work level, and the technology acceptance model (TAM) and the unified theory of acceptance and use of technology (UTAUT) supply the most validated constructs for why the people inside firms do or do not take up a new technology. Davis (1989) established that behavioral intention to use a technology is driven mainly by its perceived usefulness and its perceived ease of use, with usefulness the stronger predictor. That parsimonious account maps neatly onto the recurring narratives of AI "saving time" and being "easy to try." Venkatesh, Morris, Davis, and Davis (2003) synthesized the acceptance literature into UTAUT, positing that performance expectancy, effort expectancy, social influence, and facilitating conditions jointly determine acceptance and use, moderated by gender, age, experience, and voluntariness. Venkatesh, Thong, and Xu (2012) extended it to UTAUT2 by adding hedonic motivation, price value, and habit, which together capture the discretionary, consumer-style experimentation that marks much early generative-AI use inside organizations, where enjoyment, subscription cost, and habit formation operate alongside instrumental utility. For the monograph these constructs matter less as a model to be tested than as a vocabulary for reading the motivational grain of the interviews, in which facilitating conditions (infrastructure, skills, managerial support) appear especially decisive for resource-constrained firms.

The acceptance tradition carries a well-documented limitation, and it is what justifies subordinating these constructs to organizational and institutional theory here. Bagozzi (2007) argues that TAM and its derivatives rest on a deterministic, individual-cognition framing that neglects group, social, and cultural determinants and the gap between intention and action, which limits their explanatory depth. For an analysis centered on firms rather than individual users, and on the collective and institutional dynamics of adoption, the critique does real work. It predicts that substantial residual variation, including collective decision dynamics, legitimacy concerns, and capability constraints, will lie beyond what acceptance constructs can reach, and it warrants pairing those constructs with theories pitched at the organizational and field levels. Integrated TAM–TOE studies of firm-level AI adoption (Chatterjee, Rana, Dwivedi, & Baabdullah, 2021) concede the point implicitly by grafting organizational and environmental factors onto the acceptance core, finding that top-management support and competitive pressure operate alongside perceived usefulness. As variance-based, manufacturing-firm surveys, though, they still cannot explain *how* adoption proceeds informally without management sponsorship, which is the dynamic the qualitative evidence brings into view.

Acceptance models explain perceptions and intentions. Capability theories explain which firms can actually turn AI exposure into use and advantage. Teece's (2007) dynamic-capabilities framework reads adoption as an organizational reconfiguration problem, decomposing the firm's capacity to renew its resource base into sensing opportunities, seizing them, and reconfiguring assets, each grounded in identifiable managerial and organizational microfoundations. On this reading AI adoption is a sensing–seizing–reconfiguring sequence, and stalled or shallow adoption can be diagnosed as a breakdown at one of those stages: a firm senses the opportunity and seizes it by acquiring tools, then never reconfigures its processes to embed and evaluate the technology. Cohen and Levinthal's (1990) absorptive capacity explains why prior related knowledge gates this process. A firm's ability to recognize the value of new external knowledge, assimilate it, and apply it commercially is largely a function of what it already knows, so firms with established data, analytics, and IT competence should assimilate AI faster and more deeply than those without. The monograph's sector findings make that contrast visible, with ICT firms reaching an implementation formality others do not. Mikalef and Gupta (2021) operationalize the capability turn for AI, conceptualizing AI capability as a multidimensional construct spanning tangible, human, and intangible resources and showing it associated with organizational creativity and firm performance. Wamba, Gunasekaran, Akter, Ren, Dubey, and Childe (2017) establish the adjacent result for big-data analytics: analytic value is realized through process-oriented dynamic capabilities, not through technology possession alone. Benefit realization, in short, is capability-mediated — which is what licenses reading the measurement gap as the symptom of a missing evaluation capability rather than of absent value.

Efficiency and capability are not the whole story. Adoption is also a matter of legitimacy. DiMaggio and Powell's (1983) account of institutional isomorphism explains why organizations in a field grow structurally similar: coercive pressures (regulation, client or parent-company mandates), mimetic pressures (copying peers under uncertainty), and normative pressures (professional and consultant-driven norms) push firms toward common forms in pursuit of legitimacy more than demonstrable efficiency. Firms confronting the radical uncertainty of generative AI are uncertain about what the technology can do, what competitors are doing, and what regulators will require. Institutional theory predicts the "everyone is doing it" and "we must keep up" rationales that neither acceptance nor capability theory captures. Mimetic dynamics are especially salient in the CEE setting, where firms watching larger or Western peers face strong imitative pressure, and coercive dynamics loom through client and supply-chain requirements and the emerging EU regulatory regime.

The institutional tradition also supplies the mechanism behind the adoption–governance gap this monograph documents, and that mechanism is decoupling. Meyer and Rowan (1977) argued that organizations routinely adopt formal structures as legitimating myths and then buffer their technical core from them, so that the structures a firm displays and the practices it runs on can drift apart by design. Bromley and Powell (2012) sharpen the construct for the contemporary case, distinguishing policy–practice decoupling — where adopted policies go unimplemented — from *means–ends* decoupling, where activities are implemented but their link to the organization's substantive goals is never examined. The second variant is the one that fits AI adoption in these firms. Where uptake is driven by legitimacy-seeking imitation rather than internally generated efficiency goals, a firm can take on the appearance of participating in the AI transformation while leaving uncoupled the codified internal structures — policies, controls, measurement — that would make adoption substantively governed. Read through Meyer and Rowan (1977) and Bromley and Powell (2012), the diffuse, informal vigilance the firms exhibit in place of formal governance, and the perceived benefits they never tie to a metric, are not two

separate lapses but the policy–practice and means–ends faces of the same decoupling. That informal vigilance has a behavioral name on the individual side. The habit of double-checking AI output rather than deferring to it is a form of algorithm aversion, the well-documented tendency of people to lose confidence in an algorithm once they see it err and to want a hand on the controls (Dietvorst et al., 2015); §6.6 reads the firms' verify-the-output posture in those terms. Decoupling is the integrative thread of this chapter, and §6.7 and Chapter 10 develop it against the evidence.

Two further commitments bind these perspectives together and shape how the monograph reads its data, though the methodological chapter develops them more fully. The first is sociotechnical. Following Sarker, Chatterjee, Xiao, and Elbanna (2019), the study treats AI value as emerging from the mutual constitution of social and technical subsystems rather than from the technical artifact alone, so that adoption success should couple technological factors with organizational and work-design factors. The second is affordance-theoretic. Drawing on Markus and Silver (2008) and Strong and colleagues (2014), the study treats the value of an AI tool as residing in relational possibilities for action, not in fixed features, possibilities that have to be perceived and then *actualized* through goal-directed organizational effort. This is useful for interpreting firms that have AI access yet realize little benefit from it: their affordances stay latent, available but unrealized, for want of the complementary organizational conditions actualization requires. Across the 58 firms, then, acceptance, capability, institutional, sociotechnical, and affordance perspectives let the monograph read adoption as perception, capability, legitimacy, and actualization at once, and ask where each mechanism dominates.

3.2 AI governance, responsible AI, shadow AI, and regulation

The adoption literature establishes that firms now take up AI rapidly and informally. The governance literature establishes what it would mean to steer that uptake responsibly, and documents with growing candor why the steering so often fails. The field has matured from a first-generation debate about *which* values should govern AI into a second-order question: how do organizations actually exercise control, through structures, roles, controls, and accountability mechanisms? Mäntymäki and colleagues (2022) give that second-order question a working definition, framing organizational AI governance as the system of structures, processes, and roles through which a firm directs and controls its AI use to keep it aligned with strategy, law, and ethics. The monograph adopts that definition. It is the codified apparatus that definition names, not the underlying concern about AI risk, whose low salience the empirical chapters document. The mapping of normative convergence is a useful starting point. Jobin, Ienca, and Vayena (2019) reviewed 84 AI ethics documents and found global convergence on five recurring principles, namely transparency, justice and fairness, non-maleficence, responsibility, and privacy, alongside substantive divergence in how, and whether, those principles are interpreted and implemented. Floridi and Cowls (2019) distilled the major declarations into five core principles (beneficence, non-maleficence, autonomy, justice, and explicability), and the AI4People framework (Floridi et al., 2018) paired such principles with concrete recommendations, among them assessment procedures, governance bodies, and incentive structures, meant to make a "good AI society" actionable. Mittelstadt, Allo, Taddeo, Wachter, and Floridi (2016) had already mapped the underlying harms, identifying six recurring ethical concerns of algorithmic decision-making, from inconclusive and inscrutable evidence to unfair outcomes and the erosion of traceability, that any governance regime must address.

The pivotal contribution for the monograph's governance argument is Mittelstadt's (2019) demonstration that principles alone cannot guarantee ethical AI. Importing principlism from

medical ethics fails, he argues, because AI lacks the common aims and fiduciary duties of a profession, the proven methods to translate principles into practice, and the robust accountability mechanisms that make professional ethics binding. High-level principles therefore do not reliably produce ethical outcomes. This is the theoretical name for what the empirical chapters observe inside firms — values endorsed diffusely and rhetorically, but rarely instantiated as policies, roles, or controls. It predicts the non-codified, informal governance the evidence records. The principles-to-practice gap has since generated its own operationalization literature. Morley, Floridi, Kinsey, and Elhalal (2020) inventory the publicly available tools for translating principles into day-to-day practice and find them scarce relative to the proliferation of principles, leaving practitioners under-equipped. Morley, Kinsey, Elhalal, Garcia, Ziosi, and Floridi (2021) catalog the organizational barriers that block translation, namely unclear responsibility, lack of method, resource constraints, and weak incentives, along with the enablers, such as governance roles and pragmatic tooling, needed to overcome them. Morley, Elhalal, Garcia, Kinsey, and Mökander (2021) propose "ethics as a service," an externalizable, modular support model premised on the fact that most organizations lack the in-house expertise to convert principles into operational decisions. Díaz-Rodríguez, Del Ser, Coeckelbergh, López de Prado, Herrera-Viedma, and Herrera (2023) integrate the most recent map, linking abstract principles through concrete EU requirements and technical pillars to regulation, so that responsibility becomes demonstrable across the system lifecycle instead of merely asserted at the level of values. This body of work does two things for the monograph. It names the implementation gap the firms exhibit, and it explains why even willing firms default to informal practice: they lack the tools, the owners, and the slack that operationalization presupposes. The barrier inventory of Morley and colleagues maps closely onto what AI-ImpactSK finds.

A parallel corpus is structural and corporate, concerned with how firms institutionalize AI oversight. Cihon, Schuett, and Baum (2021) argue that effective corporate AI governance requires aligning internal mechanisms (board oversight, management structures, auditing, incentives) with external ones such as regulation, standards, and stakeholder pressure, rather than relying on voluntary principles alone. That internal–external taxonomy helps interpret why the firms in this study, which face weak external pressure and hold thin internal structures, so rarely foreground codified AI governance in their own accounts of adoption. Hickman and Petrin (2021) bridge EU trustworthy-AI policy and firm practice from a company-law perspective, showing that translating high-level requirements into accountability demands concrete director duties and internal arrangements that the guidelines do not themselves create. Raji and colleagues (2020) offer the most concrete codifiable practice, an end-to-end internal algorithmic-auditing framework that turns abstract responsibility into documented, reviewable artifacts and decision checkpoints. Zuiderwijk, Chen, and Salem (2021), reviewing AI in public governance, find a pattern that recurs throughout this monograph: benefits are frequently asserted while governance structures, accountability mechanisms, and empirical evaluation stay underdeveloped. The thread across this cluster is that governance becomes real through boards, audits, and documented controls. Those are the codified mechanisms the firms' risk talk gestures toward but seldom narrates as built.

The regulatory and standards layer translates these expectations into binding and voluntary obligation. The EU Artificial Intelligence Act (Regulation (EU) 2024/1689; European Parliament & Council, 2024) establishes a horizontal, risk-tiered legal regime that prohibits unacceptable-risk practices, imposes extensive obligations on high-risk systems, and sets transparency duties for certain systems including generative ones, applied in phases from 2025 onward, with measures such as regulatory sandboxes and proportionate documentation intended to ease the burden

on SMEs. The Act is the binding legal backdrop for every firm in the study. Its phased timeline is what makes the low salience of regulation in these firms' accounts a possible readiness gap rather than a mere descriptive curiosity. The NIST AI Risk Management Framework (Tabassi, 2023) offers the leading voluntary, outcomes-based scaffold, organized around four functions, Govern, Map, Measure, and Manage, whose very first function names the codified governance the sample's SMEs seldom foreground. Its Generative AI Profile (NIST, 2024) enumerates risks distinctive to or amplified by generative systems, including confabulation, data leakage, and information-integrity and intellectual-property harms, that employees incur informally and without controls when they use consumer-grade tools. The OECD AI Principles (Yeung, 2020) supply the intergovernmental normative floor beneath both the EU regime and national strategies. Critical legal analysis tempers any assumption that this scaffolding translates smoothly to smaller firms. Veale and Zuiderveen Borgesius (2021) argue that the Act's risk-based architecture mixes genuine protections with vague, self-assessment-heavy obligations and standardization dependencies that risk uneven and burdensome compliance, concerns that bear especially on smaller actors. Schwartmann, Keber, Zenner, and Kurth (2024) show that the AI Act and the GDPR overlap and must be read together, compounding the compliance complexity firms processing personal data face. For SMEs with minimal legal capacity, the combined effect is a regime binding in principle and opaque in practice, which predicts the low salience of regulation the interviews record.

The most direct bridge between the governance literature and the monograph's empirical pattern is the shadow-IT and shadow-AI strand. Silić and Back (2014) established the foundational construct: shadow IT, the use of systems and software inside organizations without explicit IT-department approval, is widespread and driven by employees' pursuit of productivity, generating benefits alongside significant security and governance risks. Silić, Silić, and Kind-Trüller (2025) extend the construct to the generative era, theorizing "shadow AI" as the unsanctioned employee use of generative AI that reproduces classic shadow-IT dynamics while introducing novel data-privacy, confidentiality, and governance-drift threats, and offering productivity and innovation opportunities that purely prohibitive responses forgo. Waters-Lynch, Allen, Potts, and Berg (2025) reframe the phenomenon constructively as "shadow user innovation," triggered by acute work-process pain and high concealability, which organizations can either suppress or deliberately surface and govern to renew their dynamic capabilities. This cluster is the closest empirical analogue to the monograph's headline pattern of widespread, informal, bottom-up AI use running ahead of any codified governance, and it supplies the normative reframing the monograph adopts: shadow AI is a governance-design problem to be channeled, not just a risk to be suppressed. Feuerriegel, Hartmann, Janiesch, and Zschech (2024) close the loop, establishing from an information-systems standpoint that generative AI raises managerial and governance challenges existing IT-governance practices do not fully address. The implication is that the low governance salience the firms display reflects a genuine novelty, one for which adequate practices are still being invented, as much as any organizational negligence.

3.3 AI and IT business value and the measurement problem

The third stream concerns a problem the information-systems discipline has wrestled with for three decades and that the generative-AI moment has revived in acute form: the difficulty of measuring the value that technology investment produces. The discipline established early and often that technology spending does not convert into firm value automatically. Melville, Kraemer, and Gurbaxani (2004) provided the canonical process model, in which the impact of IT resources on firm performance is mediated by business processes and moderated by

complementary organizational resources and the competitive and macro environment, so that value is realized, or dissipated, several mediating steps from the technology itself and becomes visible only if firms instrument those process-level outcomes. Schryen (2013) surveyed the resulting work and concluded that, despite hundreds of studies, IS business-value research stays fragmented by inconsistent value conceptualizations and weak measurement, and still lacks robust ways to capture how and under what conditions IS investments translate into value. A separate management strand makes the practical point sharper. Ward and Daniel (2012) argue that benefits do not fall out of a technology on their own; they have to be planned, owned, actively managed, and then verified after the fact, which is the discipline of benefits-realization management, and it is exactly the discipline firms skip when they treat acquiring a tool as equivalent to capturing its value. Together these works supply the monograph's core vocabulary, namely IT business value, benefits realization, and the value gap, and they yield its central prediction: claiming a benefit and measuring it are distinct organizational acts, the second far harder and rarer than the first.

The lineage of the monograph's measurement paradox runs directly through the IT productivity paradox. Brynjolfsson (1993) formalized Solow's observation that computers were visible everywhere except in the productivity statistics, arguing that the apparent absence of returns was plausibly explained by the mismeasurement of outputs and benefits, by redistribution, and by lags, that is, by the limits of measurement rather than an absence of value. Brynjolfsson and Hitt (1996) then dissolved the paradox at the firm level, showing with data on 367 large firms that IS spending made a substantial and statistically significant contribution to output once measurement was conducted at the firm rather than the aggregate level. Brynjolfsson and Hitt (2000) supplied the single most direct theoretical statement of the monograph's empirical pattern: organizational investments largely determine the value of IT, and "the benefits of IT investment are often intangible and disproportionately difficult to measure," so realized value depends on complementary organizational change that is itself hard to quantify. That is exactly how the firms describe the time-saving and efficiency gains they perceive but never quantify. Brynjolfsson, Rock, and Syverson (2017) re-opened the argument for AI, contending that a gap between extraordinary AI expectations and disappointing measured productivity is to be expected, because realizing AI's value requires costly, time-consuming, and largely intangible complementary investments in skills, process redesign, and new capital that depress measured benefits in the interim. Brynjolfsson, Rock, and Syverson (2021) turned that intuition into a formal time profile, the productivity J-curve, in which measured productivity first understates true value while firms accumulate unmeasured intangible complementary capital, then overstates it as those intangibles pay off. Read through the J-curve, the non-measurement AI-ImpactSK documents among early-stage adopters is a predicted feature of the adoption phase rather than simple negligence: these firms may be creating value that does not yet appear in any metric they track because the complementary intangibles, including the evaluation capability itself, have not yet been built.

Kohli and Devaraj (2003) make the mechanism explicit. In a meta-analysis of 66 firm-level studies, they found that whether a study detects a positive IT payoff is significantly determined by measurement choices, namely sample size, data source, and the dependent variable selected, and not only by the underlying technology. Payoff is in that sense endogenous to measurement: firms that do not measure, or that measure crudely, will tend to miss benefits they actually obtain. That is the mechanism behind the benefit–measurement dissociation in these firms, and the monograph carries the point from researchers' methodological choices to firms' own absent evaluation practices. The capability literature reaches the same diagnosis from the value side.

Mikalef and Gupta (2021) show that translating AI into measurable performance requires a developed, multidimensional AI capability, so benefit-claims without measurement are consistent with firms holding nascent tools but lacking the intangible and governance components of capability needed to evaluate outcomes. Enholm, Papagiannidis, Mikalef, and Krogstie (2022) confirm at review level that AI value mechanisms and their assessment remain underdeveloped across the literature, which places the measurement paradox within a recognized field-level gap, not an idiosyncratic SME failing.

One thing sets the present moment apart from earlier general-purpose information technologies. The value of generative AI has now been measured directly and consistently in rigorous field and lab-field experiments, and the contrast between those studies and ordinary firm practice is instructive. Brynjolfsson, Li, and Raymond (2025) found, among 5,172 customer-support agents, that access to a generative-AI assistant raised productivity by 15% on average, with the largest gains for less-experienced and lower-skilled workers and minimal or slightly negative effects for the most skilled. Noy and Zhang (2023) found in a preregistered experiment with 453 professionals that access to ChatGPT cut completion time on mid-level writing tasks by 40% and raised output quality by 18% while reducing inequality between workers. Peng, Kalliamvakou, Cihon, and Demirer (2023) found that developers given an AI coding assistant completed a programming task 55.8% faster than a control group. Dell'Acqua and colleagues (2026) found, among 758 consultants, that AI access raised output inside the technology's competence frontier, with 12.2% more tasks completed, 25.1% faster, and significantly higher quality; but for a task deliberately placed beyond that frontier, AI users were 19% less likely to reach the correct answer. That is the "jagged frontier" caveat — a warning of conditional and sometimes negative effects. What these studies share is methodological. They obtain credible benefit estimates only because they impose controlled measurement, namely randomization, baselines, and objective output metrics, that ordinary firms do not deploy. They serve the monograph two ways. They supply quantified benchmarks against which the firms' unquantified time-saving claims can be calibrated, indicating that the benefits left unmeasured are plausibly large and real. And they establish that quantification is feasible with modest apparatus, so that the non-measurement these firms display is not a technical impossibility, whatever its cause. The jagged-frontier result raises the stakes further. Because AI benefits are uneven and occasionally negative, the absence of measurement is not only a missed opportunity to document gains but a risk of masking quality losses on ill-suited tasks.

3.4 Generative and agentic AI and configurational approaches

The fourth stream addresses the newest objects of inquiry and the apparatus the monograph uses to make sense of them. Since the public release of ChatGPT in late 2022, generative AI has diffused into firms faster than any prior enterprise technology, and the dominant early use is text and content generation, not predictive analytics, which fits the Slovak finding that generative tools dominate adoption. Feuerriegel, Hartmann, Janiesch, and Zschech (2024) and Banh and Strobel (2023) fix the terminology and distinctive properties of generative AI, drawing the boundary between generative and discriminative or predictive AI, and among modalities such as text, code, and image, that the monograph's adopter typology relies on. Management scholarship maps the organizational implications. Kanbach, Heiduk, Blueher, Schreiter, and Lahmann (2024) show that generative AI reshapes value creation, value proposition, and value capture along distinct business-model-innovation pathways depending on how firms integrate it; Mariani and Dwivedi (2024) chart its restructuring of innovation processes; and Bankins, Ocampo, Marrone, Restubog, and Woo (2024) establish that AI's organizational impact is multilevel and contingent, operating at individual, team, and organizational levels at once. The

strongest causal-empirical studies sharpen the point. Brynjolfsson, Li, and Raymond (2025), together with the task-exposure analysis of Eloundou, Manning, Mishkin, and Rock (2023), which estimates that around 80% of the U.S. workforce could have at least 10% of their tasks affected by large language models, with exposure rising sharply when the models are embedded in additional software, show that the impact of generative AI is profoundly *heterogeneous*, concentrated among particular workers and unevenly spread across tasks and occupations. That heterogeneity is itself an argument for a configurational rather than a single-effect reading of adoption. If the technology's effects depend on the surrounding mix of workers, tasks, and complementary software, "GenAI adoption" is better treated as a space of distinct firm profiles than as one phenomenon with one average effect.

The fastest-moving body of work concerns agentic AI, autonomous LLM-based systems that plan and act with reduced human oversight, and it mirrors the monograph's finding that agentic discourse rose sharply between the two waves of fieldwork. Wang and colleagues (2024) provide the foundational survey and a unified architecture for LLM-based autonomous agents organized around profiling, memory, planning, and action modules, supplying the taxonomy that lets the monograph interpret what firms mean when they speak of "agents." Liu and colleagues (2025) extend this into a book-length map of foundation-agent architectures, multi-agent systems, and safety challenges. Against the capability optimism, Mitchell, Ghosh, Luccioni, and Pistilli (2025) supply the necessary counterweight. Delineating AI agents by autonomy level, they argue that risks to people, safety risks above all, increase as users cede more control, so that fully autonomous agents trade away human-protective values. That autonomy-risk frame matters for the Slovak evidence, in which enthusiasm for agents coexists with caution, because it recasts agent adoption as a graded, risk-laden choice along a control–autonomy continuum rather than a binary. The agentic literature is, however, almost entirely supply-side. It defines, builds, and theorizes agents and their risks while offering little systematic evidence on how organizations actually talk about, adopt, or govern them — the demand-side, discourse-level dynamic the monograph measures.

How firms obtain their AI capability is a further dimension recent theory has begun to address. AI-ImpactSK finds that sourcing spans build, buy, and adapt, and a small but pointed literature supplies the apparatus to read that variation as a continuum rather than a binary procurement choice. Klotz (2026) brings transaction-cost economics and the resource-based view up to date for the agentic era, arguing that AI-assisted "agentic coding" reshapes the canonical make-or-buy determinants and turns the "build" option from a pure hierarchy into a hybrid governance form that combines code ownership with dependency on external AI infrastructure. On this account, building becomes most compelling for commodity utilities and differentiating custom applications, while regulated and mission-critical systems stay in the "buy" domain. Retrieval-augmented generation, introduced by Lewis and colleagues (2020) as a method for coupling a generative model with a non-parametric retrieval index to produce more specific and factual output without retraining the underlying model, is the dominant technical mechanism for the "adapt" path, grounding general models in proprietary knowledge. It occupies the middle of the build/buy/adapt continuum the firms traverse, from SaaS subscriptions through retrieval-based adaptation and fine-tuning to in-house construction. Read together, these works let the monograph treat sourcing as a continuum whose selection depends on asset specificity, differentiation, and capability, the same contingencies that shape the adopter configurations.

The backbone for the monograph's central analytic move, clustering firms into coding-independent adopter configurations, comes from the configurational tradition in organization and information-systems research. That tradition holds that organizational reality comes in a

limited number of internally coherent gestalts, and that outcomes arise from combinations of elements rather than from isolated variables acting independently. Miller (1986) established the warrant, arguing that strategy and structure cohere into a small number of common, internally consistent configurations, so that organizations are better understood as integrated types than as bundles of separable variables. Short, Payne, and Ketchen (2008) supply the definitional and review scaffolding for positioning new configurational work within the established program. Fiss (2011) operationalizes the approach through a fuzzy-set logic that distinguishes core from peripheral conditions and allows for equifinality, multiple distinct paths to the same outcome, and for causal asymmetry between the presence and the absence of an outcome. It is the toolkit best suited to small-N, theory-rich data. Nickerson, Varshney, and Muntermann (2013) provide the IS-specific, replicable method for *building* taxonomies through iterative conceptual-to-empirical and empirical-to-conceptual passes against explicit ending conditions, which legitimizes constructing an adopter typology as a defensible analytic product, not an ad hoc grouping. El Sawy, Malhotra, Park, and Pavlou (2010) make the seminal IS argument for holistic, configurational thinking with IT at its center; their "digital ecodynamics" treats IT, dynamic capabilities, and environmental turbulence as a fused configuration, not separable variables with linear effects. Park, El Sawy, and Fiss (2017) supply the closest methodological template, applying fuzzy-set analysis to IT-enabled organizational agility and recovering equifinal pathways whose contribution depends on the surrounding configuration of organizational and environmental conditions. Orlikowski and Gash (1994) add the interpretive layer through "technological frames," the assumptions, expectations, and knowledge that organizational groups hold about a technology, and the observation that incongruence between the frames of key groups produces difficulty and conflict. That frame concept supplies the sensemaking mechanism behind the discursive shift toward "agents" the monograph quantifies across waves: rising agentic discourse reflects evolving and partly contested frames, and frame incongruence helps explain why some firms become vision-and-agent specialists while others hesitate. This apparatus is mature and IS-validated. It has not been applied to the combination of generative and agentic AI adoption and sourcing, which is the gap the monograph's typological work fills.

3.5 Research gaps

The four streams converge on a small set of gaps. Each corresponds to one of the study's analytic angles and to the research questions developed in the next section. Table 3.2 summarizes these gaps, while the following paragraphs substantiate them in relation to the relevant literature and the positioning of the AI-ImpactSK study.

Take adoption, the SME, and the CEE context first. The dominant frameworks, TOE (Tornatzky & Fleischer, 1990; Zhu & Kraemer, 2005) and its TAM and diffusion extensions (Chatterjee et al., 2021; Sánchez et al., 2025), are overwhelmingly operationalized as firm-level, variance-based survey designs that treat adoption as a managed, top-down decision. They largely cannot account for the informal, bottom-up, employee-initiated uptake of platform-delivered AI that the most recent SME syntheses report as the norm (Schwaeke et al., 2024; Wei & Pardo, 2022). The CEE, Slovak, and Visegrád evidence base is thin and recent, confined to single-function or single-firm case studies (Snieška et al., 2020; Dvorský, 2025; Marcineková et al., 2025) or to cross-sectional maturity and benefit-perception snapshots (Tick, 2023; Jaciow et al., 2026), with firm-size asymmetries established mainly for manufacturing Industry 4.0 (Horváth & Szabó, 2019). No existing study offers in-depth, theory-anchored, multi-firm qualitative evidence from a CEE, SME-heavy economy across two closely spaced waves. AI-ImpactSK fills the gap, bringing a longitudinal, qualitatively grounded, Slovak-specific account to the informal-adoption

phenomenon the international literature identifies but does not explain for resource-constrained firms.

Table 3.2 Research gaps addressed by the monograph[2]

Area	Gap in the literature	How the monograph addresses it
AI adoption in SMEs and CEE	Existing research is often survey-based, cross-sectional, large-firm oriented, or focused on single functions and cases	Provides a two-wave, qualitative, multi-firm account of AI adoption in Slovak enterprises
Governance, shadow AI, and regulation	Responsible-AI work is often normative or framework-oriented, with limited evidence on how SMEs actually articulate governance	Examines how rarely codified governance surfaces in firms' own accounts and how informal vigilance substitutes for formal control
AI value and measurement	Generative-AI benefits can be measured in controlled studies, but little is known about why ordinary firms perceive benefits without measuring them	Documents the measurement paradox as adoption without systematic evaluation
Generative/agentic AI and configurations	Agentic-AI work is largely technical and supply-side; configurational approaches have rarely been applied to GenAI/agentic adoption	Captures the agentic turn and develops indicative adopter configurations
Theory and method	Prior work has not, to the authors' knowledge, integrated TOE, institutional and capability perspectives with AI-assisted qualitative analysis	Builds an integrated conceptual framework and uses AI-assisted coding with computational triangulation

On governance, shadow AI, and the EU AI Act, the normative literature (Jobin et al., 2019; Mittelstadt, 2019; Floridi & Cowls, 2019) and its operationalization successors (Morley et al., 2020, 2021a, 2021b; Díaz-Rodríguez et al., 2023) establish what responsible AI requires and why principles alone fail, while the regulatory and standards corpus (European Parliament & Council, 2024; Tabassi, 2023; NIST, 2024; Yeung, 2020; Veale & Zuiderveen Borgesius, 2021; Schwartmann et al., 2024) specifies the binding and voluntary scaffolding. Both strands are overwhelmingly normative, doctrinal, or framework-proposing, and both are calibrated to large, AI-developing, largely Western organizations. The shadow-AI strand (Silić & Back, 2014; Silić et al., 2025; Waters-Lynch et al., 2025) is the closest empirical analogue but remains young and is not anchored in a single small EU economy or a longitudinal SME design. The gap is therefore empirical evidence on how rarely codified governance surfaces in SMEs' own accounts of adoption under an emerging EU regime, with base rates for how seldom governance and regulation are made explicit when adoption is bottom-up, and on the character of governance-as-informal-vigilance, where firms substitute diffuse output-checking and data-protection caution for codified policy. AI-ImpactSK supplies that firm-level, SME-heavy, two-wave evidence, testing whether the surface-and-govern alternative is feasible where firms lack the audit, role-ownership, and ethics-as-a-service capacity the literature presupposes.

On value and measurement, the IS-value and productivity-paradox tradition (Brynjolfsson, 1993; Brynjolfsson & Hitt, 1996, 2000; Melville et al., 2004; Schryen, 2013; Ward & Daniel, 2012; Brynjolfsson, Rock, & Syverson, 2017, 2021) has argued for thirty years that technology benefits are real but intangible and disproportionately difficult to measure, and that the detection of value is endogenous to measurement design (Kohli & Devaraj, 2003). The 2023–2026 generative-AI experiments (Noy & Zhang, 2023; Peng et al., 2023; Brynjolfsson, Li, & Raymond, 2025; Dell'Acqua et al., 2026) prove that generative-AI benefits can be quantified precisely, but only inside controlled designs ordinary firms do not deploy. Almost no work examines, with primary

[2] **Source:** Authors' synthesis based on the literature reviewed in this chapter.

data, why firms that readily perceive AI benefits nonetheless do not measure them, especially in resource-constrained, SME-heavy economies absent from both the macro-paradox and the elite-firm experimental literatures. AI-ImpactSK gives the gap an empirical face, documenting and quantifying the adoption-without-evaluation pattern, and it connects the two literatures by showing that the generative re-run of the productivity paradox is, at the firm level, primarily a measurement and evaluation-capability deficit (Mikalef & Gupta, 2021; Wamba et al., 2017; Enholm et al., 2022).

On generative and agentic AI and configurations, four sub-literatures each leave a distinct opening. The enterprise-GenAI cluster (Feuerriegel et al., 2024; Kanbach et al., 2024; Bankins et al., 2024) documents rapid, heterogeneous impact but treats adoption largely as an average effect or a forward-looking agenda, with little firm-level evidence from CEE transition economies and no two-wave view of how use evolves. The agentic cluster (Wang et al., 2024; Mitchell et al., 2025) is technically rich yet almost entirely supply-side. The sourcing cluster (Klotz, 2026; Lewis et al., 2020) updates make-or-buy theory and supplies the "adapt" mechanism but remains conceptual and untested against real firm choices. The configurational and frames apparatus (Fiss, 2011; Park, El Sawy, & Fiss, 2017; Orlikowski & Gash, 1994) is mature but has not been applied to the combination of generative and agentic AI adoption and sourcing. The gap is empirical, near-real-time evidence of the agentic turn and of adopter configurations in a non-frontier economy. AI-ImpactSK contributes rare two-wave evidence on generative *and* agentic discourse, a theoretically grounded adopter typology built with established configurational and taxonomy methods, and an integration of configurations with build/buy/adapt sourcing and technological-frame shifts.

Finally, on theory and method, the gap is integrative. No prior study supplies a combined TOE, institutional, and capabilities reading of *why* governance and measurement lag adoption, nor a demonstration of AI-assisted framework coding triangulated with computational text analysis and disclosed coding-depth transparency (Tornatzky & Fleischer, 1990; Venkatesh et al., 2003; DiMaggio & Powell, 1983; Meyer & Rowan, 1977; Bromley & Powell, 2012; Teece, 2007). The monograph's theoretical contribution is to *apply* these lenses to a novel dataset and setting and to return testable propositions from them, with institutional decoupling (Meyer & Rowan, 1977; Bromley & Powell, 2012) as the integrative mechanism that reads the governance and measurement gaps as the policy–practice and means–ends faces of one phenomenon rather than two unrelated deficits. It is application and proposition-generation, not a claim to have extended four traditions at once. On method, and as the methodological chapter details, it offers AI-assisted, human-in-the-loop coding triangulated with computational evidence as a replicable template.

3.6 Conceptual framework

The framework that integrates these four streams can be stated as one organizing model, built around a single through-line: AI adoption now outpaces the structures that would govern it and the instruments that would measure it, and the form adoption takes is configurational, not uniform. The model has three layers, namely antecedents, the adoption event, and three realization dimensions, and it locates the study's contribution in the third layer, which the existing literature has examined comparatively little. Figure 3.1 summarizes this layered logic before the text unpacks each component.

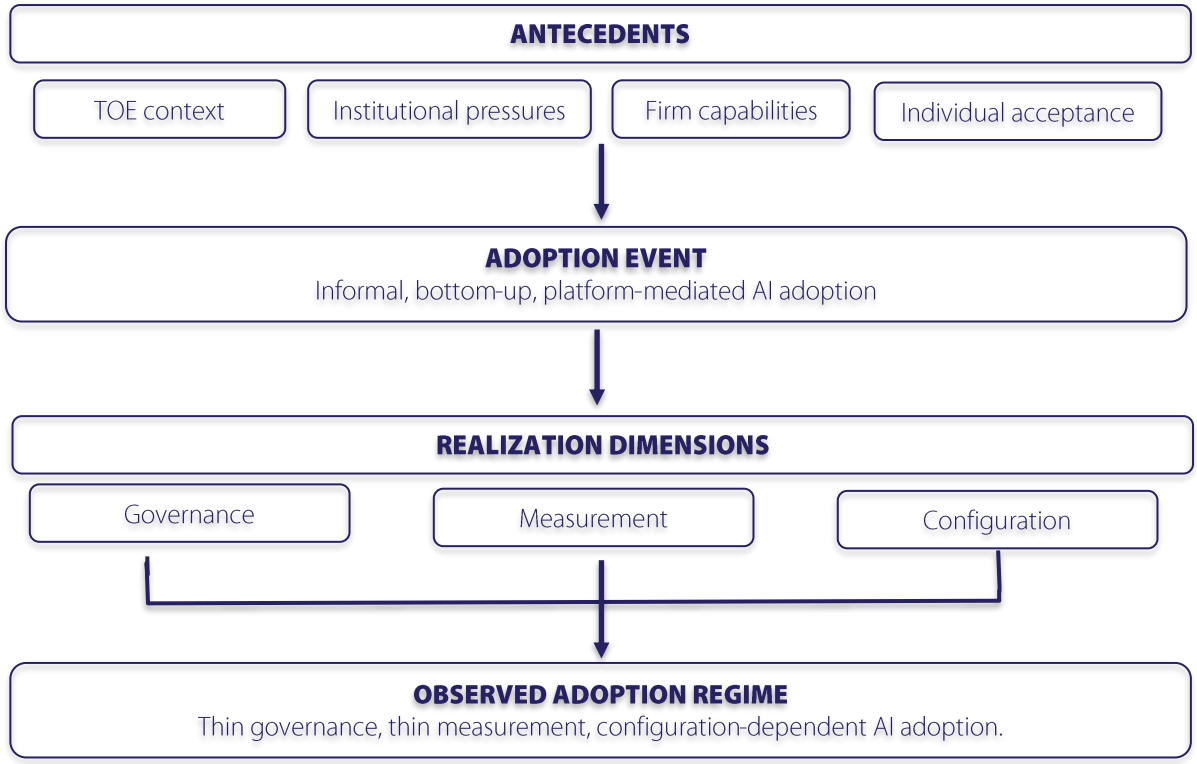


Figure 3.1 Conceptual framework of AI adoption, governance, measurement, and configuration[3]

Note: The framework links adoption antecedents to an informal, bottom-up adoption event and to three realization dimensions — governance, measurement, and configuration — leading to the observed adoption regime of thin governance, thin measurement, and configuration-dependent AI adoption. Temporal change and methodological reflexivity are treated in the accompanying text as cross-cutting elements rather than as separate boxes in the figure.

At the first layer, *antecedents*, adoption is shaped by three forces drawn from the adoption literature. The TOE contexts (Tornatzky & Fleischer, 1990; Zhu & Kraemer, 2005) supply the technological, organizational, and environmental conditions under which adoption becomes possible and takes its initial shape. Institutional pressures (DiMaggio & Powell, 1983) supply the coercive, mimetic, and normative forces that drive legitimacy-seeking uptake under uncertainty, which accounts for adoption that outpaces internally generated efficiency goals and that can decouple from substantive internal structures (Meyer & Rowan, 1977; Bromley & Powell, 2012). Firm capabilities, namely dynamic capabilities (Teece, 2007), absorptive capacity (Cohen & Levinthal, 1990), and the AI-specific capability construct (Mikalef & Gupta, 2021), determine which firms can convert exposure into deep, embedded, evaluated use. Individual-level acceptance factors (Davis, 1989; Venkatesh et al., 2003, 2012), working through the multilevel logic recent work foregrounds (Tiago & Almeida, 2026), feed the bottom-up channel through which much generative-AI adoption now enters the firm, beneath the threshold of managerial decision. These antecedents do not fix a binary adoption outcome. Consistent with the adoption-versus-usage distinction (Zhu & Kraemer, 2005) and affordance theory (Markus & Silver, 2008; Strong et al., 2014), they condition whether the affordances of an adopted technology are actualized or left latent.

[3] **Source:** Authors' own elaboration.

At the second layer, the evidence reconceives the *adoption event* itself. Less a discrete managed decision than an ongoing, frequently informal, often employee-initiated practice, it is the shadow-AI dynamic (Silić et al., 2025; Waters-Lynch et al., 2025) in which platform-delivered tools (Wei & Pardo, 2022) enter the firm with low friction and high concealability. This is the hinge of the framework. The informal, bottom-up character of adoption is what explains why the three realization dimensions are so often left thin.

At the third layer, adoption is realized, or fails to be, along three dimensions the monograph treats as its principal objects of study and that prior work has examined far less than adoption itself. The first is *governance*: whether firms codify the steering of AI through policies, roles, and controls, or substitute diffuse, informal vigilance for it. The governance literature predicts, and the principles-to-practice mechanism (Mittelstadt, 2019; Morley et al., 2021) explains, that under bottom-up adoption and weak external pressure codified governance will rarely be foregrounded in firms' own accounts. The second is *measurement*: whether firms instrument and evaluate the value they perceive, or assert benefits they never quantify. The productivity-paradox and benefits-realization literatures (Brynjolfsson & Hitt, 2000; Ward & Daniel, 2012; Brynjolfsson, Rock, & Syverson, 2021; Kohli & Devaraj, 2003) predict, and the recent experimental evidence (Brynjolfsson, Li, & Raymond, 2025; Dell'Acqua et al., 2026) calibrates, the dissociation between articulated and measured benefits, the measurement paradox. The third is *configuration*: the form adoption takes, including the modalities of AI deployed, the breadth of the AI portfolio, the degree of implementation formality, and the build/buy/adapt sourcing posture. The configurational and sourcing literatures (Fiss, 2011; Park, El Sawy, & Fiss, 2017; Klotz, 2026; Lewis et al., 2020) predict, and the monograph's typology operationalizes, that adoption sorts into a limited number of internally coherent adopter profiles rather than scattering across continuous dimensions.

These three dimensions are not independent. Implementation formality tends to co-vary with measurement practice, and configuration conditions both governance and measurement, with ICT-intensive, multi-modal configurations realizing adoption more formally than generative-only minimalist ones. Read through institutional theory, the first two dimensions have a common name. The thin codification of governance is policy–practice decoupling, and the gap between perceived and measured value is means–ends decoupling: an activity is implemented, but its link to the firm's substantive goals is never examined or evidenced (Meyer & Rowan, 1977; Bromley & Powell, 2012). Treating the governance and measurement gaps as two faces of one decoupling, rather than as separate failings, is the framework's integrative core, and the discussion chapter takes it up.

Two cross-cutting elements complete the framework. The first is *temporal change*. The two-wave design situates the whole model in a turbulent 2025–2026 window in which technological frames are shifting (Orlikowski & Gash, 1994) and agentic AI is rising (Mitchell et al., 2025; Wang et al., 2024), so the antecedents, the adoption event, and the realization dimensions are all observed at two points, and the maturation of discourse toward agents and "practical, verified production" becomes itself an object of analysis. The second is *methodological reflexivity*, which runs alongside the substantive model, not within it. Because the analysis depends on AI-assisted, human-in-the-loop framework coding triangulated with computational text analysis, the framework carries the coding-depth transparency the methodological chapter develops, so that claims about change over time stay separate from artifacts of analytic intensity.

The study's four research questions map onto this framework as a coordinated set. RQ1 addresses the *configuration* dimension and its antecedents, asking how AI adoption varies in

form and structure across the firms and what shapes it, engaging TOE, dynamic capabilities, and configurational theory to interpret the adopter typology and the sector-linked formality differential. RQ2 addresses the *measurement* dimension, asking whether firms that claim benefits actually measure them, and reading the resulting paradox through the productivity-paradox and benefits-realization literatures and the endogeneity of payoff to measurement. RQ3 addresses the *governance* dimension, asking how often explicit governance surfaces and whether its salience is associated with firm traits, and reading the pattern through institutional theory, decoupling, and the principles-to-practice mechanism. RQ4 addresses the *temporal* cross-cutting element, asking how the discourse of adoption changed across the two waves net of coding depth, and interpreting the agentic turn through technological frames and the autonomy-risk literature. The framework binds four literatures that currently run in parallel into one account: AI adoption, shaped by TOE, institutional pressure, and firm capability and increasingly arriving bottom-up, is realized unevenly along the three under-studied dimensions of governance, measurement, and configuration. That is the account the rest of the monograph tests against the AI-ImpactSK evidence.

4 Research Design and Methods

This chapter explains how the AI-ImpactSK interview evidence was generated, prepared, coded, triangulated, and analyzed. It presents the study as a qualitative-dominant mixed-methods design, describes sampling and recruitment, sets out the interview instrument, and details transcription, translation, anonymization, framework-based thematic analysis, AI-assisted coding, computational triangulation, and quantitative procedures. Because the study is purposive, modest in size, and ICT-skewed, the chapter treats numerical tests as structured probes of robustness rather than as a basis for population inference. It also makes explicit the main methodological caveats that condition the interpretation of the findings: uneven elicitation of measurement and governance, partial size information, and the coding-depth difference between the two waves.

4.1 Research Philosophy and Overall Design

This study adopts a pragmatist stance toward inquiry, treating the research questions rather than any prior commitment to a single paradigm as the arbiter of method. Pragmatism is well suited to a problem that is at once interpretive and quantifiable: understanding *how* Slovak enterprises adopt artificial intelligence requires close attention to the meanings that managers and practitioners attach to their own practices, while assessing *whether* benefits are measured, *how prevalent* formal governance is, and *whether* discourse changed across a year demands numerical estimation and hypothesis testing. The pragmatist commitment to methodological pluralism licenses a design in which qualitative evidence carries the primary interpretive load and quantitative procedures discipline, summarize, and stress-test the patterns that the qualitative work surfaces. Throughout, the study treats its sociotechnical object (AI as people, structures, and tasks rather than as an artifact in isolation) in the spirit of Sarker et al. (2019), and it reads adoption simultaneously as perception, capability, legitimacy, and actualization (Davis, 1989; Teece, 2007; DiMaggio & Powell, 1983; Strong et al., 2014), as set out in Chapter 3.

The concrete realization is a qualitative-dominant mixed-methods design organized as a two-wave panel of semi-structured interviews. The qualitative component is dominant in the sense that the corpus, the analytic engine, and the principal contributions are interpretive; the quantitative component is integrated rather than parallel, because it operates on quantities *derived from* the coded qualitative material (theme coverage, formality indices, lexical counts) and on coding-independent text statistics extracted from the same transcripts. The two waves were timed to bracket a period of unusually rapid change in the diffusion of generative and agentic AI, with Wave 1 conducted in late 2025 and Wave 2 in the spring of 2026. The two-wave structure was intended less as a formal longitudinal panel — most participants differ across waves — than as a near-real-time window onto an evolving field, allowing the analysis to distinguish, where possible, genuine maturation in enterprise AI discourse from artifacts of the measurement process itself. The complete dataset comprises 58 interviews and is cited throughout as (Štetka et al., 2026).

It bears emphasizing at the outset that the design is exploratory-confirmatory rather than confirmatory in the strict inferential sense. The sample is purposive and modest (N = 58), recruited from a single national economy, and skewed toward the information and communication technology (ICT) sector. The quantitative procedures are therefore framed as a structured way to characterize, summarize, and probe the robustness of qualitative patterns

within this sample, not as a basis for population inference. This caveat is repeated wherever an inferential test is reported.

4.2 Sampling and Recruitment

Recruitment was contracted to a professional market-research agency, which conducted computer-assisted telephone interviewing (CATI). Delegating fieldwork to an agency afforded access to a curated business-contact frame, professional interviewer capacity, standardized call protocols, and per-recording documentation of outcomes. For an academic team studying time-pressed managers, these advantages materially improved reach. They also introduced dependencies on an external frame and on agency call dispositions, which are disclosed below.

Table 4.1 Fieldwork and recruitment by wave[4]

Fieldwork indicator	Wave 1	Wave 2
Fieldwork period	28 Aug–9 Oct 2025	10 Apr–2 Jun 2026
Unique numbers dialled	135	435
Connected calls	81	377
Interviews completed in the field	30	28
Cases included in the dataset	29	29
Response rate on numbers dialled	22.2%	6.4%
Response rate on connections	37.0%	7.4%
Main caveat	One cross-wave duplicate excluded	One interview covered two companies in one transcript

The two waves differed sharply in their dialling effort and in the yield that effort produced, and these differences are themselves a finding about the changing willingness of firms to participate. In Wave 1 (fieldwork 28 August–9 October 2025), 135 unique numbers were dialled, of which 81 connected and 30 interviews were completed in the field; one of these was subsequently excluded as a cross-wave duplicate, yielding 29 cases in the dataset. The corresponding response rates were 22.2% on numbers dialled and 37.0% on connections. In Wave 2 (fieldwork 10 April–2 June 2026), 435 numbers were dialled and 377 connected, with 28 interviews completed per the agency's per-recording count; the dataset nonetheless contains 29 Wave 2 transcripts because one interview (designated F2.02) covered two companies and was rendered as a single combined transcript. The Wave 2 response rates were 6.4% on numbers dialled and 7.4% on connections. The two waves are thus balanced at 29 cases each, for a combined corpus of 58.

The most conspicuous fieldwork pattern is the collapse in the response rate between waves: from 22.2% to 6.4% on dialled numbers, and from 37.0% to 7.4% on connections. Several non-exclusive interpretations are plausible — survey fatigue around AI topics, a larger and less pre-qualified Wave 2 frame, or seasonal effects — and the design cannot adjudicate among them. What is methodologically essential is that Wave 2 reached its 29 completed interviews only by dialling a much larger frame (435 versus 135 numbers), so that the second wave's respondents were drawn from a substantially wider and more thinly converting pool. This is treated as

[4] **Source:** Authors' own processing based on AI-ImpactSK fieldwork documentation and interview dataset

fieldwork and selection context bearing on any cross-wave comparison, and it is revisited in the reflexivity discussion (§4.9) and in the limitations chapter.

Two minor independence caveats are noted for completeness. One participant (F2.28) took part in both waves, and one Wave 2 interview (F2.02) combined two companies into a single transcript. Each concerns a single case out of 58 and is flagged where relevant rather than treated as a threat warranting case exclusion.

Sampling was, in formal terms, purposive and convenience-based: the agency frame targeted Slovak organizations with plausible exposure to AI, and participation was self-selected among those willing to be interviewed. This recruitment logic produces a sample weighted toward organizations and respondents already engaged with AI. That is appropriate for a study of adoption *practice*, but it is not representative of the Slovak enterprise population. The resulting composition is described fully in Chapter 5; for present purposes the decisive feature is the sector skew. The information and communication technology sector accounts for 26 of 58 cases (44.8%), making the sample distinctly ICT-heavy and, by extension, biased toward early and sophisticated adopters. A further 21 firms could not be assigned a confident size class, leaving firm size only partly observed. These features (ICT over-representation, early-adopter self-selection, and partial size information) bound the transferability of every claim and are carried forward explicitly into the inferential framing, the configurational analysis, and the limitations chapter.

4.3 The Interview Instrument

Data were generated through a semi-structured interview protocol administered by telephone, with each interview lasting approximately 30 to 35 minutes. The semi-structured form was chosen deliberately: it imposes enough common structure to support cross-case comparison and framework charting, while leaving room for respondents to narrate their own adoption trajectories, name the tools and use cases that mattered to them, and raise concerns the researchers had not anticipated. This balance between comparability and openness is precisely the configuration that the framework method presupposes (Gale et al., 2013).

The protocol was organized into six thematic blocks, labeled A through F, that move from the origins of adoption to its prospective future. Block A addressed the *trigger and motivation* for adopting AI: the initiating event, the actors who initiated it, the rationale for engagement, and whether the firm had framed its goal in measurable terms. Block B turned to the firm's *goals and intended outcomes*. Block C explored the *types of AI* in use, distinguishing generative tools, agentic systems, natural-language and optical-character processing, computer vision, predictive analytics, and proprietary or open-source solutions. Block D covered *implementation*: the approach and phasing, the teams involved, and any testing. Block E elicited the *challenges* encountered, spanning technical, organizational, legal-ethical, and adoption-resistance dimensions, and it asked directly how the firm measured success, which key performance indicators it tracked, and whether it could put concrete figures on the benefits it claimed. Block F concerned *results and practices*, including perceived benefits, the measurement of those benefits, and emergent best practices; a forward-looking strand on future plans, including whether respondents expected AI regulation to affect them, rounded out the conversation. This block architecture maps directly onto the a priori coding scaffold described next, and it operationalizes the multi-lens theoretical frame of Chapter 3 by inviting accounts of perceived usefulness and ease of use (Davis, 1989), facilitating conditions and adoption drivers (Venkatesh

et al., 2003), and the sensing–seizing–reconfiguring sequence of capability deployment (Teece, 2007).

One feature of this design conditions how the later results should be read, and it is best stated plainly here. The protocol probed measurement and governance very differently. Measurement was asked about directly, and twice: the goal-measurability question in Block A and the explicit success-metrics question in Block E (how do you measure success, which KPIs, what concrete figures). Governance and regulation had no question of their own. They could enter the record in only two ways: when a respondent volunteered them under the legal and ethical challenges prompt in Block E, or through the forward-looking question about expected AI regulation. The two topics were therefore elicited on an uneven footing, a point we return to when interpreting their respective coverage rates. We should also note that the measurement probe, although part of the standard instrument, was administered somewhat unevenly across the corpus, with most of the gaps falling in Wave 1; the figures it yields are best read as near-lower bounds rather than as a complete census of measurement practice.

4.4 Data Preparation: Transcription, Translation, and Anonymization

Interviews were conducted in Slovak, the working language of the participants, and were audio-recorded with consent. Two transcript artifacts were produced for each interview: a verbatim transcript faithful to the spoken record, and a lightly cleaned transcript in which disfluencies and transcription noise were smoothed to support analysis without altering substantive content. Transcription and the subsequent Slovak-to-English translation were AI-assisted and then author-reviewed. Consistent with cross-language research guidance (Squires, 2009; van Nes et al., 2010), the Slovak text was treated as the source of record and the locus of interpretive authority, with English renderings produced for analysis and reporting and checked by the authors against the Slovak originals. This ordering (analyze in the original language as far as possible, translate deliberately, and document the translation choices) is a trustworthiness safeguard against the meaning slippage that the cross-language literature identifies at the translation interface. All verbatim quotations reproduced in this monograph are the author-reviewed English renderings, tagged to their case, wave, and respondent profile.

Anonymization was applied during preparation. Personal names, company names, and other directly identifying details were replaced with placeholders, and each interview was assigned a stable case identifier (of the form F1.## for Wave 1 and F2.## for Wave 2) used consistently across the dataset and the text. Processing complied with the General Data Protection Regulation: participation was on an informed, voluntary basis; the lawful basis and retention were specified in the project's ethics documentation; and the original audio recordings are not shared. The anonymized transcript dataset is released under a Creative Commons Attribution 4.0 (CC BY 4.0) licence to support reuse and verification, while the audio remains restricted to protect respondent identity.

4.5 Analytic Approach: Framework-Based Reflexive Thematic Analysis

The analytic engine of the study is a framework-based reflexive thematic analysis. The approach draws its phased procedure and quality standards from reflexive thematic analysis (Braun & Clarke, 2006, 2019, 2021) and its matrix-based charting structure from the framework method (Gale et al., 2013). The two traditions are complementary here. Reflexive thematic analysis supplies the epistemological stance — themes are actively generated by theoretically informed researchers rather than passively "emerging," and analytic quality is judged by coherence with stated assumptions rather than by inter-rater reliability coefficients (Braun & Clarke, 2019, 2021).

The framework method supplies the apparatus for organizing a partly deductive, category-driven analysis of many cases into an auditable case-by-code matrix, which is exactly what a study with a predefined block structure and a multi-category codebook requires (Gale et al., 2013).

The codebook comprises eight categories, labeled A through H, subdivided into 24 themes. The categories span context and motivation (A), use areas and data (B), technology and sourcing (C), implementation (D), challenges (E), results and practices (F), governance and regulation (G), and future orientation (H). Application of this codebook to the 58 transcripts produced 629 coded segments. A segment is the unit of coding, a passage of respondent speech assigned to one theme, and *coverage* is defined at the case level as the proportion of interviews (out of 58) in which a given theme or category appears at least once. Coverage, rather than raw segment counts, is the primary descriptive quantity reported in subsequent chapters, because it speaks to how widely a practice or concern is shared across firms rather than to how voluminously any one respondent spoke. These theme-presence rates are best understood as articulated or coded rates: lower bounds on what firms actually do, not censuses of practice. A topic that a respondent did not raise in a roughly thirty-minute conversation may still be present in the firm, so a low coverage figure marks the limit of what was said, not proof of what was absent. This reading matters most for the topics the instrument did not probe directly. As an internal validation, the analytic pipeline reproduces the authors' published theme-coverage table exactly, lending confidence that the charted matrix faithfully represents the underlying coding.

The adequacy of the 58-interview, two-wave corpus for framework-based thematic mapping is assessed in relation to saturation benchmarks. The empirical benchmark literature indicates that the great majority of codes and core themes in reasonably homogeneous samples are typically identified within roughly the first dozen interviews (Guest et al., 2006), a threshold the present corpus comfortably exceeds. The study is careful, however, to specify which saturation it claims. Following the conceptual distinctions of Saunders et al. (2017) and Hennink et al. (2017), it treats code saturation, the identification of the range of codes, as securely achieved, and positions the two-wave design as a means of moving beyond mere code capture toward richer meaning saturation for the core themes. It declines to over-claim saturation for the sparsely populated categories, notably governance, where coverage is low. For governance that low coverage carries a specific meaning, set by the elicitation asymmetry described in §4.3: because the topic was never asked about as such, its coverage records how often firms raised it on their own, not how often they actually do it.

4.6 AI-Assisted Coding and the Human-in-the-Loop

Because the study used AI assistance not only for transcription and translation but also for a first pass of thematic coding, it engages directly with the emerging methods literature on large-language-model-assisted qualitative analysis. The closest published analogue to the present workflow is codebook-anchored deductive coding with an LLM (Xiao et al., 2023; Chew et al., 2023), in which a model guided by an expert codebook proposes segment-to-theme assignments that human analysts then review. This is precisely the configuration adopted here: the codebook of eight categories and 24 themes was the controlling instrument, the model's role was to accelerate the labeling of 629 segments across 58 transcripts, and every assignment was author-reviewed rather than accepted automatically. The justification for AI assistance is the documented efficiency and scalability gain it offers for deductive coding of sizable corpora (Chew et al., 2023), set against the equally documented requirement that outputs be validated by human analysts.

The cautions in this literature are taken seriously and are designed into the procedure. Studies of inductive and deductive LLM coding consistently report that models can approximate human thematic work and do so far faster, yet miss nuance and context, vary in consistency, and are prone to fabrication if unsupervised (De Paoli, 2023; Mathis et al., 2024; Prescott et al., 2024). The human-in-the-loop design responds to each concern: first-pass machine assignments were reviewed and corrected by the authors against the source text; interpretive authority over theme generation and over contested or subtle passages rested with the human analysts in keeping with the reflexive stance (Braun & Clarke, 2019); and, crucially, the coding is not allowed to stand alone as the sole warrant for the study's central claims. Instead, the key quantitative patterns are triangulated against coding-independent computational text analysis, described next, so that conclusions do not rest solely on AI-influenced labels. This positions the contribution as a demonstration of AI-assisted framework coding *disciplined by* human review and computational triangulation, rather than as an endorsement of unsupervised machine coding.

4.7 Computational Triangulation

To provide a validation pathway independent of the manual codebook, the study conducted a parallel computational text analysis on the cleaned English transcripts. The transcripts were parsed to separate respondent turns from interviewer turns, so that the analysis operates on what participants actually said rather than on the questions put to them; this respondent-only text formed the basis for a corpus of 58 cleaned English transcripts. On this corpus the study computed several coding-independent measures: lexical analyses keyed to substantive vocabularies (for example, governance and key-performance-indicator lexicons); keyness analysis using the log-likelihood statistic to identify words distinctively associated with each wave's respondent speech; counts of named-tool mentions; and counts of agentic-AI vocabulary (terms such as *agent*, *agentic*, *autonomous*, and *RAG*).

The purpose of these computations is explicitly *coding-independent validation*, not the generation of new themes. Where a manual code aligns with an independent lexical signal, confidence in the code's substantive meaning is strengthened; where it does not, the discrepancy is itself informative. Two illustrative checks anticipate the results chapters. The manual measurement code aligns with the key-performance-indicator lexicon (mean lexical hits of 2.58 when the code is present versus 1.35 when absent; point-biserial $r = 0.27$, $p = .041$), supporting the validity of the measurement coding. By contrast, the coding-independent governance lexicon does *not* align with the formal governance code (point-biserial $r = -0.20$, not significant) and is flat across waves (Wave 1 1.0 versus Wave 2 0.9; Mann–Whitney $p = .95$). When firms did speak to governance, then, they did so in the vocabulary of diffuse, informal vigilance rather than codified practice, a reading the manual code alone could not have established. This independent check matters all the more given the elicitation asymmetry of §4.3: governance was never asked about as such, so a convergent text signal does some of the work a dedicated question would otherwise have done.

4.8 Quantitative Procedures

The quantitative procedures summarize coded and computational quantities and probe the robustness of qualitative patterns; all are interpreted within the exploratory-confirmatory frame appropriate to a purposive sample of $N = 58$. Theme and category coverage are reported as proportions, with uncertainty expressed through Wilson score intervals, which behave well for the small samples and near-boundary proportions that several themes produce. The central analysis of the measurement paradox uses McNemar's exact test for the paired contrast between

articulating a benefit and measuring it, the natural test for two binary attributes observed on the same firms. Bivariate associations between a binary outcome (for example, presence of any governance) and a categorical predictor (for example, firm-size class or ICT membership) are assessed with Fisher's exact test and the chi-square test, always accompanied by an effect-size estimate, the odds ratio for two-by-two contrasts and Cramér's V for the strength of association, so that statistical significance is never reported without magnitude. Comparisons of continuous or index quantities between groups (for example, implementation formality by sector, or respondent word counts by wave) use the Mann–Whitney U test, reported with rank-biserial correlation and Cliff's delta as non-parametric effect sizes. Monotonic associations between ordinal or continuous measures (for example, formality and AI-technology breadth) are assessed with Spearman's rank correlation.

Two multivariate procedures support the configurational analysis of adoption. An indicative adopter typology is derived by k-modes clustering on coding-independent metadata and AI-profile attributes, with cluster stability assessed by the bootstrap adjusted Rand index (ARI); the resulting ARI of 0.35 is reported as evidence that the configurations are indicative rather than definitive, and they are presented with corresponding humility. Multiple correspondence analysis (MCA) is used to visualize the associational structure among categorical adoption attributes. Because the analysis entails a family of related significance tests, the confirmatory family is corrected for multiple comparisons using the Benjamini–Hochberg false-discovery-rate procedure, and both nominal and FDR-adjusted values are reported so that readers can see which results survive correction. Throughout, association is paired with effect size, causal language is avoided, and the modest, purposive, single-country, ICT-skewed character of the sample is treated as a standing constraint on interpretation rather than a footnote.

Table 4.1 summarizes the principal statistical tests, the quantities they operate on, and their analytic purpose. All numerical values cited in the table and in the surrounding text are drawn from the study's findings; the detailed results are presented in Chapters 5 through 9.

Table 4.2 Statistical tests, targets, and purpose[5]

Procedure	Applied to	Effect size reported	Analytic purpose
Wilson score intervals	Theme and category coverage proportions	Interval estimate	Quantify uncertainty around coverage (e.g., benefits articulated 65.5% [52.7–76.4]; benefits measured 20.7% [12.3–32.8])
McNemar's exact test	Paired benefit-articulation vs. benefit-measurement	— (paired discordance)	Test the measurement paradox (statistic 8.0, $p = 6.9\times10^{-5}$; survives BH-FDR q = .0007)
Fisher's exact test	Governance presence by size / ICT	Odds ratio	Test whether low governance salience is subgroup-specific (size OR 2.93, p = .327, q = .41; ICT OR 0.87, p = 1.0); no subgroup difference detected, but underpowered (Large n = 9)
Chi-square test	Categorical adoption associations	Cramér's V	Assess strength of association (e.g., measurement × size V = .03; × ICT V = .14)
Mann–Whitney U test	Formality, word counts, lexical/agentic counts by group	Rank-biserial; Cliff's δ	Compare distributions across sector or wave (e.g., agentic mentions W1 0.72 vs. W2 3.14, p = .0022, rank-biserial 0.44)
Spearman's rank correlation	Breadth ↔ formality	ρ	Test monotonic association (ρ = 0.20, p = .14, ns)

[5] **Source:** Authors' own processing based on the AI-ImpactSK dataset

Procedure	Applied to	Effect size reported	Analytic purpose
k-modes clustering	Coding-independent metadata + AI profile	Bootstrap ARI	Derive indicative adopter configurations (four clusters; ARI = 0.35 → indicative)
Multiple correspondence analysis	Categorical adoption attributes	— (geometry)	Visualize associational structure among adoption attributes
Benjamini–Hochberg FDR	Confirmatory family of tests	q-values	Control false-discovery rate across multiple comparisons
Point-biserial correlation	Manual codes vs. independent lexicons	r	Triangulate codes against coding-independent text (measurement r = 0.27, p = .041; governance r = –0.20, ns)

4.9 Reflexivity and Methodological Transparency: The Coding-Depth Artifact

A central commitment of this monograph is to disclose, rather than to conceal, a methodological artifact that materially conditions cross-wave comparison. Reflexive thematic analysis treats the analyst as an instrument whose decisions shape the data (Braun & Clarke, 2019), and the most consequential such decision here concerns coding *depth* — how finely the same speech is segmented into coded units — which differed systematically between the two waves.

The evidence for the artifact is direct. Wave 2 interviews were coded substantially more densely than Wave 1 interviews: the mean number of coded segments per interview rose from 8.21 in Wave 1 to 13.48 in Wave 2 (overall 10.84), an increase of roughly 64%, with the difference highly significant on a Mann–Whitney test ($p < 1\times10^{-7}$). Part of this reflects the interviews themselves rather than the coder: Wave 2 respondents simply spoke more, averaging 1,123 words per interview against 696 in Wave 1 (Mann–Whitney $U = 203$, $p = .0007$, rank-biserial 0.52), so that Wave 2 participants talked roughly 1.6 times as much. The two facts are linked — a formality index correlates with segments per interview at $r = 0.56$ — and together they imply that any naive cross-wave rise in theme coverage partly reflects denser coding and longer answers, not only real change in the underlying practice. The intensity coding is even more severely wave-confounded: the share of segments coded "high" intensity stands at 0.43 in Wave 1 against 0.91 in Wave 2. For this reason the study declines to use intensity for any cross-wave substantive claim, and it treats cross-wave coverage shifts as hypothesis-generating signals to be triangulated against coding-independent text (per §4.7) rather than as settled facts.

Two consequences follow for how results are reported. First, every cross-wave coverage and intensity comparison in this monograph is hedged accordingly, and wherever a substantive temporal claim is advanced it is anchored in coding-independent evidence. The clearest case is the rise in agentic-AI discourse, which is visible in respondents' own word counts (Wave 1 0.72 versus Wave 2 3.14 agentic mentions; Mann–Whitney $U = 234$, $p = .0022$, rank-biserial 0.44) and is corroborated by the agentic code's strong alignment with the text (point-biserial $r = 0.55$, $p < .0001$). Second, and reassuringly, the denser Wave 2 coding did *not* mechanically inflate the two patterns the study leans on. Had coding depth been driving the findings, the measured-benefit rate and the governance-salience rate should both have risen in the more densely coded wave. Instead they fell. The presence of any measurement was 24.1% in Wave 1 versus 17.2% in Wave 2, and the share of firms that raised governance on their own was likewise 24.1% in Wave 1 versus 17.2% in Wave 2. Neither pattern is a coding-depth artifact, which is a robustness check worth having precisely because the two rates carry such different evidential weight: the measured-benefit rate rests on a question that was directly put to respondents, whereas the

governance rate is a salience measure shaped by the asymmetry of §4.3. We characterize the combined picture as adoption running ahead of governance and measurement, and treat that phrase as our interpretation of the evidence rather than as a direct observation of what firms do behind the interview.

4.10 Rigor, Trustworthiness, and Ethics

The study's claims to rigor are organized around the trustworthiness criteria for naturalistic inquiry — credibility, transferability, dependability, and confirmability (Lincoln & Guba, 1986). Credibility is pursued chiefly through triangulation: the manual framework coding is cross-checked against coding-independent computational text analysis, so that convergent signals reinforce one another and divergences are surfaced and interpreted rather than suppressed. Transferability is supported by thick description of the sample and context (Chapter 5) together with the candid statement of its purposive, ICT-skewed boundaries, so that readers can judge the reach of the findings to other settings. Dependability and confirmability rest on an audit trail: a fixed codebook of eight categories and 24 themes, a reproducible analytic pipeline that recovers the authors' published coverage table exactly, documented translation and anonymization procedures, and the release of the anonymized transcript dataset under CC BY 4.0. Reporting follows the consolidated criteria for reporting qualitative research (COREQ), the 32-item checklist of Tong et al. (2007), which structures disclosure of the research team, the study design, and the analysis. Quality in the thematic component is judged, per Braun and Clarke (2021), by the coherence of the analysis with its stated reflexive and framework assumptions rather than by positivist reliability metrics, and the coding-depth disclosure of §4.9 is offered as a concrete enactment of that reflexive standard.

The study was conducted under formal ethical approval. Participation was voluntary and based on recorded informed consent obtained at the start of each recorded interview. The protocol received institutional ethics approval (EKEUBA 5/2025, granted 27 April 2025) from the relevant committee of the Bratislava University of Economics and Business. Data handling complied with the General Data Protection Regulation: identifying details were replaced with placeholders during preparation, original audio recordings are not shared, and the lawful basis and retention arrangements were specified in the approved protocol. The anonymized transcript dataset is published under a Creative Commons Attribution 4.0 licence to enable verification and reuse consistent with these protections. The research was carried out within the AI-ImpactSK project, funded by the European Union–NextGenerationEU through the Recovery and Resilience Plan for Slovakia (project No. 09I05-03-V02-00003/2025/VA).

5 The Landscape of AI Adoption: Descriptive Findings

Fifty-eight Slovak organizations participated in the interviews, and before any of the deeper arguments can be made, the data themselves have to be set out. This chapter is descriptive and orienting rather than argumentative. It establishes who these firms are, what artificial intelligence they report using and to what end, and how the interview corpus is distributed across the analytic frame of eight categories and twenty-four themes. The harder interpretive claims are deferred: the governance question to Chapter 6, the measurement paradox to Chapter 7, and the configurations of adoption to Chapter 8. The aim here is narrower: to lay out the descriptive structure clearly enough that those later arguments rest on visible ground, and to flag early the methodological feature that conditions every cross-wave reading to come, namely the difference in coding depth between the two fieldwork waves. All numbers reported below are drawn from the AI-ImpactSK dataset (Štetka et al., 2026). The sample is purposive and modest (N = 58), so these distributions characterize this corpus, not the Slovak enterprise population.

One pattern runs through the chapter and the three that follow it. The most heavily populated themes are those that narrate adoption itself: where AI is applied, why firms turned to it, which technologies they use, how they describe the benefits, and what they plan next. The themes that would record disciplined oversight and disciplined evaluation are thin. Formal governance, regulation, success measurement, and systematic testing are reported far less often, and in two cases barely at all. So the data are dominated by an optimistic adoption account, with governance and measurement at the edges of what firms raised. One caveat travels with this from the outset. These are articulated and coded rates, not a census of practice; a theme that did not surface in roughly half an hour of conversation may still describe something the firm does. The chapter shows the pattern in the descriptive structure. The chapters that follow probe why it holds and what it means.

5.1 Sample composition

The corpus comprises fifty-eight semi-structured computer-assisted telephone interviews conducted in two waves, evenly split with twenty-nine interviews in each (Wave 1, late August to early October 2025; Wave 2, April to June 2026). The even wave balance is a deliberate design feature that supports the temporal comparison developed, with appropriate caution, in Chapter 9. The recruitment details and the steep wave-to-wave decline in response rates are documented in the methods chapter and are not rehearsed here. Figure 5.1 presents the full composition graphically; the discussion below draws out its salient contours.

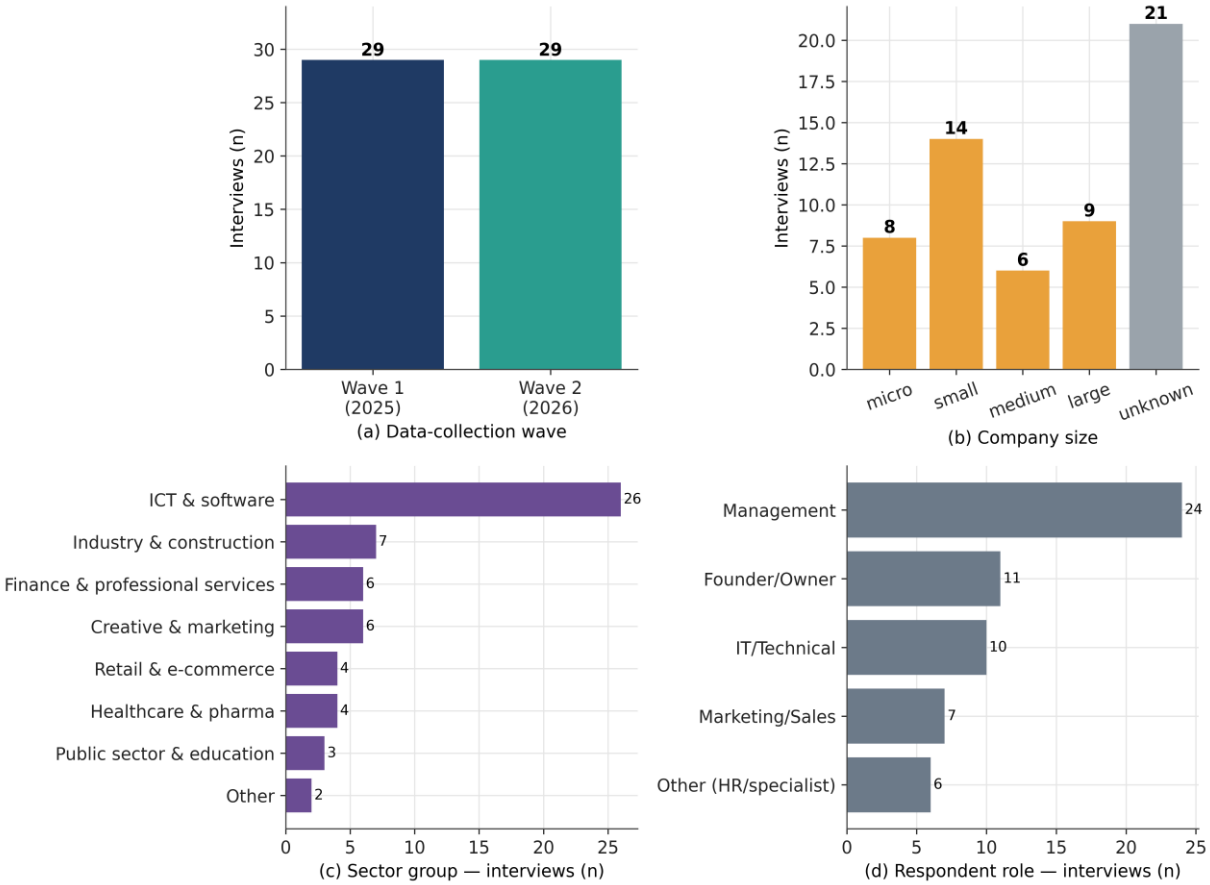


Figure 5.1 Composition of the AI-ImpactSK sample (N=58): (a) wave, (b) company size, (c) sector group, (d) respondent role.[6]

The most consequential feature of the sample is its sectoral skew. Organizations in information and communication technology and software constitute the single largest group at 26 of 58 firms (44.8%), with the remainder distributed across industry and construction (7), finance and professional services (6), creative and marketing (6), retail and e-commerce (4), healthcare and pharmaceuticals (4), public sector and education (3), and a residual "other" (2). This ICT concentration is not incidental to a study of AI adoption. It reflects the reality that technology-intensive firms are both more reachable for a study of this kind and more likely to have something to say about AI. It is, equally, a limitation to carry forward explicitly. An ICT-heavy sample tilts toward early and sophisticated adopters, and several of the descriptive patterns reported below, including breadth of AI use and implementation formality, are partly a function of who is in the room. The sectoral split also furnishes the contrast that Chapter 8 develops, between the ICT firms and the broader-economy adopters that make up the rest of the sample.

By size, the picture is dominated by small and medium-sized enterprises, consistent with the structure of the Slovak economy and with the broader literature on AI in SMEs (Schwaeke et al., 2024). Grouped, the sample contains twenty-eight SMEs and nine large firms (250 or more employees), with twenty-one firms of unknown size. Disaggregated, the SME band comprises fourteen small firms (10–49 employees), eight micro firms (fewer than 10), and six medium firms (50–249). The large block of size-unknown cases is a genuine data limitation rather than an

[6] **Source:** Authors' own processing based on the AI-ImpactSK interview dataset and fieldwork metadata

analytic category, and it is treated as such throughout: where size is used in later inferential comparisons, the unknown group is reported alongside but not silently absorbed into the SME or large bands.

Most respondents were positioned to speak with first-hand authority about AI in their organizations. Roles cluster in management (24), founders or owners (11), and IT or technical staff (10), with smaller numbers in marketing and sales (7), human resources and administration (3), specialist roles (2), and one other. More telling than role is reported involvement with the AI initiatives discussed. Forty-three of the fifty-eight respondents had directly implemented the AI in question, with the remainder split among users (6), project managers (5), and those closely informed (4). This high degree of first-hand involvement strengthens the evidentiary standing of the interview accounts, since these are mostly the people who did the adopting. It also reinforces the practitioner, inside-the-firm vantage point from which the entire dataset is constructed.

5.2 AI technology types and use areas

What kinds of artificial intelligence do these firms actually use? The interview protocol captured AI technology types as a multi-label attribute, allowing each firm to be recorded under as many categories as applied; Figure 5.2 displays the resulting prevalence. The dominant fact is the near-universality of generative AI, reported by 52 of 58 firms (89.7%). This is consistent with the wider characterization of generative AI as a general-purpose technology whose low barrier to entry has driven rapid, broad uptake (Feuerriegel et al., 2024; Banh & Strobel, 2023), and it sets the tone for the entire dataset. When firms in this corpus speak of "using AI," they are most often speaking of generative tools.

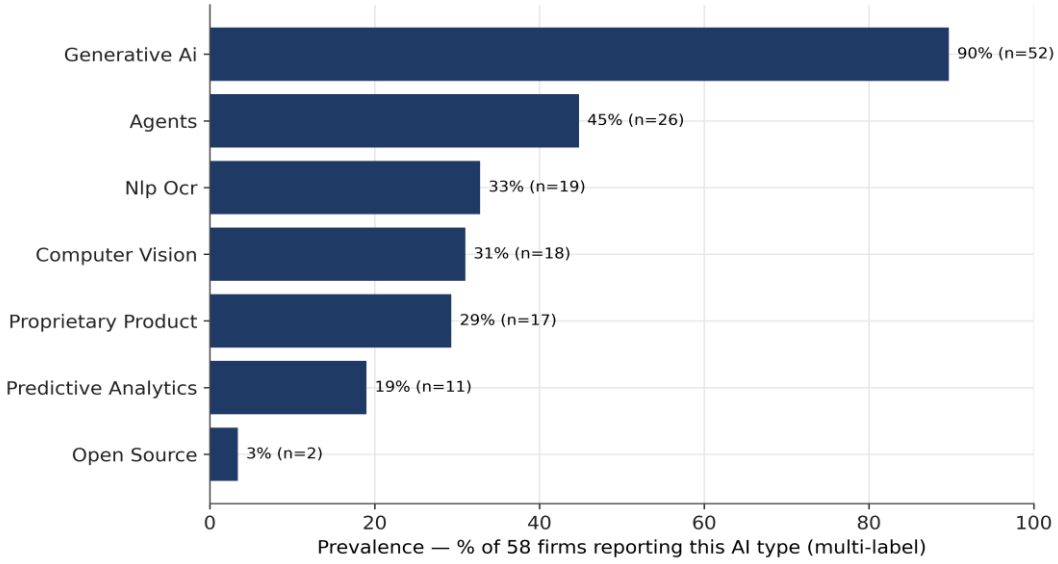


Figure 5.2 Prevalence of AI technology types across 58 firms.[7]

Note: Categories are multi-label; one firm could report more than one AI technology type. Generative AI dominates, while agentic AI is the second most common type.

[7] **Source:** Authors' own processing based on the AI-ImpactSK dataset

Beyond the generative baseline, the technology profile thins markedly. AI agents are the second most prevalent category at 26 of 58 firms (44.8%), a figure that, as Chapter 9 shows, sits at the leading edge of a genuine discursive shift toward agentic systems over the study window (Wang et al., 2024). Natural-language processing and optical character recognition follow at 19 firms (32.8%), computer vision at 18 (31.0%), and proprietary AI products at 17 (29.3%). Predictive analytics, the modality most associated with the earlier "classical" machine-learning era, registers at only 11 firms (19.0%), and open-source AI at a mere 2 (3.4%). The mean number of technology types per firm is 2.5 (median 2). The typical organization here combines a small handful of modalities, most often generative AI plus one or two others, rather than deploying a wide technological portfolio. This modest breadth, and its uneven distribution across the sample, becomes one of the dimensions along which Chapter 8 distinguishes its four adopter configurations.

The use areas to which these technologies are applied are captured in the thematic structure itself, and are addressed in the next section through the coverage of theme B1 (application areas) and the related themes in Category B. The headline is that application is the single best-attested fact in the entire corpus. More firms talk concretely about where they apply AI than about anything else. That application is overwhelmingly oriented toward efficiency and assisted decision-making rather than toward measured transformation. A respondent in retail captured the characteristic emphasis when describing how a generative tool "quickly shows options and lists pros and cons, on the basis of which one can decide better" (F1.01, Wave 1, retail, sales representative). The distance between this fluent articulation of use and benefit, on one hand, and the sparse evidence of formal evaluation, on the other, is the subject of Chapter 7. Here it is enough to note that the descriptive weight of the corpus falls heavily on the side of articulated use.

5.3 The thematic structure: twenty-four themes

The interview corpus was analyzed through framework thematic analysis (Braun & Clarke, 2006; Gale et al., 2013), with AI-assisted coding subsequently reviewed by the authors (Xiao et al., 2023; Mathis et al., 2024). The frame comprises eight categories (A–H) and twenty-four themes, against which 629 coded segments were assigned. One point of analytic reassurance precedes the substantive structure. The computational pipeline used to generate the coverage figures reproduced the independently documented theme-coverage table exactly, theme for theme and count for count. This concordance confirms that the reproducible, code-driven tabulation and the hand-built analytic table describe the same corpus without drift, which licenses confidence in the descriptive numbers that the rest of this chapter and the three that follow rely upon.

Table 5.1 reproduces the full twenty-four-theme coverage. For each theme it reports the parent category, the number of interviews in which the theme appears (n), and the corresponding coverage as a percentage of the fifty-eight interviews. The themes are ordered, as in the source table, from most to least prevalent, so that the descriptive contour of the corpus is legible at a glance. One reading caution applies throughout. Coverage records how often a theme was articulated and coded, not whether the underlying practice exists; a firm that never mentioned testing or measurement in a roughly thirty-three-minute interview may still test or measure. The coverage percentages are therefore best read as lower bounds on practice, most sharply for the themes that depend on a respondent volunteering them.

Table 5.1 Thematic coverage across the twenty-four themes (N = 58 interviews)[8]

Theme code	Theme	Category	n	Coverage %
B1	Application areas	B — Use areas & data	48	82.8
H1	Future plans	H — Future	47	81.0
C1	AI technologies used	C — Technology & sourcing	45	77.6
D1	Approach & phases	D — Implementation	40	69.0
A2	Motivation	A — Context & motivation	39	67.2
F1	Benefits	F — Results & practices	38	65.5
A1	Trigger & initiator	A — Context & motivation	32	55.2
A3	Goals	A — Context & motivation	30	51.7
D2	Teams	D — Implementation	29	50.0
B3	Data & data security	B — Use areas & data	28	48.3
C2	Build/buy/adapt	C — Technology & sourcing	28	48.3
E5	Employee adoption & resistance	E — Challenges	27	46.6
E1	Technical challenges	E — Challenges	24	41.4
F3	Best practices	F — Results & practices	22	37.9
E4	Barriers (general)	E — Challenges	20	34.5
E3	Legal & ethical	E — Challenges	16	27.6
B2	Profile of use	B — Use areas & data	13	22.4
F2	Success measurement & KPIs	F — Results & practices	12	20.7
E2	Organizational challenges	E — Challenges	10	17.2
G1	Governance	G — Governance & regulation	8	13.8
D3	Testing	D — Implementation	5	8.6
G2	Regulation	G — Governance & regulation	4	6.9
C3	Selection criteria & supplier	C — Technology & sourcing	4	6.9
C4	Licensing & open-source	C — Technology & sourcing	2	3.4

The vertical structure of Table 5.1 is itself the chapter's central descriptive finding. The most prevalent themes are application areas (82.8%), future plans (81.0%), AI technologies used (77.6%), implementation approach (69.0%), motivation (67.2%), and benefits (65.5%). All of them narrate the adoption project: what firms do with AI, why they started, and what they expect of it. The disciplines of oversight and evaluation sit far down the list, and there the coverage collapses. Success measurement and KPIs reach 20.7%, governance 13.8%, testing 8.6%, and regulation 6.9%. The single best-attested fact in the corpus, where AI is applied, and one of the least-attested, whether its benefits are measured, are separated by sixty-two percentage points. Two qualifications keep this honest. The measurement figure rests on a direct probe and is well-grounded as a near-lower-bound; the governance and regulation figures record how often respondents raised those topics on their own, since the protocol carried no standalone governance question, and they are better read as measures of salience than of whether oversight practices exist. With that distinction in place, the table gives the clearest single view of what this study characterizes as adoption running ahead of governance and measurement. Figure 5.3 renders the same coverage as a category-colored bar chart, making the contrast between the populated upper themes and the sparse governance-and-measurement tail visible at a glance.

[8] **Source:** Authors' own processing based on the AI-ImpactSK interview dataset and fieldwork metadata

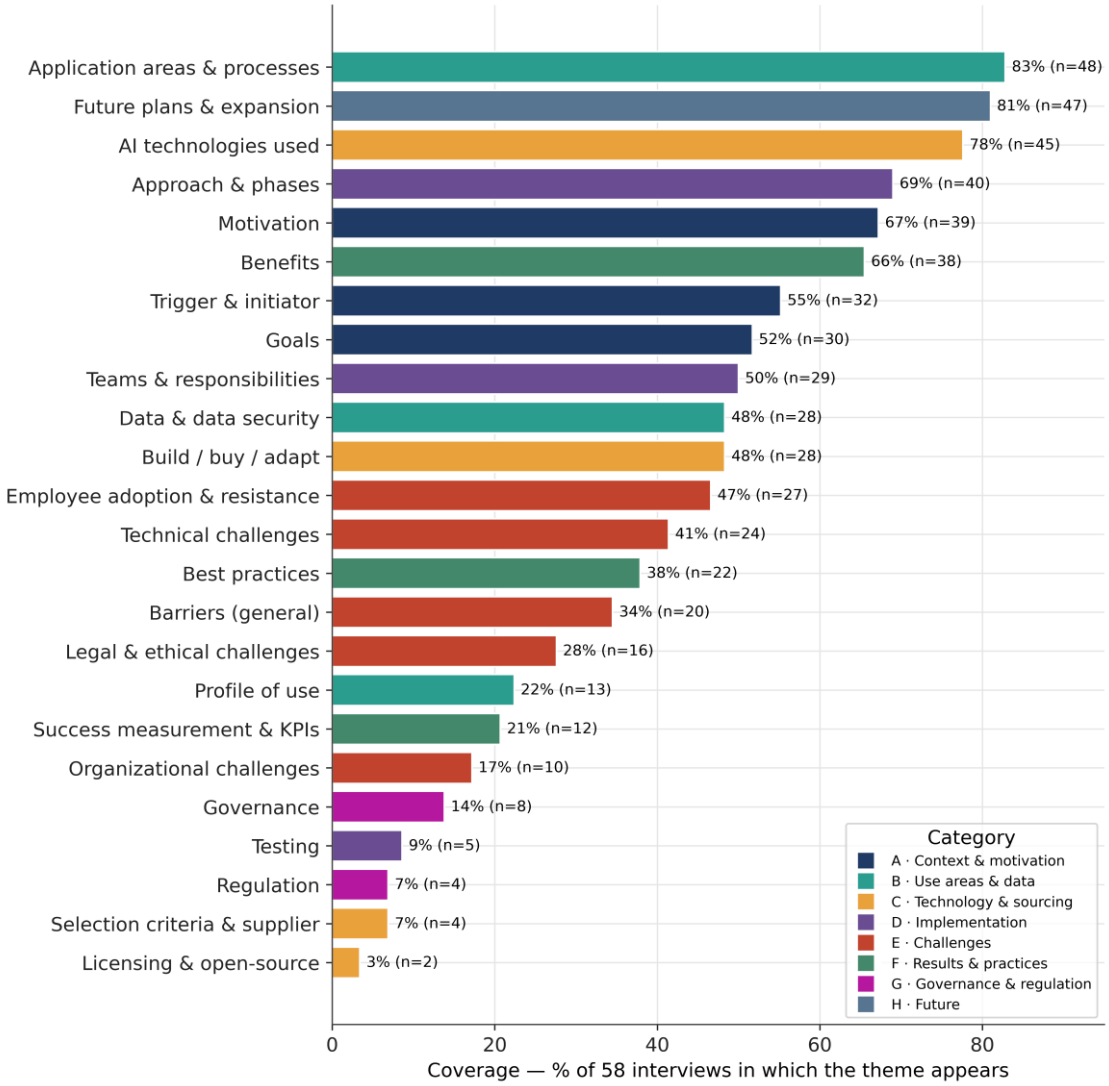


Figure 5.3 Coverage of the 24 themes across 58 interviews, coloured by codebook category. Coverage = share of interviews in which a theme appears at least once.[9]

5.4 The category structure

Aggregating themes to their eight parent categories yields a complementary view, expressed in coded segments rather than interview counts, and displayed in Figure 5.4. By share of all 629 segments, the corpus is led by Category B (use areas and data) at 119 segments (18.9%), Category E (challenges) at 110 (17.5%), and Category A (context and motivation) at 107 (17.0%). A middle band follows: Category C (technology and sourcing) at 81 segments (12.9%), Category D (implementation) at 76 (12.1%), and Category F (results and practices) at 74 (11.8%), with Category H (future) at 50 segments (7.9%). The head of this distribution reinforces the thematic reading. Discussion concentrates on use, on the difficulties of using, and on the motivations for using.

[9] **Source:** Authors' own processing based on the AI-ImpactSK framework-coded interview dataset

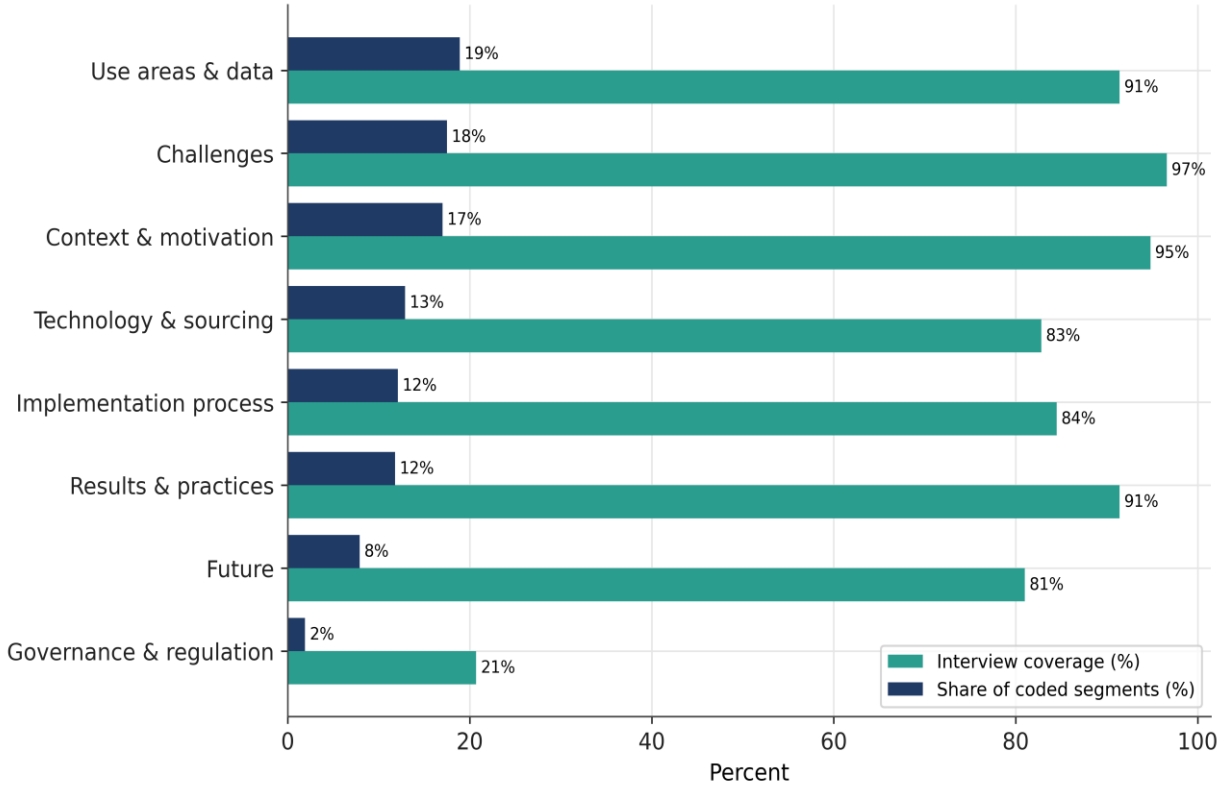


Figure 5.4 Eight codebook categories by interview coverage and share of the 629 coded segments. Governance & regulation (G) is marginal on both measures.[10]

The tail is where the decisive observation lies. Category G (governance and regulation) accounts for just 12 of 629 segments, 1.9% of the entire coded corpus. By interview coverage the marginality is equally stark. Categories A through F and H each appear in roughly forty-seven to fifty-six of the fifty-eight interviews; Category G is touched in only twelve. As a region of the analytic frame, governance was seldom foregrounded in firms' own accounts of adoption. It surfaced in fewer than a quarter of interviews and contributed fewer than one segment in fifty. The reason matters and is taken up directly in Chapter 6: governance and regulation entered the record only when a respondent volunteered them, since the protocol contained no question that asked about them head-on. The low count is therefore a fact about salience and discourse, about what firms chose to talk about unprompted, not a finding that oversight practices are absent. Chapter 6 examines that governance gap, situates it against the principles-to-practice problem (Mittelstadt, 2019), and asks whether the low salience is confined to any particular firm subgroup. For present purposes the point is narrower: in both the theme table and the category aggregation, governance and regulation are the topics these firms least often raised.

5.5 Theme co-occurrence: the adoption-narrative core

The coverage tables describe how often each theme appears; they do not describe how themes travel together. To see the relational structure of the corpus, theme co-occurrence was examined at the level of individual participants and rendered as a radial map (Figure 5.5), with the strength of association between any two themes summarized by a participant-level Jaccard coefficient. The map reveals a dense, mutually reinforcing cluster that may be called the

[10] **Source:** Authors' own processing based on the AI-ImpactSK framework-coded interview dataset

adoption-narrative core: application areas (B1), future plans (H1), AI technologies used (C1), motivation (A2), implementation approach (D1), and benefits (F1) co-occur tightly and repeatedly. The strongest links bind the themes one would expect a firm telling a coherent adoption story to invoke together: application areas with future plans (both n = 40, Jaccard .73), application areas with benefits (n = 36, Jaccard .72), motivation with implementation approach (Jaccard .72), application areas with AI technologies used (n = 38, Jaccard .69), and AI technologies used with future plans (n = 37). These firms narrate a single connected story. Here is what we use, here is why we adopted it, here is the benefit we perceive, and here is what we plan next.

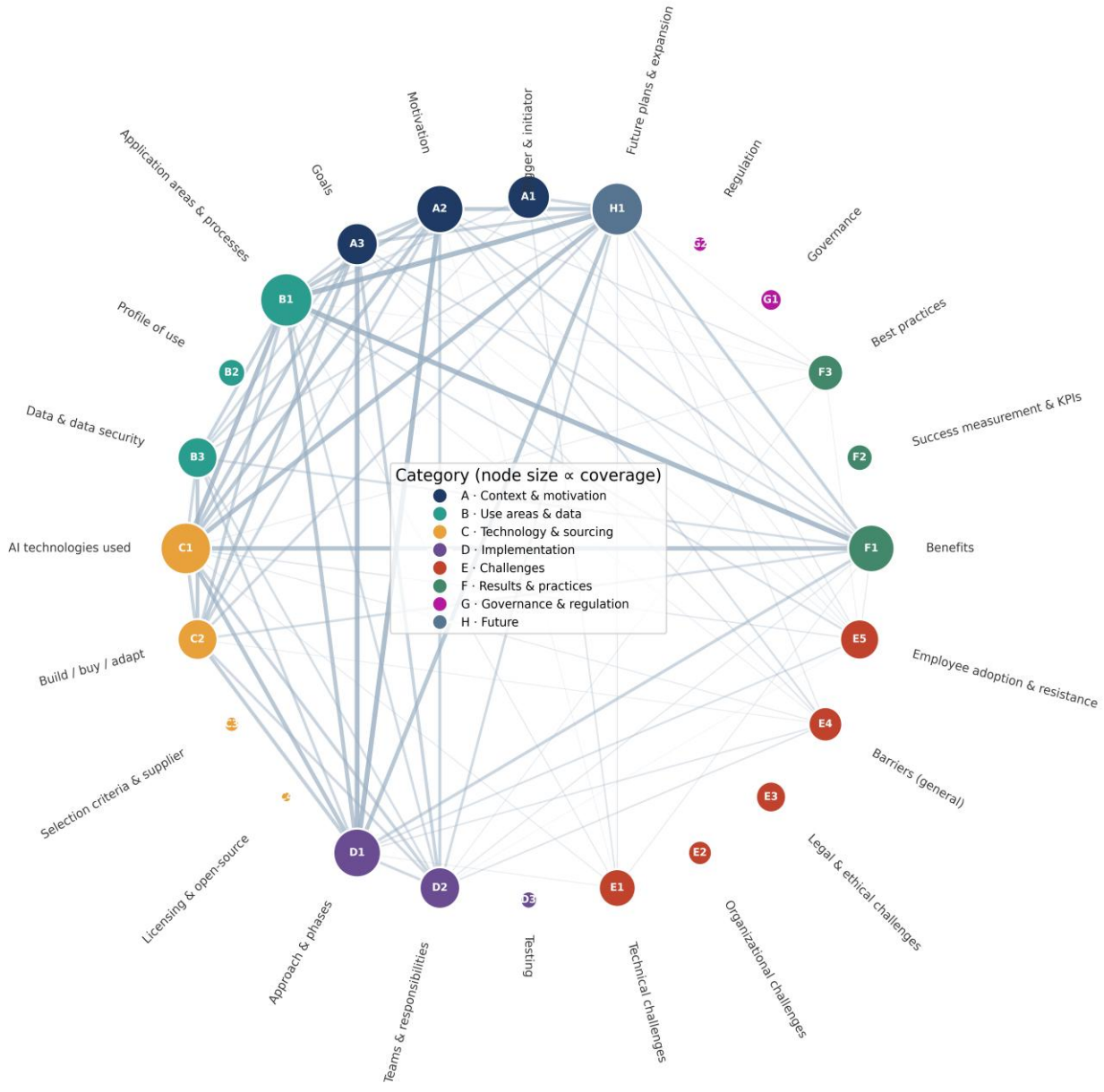


Figure 5.5 Radial co-occurrence map of the 24 themes[11] (chords where participant-level Jaccard ≥ 0.30; node size ∝ coverage). A densely connected adoption-narrative core (B1, H1, C1, A2, D1, F1) contrasts with near-isolated governance, regulation, testing and KPI themes.

[11] **Source:** Authors' own processing based on the AI-ImpactSK framework-coded interview dataset

Note: Chords are shown for participant-level Jaccard coefficients ≥ 0.30; node size is proportional to coverage.

What stands outside that core is as informative as what lies within it. The governance, regulation, testing, and success-measurement themes are near-isolated on the radial map, with few or no strong links to the dense cluster or to one another. They are not woven into the adoption narrative; they sit at its edge, invoked sporadically and in relative disconnection. The relational picture thus converges with the coverage picture from a different direction. The optimistic adoption account is not only the most frequent material in the corpus but also its most tightly integrated, while oversight and evaluation are both infrequently raised and structurally peripheral. This is the descriptive foundation on which Chapters 6 through 8 build: the core that Chapter 8 reads as an "informality-as-default" configuration, and the periphery that Chapters 6 and 7 examine as the governance gap and the measurement paradox respectively.

5.6 Coded-segment totals and the coding-depth difference

A final descriptive feature must be introduced here, plainly and prominently, because it conditions every comparison the monograph later draws between the two waves. The 629 coded segments are not evenly distributed: 238 fall in Wave 1 and 391 in Wave 2. Expressed per interview, Wave 1 averages 8.21 segments (SD 0.73; median 8; range 7–10) while Wave 2 averages 13.48 (SD 1.18; median 14; range 10–15), for an overall mean of 10.84; a Mann–Whitney comparison confirms the Wave 2 excess at $p < 1 \times 10^{-7}$. Wave 2 interviews were coded roughly sixty-four percent more densely than Wave 1 interviews. Part of this difference traces to the respondents themselves. Wave 2 participants spoke considerably more, a mean of 1,123 words per interview against 696 in Wave 1 (Mann–Whitney $U = 203$, $p = .0007$, rank-biserial 0.52), so later respondents simply offered more codable material. The consequence for analysis is the same whatever the cause.

The implication is methodological and is flagged here for cautious use throughout. Because coverage is built from coded segments, any naive comparison that shows a theme "rising" from Wave 1 to Wave 2 may be registering nothing more than the denser coding of the second wave rather than a real change in firm practice or discourse. Cross-wave coverage shifts are therefore to be treated as hypothesis-generating, and Chapter 9 takes up the temporal question only after separating genuine signal from coding-depth artifact. The most important piece of genuine signal is the rise of agentic-AI discourse, which survives triangulation against coding-independent text. The same caution applies, with even greater force, to the intensity of coding, which is so strongly confounded with wave that it is not used for any cross-wave substantive claim in this monograph.

One robustness observation belongs alongside the caution, because it protects the two findings the monograph most depends upon. If denser coding mechanically inflated coverage, it would inflate it everywhere, including in the very themes that record measurement and governance. It does not. The articulation of measurement was, if anything, slightly lower in the more densely coded Wave 2 (17.2%) than in Wave 1 (24.1%), and any governance theme shows the identical pattern (24.1% in Wave 1 against 17.2% in Wave 2). The thin coverage of governance and measurement is thus not an artifact of how the two waves were coded. It is a stable feature of the corpus that persists precisely where the coding-depth difference would have worked against it. That is reassuring for what follows. The peripheral position of oversight and evaluation, established descriptively in this chapter, is robust, whereas claims of cross-wave growth are not to be advanced without the coding-depth caveat firmly attached.

5.7 The landscape in summary

Read together, the descriptive findings sketch a coherent picture. The sample is ICT-skewed, SME-heavy, and reported by people who for the most part did the adopting themselves. Generative AI is all but universal among them, layered in the typical firm with one or two further modalities to a mean breadth of 2.5 technology types. The thematic structure, verified against the authors' documented table, is heavily weighted toward the narration of adoption: application, motivation, technology, benefit, and future plans dominate both the coverage tables and, relationally, the dense co-occurrence core. Governance, regulation, testing, and success measurement are sparse in coverage, marginal in segment share, and near-isolated in the co-occurrence map. Two of these gaps are not alike, and the difference is worth keeping in view. Measurement was probed directly, so its 20.7% figure is a near-lower-bound on a practice that was asked about; governance was not, so its scarcity is best read as low salience in firms' unprompted accounts rather than evidence that oversight is absent.

That through-line sets up the three deep-dive chapters without prejudging them. The thin, low-salience governance region becomes the subject of Chapter 6, which asks how prevalent explicit governance really is, what its scarcity in the record reflects, and whether that scarcity varies across firm types. The 65.5% articulation of benefits against the 20.7% measurement of them becomes the paradox dissected in Chapter 7. And the tightly integrated adoption-narrative core, together with the uneven breadth and formality across the sample, becomes the configurational analysis of Chapter 8. What the firms in this sample describe is an optimistic adoption account with governance and measurement at its edges. The chapters that follow ask what that arrangement costs, why it persists, and how it varies across the firms that report it.

6 The Governance Gap

This chapter examines the first of the monograph's three central empirical gaps: the low salience of explicit, codified AI governance in firms' own accounts of adoption. It asks not whether governance practices are entirely absent from Slovak enterprises, but how often firms themselves foreground policies, assigned responsibility, oversight mechanisms, regulation, or documented controls when describing their use of AI. This distinction is essential, because the interview protocol did not contain a standalone governance question. Governance and regulation entered the record only when respondents raised them under adjacent prompts, especially legal and ethical challenges or expected future regulation. The chapter therefore treats the governance figures as salience rates rather than as a census of practice.

The analysis proceeds in three steps. First, it documents the low coverage of formal governance and regulation in the interview corpus. Second, it tests whether this low salience is concentrated in particular firm groups or instead appears across the sample. Third, it examines what takes the place of codified governance, showing that firms often voice data-security, legal, ethical, and output-verification concerns, but usually as diffuse informal vigilance rather than as formal organizational control. The chapter argues that these firms are not indifferent to AI risk; rather, their concern has not yet been converted into the policies, roles, and routines that responsible-AI frameworks would recognize as governance.

6.1 Research question, hypotheses, and the shape of the argument

Governance almost never came up on its own. Across 58 interviews about how firms adopt and use AI, codified policy, assigned responsibility, and regulatory awareness surfaced only when a respondent reached for them unprompted. That fact governs everything in this chapter, and it must be stated before any number is. The interview protocol carried no standalone governance or policy question. Governance and regulation entered the record only when a respondent volunteered them, typically under the prompt about legal and ethical challenges or the forward-looking question about expected AI regulation. The figures reported here therefore measure how often governance was spontaneously raised, not whether governance practices exist. They are salience rates, or discourse rates, and the chapter reads them as such throughout.

Chapter 5 established that AI has diffused widely and informally through these firms, entering from the bottom up and settling into everyday work ahead of any deliberate organizational scaffolding. This chapter narrows that observation to governance. It asks the third research question, RQ3, in two parts: how often does explicit, codified AI governance surface in firms' own accounts of adoption, and is that salience associated with firm characteristics such as size or sector? The question matters because the firms in the sample operate under an emerging European governance regime, are studied within an EU-funded research programme, and have adopted precisely the generative and agentic tools that the responsible-AI and risk-management literatures single out as most in need of structured oversight (Feuerriegel et al., 2024; National Institute of Standards and Technology, 2024). Where governance rarely surfaces and adoption is everywhere, that asymmetry is worth taking seriously as a feature of the adoption regime rather than as a peripheral curiosity. Codified organizational AI governance, in the sense used here, denotes the structures, processes, and roles through which a firm directs and controls its AI use (Mäntymäki et al., 2022); it is that codified apparatus, not the underlying concern, whose low salience the chapter documents.

Three hypotheses give the question analytic shape, and the chapter tests each in turn. The first, H3a, holds that explicit governance and regulation are rarely foregrounded in the corpus: when firms describe adopting and using AI, formal policy, codified responsibility, and regulatory awareness are seldom part of the account. The second, H3b, asks whether that low salience is associated with subgroup membership, in particular whether it concentrates among small firms or non-technical sectors. The third, H3c, holds that governance, where it surfaces at all, appears as diffuse and informal vigilance rather than as codified practice: the responsible-AI values of data care, output verification, and caution circulate in firms' discourse as scattered attention rather than crystallizing into the policies, roles, and controls that governance frameworks treat as the substance of the function. Across the three, the chapter advances a single reading, stated in full in §6.8: in these firms a recognizable ethic of caution circulates informally, while the codified apparatus of AI governance rarely surfaces in their own narration of adoption.

The evidential basis is the purposive, ICT-skewed sample of 58 interviews conducted in two waves (Štetka et al., 2026), and the analysis is exploratory-confirmatory rather than confirmatory in the strict inferential sense. The N = 58 caveat is repeated wherever an inferential test is reported, and the coding-depth artifact disclosed in Chapter 4, the fact that Wave 2 was coded roughly 64% more densely than Wave 1, is held in view wherever a cross-wave comparison appears. Coverage, defined at the case level as the proportion of interviews (out of 58) in which a theme appears at least once, is again the primary descriptive quantity. Because governance was not probed directly, each coverage figure is best read as an articulated rate, a lower bound on the underlying practice: a firm that did not raise governance in a roughly 33-minute interview may still have some in place, so absence of mention is not evidence of absence. Each estimate is paired with a 95% Wilson score interval so that the precision of these small-sample proportions is visible rather than implied.

6.2 The low salience of codified governance (H3a)

The first and most direct finding is one of low salience. Across the 58 firms, an explicit governance theme, coded where a respondent described a defined policy, an assigned owner, an oversight mechanism, or a documented control for AI use, appears in 8 interviews, a coverage of 13.8% (95% Wilson score interval 7.2–24.9). Explicit regulation, coded where a respondent engaged with the legal-regulatory environment for AI as something bearing on their own practice, is rarer in the talk, surfacing in 4 interviews, or 6.9% (95% Wilson score interval 2.7–16.4). At the level of the coded corpus the figure is smaller again: the entire governance-and-regulation category (category G) accounts for just 12 of the 629 coded segments, or 1.9% of all coding. In a corpus where use areas and data draw 18.9% of segments, challenges 17.5%, and context and motivation 17.0%, the category charged with capturing how firms steer and constrain their AI is the smallest of the eight by a wide margin. The intervals are correspondingly wide, as is unavoidable for proportions built on single-digit counts in a sample of 58. Read as discourse rather than census, the message is that governance was seldom the thing firms chose to talk about when describing their adoption, even where the protocol opened a door to it.

A governance index built to summarize the codified signals a firm volunteered tells the same story in one number. Ranging from 0 to 4 across its component governance and oversight signals, the index has a sample mean of 0.97 out of 4. The average firm, on this measure, articulated something close to one of the four codified signals the index can detect, and many articulated none. Two readings of that number are available and the data cannot adjudicate between them: a firm may genuinely lack the apparatus, or it may hold some and simply not have surfaced it in a half-hour conversation that never asked. Hypothesis H3a, which predicts

that explicit governance and regulation are rarely foregrounded, is supported on every measure available, theme coverage, segment share, and the summary index. Figure 6.1 presents the governance, ethics, and security coverage estimates with their intervals, and it is to the contrast in that figure that the argument turns.

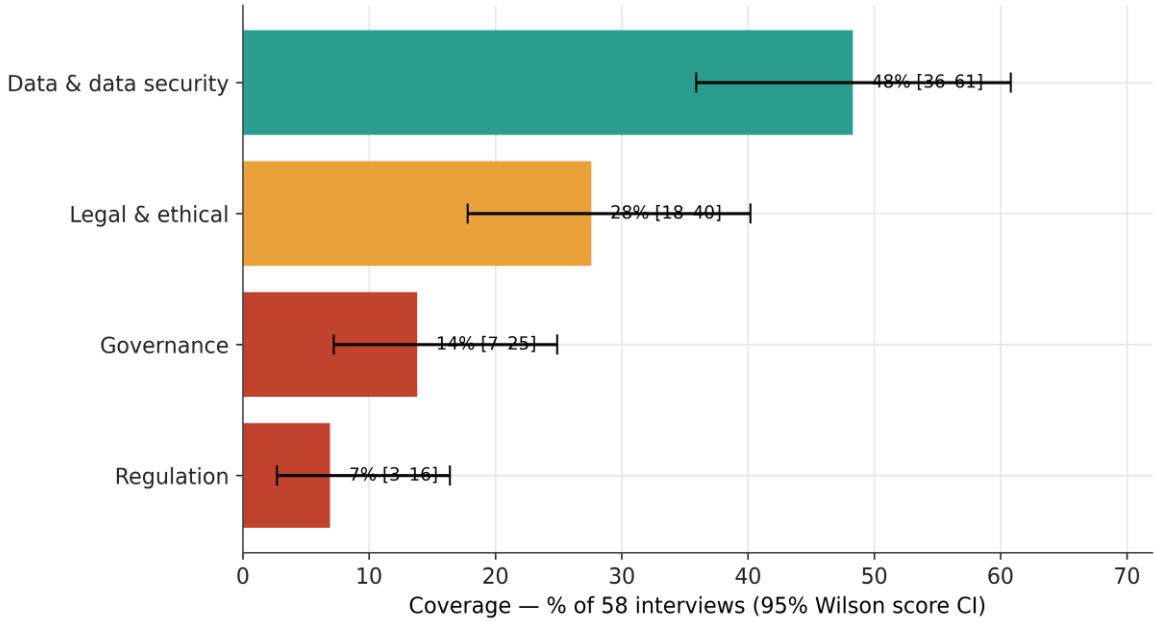


Figure 6.1 Coverage of formal governance (G1), regulation (G2), legal/ethical (E3), and data-security (B3) themes with 95% Wilson score intervals. Formal governance and regulation rarely surfaced; concern was voiced mostly as data security.[12]

6.3 Risk talk without a spine: Ethics and security against governance

What surrounds governance in the talk is more telling than governance alone. Two adjacent themes, both concerned with the harms that governance would ordinarily address, are far more present in the corpus. A legal-and-ethical theme (E3), capturing respondents' engagement with the legal and ethical dimensions of AI use, appears in 16 interviews, a coverage of 27.6% (95% Wilson score interval 17.8–40.2), roughly double the reach of explicit governance. A data-and-data-security theme (B3) is more prevalent again, surfacing in 28 interviews, or 48.3% (95% Wilson score interval 35.9–60.8), which makes data security one of the more widely shared concerns in the study and the most common point of contact between these firms and the risk surface of AI. Figure 6.1 sets the four quantities side by side: data security at 48.3%, legal-ethical at 27.6%, governance at 13.8%, and regulation at 6.9%. That descending staircase is the empirical heart of the chapter.

The firms are not oblivious to the risks of AI. Nearly half raise data security and more than a quarter raise legal and ethical questions. What rarely surfaces alongside that awareness is its conversion into structure. Firms name the harms that governance exists to manage, the exposure of sensitive data, the legal and ethical exposure of automated outputs, but in the main they do not narrate the codified mechanisms that would constitute governance proper: written policy, assigned ownership, documented review, auditable control. The concern is articulated; the structure is not narrated alongside it. This is the pattern Mittelstadt et al. (2016) anticipate

[12] **Source:** Authors' own processing based on the AI-ImpactSK framework-coded interview dataset

when they map algorithmic decision-making onto a set of recurring ethical concerns that organizations gesture toward, and it is the pre-codification stage that the corporate-governance and internal-audit literatures try to move organizations beyond (Cihon et al., 2021; Raji et al., 2020). The decoupling of espoused concern from organizational structure is also the mechanism that institutional sociology has long described, in which formal policy and day-to-day practice drift apart (Meyer & Rowan, 1977; Bromley & Powell, 2012); §6.7 develops that reading. On present evidence the firms sit at the pre-codification stage in their own accounts: aware enough to worry aloud, not visibly structured enough to govern. The verbatim record in §6.6 shows what that awareness sounds like; first the chapter asks whether the low salience is associated with who happens to populate the sample.

6.4 No detected subgroup difference (H3b)

A natural objection to the headline finding is that it might be an aggregation artifact. If volunteered governance were concentrated in large firms with the slack to formalize it, or in technical firms fluent in the relevant standards, then the low overall coverage might reflect a sample weighted toward small, bottom-up adopters, and the right move would be to condition on firm characteristics. Hypothesis H3b examines whether the presence of any explicit governance or regulation (a gov-any indicator combining the two themes) is associated with firm size or sector.

No difference was detected. Among firms whose size could be classified, large organizations (250 or more employees) show gov-any in 3 of 9 cases, or 33.3%, against 4 of 28 small-and-medium enterprises, or 14.3%. The direction is the intuitive one, larger firms somewhat more likely to have articulated something codified, and the odds ratio of 2.93 is not trivial in magnitude. But the comparison is badly underpowered: Fisher's exact test returns $p = .327$, and the result does not survive false-discovery-rate correction ($q = .41$). Critically, this null is not evidence that the salience gap is uniform across firm types. With only nine large firms, the test has little chance of detecting even a real difference of this size, so the data cannot establish that low governance salience is evenly distributed; they can only say that no subgroup difference was detected here. The point is not that size is irrelevant in principle, the governance literature gives good reasons to expect larger firms to govern more, but that within these 58 firms size does not account for how rarely governance surfaced. The sector comparison is flatter still. Comparing ICT firms against the rest, the odds ratio for gov-any is 0.87 with $p = 1.0$: the technically sophisticated, early-adopting ICT firms that dominate the sample are no more likely than others to have foregrounded codified AI governance, and the point estimate tilts marginally the other way.

Two implications follow. First, H3b finds no detected subgroup difference: the low salience of governance is not localized in a subgroup that a covariate could isolate, though the size comparison is too small (Large $n = 9$) to license a stronger claim that the gap is uniform. Second, the absence of a visible sector effect is itself worth recording. One might expect ICT firms, closest to the technical frontier, most fluent in the language of standards and controls, best resourced in technical talent, to be the ones who foreground AI governance. In this sample they do not. That the firms most capable of operationalizing governance are no more likely to have articulated it is at least consistent with the reading that the binding constraint is not capability but the absence of a felt demand for codified governance. This anticipates the institutional interpretation in §6.7. For now it is enough that no subgroup correction the data permit alters the basic picture.

6.5 Governance as diffuse vigilance, not codified practice (H3c)

If governance rarely surfaces, and does so without an obvious subgroup pattern, a further question presses: what is present where codified governance is not narrated? Hypothesis H3c proposes that the responsible-AI impulse has not vanished from these firms but has taken a different form, diffuse and informal vigilance rather than codified practice, and the study tests this through a triangulation that steps outside the coding frame. The manual G1 governance code is an interpretive judgment by the analysts, so it can in principle be checked against an independent, coding-agnostic measure of governance talk built directly from the transcript text.

That measure is a governance lexicon: a fixed vocabulary, governance, policy, compliance, GDPR, audit, oversight, and their cognates, counted in respondents' own speech without reference to how any passage was coded. Two results emerge. First, the lexicon does not align with the formal governance code. The correlation between a firm's governance-lexicon intensity and whether it received the explicit G1 code is a point-biserial r of −0.20, not statistically distinguishable from zero and running in the negative direction. Where the manual code captures something the analysts judged to be codified governance, the raw density of governance vocabulary does not track it. Firms invoke the vocabulary of oversight diffusely, across contexts that do not correspond to the presence of an actual governance apparatus, so lexical density and codified practice come apart. That is what one would expect if governance, for these firms, surfaces as scattered concern rather than as identifiable structure, and what one would not expect if the governance code were tracking vocabulary-dense governance documents and policies.

Second, the governance lexicon is flat across waves. Its mean intensity is 1.0 in Wave 1 and 0.9 in Wave 2, a difference a Mann–Whitney test renders non-significant ($p = .95$). This flatness matters twice over. Substantively, the informal vigilance the lexicon captures did not intensify over the year separating the waves: even as agentic AI discourse rose sharply across the same interval, a genuine, coding-independent shift documented in Chapter 9, governance talk held constant at a low level. The discursive maturation the study detects elsewhere did not extend to governance. Methodologically, the flatness is a quiet control. Because Wave 2 was coded roughly 64% more densely than Wave 1, any measure drawn from the coding frame is open to the suspicion that apparent change reflects coding effort rather than substance. The governance lexicon, drawn from text rather than codes, is immune to that suspicion, and it shows no change. Figure 6.2, which assembles the study's code-versus-text triangulations, situates this governance result alongside its counterparts for measurement and agentic discourse. The contrast is the informative part: where the measurement code aligns with its lexicon and the agentic code aligns with agentic text, the governance code stands apart, the one domain where the formal code and the underlying talk diverge. H3c is supported.

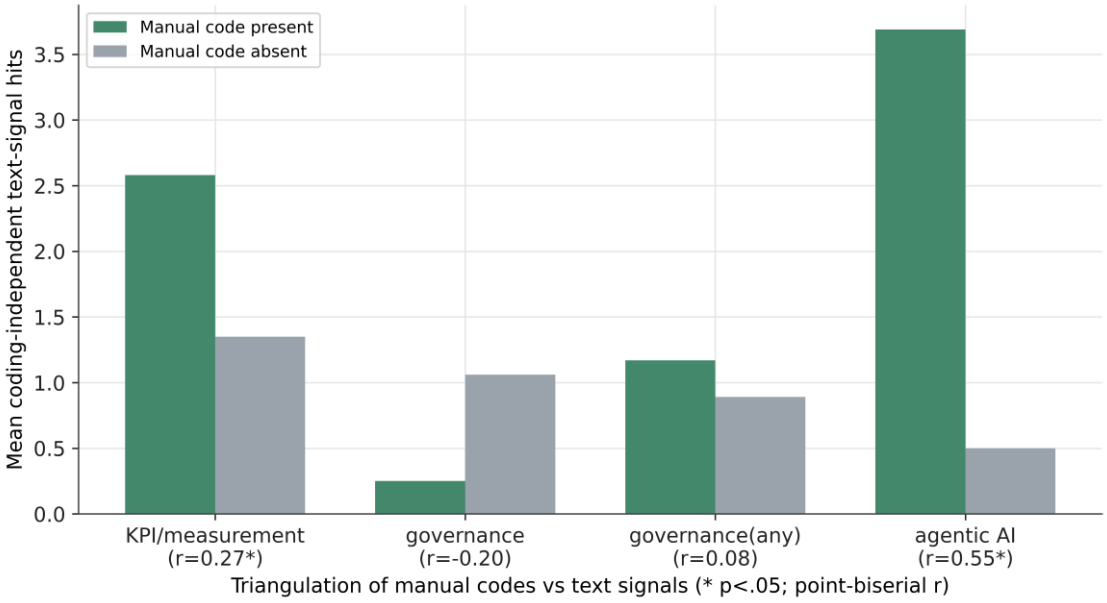


Figure 6.2 Validation of manual coding against independent lexical signals. Measurement and agentic codes are text-valid; governance talk is diffuse and weakly tied to the formal code.[13]

6.6 The texture of informal vigilance

The quantitative anatomy of the gap takes on meaning in the voices behind it, and the curated verbatim record makes the texture of informal vigilance legible. The most economical statement comes from a Wave 2 ICT firm: "We do not have a precise policy on how to use AI" (F2.02, Wave 2, ICT, SME). The admission is matter-of-fact rather than anxious, offered as a description of how things stand rather than a confession of failure. What sharpens it is the firm's own situation. This is not an organization indifferent to formal controls: it holds ISO 27001 certification and is pursuing ISO 45001. So the point is not that it lacks any policy apparatus but that it has no AI-specific policy despite holding a formal information-security certification. The codified governance of AI is a step this certified firm has not yet taken, or at least did not narrate, and the contrast between general security formality and AI-specific informality is the more accurate way to read the quote.

The same firms that lack an AI policy are not careless. What appears in place of codified governance is a set of personal, improvised practices of caution, vigilance enacted by the individual user rather than mandated by the organization. A Wave 1 ICT firm describes both halves of it. On the reliability of outputs: "I always check the outputs because they are not one hundred percent reliable" (F1.06, Wave 1, ICT, SME). On data exposure: "We are careful that our data is not used for model training" (F1.06). Two of the central concerns of AI governance, output verification and data protection, are being actively managed, but managed as the habits of a careful practitioner rather than as organizational controls. The verb tenses are revealing. "I always check," "we are careful": these are dispositions, not procedures, describing how a conscientious person behaves rather than how a governed organization is structured. The values are present and operative; the apparatus that would make them durable, transferable, and auditable is not narrated.

[13] **Source:** Authors' own processing based on the AI-ImpactSK framework-coded interview dataset and computational text analysis

The data-security concern that registers in nearly half the sample acquires concrete stakes in the words of a larger healthcare firm: "The biggest challenge is working with sensitive data" (F2.27, Wave 2, healthcare, Large). For an organization handling health information, the data-protection surface of AI is acute, and framing it as "the biggest challenge" signals genuine salience. What stands out, against §6.4, is that even this acute, sector-specific exposure, in a large firm, in a regulated domain where one would most expect codified controls, surfaces as a recognized challenge rather than as a described governance response. The concern is named with clarity; the structure to manage it does not appear beside it. Whether that structure exists and went unmentioned, or does not exist, the interview cannot say. The aggregate pattern, awareness outrunning narrated structure, is visible at the level of the individual case, and it holds even where the stakes are highest.

A related strand of vigilance concerns the trustworthiness of AI outputs, and here the firms converge on a single technical anxiety. "The biggest problem is so-called hallucination," reports one Wave 1 ICT firm (F1.03, Wave 1, ICT), while another Wave 1 firm states the operative norm bluntly: "You must not blindly trust its outputs" (F1.01, Wave 1). A year later the concern persists in identical terms: "Hallucination is thus the primary problem" (F2.25, Wave 2, ICT, Large). That the same problem is named as "the biggest" and "the primary" across both waves, by different firms, underscores its salience and the constancy documented lexically in §6.5. Hallucination is exactly the risk the generative-AI risk literature foregrounds, enumerated among the confabulation and information-integrity harms distinctive to generative systems (National Institute of Standards and Technology, 2024). These firms perceive the risk and respond to it, but they respond through the individual discipline of checking outputs ("I always check the outputs," F1.06) rather than through an organizational mechanism of validation, review, or sign-off. This verify-the-output posture has a name in behavioural research: it is the vigilance of users who have learned not to defer to an imperfect system, the kind of human override that follows from seeing algorithms err (Dietvorst et al., 2015). Across these quotations a consistent grammar emerges. The firms manage AI the way a careful individual manages a powerful but unreliable tool, through attention and habit, rather than the way an organization manages a material risk, through policy and control. That grammar is the qualitative substance of what the salience figures point to.

6.7 Interpretation: Principles without practice, shadow without sanction

The pattern has a name in the responsible-AI literature, and naming it correctly is the first step in interpretation. Mittelstadt (2019) argues that high-level ethical principles, imported into AI from the model of medical ethics, cannot by themselves guarantee ethical outcomes, because AI lacks the common aims, professional duties, proven translation methods, and robust accountability structures that would convert principles into reliably ethical practice. His argument predicts that organizations will internalize the values of responsible AI while failing to instantiate them as binding structures, so that principles float free of practice. The firms here are consistent with that prediction, observed bottom-up at the level of the individual enterprise. They articulate the values: data care ("we are careful that our data is not used for model training"), output verification ("I always check the outputs"), epistemic humility before unreliable systems ("you must not blindly trust its outputs"). What they do not narrate is the apparatus that would bind those values, the policy, the assigned responsibility, the documented control. This is the principles-to-practice gap of the responsible-AI literature (Jobin et al., 2019; Morley et al., 2020), appearing here not as a shortfall of tools available to willing adopters but as the prior condition

of firms that have not yet, in their own telling, reached for tooling at all. Where the operationalization literature inventories the methods that would close the gap and laments their thin uptake (Morley et al., 2021), the present evidence locates these firms one step further back, with the move from value to structure largely not yet evidenced, whether because it has not been attempted, lies beyond current capability, or simply has not yet been undertaken.

The second frame is shadow AI, which captures the demand-side dynamic the principles-to-practice gap leaves unexplained. The descriptive chapters established that adoption here is overwhelmingly bottom-up: tools enter through individual employees solving task-level problems, ahead of and independent of organizational sanction. This is the structure of shadow IT (Silić & Back, 2014) transposed to the generative-AI era as shadow AI, the unsanctioned use of consumer-grade generative tools at work, driven by acute task pain and high concealability, carrying novel data-privacy, confidentiality, and governance-drift risks even as it delivers genuine bottom-up productivity (Silić et al., 2025; Waters-Lynch et al., 2025). The governance gap documented here is the natural companion of that adoption structure: when AI arrives from below, the codified governance that would ordinarily accompany a sanctioned, top-down deployment has no obvious occasion to be created. The individual vigilance these firms exhibit, the careful checking, the wariness about data, is the self-protective behaviour of users operating in a shadow regime without an organizational backstop. The framing matters for what follows, because the shadow-AI literature reframes the gap as a governance-design problem rather than only a deficiency to deplore. Waters-Lynch et al. (2025) argue that covert generative-AI use can be deliberately surfaced and governed rather than suppressed, turning shadow user innovation into a renewal of dynamic capabilities; the implication, developed in the implications chapter, is that the right response is to channel the informal vigilance that already exists into lightweight structure, not to demand a heavyweight apparatus these firms will not build.

The third frame is institutional, and it speaks to why governance is uniformly low in salience across the sample. Institutional theory holds that organizations adopt structures less for their efficiency than for the legitimacy they confer, converging on common forms through coercive, mimetic, and normative pressures (DiMaggio & Powell, 1983). Its diagnostic value here lies in what it predicts when those pressures are weak. Coercive pressure to govern AI, binding regulation actually enforced against these firms, was during the study window still nascent: the EU AI Act applies in phases from 2025 onward (European Parliament & Council of the European Union, 2024), and its obligations had not yet hardened into a felt compliance demand, a reading consistent with the marginal 6.9% regulation coverage. Mimetic pressure, the imitation of visible governance exemplars, has little to work on when codified AI governance is itself rarely visible across the field: there is no modal governed firm to copy. Normative pressure, professional and consultant-driven norms prescribing AI governance as standard practice, has not yet diffused to the Slovak SME shop floor, where the internal-audit functions, board-level ownership, and ethics-as-a-service capacity the governance literature presupposes are largely absent (Cihon et al., 2021; Morley et al., 2021). A complementary mechanism sharpens the reading. Even where a firm holds formal certifications, as the ISO 27001-certified firm in §6.6 does, the institutional account predicts that espoused policy and operational practice can decouple, with formal structure adopted for legitimacy while day-to-day work proceeds on its own logic (Meyer & Rowan, 1977; Bromley & Powell, 2012). The low salience of AI-specific governance, even amid general security formality, is what means-ends decoupling looks like in this setting. With external pressure weak and the means-ends link loose, the prediction is not a few laggards against a governed norm but field-wide low salience of codified governance, because little in the firms' institutional environment yet compels its construction or its narration. That is also why the ICT

firms are no more likely to foreground governance than the rest (§6.4): capability is not the binding constraint when institutional demand is itself missing.

These frames converge when read against the governance frameworks the firms have not adopted. The NIST AI Risk Management Framework organizes AI risk management around four functions, Govern, Map, Measure, and Manage, and places Govern first as the cultural and structural foundation on which the others rest (Tabassi, 2023). With a governance index averaging 0.97 out of 4 and explicit governance surfacing in 13.8% of cases, the firms sit at the bottom of the Govern function as their accounts present it: the policies, roles, and accountability the function names as its substance rarely appear. The European regulatory regime that would, in time, supply the coercive pressure to climb that function is real and on a known timeline, and its design has long been flagged as bearing disproportionately on smaller actors, with self-assessment-heavy obligations and standardization dependencies that the Act's critics argue render compliance opaque and onerous for SMEs (Veale & Zuiderveen Borgesius, 2021), compounded by the combined data-and-AI duties that arise where the AI Act and the GDPR overlap (Schwartmann et al., 2024). The marginal regulation coverage observed here is the empirical complement of that concern: the obligations exist on paper and on a phased timetable, yet awareness and readiness among the firms most burdened by them rarely surfaced. The Act's own provisions for SME proportionality, regulatory sandboxes and lighter documentation pathways (European Parliament & Council of the European Union, 2024), are in this light the instruments through which the coercive pressure to govern might be made bearable for firms that, on present evidence, are unlikely to build the apparatus unprompted. Whether that proportionality suffices is a policy question the study cannot resolve. What the study can offer is a sense of how far down the Govern function these firms begin.

6.8 Robustness and the governance-salience thesis

One robustness concern must be retired first, because it bears on every coverage figure here. The coding-depth artifact disclosed in Chapter 4, Wave 2 coded roughly 64% more densely than Wave 1, raises the worry that any governance signal might be an effect of coding effort, and, conversely, that the low coverage might be an artifact of under-coding. The data settle it. Denser Wave 2 coding did not raise governance: the gov-any indicator stands at 24.1% in Wave 1 against 17.2% in Wave 2, so the wave coded far more densely shows lower governance coverage, not higher. The same holds for measurement, the parallel theme examined in Chapter 7 (has-measurement 24.1% in Wave 1, 17.2% in Wave 2). Governance did not rise on the back of heavier coding, so the low salience documented throughout cannot be dismissed as an under-coding artifact; it persists, and is marginally more pronounced, in the wave where the analysts coded most aggressively. Read with the lexical flatness of §6.5, governance talk constant across waves in coding-independent text, the robustness picture is clear. What the chapter measures is stable and independent of how the corpus was coded. It is a feature of the talk, not an artifact of method.

The thesis can now be stated, and a boundary stated with it. The interview protocol had no standalone governance question, so the chapter measures the salience of governance in firms' own accounts of adoption, not a census of governance practice. Within that boundary the evidence converges from three directions, one per hypothesis. Explicit governance surfaces in 13.8% of firms and regulation in 6.9%, with the governance category drawing 1.9% of all coding and the index averaging 0.97 out of 4: governance is rarely foregrounded (H3a). No subgroup difference was detected, by size (odds ratio 2.93, Fisher p = .327) or by sector (odds ratio 0.87, p = 1.0), though the size comparison is underpowered (Large n = 9) and cannot establish that the

low salience is uniform (H3b). And where governance surfaces it does so as diffuse, informal vigilance, a governance lexicon decoupled from the formal code (r = –0.20, ns) and flat across waves (1.0 versus 0.9, p = .95), rather than as codified practice (H3c). These firms are not reckless. Nearly half raise data security, more than a quarter raise legal and ethical questions, and individual users discipline themselves with genuine care. What rarely appears in their narration is the conversion of that care into the policies, roles, and controls the responsible-AI, risk-management, and regulatory literatures treat as the substance of governance. We characterize the resulting pattern as adoption that has raced ahead of its own scaffolding, with an authentic ethic of caution sustained informally and from below, ahead of an institutional environment that has not yet made codified governance the expected or the enforced thing to do. Whether the apparatus is genuinely absent or merely unspoken, the data cannot finally say; what they show is how seldom it surfaced when firms were free to raise it. It is the first of the monograph's three faces, and it sets up the measurement paradox of Chapter 7, where, unlike governance, the protocol probed directly.

6.9 Governance-related theme coverage at a glance

Table 6.1 consolidates the coverage evidence, ordering the governance-relevant themes from most to least prevalent and pairing each with its 95% Wilson score interval. The descending sequence, data security, then legal-ethical, then governance, then regulation, is the staircase of §6.3 in numbers, and the corpus-level and index figures locate the gap within the study as a whole. The table is descriptive; the subgroup tests of H3b and the triangulation of H3c appear in the body of the chapter, since they concern associations and text-to-code alignment rather than coverage. Every figure in it is an articulated rate, conditional on a respondent raising the theme unprompted, and so a lower bound on the underlying practice.

Table 6.1 Coverage of governance-related themes (N = 58), with 95% Wilson score intervals[14]

Theme (code)	Firms (n / 58)	Coverage	95% CI	Reading
Data and data security (B3)	28	48.3%	35.9–60.8	Most prevalent risk-surface concern
Legal and ethical (E3)	16	27.6%	17.8–40.2	Roughly double governance coverage
Governance (G1)	8	13.8%	7.2–24.9	Codified governance rarely foregrounded
Regulation (G2)	4	6.9%	2.7–16.4	Regulatory engagement seldom raised
Governance category (G, segment share)	—	1.9%	—	Smallest of the eight categories
Governance index (mean, 0–4)	—	0.97	—	Average firm near one codified signal

Note. Coverage is the proportion of the 58 interviews in which a theme appears at least once; intervals are 95% Wilson score intervals. Because the protocol contained no standalone governance question, these are articulated (salience) rates, lower bounds on underlying practice, not a census. The governance category's 1.9% is its share of the 629 coded segments, not a coverage proportion. The governance index ranges from 0 to 4 over its component governance and oversight signals; its sample mean is 0.97. Subgroup tests of whether any explicit governance (gov-any) is associated with firm size or sector detect no difference (size: odds ratio 2.93, Fisher p = .327; sector: odds ratio 0.87, p = 1.0); the size comparison is underpowered (Large n = 9), so the null does not establish that the low salience is uniform. A coding-independent governance lexicon is decoupled from the formal governance code (point-biserial r = –0.20, ns) and flat across waves (1.0 versus 0.9, Mann–Whitney p = .95). Denser Wave

[14] **Source:** Authors' own processing based on the AI-ImpactSK framework-coded interview dataset and computational text analysis

2 coding did not raise governance coverage (gov-any 24.1% in Wave 1 versus 17.2% in Wave 2), confirming that the low salience is not a coding-depth artifact. Source: AI-ImpactSK dataset (Štetka et al., 2026); see Figure 6.1 for the coverage estimates with intervals and Figure 6.2 for the code-versus-text triangulation. The sample is purposive, ICT-skewed, and modest (N = 58); estimates characterize this sample and are not a basis for population inference.

7 The Measurement Paradox

This chapter examines the second central empirical gap of the monograph: the dissociation between perceived AI benefits and the measurement of those benefits. Whereas Chapter 6 showed that AI adoption often proceeds without codified governance, this chapter shows that it also proceeds without systematic evaluation. The question is not whether firms experience value from AI; many respondents describe concrete gains in time, efficiency, output quality, or decision support. The question is whether those gains are made visible through indicators, baselines, key performance measures, return-on-investment calculations, or other evaluative routines.

The chapter argues that the answer is usually no. Across the sample, benefit articulation is common, but benefit measurement is marginal. This pattern is treated as a measurement paradox: firms perceive and claim AI-related value more readily than they instrument or verify it. The analysis first documents the core contrast between benefit articulation and measurement, then tests whether the gap is confined to particular firm types, examines the role of implementation formality, triangulates the measurement code against independent lexical evidence, and finally interprets the pattern through the information-systems productivity-paradox and benefits-realization literatures. The chapter concludes by framing the finding as adoption without evaluation: a regime in which AI value may be real, but remains organizationally fragile because it is not made measurable.

7.1 The question and the core results

Firms in this sample assert benefits from artificial intelligence that they have never measured. That is the finding this chapter documents with primary field data, and the authors regard it as the flagship empirical contribution of the monograph. Where Chapter 6 concerned what firms do not codify before they act, the measurement paradox concerns what they do not learn after they act. The two are companion symptoms: adoption running ahead of the organizational disciplines that would make it manageable, defensible, and improvable.

Research Question 2 asks whether firms that claim benefits from AI actually measure them. The question is simple to pose, but it reaches a long-standing tension in the information-systems literature between perceiving value and evidencing it (Schryen, 2013; Melville et al., 2004; Ward & Daniel, 2012). Three hypotheses are stated in advance and tested in turn, within the exploratory-confirmatory frame appropriate to a purposive sample of fifty-eight firms. H2a holds that the articulation of benefits is far more prevalent than the measurement of benefits, so that asserting value and quantifying it are distinct organizational acts, the former far more common than the latter. H2b holds that this dissociation is broad rather than concentrated, present across firm types rather than confined to the smallest or least sophisticated firms. H2c holds that whatever measurement does occur is associated with implementation formality, so that the firms which measure are the firms which approach AI as a structured undertaking. The first hypothesis is the heart of the paradox; the second establishes its scope; the third begins to identify its correlates. As throughout the monograph, every inferential statement is bounded by the modest, single-country, ICT-skewed character of the sample, and the dataset is cited as (Štetka et al., 2026).

The central finding can be stated in a single sentence and is illustrated in Figure 7.1. Across the fifty-eight firms, concrete benefits from AI were articulated by 38, or 65.5% (95% Wilson score interval 52.7–76.4), whereas the measurement of those benefits, through any key performance

indicator, return-on-investment calculation, or systematic evaluation, was present in only 12 firms, or 20.7% (95% Wilson score interval 12.3–32.8). Almost two-thirds of the sample could name a benefit; barely one-fifth could show one. The benefit theme (F1) is among the better-populated themes in the entire codebook, sitting in the dense narrative core of the corpus alongside application areas and future plans. The success-measurement-and-KPIs theme (F2) sits far down the coverage distribution, near the sparsely populated governance themes. Articulation is mainstream; measurement is marginal.

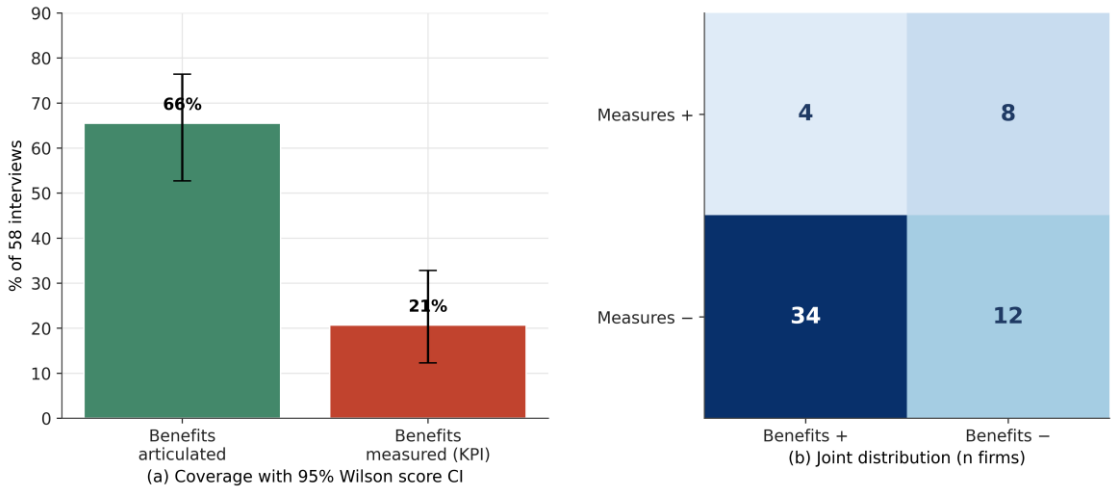


Figure 7.1 The measurement paradox. (a) 65.5% of firms articulate benefits but only 20.7% measure them (95% Wilson score CIs). (b) Only 4 firms both articulate and measure benefits; McNemar p<.001.[15]

Note: Panel (a) reports theme coverage with 95% Wilson score intervals. Panel (b) reports the joint distribution of benefit articulation and benefit measurement across firms.

The full structure of the relationship is more revealing than the two marginal rates alone, because what matters is not only that fewer firms measure than articulate but *which* firms do each. Treating benefit articulation and benefit measurement as two binary attributes observed on the same firms yields the two-by-two joint distribution reported in Table 7.1. Of the 58 firms, only 4 both articulated and measured benefits; 34 articulated a benefit but did not measure it; 8 measured something without articulating a corresponding benefit in the interview; and 12 did neither.

Table 7.1 Joint distribution of benefit articulation and benefit measurement across the 58 firms (AI-ImpactSK; see Figure 7.1)[16].

Articulation status	Measures benefits	Does not measure	Row total
Articulates benefits	4	34	38
Does not articulate	8	12	20
Column total	12	46	58

The decisive quantity in this table is the conditional probability of measurement given articulation. Among the 38 firms that articulated a benefit, only 4 also measured it, so the probability of measurement conditional on having articulated a benefit is just 4 in 38, or 10.5%.

[15] **Source:** Authors' own processing based on the AI-ImpactSK framework-coded interview dataset

[16] **Source:** Authors' own processing based on the AI-ImpactSK dataset

Put plainly: when a firm in this sample tells us that AI has delivered a concrete benefit, the chance that it can point to any metric, indicator, or evaluation behind that claim is roughly one in ten. For the overwhelming majority of these firms, the benefit claim and its evidentiary basis are simply detached.

One caveat belongs here, and it sharpens rather than softens the result. The 20.7% figure is an articulated or coded rate. It counts the firms for which measurement surfaced clearly enough in a roughly thirty-three-minute interview to be coded, and it should be read as a near-lower bound on what firms actually do, since a practice that did not come up may still exist. Two features qualify the number in opposite directions. The measurement probe was administered unevenly across the corpus, with most of the gaps falling in Wave 1, so the headline rate understates measurement somewhat. Yet the direction of the paradox does not depend on the exact figure, and neither does the McNemar result reported next, which rests on the within-firm contrast rather than on the marginal rate.

The asymmetry is not an artifact of small numbers or of how the two attributes happen to balance overall. The natural test for a paired contrast between two binary attributes observed on the same units is McNemar's exact test, which compares the two discordant cells: the firms that articulate but do not measure (34) against the firms that measure but do not articulate (8). The exact statistic is 8.0 with an exact p-value of 6.9×10^{-5}, far below conventional thresholds. Because the confirmatory analyses across the monograph are corrected as a family using the Benjamini–Hochberg false-discovery-rate procedure, it is worth recording that this result comfortably survives correction, with an adjusted q-value of .0007. Firms are, with very high statistical confidence, far more likely to articulate a benefit than to measure one. Hypothesis H2a is strongly supported.

A complementary way to summarize the pattern is the dissociation rate, the share of all firms that claim a benefit without measuring it. This is the upper-right cell of Table 7.1 expressed as a proportion of the whole sample: 34 of 58, or 58.6% (95% Wilson score interval 45.8–70.4). More than half of the entire sample, not merely more than half of the benefit-claimers, shows the specific pattern at issue: a perceived gain that has never been subjected to evaluation. It is this figure, more than any other, that justifies treating adoption without evaluation as a structural feature of the regime rather than an incidental shortfall in a handful of firms.

Because the core result rests on coded evidence, the measurement finding was checked against a coding-independent lexical signal. This triangulation is reported here, rather than in a separate methodological subsection, because it directly supports the validity of the chapter's central contrast between benefit articulation and benefit measurement.

Because the study's central claims rest in part on AI-assisted coding, the measurement finding is deliberately not allowed to stand on the manual code alone. As set out in Chapter 4, each key quantitative pattern is checked against a coding-independent computational signal extracted from the respondent-only text, so conclusions do not depend solely on labels that an AI-assisted pass may have shaped. For the measurement paradox the relevant check, summarized in Figure 6.2, compares the manual measurement code against an independent key-performance-indicator-and-measurement lexicon, vocabulary such as metrics, KPI, ROI, measurement, and their cognates, counted directly in what respondents said.

The two align. In firms where the measurement code is present, the independent lexicon shows a mean of 2.58 hits, against 1.35 hits in firms where it is absent; the point-biserial correlation between the manual code and the lexical signal is 0.27, with a p-value of .041. The measurement

code is text-valid: firms that the coders judged to be measuring are also the firms that independently talk in the vocabulary of measurement, at roughly twice the lexical rate of the others. This is the corroboration that the governance code, by contrast, failed to obtain in Chapter 6, where the governance lexicon did not align with the formal governance code. The asymmetry between the two triangulations is itself instructive. The low measurement rate is not an artifact of coders overlooking diffuse, informal evaluation that the text would reveal, because the independent lexical signal tracks the code rather than contradicting it. When these firms do not measure, the silence is audible in their own words and not only in the coders' judgments.

7.2 The Breadth and Correlates of the Gap

A natural first objection to the core result is that it might be a small-firm story, that the measurement gap reflects the limited administrative capacity of micro and small enterprises and would dissolve among larger, better-resourced firms with finance functions and reporting routines. The evidence does not support that reading. Decomposing the three quantities by firm-size class, as reported in Table 7.2 and visualized in Figure 7.2, shows that the dissociation is broad rather than concentrated.

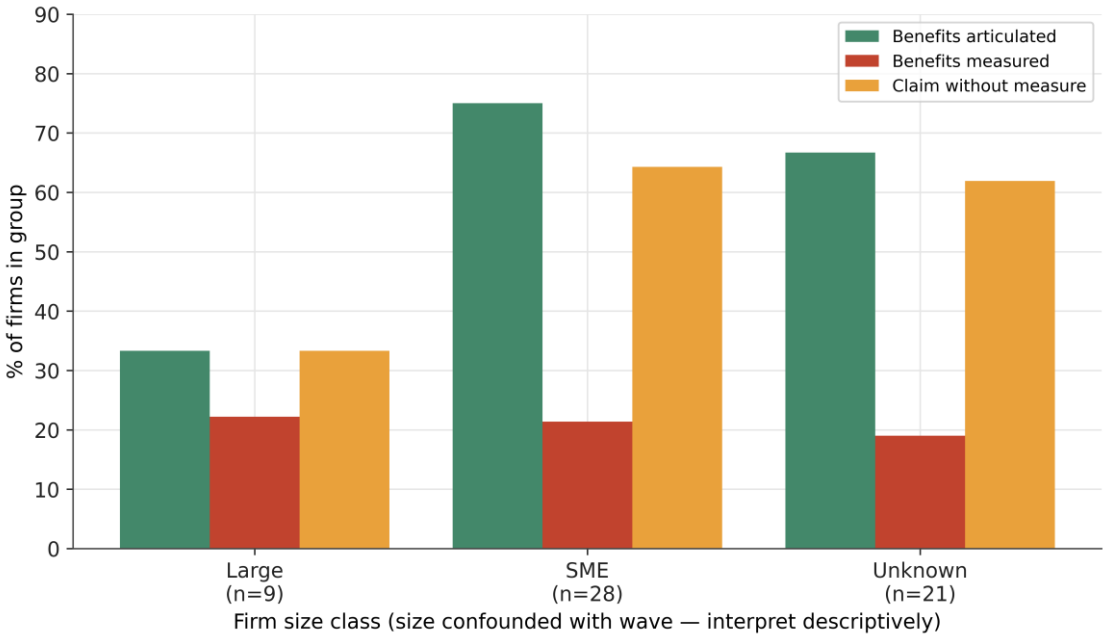


Figure 7.2 The benefit–measurement dissociation across size classes. The gap is broad, not confined to small firms (size class is confounded with wave; read descriptively)[17].

Note: Firm size is confounded with wave in this sample; the size-class breakdown is therefore interpreted descriptively rather than as an estimate of a size effect.

Read across the columns of Table 7.2, the measurement rate is strikingly stable: 21.4% among the small and medium-sized enterprises, 22.2% among the large firms, and 19.0% among the firms whose size could not be confidently classified. Whatever else differs between a nine-person studio and a multinational subsidiary, the propensity to measure AI benefits does not. It hovers near one-fifth regardless. What differs across the classes is the rate of articulation, which is markedly higher among the SMEs (75.0%) and unknown-size firms (66.7%) than among the

[17] **Source:** Authors' own processing based on the AI-ImpactSK dataset

large firms (33.3%). The consequence is that the dissociation rate is highest precisely where articulation is most exuberant: 64.3% among SMEs and 61.9% among unknown-size firms, against 33.3% among large firms. The large firms are not conspicuously better at measuring; they are merely more reticent about claiming, so fewer of their claims are left hanging. This supports H2b: the dissociation is observed across every size class rather than localized to the smallest firms.

Table 7.2 Benefit articulation, benefit measurement, and dissociation by firm-size class (AI-ImpactSK; descriptive — size is confounded with wave; see Figure 7.2)[18].

Size class	Articulate benefits	Measure benefits	Dissociation (claim, no measure)
SME	75.0%	21.4%	64.3%
Large	33.3%	22.2%	33.3%
Unknown size	66.7%	19.0%	61.9%

Two interpretive cautions must accompany Table 7.2, in keeping with the methodological commitments set out in Chapter 4. First, firm size is confounded with wave in this sample, because the small and medium-sized enterprises are concentrated in Wave 2, which was coded substantially more densely than Wave 1. The size breakdown is therefore offered descriptively, as a check on the breadth of the paradox, not as an estimate of a size effect on measurement. Second, the denser Wave 2 coding did not inflate the measurement signal: the share of firms recording any measurement is actually marginally lower in Wave 2 than in Wave 1, so the low measurement rate is not an absence that fuller coding would have filled in. The robustness anchor cuts the right way.

This breadth is corroborated by formal tests of association that fail to find one. Benefit measurement is not associated with firm size, where the strength of association is negligible (Cramér's V = .03), nor with membership in the ICT sector, where the association is weak and statistically non-significant (Cramér's V = .14). Neither of the two structural variables that might most plausibly explain away the measurement gap as a subgroup phenomenon actually does so. Within this sample, the paradox is a property of the population of adopters rather than the signature of a particular kind of firm.

If size and sector do not separate the firms that measure from those that do not, something else might. The third hypothesis directs attention to implementation formality, the degree to which a firm approaches AI as a structured project with defined phases, dedicated teams, and testing, captured in the study's composite formality index. Here the evidence is suggestive though not decisive. The 12 firms that measured benefits scored a mean formality of 2.83, against 1.91 for the firms that did not, a gap of nearly a full index point on the study's scale. Expressed as a non-parametric effect size suited to the ordinal, modestly sized data, Cliff's delta is 0.41, a moderate-to-large separation, with a nominal *p*-value of .026.

The appropriate degree of confidence in this association is tempered by the multiple-comparison correction. Under the Benjamini–Hochberg procedure the adjusted *q*-value is .053, fractionally above the conventional .05 threshold and thus best described as borderline. The honest reading is that measurement and formality co-occur, that the firms which instrument their AI use are the firms which organize it deliberately, but that in a sample of this size, with the confirmatory family appropriately corrected, the association is weak support rather than firm proof. Hypothesis H2c is recorded as receiving weak support. The direction is clear and

[18] **Source:** Authors' own processing based on the AI-ImpactSK dataset

theoretically expected; the magnitude is moderate; the statistical warrant is suggestive. What the association does establish, even at this modest strength, is that measurement is not a free-standing habit but one facet of a broader posture of formality. Where firms treat AI as something to be planned, staffed, and tested, they are also more likely to treat its results as something to be evaluated. The configurational chapter that follows develops the implication: measurement is part of a coherent adopter stance, not an isolated technique that firms might simply be exhorted to add.

7.3 Non-Measurement in the Respondents' Own Words

The statistics establish the shape of the paradox; the interviews show how it sounds from inside the firms. What the qualitative record reveals, again and again, is not indifference to value but an unembarrassed candor about its non-quantification, frequently coupled with a sense that measurement is a future aspiration rather than a present practice.

The most direct statements simply name the absence. A finance-sector respondent from a small firm, asked about the benefits of an AI deployment, explained that "I cannot quantify it because we did not track it" (F1.12). The clause captures the logic of the whole paradox, since the inability to quantify follows not from the elusiveness of the benefit but from the prior fact that it was never instrumented. An ICT respondent was equally forthright: "Completely honestly, I did not quantify it" (F2.02), the adverb signaling an awareness that quantification is what one might have been expected to do. These are not evasions. They are admissions offered without prompting, and their frankness is part of the finding.

A second register frames non-measurement as a stage rather than a settled state, in which benefits are presently felt and only prospectively measured. One ICT respondent described the basis for the benefit claim as impressionistic while gesturing toward an intended future of evaluation: "At the moment it is more of a gut feeling, but I believe we will gradually get to measurement" (F1.03). The phrase "gut feeling" is almost a verbatim instantiation of the intangible, perceived-but-unmeasured benefit that the productivity-paradox literature has theorized for three decades, and the projected trajectory, with measurement deferred to a later phase of maturity, is echoed by others. A small-firm ICT respondent put the current absence in similar terms: "That is also why we are not measuring things dramatically at this stage" (F2.23), locating non-measurement in the early, exploratory phase of adoption.

A third register attributes the difficulty to the domain, holding that the benefits in question are intrinsically hard to attribute. A marketing respondent observed that the gains are "difficult to measure in marketing at the level of a single campaign" (F1.27). It would be a mistake to read this firm as a non-measurer: it does measure digital-campaign reach and impressions, and it spoke about those figures readily. Its point is narrower and more credible. Single-campaign attribution is genuinely hard, because the outcome is diffuse and multi-causal and resists clean isolation. The lesson is that measurement is feasible and is done where the firm chooses to do it, and that the obstacle here is attribution at the level of one campaign rather than measurement as such. Across the three registers, the plain admission, the deferral to a later stage, and the appeal to attribution difficulty, the picture is of a population that perceives value vividly, articulates it readily, and evaluates it sparingly.

7.4 Interpretation: The Productivity Paradox and Measured GenAI Gains

The pattern documented here has a long lineage, and naming it is the first step in interpreting it. The information-systems discipline has spent three decades establishing that technology spending does not convert into measurable value automatically, and that the hardest part of value creation is not generating benefits but identifying, measuring, and appropriating them (Melville et al., 2004; Schryen, 2013; Ward & Daniel, 2012). The canonical statement of the difficulty is the IT productivity paradox: Solow's observation that computers were visible everywhere except in the productivity statistics, formalized and then progressively dissolved at the firm level by Brynjolfsson and Hitt (1996, 2000). The resolution of that paradox was itself a measurement story. Returns to information technology appeared once analysis moved to the firm level and once the intangible organizational complements of technology were taken into account; as Brynjolfsson and Hitt (2000) put it, "the benefits of IT investment are often intangible and disproportionately difficult to measure." The firms' "we save time but cannot quantify it" is the micro-level echo of exactly this argument, now voiced by individual managers rather than inferred from aggregate statistics.

The contemporary re-run of the paradox for AI sharpens the parallel. Brynjolfsson, Rock, and Syverson (2017, 2021) argue that a gap between extraordinary AI expectations and disappointing measured productivity is expected in the early years of a general-purpose technology, because realizing AI's value requires costly, time-consuming, and largely intangible complementary investments, including skills, process redesign, and new organizational capital, that depress or obscure measured benefits in the interim. Their productivity J-curve makes the mechanism precise: measured productivity initially understates true value while firms accumulate unmeasured intangible complementary capital, and only later overstates it as those intangibles pay off. On this reading, the 65.5%-versus-20.7% split observed here is consistent with a predictable feature of an early adoption phase as much as with any managerial failing. The firms are, for the most part, at the front of the J-curve, sensing value before they possess either the complementary capabilities or the measurement apparatus that would let them evidence it. The interview record, with its recurrent framing of measurement as a later-stage aspiration ("we will gradually get to measurement"), is firm-level micro-evidence of precisely the missing intangible complements that this macro literature hypothesizes but rarely observes directly, least of all in a resource-constrained, SME-heavy economy.

Yet the productivity-paradox framing, taken alone, risks being too forgiving, and the study's evidence licenses a sharper claim. The key theoretical move is to recognize that the detection of value is endogenous to the act of measurement. In their meta-analysis of firm-level IT-payoff studies, Kohli and Devaraj (2003) showed that whether a study detects a positive payoff is significantly determined by its measurement choices, including sample, data source, and dependent variable, and not only by the underlying technology. The implication for firms is stark. An organization that does not measure, or that measures crudely, will systematically fail to see benefits it actually obtains. Measurement is not a passive recording of value that exists independently; it is the instrument through which value becomes visible, governable, and improvable at all. A firm without measurement is not a firm with merely unrecorded gains. It is a firm that cannot distinguish a real gain from an imagined one, cannot tell a deteriorating deployment from an improving one, and cannot allocate effort toward the uses that pay and away from those that do not. Seen this way, the 58.6% dissociation rate is a managerial blind spot, not a benign waiting period.

A body of evidence that did not exist when the original productivity paradox was framed now throws the sample's pattern into sharp relief. A wave of rigorous field and field-experimental studies between 2023 and 2026 has, for the first time, measured generative-AI productivity directly, and found large, consistent effects. In a study of 5,172 customer-support agents, access to a generative-AI assistant raised productivity, measured as issues resolved per hour, by 15% on average, with the largest gains accruing to less-experienced and lower-skilled workers and minimal effects for the most skilled (Brynjolfsson, Li, & Raymond, 2025). In a preregistered experiment with 758 management consultants, AI access raised output inside the technology's competence, with 12.2% more tasks completed, 25.1% faster, at significantly higher quality, while, tellingly, for a task deliberately placed beyond that competence, AI users were 19% *less* likely to reach the correct answer, the now-canonical "jagged frontier" (Dell'Acqua et al., 2026).

The decisive point about these studies is methodological. They obtained credible benefit estimates only because they imposed deliberate measurement: randomization, baselines, and objective output metrics. The very gains that the firms in this sample assert qualitatively, faster support, quicker drafting, time saved on routine knowledge work, are in these studies quantified precisely, with apparatus no larger than a defined task and a counterfactual. The benefits the respondents describe as a gut feeling are, in the experimental literature, point estimates with confidence intervals. The contrast isolates the object of the present study with unusual clarity. The binding constraint among these firms is not the absence of measurable value but the absence of measurement. Quantification is demonstrably feasible; the firms simply do not perform it.

The jagged-frontier result adds a consequential twist that raises the stakes of non-measurement from a missed opportunity to a latent risk. If AI's benefits are real but task-contingent, and can turn negative on tasks beyond the model's competence (Dell'Acqua et al., 2026), then unmeasured adoption does not only leave gains unrecorded; it leaves losses undetected. A firm that deploys a generative tool across a range of tasks without instrumenting outcomes cannot know which of those tasks sit inside the frontier and which fall outside it, and will be unable to distinguish the uses that help from the uses that quietly harm. The absence of measurement is most dangerous precisely where benefits are uneven and cannot be assumed, which, on the current evidence, is the general case. The experiments establish that the sample firms' non-measurement is neither technically necessary nor prudently safe.

7.5 Adoption Without Evaluation: Consequences and Synthesis

The measurement paradox is best understood not as a discrete deficiency but as the evaluative face of a broader regime that this monograph characterizes as adoption without evaluation. The firms in this sample have, in the main, adopted AI bottom-up and informally, as the configurational analysis in the next chapter details and as the governance analysis in the previous one foreshadowed; the thin measurement documented here co-occurs with that informality. Where adoption is improvised rather than planned, there is no project to which metrics would attach, no baseline against which a gain could be assessed, and no owner accountable for demonstrating value. Measurement, on the evidence of the formality association in §7.4, is part of a posture of deliberateness; its scarcity is part of a posture of informality.

The consequences of this regime are concrete, and they follow from the endogeneity of value detection to measurement (Kohli & Devaraj, 2003). A firm that does not measure cannot manage what it has adopted: it cannot tune deployments toward the tasks that pay, retire those that do

not, or detect the silent quality losses that the jagged frontier predicts. It cannot defend what it has adopted: when budgets tighten or a sponsor departs, an AI initiative justified only by a gut feeling has no evidentiary basis to survive scrutiny, and is vulnerable to being cut precisely because its value was never made legible. And it cannot scale what it has adopted: a benefit that has never been quantified in one unit cannot be projected, costed, or replicated across others with any confidence, so promising local uses remain stranded as anecdotes rather than diffusing as demonstrated practice. The three failures share a single root. Value perceived but never measured remains, organizationally, value that the firm does not actually possess in any usable form.

There is a more hopeful corollary, worth stating because the diagnosis points toward the remedy. Because the constraint is measurement rather than value, and because, as the experimental literature shows, the gains are genuinely there and genuinely quantifiable, the paradox is in principle tractable. The firms are not failing to obtain benefits; they are failing to evidence benefits they very likely obtain. This reframes the managerial and policy task, taken up in the implications chapter, as one of measurement-by-design rather than value-creation from scratch. Lightweight baselines, a small number of fit-for-purpose indicators, and the simple discipline of instrumenting a deployment before rather than after the fact would move a substantial share of the dissociated firms from assertion to evidence. The productivity J-curve will, on the macro account, eventually surface these gains in the aggregate (Brynjolfsson, Rock, & Syverson, 2021). But for any individual firm, the difference between sensing value and managing it is the act of measurement, and that act remains, for now, the exception rather than the rule.

8 Configurations of Adoption

Two of the six markers of formal implementation. That is what the average firm in this sample musters around its AI use. The two preceding chapters described what these firms do *not* do well: they assert benefits they rarely measure (Chapter 7) and run adoption with little codified governance (Chapter 6). Here the question is the form adoption actually takes. How does AI enter the organization, who initiates it, how is it sourced, how broadly does it span technologies, and how much scaffolding surrounds it? The short answer is that adoption in this sample defaults to informal: bottom-up, individual, voluntary. The modest formality that does appear is not random. It varies with sector and with the breadth of AI technologies a firm has taken up. From that patterning we read off, cautiously, four indicative adopter groupings.

8.1 Research question and the informal default

This chapter addresses the first research question of the monograph (RQ1): *How does AI adoption vary in form and structure across Slovak firms, and what shapes that variation?* The question is configurational in spirit. Rather than asking whether a firm has adopted AI — a near-universal "yes" in this purposive sample — it asks how the adoption is organized and what combinations of conditions co-occur. Three hypotheses sharpen the inquiry. H1a holds that firms in the information and communication technology (ICT) sector implement AI more formally than non-ICT firms, on the expectation that technical firms possess the absorptive capacity (Cohen & Levinthal, 1990) and the prior digital routines to wrap AI use in defined approaches, teams, testing, and data-security practices. H1b holds that ICT firms also deploy a *broader* range of AI technologies than non-ICT firms. H1c proposes a within-firm link: that technological breadth and implementation formality move together, such that firms doing more with AI also do it more formally.

The verdicts, detailed below, are uneven. H1b finds clear support. H1a draws only weak, suggestive support once the wave confound is taken seriously. H1c is not borne out. The four-configuration grouping that follows the tests is exploratory, not confirmatory. It reads how adopter profiles fall out, and it is not a validated taxonomy. The N = 58 purposive, ICT-skewed sample (Štetka et al., 2026) bounds every claim. These groupings describe this corpus and are put forward as hypotheses for representative testing, not as population structure.

The dominant fact about AI adoption in the AI-ImpactSK corpus is its informality. Across the interviews, AI enters the firm from the bottom up, through individual initiative, on a voluntary basis. There is typically no defined project, no pilot with entry and exit criteria, and no key performance indicators against which use is assessed. Two of the sparsest themes in the entire coding structure corroborate the pattern: structured *testing* of AI (theme D3) is coded for only 8.6% of firms, and *success measurement and KPIs* (theme F2) for only 20.7% (Štetka et al., 2026). Adoption races ahead of the apparatus that would normally accompany a deliberate organizational change.

Respondents put it more plainly than any index can. A retail sales representative described diffusion as something that simply happens without managerial prompting: "We do not motivate people to use it; it happens organically," and, in the same interview, "We started using the technology individually, each person in their own way" (F1.01, Wave 1, retail, sales rep). The same individualized pattern recurs in a different sector and wave: "Each person uses it independently" (F2.04, Wave 2, healthcare, SME). That it shows up across settings suggests it is not an artifact of one firm. And where a manager might be expected to declare a policy or a

programme, respondents described its absence instead: "Some we may be using, but it is not formally defined" (F2.01, Wave 2, ICT, SME). These accounts point to adoption as accretion, assembled tool by tool and user by user, rather than as a governed initiative.

This default mode aligns with the emerging consensus that AI adoption in small and medium-sized enterprises has been normalized but remains predominantly informal, tool-level, and experimental, with structured implementation the exception rather than the rule (Schwaeke et al., 2024). The platform-mediated, low-friction character of contemporary AI access (Wei & Pardo, 2022) helps explain why the entry point is so often the individual employee. A subscription or a free web interface lowers the technical and financial barrier to the point where adoption needs no organizational decision at all. The evidence here extends that picture into a Central and Eastern European, SME-heavy economy, and it supplies the base rates that quantify how thin the surrounding scaffolding is: single-digit testing, one-in-five measurement.

8.2 Formality, technological breadth, and the ICT effect

To move from description to comparison, the analysis operationalizes the degree of structure surrounding AI use as an *implementation formality index*. The index runs from 0 to 6 and sums six coding-independent components, each marking the presence of a recognizable element of deliberate implementation: a defined approach or phasing (theme D1), the existence of dedicated teams or roles (D2), structured testing (D3), success measurement or KPIs (F2), formal governance (G1), and articulated data-security practice (B3). Across the 58 firms the index has a mean of 2.10 (SD 1.13), confirming numerically what the quotations convey qualitatively: the typical firm exhibits roughly two of the six markers of formal implementation, and several firms exhibit fewer. Formality is present but thin.

The first hypothesis predicts that this thin formality is unevenly distributed by sector. At the headline level, it is. ICT firms score 2.50 on the index against 1.78 for non-ICT firms, a moderate effect size (Cliff's $\delta = 0.36$) that survives correction for multiple comparisons ($p = .014$; FDR $q = .036$). Figure 8.1 displays this contrast. But that headline number cannot be taken at face value, because firm size is entangled with the design. Smaller firms are concentrated in Wave 2, and Wave 2 was coded substantially more densely than Wave 1 (the coding-depth artifact discussed in the methods chapter and revisited in Chapter 9). Any apparent association between firm size and formality is therefore confounded with wave and with coding depth, and the monograph does **not** interpret a size-to-formality relationship at all. The formality index correlates with segments per interview ($r = 0.56$), so size-based formality differences cannot be cleanly separated from how richly each wave was coded.

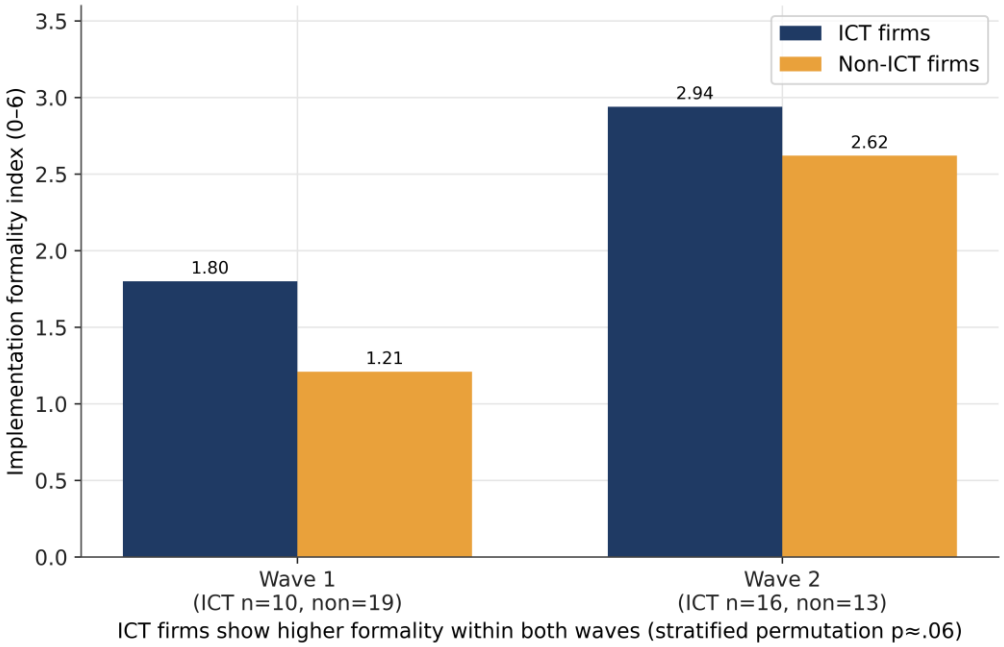


Figure 8.1 Implementation formality index for ICT vs non-ICT firms within each wave, isolating the firm effect from the coding-depth/wave confound.[19]

The one test that holds coding depth approximately constant is the within-wave comparison, and it is the only confound-clean reading available. Inside Wave 1, ICT firms score 1.80 against 1.21 for non-ICT firms ($\delta = 0.35$); inside Wave 2, 2.94 against 2.62, where δ falls to 0.21. The direction is the same in both waves, which is reassuring. The magnitude is not. A stratified permutation test that respects the wave structure yields $p = .059$, short of conventional significance (Štetka et al., 2026). So the within-wave evidence is weaker than the pooled 2.50-versus-1.78 gap implies: a smaller effect in Wave 2, and a permutation test that does not clear the .05 threshold. The fair characterization is a modest, directionally consistent tendency, suggestive rather than established. On that basis H1a earns weak, suggestive support, no more. We report the within-wave check precisely so the cross-wave headline is not over-read.

The second hypothesis concerns not how formally firms implement AI but how widely they reach across AI technologies. Breadth is measured as the count of distinct AI technology types a firm reports — generative AI, AI agents, natural-language processing or optical character recognition, computer vision, predictive analytics, proprietary AI products, and open-source models — with a sample mean of 2.5 types per firm. Here the sector contrast is again clear and, in this case, unambiguously significant. ICT firms report a mean breadth of 2.92 against 2.16 for non-ICT firms, a moderate effect (Cliff's $\delta = 0.37$) that survives correction ($p = .014$; FDR $q = .036$). H1b is supported: technical firms not only formalize AI somewhat more, they also span more of the AI technology space, consistent with their greater absorptive capacity and their position closer to the technological frontier.

The third hypothesis is not borne out. If breadth and formality were two expressions of the same underlying sophistication, they should correlate within firms. Across the 58 firms they do not: the rank correlation between technological breadth and the formality index is weak and statistically indistinguishable from zero (Spearman $\rho = 0.20$, $p = .14$). H1c is not supported. The

[19] **Source:** Authors' own processing based on the AI-ImpactSK framework-coded interview dataset

null is substantively interesting. Doing *more* with AI is not the same as doing AI more *formally*. A firm can assemble a wide repertoire of AI technologies while leaving testing, measurement, and governance largely unbuilt. Breadth is an acquisition; formality is an organizational investment, and in this sample the two come apart. The decoupling sets up the configurations below, where the broadest adopters tend to be the ones that never measure.

8.3 Four indicative adopter configurations and sourcing choices

To characterize how these dimensions combine at the level of whole firms, the analysis applies a k-modes clustering to coding-independent metadata and the AI-technology profile, deliberately excluding the coded thematic content so that the resulting groups are not circular with the outcomes they are then used to interpret. Two caveats have to come first, because they bound everything that follows. The internal validity diagnostics did not point to four groups at all. A two-cluster split scored best on the silhouette criterion (0.369), so the four-way partition is a choice made for descriptive richness, not a structure the data demanded. And the four-way solution separates only weakly: a bootstrap adjusted Rand index of 0.35 means the cluster assignments shift under resampling and the boundaries between groups are soft, not sharp. What follows is therefore a heuristic descriptive partition. The four configurations are *indicative groupings*, read in the configurational tradition that organizations cohere into a limited number of internally consistent gestalts (Miller, 1986; Short, Payne, & Ketchen, 2008). They are not validated types. Taxonomy development in information systems is properly iterative and replicable (Nickerson, Varshney, & Muntermann, 2013); this is a first, exploratory pass on a single purposive sample, to be tested rather than trusted. Figure 8.2 plots the four groupings by their AI-technology profiles.

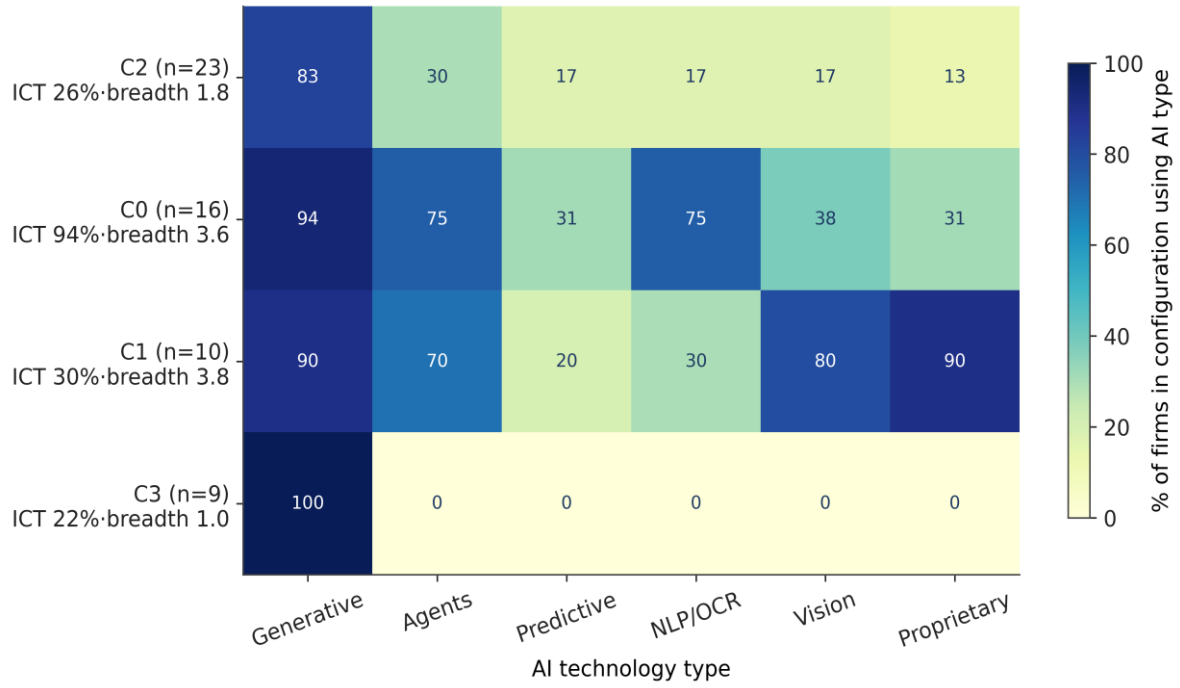


Figure 8.2 Four coding-independent adopter configurations (k-modes on AI-profile and firm metadata) by share using each AI type.[20]

[20] **Source:** Authors' own processing based on the AI-ImpactSK AI-profile metadata and clustering analysis

The table below summarizes the four groups on the dimensions that distinguish them.

Table 8.1 Indicative adopter configurations and distinguishing dimensions[21]

Configuration	n	ICT %	Key AI types	Breadth (mean)	Formality (mean)	Measurement %	Governance %
C0 — Multi-modal technical adopters	16	94	GenAI (94%), agents (75%), NLP/OCR (75%), vision (38%)	3.56	2.50	31	19
C1 — Vision/agent specialists	10	30	GenAI (90%), agents (70%), vision (80%)	3.80	2.30	0	20
C2 — Generative mainstream / broad-economy	23	26	GenAI (83%), agents (30%)	1.78	2.04	17	26
C3 — Generative-only minimalists	9	22	GenAI (100%), all else 0%	1.00	1.33	33	11

Note. Percentages are within-configuration shares; breadth and formality are means on the respective indices (Štetka et al., 2026). Bootstrap ARI = 0.35; configurations are indicative, not definitive.

C0 — Multi-modal technical adopters (n = 16). These are the sample's technical builders. The grouping is almost entirely ICT (94%), and its firms combine generative AI (94%) with AI agents (75%) and NLP/OCR (75%), reaching a breadth of 3.56 technologies. Its formality (2.50) is the equal-highest among the four groups; roughly a third of these firms measure outcomes (31%) and about a fifth report governance (19%). C0 is adoption as engineering: firms with the absorptive capacity to integrate several modalities and to surround them with at least some testing, measurement, and data-security routine. In dynamic-capabilities terms (Teece, 2007), their sensing of new AI possibilities is matched by a partial capacity to seize and reconfigure around them, though even here the formal apparatus is incomplete.

C1 — Vision/agent specialists (n = 10). The C1 firms are the technically broadest of all, with a breadth of 3.80, the highest in the sample, combining generative AI (90%), AI agents (70%), and computer vision (80%). Their formality is middling (2.30) and governance sits at 20%. The striking number is measurement: across the ten firms, none was coded as measuring outcomes. That 0% is worth flagging, but it rests on ten firms in a grouping whose boundaries are soft, so it should be read as a vivid illustration rather than an established rate. The descriptive contrast with C3, where three of nine minimalist firms do measure, is suggestive of the broader decoupling, not proof of a pattern. With that caution in place, C1 fits the H1c null: a firm can run vision pipelines and multiple agents and still leave the question "what did it achieve?" unasked. It is the measurement paradox of Chapter 7 in a small, sharp form.

C2 — Generative mainstream / broad-economy adopters (n = 23). The largest grouping is also the most ordinary. C2 firms are predominantly non-ICT (only 26% ICT) and sit across industry, finance, and the broader economy. Their AI use centers on generative AI (83%) with a minority using agents (30%), and their breadth is low (1.78). Formality is near the sample mean (2.04) and the measurement rate is modest (17%). Notably, they post the *highest* governance share of any

[21] **Source:** Authors' own processing based on the AI-ImpactSK framework-coded interview dataset and AI-profile metadata

group (26%). This is the grouping that best represents AI adoption in the everyday economy: firms reaching for generative tools to meet concrete operational needs, technically narrow but, relative to the technical specialists, marginally more attentive to governance. That the broad-economy adopters govern somewhat more than the technical virtuosos is worth flagging. It cautions against assuming that technical sophistication and responsible-use practice rise together, though with cells this small the gap should be held lightly.

C3 — Generative-only minimalists (n = 9). The smallest grouping is also the lightest. C3 firms use generative AI and nothing else (100% generative, 0% on every other modality), giving them the minimum possible breadth of 1.00 and the lowest formality of the four groups (1.33). These are single-tool users, often in the public sector or other non-technical settings, for whom "adopting AI" means using one generative assistant for discrete tasks. Their governance is the lowest (11%), consistent with their minimal footprint. Their measurement rate, three of nine firms (33%), is nominally the highest in the sample. The figure rests on a handful of firms, so it is best read as descriptive rather than as a reliable ordering; a plausible reading is that a single, well-bounded use case is easier to assess than a sprawling multi-modal portfolio. C3 marks the floor of AI adoption: a single generative tool, lightly used and barely structured.

Read together, the four groupings make the chapter's central analytic point. They do not array along a single ladder from "less" to "more" adoption. They suggest instead that breadth, formality, measurement, and governance vary semi-independently. In this sample the broadest adopters (C1) measure least, the most ordinary adopters (C2) govern most, and the lightest adopters (C3) measure most, though each of those orderings rests on a small cell and should be treated as descriptive rather than settled. The pattern is consistent with equifinality in the configurational sense (Fiss, 2011; Park, El Sawy, & Fiss, 2017): different combinations of conditions can coexist rather than collapse onto one dimension, which is why a configurational reading fits the data better than any single adoption score. The soft boundaries (ARI = 0.35) mean these profiles should be held loosely. Their value is heuristic, naming recognizable adopter groupings whose existence and prevalence a representative study could confirm or revise.

Adoption configurations also differ in how firms *source* their AI capability. Sourcing is not a binary make-or-buy choice in these interviews but a continuum: buying off-the-shelf software at one end, building bespoke systems in-house at the other, and adapting general models to proprietary knowledge in between. The relevant theme, build/buy/adapt (theme C2 in the codebook), is coded for 48.3% of firms, one of the more salient topics in the corpus and a sign that procurement strategy is an active, articulated concern.

At one end of the continuum sits straightforward consumption of software-as-a-service. One respondent captured the buyer's menu plainly: "A user can pay for the application, buy a subscription, or test it" (F2.05, Wave 2, ICT, SME). The model is treated as a utility accessed by subscription. At the other end sit firms that build. Some maintain their own systems on top of foundation technologies, as one put it: "We also have our own solutions built on AI technologies" (F2.28, Wave 2, ICT, SME). A finance firm reported training a bespoke model for a regulated, knowledge-intensive task: "We are also training our own model for responding to regulators" (F1.29, Wave 1, finance). Between pure buy and pure build lies the *adapt* path, in which a firm grounds a general model in its own documents and data without retraining it from scratch. This middle path is realized above all through retrieval-augmented generation, the canonical mechanism for coupling a general model with a firm-specific knowledge index to obtain more specific and factual outputs without full in-house training (Lewis et al., 2020). RAG and fine-tuning are what make "adapt" a genuine third option.

The interpretation of this variation draws on the make-or-buy literature as updated for the agentic era. Klotz (2026) argues that AI-assisted ("agentic") development is reshaping the classical make-or-buy determinants and is transforming the "build" option itself from a pure hierarchy into a hybrid governance form, code ownership combined with dependence on external AI infrastructure. On this account, building becomes most compelling for commodity utilities and for differentiating custom applications, while regulated or mission-critical systems tend to remain in the "buy" domain. The cases here fit the logic. The finance firm builds precisely where the task is differentiating and knowledge-intensive (regulatory response), while ordinary operational needs are met by subscription. Sourcing choice tracks asset specificity, differentiation, and capability, the same contingencies that shape the adopter groupings: the multi-modal technical adopters (C0) and vision/agent specialists (C1) hold the in-house capability to build and adapt, whereas the generative mainstream (C2) and minimalists (C3) more often buy.

8.4 Interpretation: TOE, dynamic capabilities, and a configurational reading

Three complementary frameworks help make sense of the findings. The technology–organization–environment framework (Tornatzky & Fleischer, 1990) locates the determinants of adoption form in three interacting contexts, and the chapter's results populate two of them with unusual clarity. The *technological* context is captured by breadth and sourcing, what AI a firm takes up and how it obtains it. The *organizational* context is captured by the formality index and by the bottom-up, individual mode of entry that the framework, written for managed industrial innovation, did not anticipate. The most striking organizational fact here is that adoption frequently bypasses the organizational decision altogether, arriving through employees rather than through management. The TOE "organizational" context has to be populated with the informal, shadow-style dynamics the original framework never modeled. The *environmental* context, the emerging EU AI governance regime examined in Chapter 6, presses only weakly on these firms, which helps explain why governance stays thin across all four groupings.

The dynamic-capabilities perspective (Teece, 2007) supplies a mechanism for the sector tendency and the differences across groupings. Implementing AI formally, spanning multiple modalities, and building or adapting bespoke systems are all expressions of the capacity to sense AI opportunities, seize them, and reconfigure organizational assets around them. ICT firms' greater formality and breadth are consistent with a deeper stock of these capabilities and of the absorptive capacity (Cohen & Levinthal, 1990) that lets prior digital knowledge accelerate AI assimilation. The groupings can be read as differing capability endowments. C0 firms sense and partially seize across modalities. C1 firms sense broadly but, on this evidence, do not build the evaluative routines that would let them learn from what they seize. C2 and C3 firms exercise a narrower, more consumption-oriented capability set.

The configurational tradition in information systems (Fiss, 2011; El Sawy, Malhotra, Park, & Pavlou, 2010; Park, El Sawy, & Fiss, 2017) supplies the overarching stance. Its central claim is that organizational outcomes arise from *combinations* of elements rather than from variables acting in isolation, and that multiple distinct configurations can be internally coherent (equifinality). The semi-independence of breadth, formality, measurement, and governance documented here is consistent with that claim. The failure of H1c is itself a finding: the dimensions of AI adoption do not bundle into a single factor, which is exactly the condition under which configurational rather than variance-based theorizing is warranted.

The claims here are deliberately bounded. The four groupings rest on a weakly separated clustering solution (bootstrap ARI = 0.35) that the internal diagnostics did not even favor over a two-cluster split, applied to a purposive, ICT-skewed sample of 58 firms (Štetka et al., 2026); the size-to-formality relationship is left uninterpreted because it is confounded with wave and coding depth. What can be defended is narrow. Informality is the default mode of AI adoption in these firms. ICT firms reach somewhat more broadly across technologies, and they show a modest, suggestive tendency toward more formal implementation that the confound-clean test does not establish. Breadth and formality are decoupled. Sourcing spans a genuine build–buy–adapt continuum. Adopter profiles fall into a small number of recognizable, internally coherent groupings. These are offered as a structured description of one corpus and as hypotheses for representative testing, not as a settled account of how Slovak enterprises, still less enterprises elsewhere, configure their adoption of AI.

Having characterized the form of adoption in cross-sectional and configurational terms, the monograph now turns to the temporal question: what, if anything, changed between the two waves of fieldwork, and which apparent changes survive the coding-depth caveat.

Having characterized the form of adoption in cross-sectional and configurational terms, the monograph now turns to the temporal question: what, if anything, changed between the two waves of fieldwork, and which apparent changes survive the coding-depth caveat.

9 The Temporal Shift, 2025 to 2026

This chapter examines the temporal dimension of the AI-ImpactSK study: what changed in firms' AI discourse between the first wave of interviews in autumn 2025 and the second wave in spring 2026. Its purpose is deliberately narrower than a simple before-and-after comparison. Because Wave 2 was coded substantially more densely than Wave 1, and because respondents in Wave 2 also spoke more, the chapter treats apparent cross-wave change with methodological caution. It asks which signals survive the coding-depth artifact and which should be read only as descriptive or hypothesis-generating.

The chapter argues that one temporal shift is robust: firms increasingly spoke in the language of agentic AI, practical solutions, verification, and production. At the same time, governance and measurement did not show a corresponding rise. The result is therefore not a broad story of organizational maturation, but a narrower and more consequential one: firms entered 2026 talking more like operators of autonomous systems, while remaining no more visibly equipped to govern or evaluate them.

9.1 Research question and the burden of methodological caution

RQ4 asks how the discourse of the firms in this sample about artificial intelligence changed between the autumn of 2025 and the spring of 2026. It also asks something sharper — how much of any apparent change survives once the analyst has stripped away the influence of the coding process itself. This is not a simple "before and after" comparison. The two waves of the AI-ImpactSK corpus were not coded with equal intensity, and that inequality contaminates the most obvious temporal measures. The chapter treats the contamination as its organizing principle. It separates, as cleanly as the data permit, what is an artifact of how the material was analyzed from what is a real shift in how respondents spoke and what they spoke about.

This is the most cautious of the four results chapters, and deliberately so. The temporal angle is a secondary and heavily caveated contribution of the monograph Štetka et al., 2026), not a flagship claim on the order of the measurement paradox (Chapter 7) or the governance gap (Chapter 6). The sample is small: twenty-nine interviews in each wave, fifty-eight in total. It is purposive rather than representative, and skewed toward the information and communication technology (ICT) sector. Any reading of change across two closely spaced waves drawn from such a sample must be hypothesis-generating rather than confirmatory, and the inferential statistics below should be understood in that spirit. What makes the chapter worth writing is the discipline it models. By refusing to read the naive cross-wave numbers at face value, and by anchoring every substantive claim in a coding-independent signal, it shows how a transparent temporal analysis can be conducted when the analytic instrument is known to have drifted between measurements.

9.2 The coding-depth artifact

The single most important fact about the two waves is that the second was coded far more densely than the first. Wave 1 yielded 238 coded segments and Wave 2 yielded 391, but this gap is not evenly distributed across a stable per-interview baseline. The mean number of coded segments per interview rose from 8.21 in Wave 1 (SD 0.73, median 8, range 7 to 10) to 13.48 in Wave 2 (SD 1.18, median 14, range 10 to 15) — against an overall mean of 10.84. A Mann–Whitney U test confirms that Wave 2 interviews were coded more densely at a level that leaves essentially no doubt about the difference itself ($p < 1 \times 10^{-7}$). Put plainly, Wave 2 was coded

approximately 64 percent more densely than Wave 1. Figure 9.1 displays this divergence in coding depth alongside the respondent-volume measures discussed below.

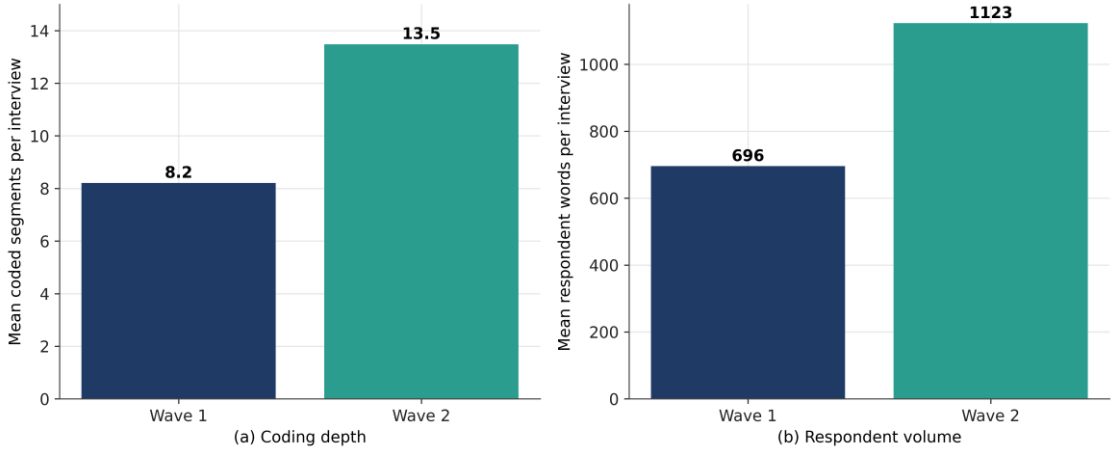


Figure 9.1 Wave 2 interviews were coded ~64% more densely and respondents spoke ~1.6x more, confounding cross-wave coverage comparisons (a methodological caution)[22].

Denser coding is not, on its own, evidence of anything substantive. Two interpretations are possible, and they carry very different consequences for how the wave comparison should be read. The first is that respondents in Wave 2 genuinely said more codeable things, so that the underlying discourse became richer and denser coding faithfully records a real change. The second is that the coding process became more granular, the codebook more familiar, the analyst more attuned to the material, so that the same underlying talk was simply parsed into more segments. The data point to a mixture of the two. The analyst's task is to disentangle them rather than to assume either.

The respondent-volume measure does some of that disentangling. The number of words spoken by respondents per interview rose from a mean of 696 in Wave 1 to a mean of 1,123 in Wave 2, a difference that a Mann–Whitney test renders unlikely to be noise ($U = 203$, $p = .0007$), with a rank-biserial correlation of 0.52 indicating a large effect. Respondents in Wave 2 spoke roughly 1.6 times as much as respondents in Wave 1. This is itself a coding-independent fact. It is measured on the raw transcript text, not on the coded segments, and it tells us that part of the denser coding rests on a real difference in how much was said. It also delivers a warning. If respondents simply talked more, then any count of coded segments, any measure of thematic coverage that depends on whether a theme was coded at all, and any tally that scales with sheer verbal volume will rise mechanically in Wave 2 — not because the phenomenon being counted became more prevalent in the firms, but because there was more text in which to find it and a more practiced eye looking for it.

The relationship between coding depth and the study's formality construct makes this concrete. The implementation formality index correlates with segments per interview at $r = 0.56$, so a substantial share of the variance in measured formality co-moves with the simple density of coding. A firm coded more granularly will, almost by construction, appear to display more of the components that the formality index aggregates. This is why the naive cross-wave story cannot be told from the coverage numbers alone. That story runs: firms became more formal, more

[22] **Source:** Authors' own processing based on the AI-ImpactSK transcript corpus and framework-coded interview dataset

sophisticated, more thoroughly engaged between 2025 and 2026. Some of that apparent maturation is the coding instrument changing its own resolution between the two measurements.

The clearest casualty of the artifact is the intensity coding. The share of segments coded at "high" intensity rose from 0.43 in Wave 1 to 0.91 in Wave 2. In absolute terms, Wave 1 contained 102 high-intensity, 126 medium-intensity, and 10 low-intensity segments, whereas Wave 2 contained 356 high, 34 medium, and a single low-intensity segment. Taken at face value, this would suggest that the firms in the sample became dramatically more emphatic, more committed, more intense in their engagement with AI over seven months. But intensity is wave-confounded to the point of being uninterpretable for cross-wave substantive claims. The near-collapse of the medium and low categories in Wave 2 is the signature of a coding convention that shifted between waves, not of a sudden surge of organizational conviction. Intensity is therefore set aside entirely as a basis for any temporal claim in this chapter. To read it as substantive change would be to mistake the behavior of the coders for the behavior of the firms.

The implication is general. Because Wave 2 was coded approximately 64 percent more densely, and because formality and coverage measures scale with coding depth, every naive cross-wave rise in theme coverage is partly an artifact and must be reported with that caveat attached. Wave comparisons built on coverage are, at best, hypothesis-generating. They acquire evidential weight only when triangulated against text-based measures that do not depend on the coding process.

A robustness check guards against over-correcting in the opposite direction. If denser coding inflated everything, one might worry that the entire Wave 2 picture is an artifact and that nothing can be salvaged. That worry is misplaced, and the governance and measurement findings show why. Denser Wave 2 coding did not raise the measurement or governance signals. The proportion of interviews exhibiting any measurement of benefits was 24.1 percent in Wave 1 and 17.2 percent in Wave 2, and the proportion exhibiting any formal governance was likewise 24.1 percent in Wave 1 against 17.2 percent in Wave 2. Had the artifact simply lifted all boats, these too would have risen with coding density. Instead they held flat or edged downward. Two things follow. First, the governance and measurement gaps documented in Chapters 6 and 7 are not creatures of coding depth; they are robust to it, and if anything the denser instrument failed to find more governance where little exists. Second, the coding-depth artifact is selective rather than uniform, which means that a signal surviving the artifact is genuinely informative. The analyst's job is to identify which signals those are.

9.3 Coding-independent signals that are robust

Against the cautionary backdrop of the artifact, three families of measure stand on coding-independent ground because they are computed on the raw respondent text rather than on the coded segments. These are the measures from which the chapter draws its substantive temporal claims.

The agentic-AI rise

The most robust and most consequential temporal signal is the rise of agentic-AI discourse. Counting respondents' own mentions of a lexicon of agentic terms ("agentic," "agent," "autonomous," "RAG"), the mean per interview rose from 0.72 in Wave 1 to 3.14 in Wave 2. A Mann–Whitney test places this difference well outside the range of plausible noise ($U = 234$, $p = .0022$), with a rank-biserial correlation of 0.44 indicating a moderate-to-large effect. The shift is

equally visible in the breadth of uptake: agentic language was present in 37.9 percent of Wave 1 interviews and in 69.0 percent of Wave 2 interviews. A minority idiom in the autumn of 2025 had become the majority idiom by the spring of 2026. Figure 9.2 displays the agentic-discourse rise together with the tool-breadth measure discussed below.

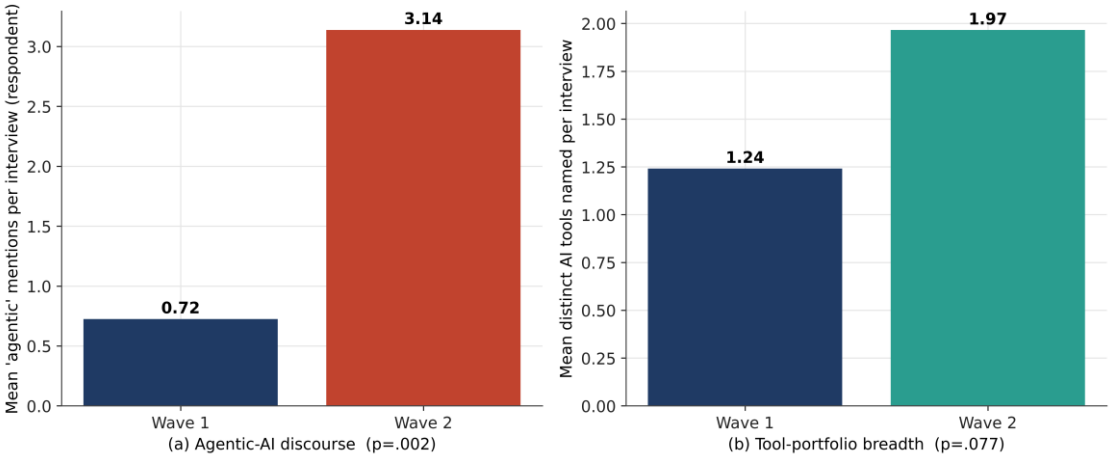


Figure 9.2 Coding-independent text signals: agentic-AI mentions rose sharply (p=.002) and named-tool breadth widened (p=.077) between waves — a real maturation robust to coding depth[23].

What makes this signal trustworthy is that it triangulates across two independent operationalizations. The text-based mention count just described is one. The other is the manually assigned "AI agents" code applied during framework analysis. These two were constructed by entirely different procedures, one a computational tally on raw speech, the other a human coding judgment, yet they converge. The manual code validates against the text at a point-biserial correlation of $r = 0.55$ ($p < .0001$), a strong association indicating that the interviews a human coder flagged as involving AI agents are also the interviews in which agentic vocabulary actually appears in the transcript. The rise is visible in a coding-independent measure and corroborated by a coding-dependent one that itself validates against the text. So the agentic turn cannot be dismissed as an artifact of denser coding. Within this sample, it is a genuine maturation in what the firms were talking about. The qualitative texture matches the quantitative signal. By Wave 2, ICT respondents describe agents as a working modality rather than a curiosity: "In programming, I use AI agents that write code directly" and "I also use multiple agents simultaneously" (F2.20, ICT), "We use primarily AI agents" (F2.18, ICT), and, of one firm's product, "Their core functionality consists of agents" (F2.06, ICT) (Štetka et al., 2026).

Keyness: the lexical signature of maturation

The second robust signal comes from a keyness analysis, a log-likelihood comparison of word frequencies in respondent speech across the two waves that identifies the terms statistically distinctive of each. Because keyness is computed on raw text, it is immune to the coding-depth artifact, and it offers a vocabulary-level portrait of how the conversation moved. The terms distinctive of Wave 1 are insurance, assessment, learning, triggering, generative, and machine. The terms distinctive of Wave 2 are solutions, verification, screening, practical, context, production, claude, and agents. Figure 9.3 presents this lexical keyness by wave.

[23] **Source:** Authors' own computational text analysis based on the AI-ImpactSK respondent-only transcript corpus

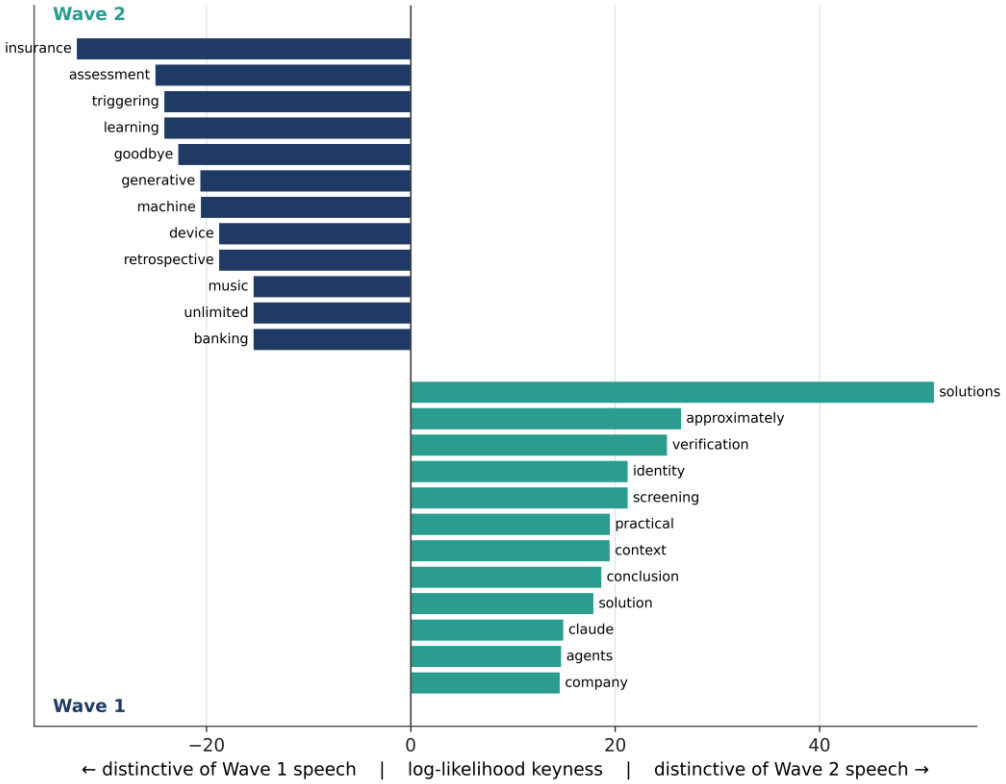


Figure 9.3 Keyness (log-likelihood) of respondent vocabulary by wave. Wave 1 foregrounds the generative/ML novelty and trigger; Wave 2 foregrounds practical solutions, verification, and production[24].

Read together, the two lists trace a coherent story. The Wave 1 vocabulary is the language of arrival and novelty. "Triggering" and "learning" evoke the initiating impulse and the climbing of a learning curve; "generative" and "machine" name the technology in its broad, still-recently-encountered categories; "assessment" suggests appraisal of an unfamiliar capability. The Wave 2 vocabulary is the language of operation and consolidation. "Solutions," "production," and "practical" speak of deployment into working systems rather than exploration; "verification" and "screening" speak of the routinized checking that accompanies real use; "context" points to the situated tailoring of tools to specific tasks; "claude" and "agents" name the concrete instruments and modality of the agentic turn. The movement runs from novelty, trigger, and generative machine-learning framing toward practical solutions, verification, and production. It is a maturation legible in the words themselves, independent of how those words were later coded.

Tool-portfolio breadth and the tool landscape

The third signal concerns the breadth of the named-tool portfolio. The number of distinct named tools per interview rose from a mean of 1.24 in Wave 1 to 1.97 in Wave 2, a difference that a Mann–Whitney test renders only marginal ($p = .077$). The count of total tool mentions did not separate the waves at conventional levels ($p = .15$, not significant). This signal is therefore weaker than the agentic-discourse rise and is reported as suggestive rather than established. Its direction is consistent with the maturation story, since firms in 2026 referenced a somewhat wider set of named tools than firms in 2025, but the marginal p-value warrants restraint. The

[24] **Source:** Authors' own keyness analysis based on the AI-ImpactSK respondent-only transcript corpus

breadth claim should not be pressed beyond a tendency. Figure 9.3 shows tool breadth alongside the agentic measure; Figure 9.4 details the tool landscape.

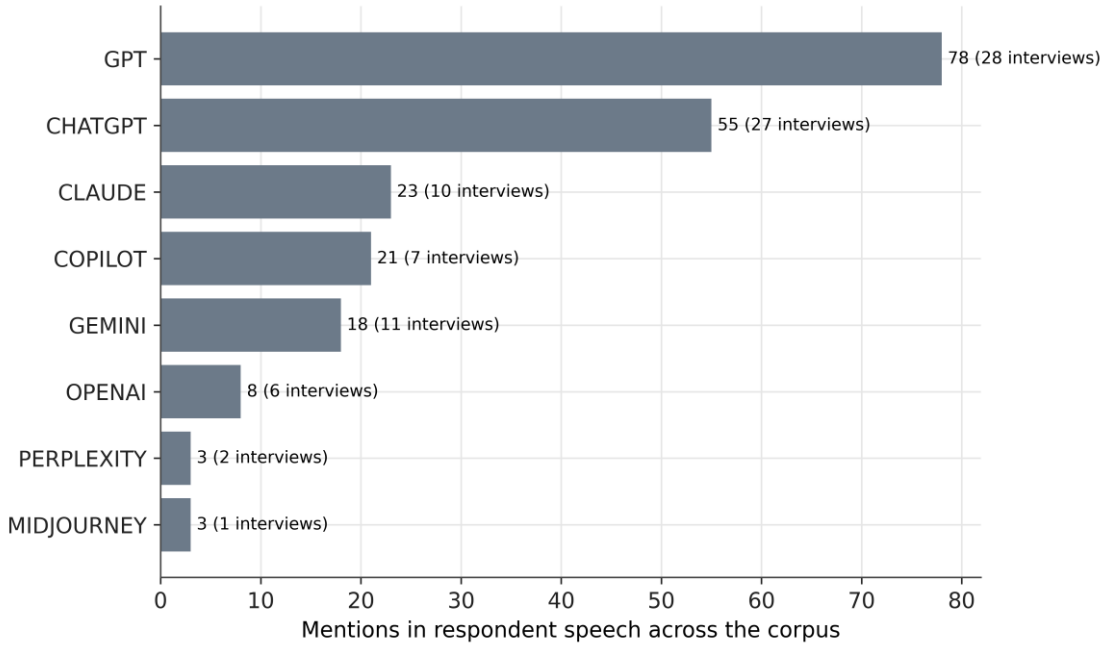


Figure 9.4 Most-mentioned AI tools in respondent speech (GPT/ChatGPT dominate; Claude, Copilot, Gemini follow)[25].

The aggregate tool landscape, pooled across both waves of respondent speech, is dominated by a small number of instruments. GPT was mentioned 78 times (across 28 interviews) and ChatGPT 55 times (27 interviews), making the OpenAI family by a wide margin the most frequently named. Claude followed with 23 mentions (10 interviews), then Copilot with 21 (7 interviews) and Gemini with 18 (11 interviews), with OpenAI named explicitly 8 times (6 interviews) and Midjourney and Perplexity trailing at 3 mentions each. ChatGPT was named in roughly 48 percent of interviews in both waves, a stable anchor of the conversation across the temporal window even as the agentic vocabulary grew around it. The picture is a consolidated mainstream, a generative-text core that did not change much, overlaid by a widening agentic and multi-tool frontier that did.

Cross-wave coverage change, with the caveat attached

It remains to address the cross-wave coverage figures directly. They are the measures most exposed to the coding-depth artifact and therefore the ones most in need of explicit caveat. Figure 9.5 presents the change in theme coverage across waves. The naive reading is that coverage rose broadly, that more themes were touched by more interviews in Wave 2 than in Wave 1, and this is descriptively true. But as Section 9.2 established, broad coverage rises are partly a mechanical consequence of denser coding and greater respondent verbal volume, not solely a real broadening of organizational engagement. The coverage rises are reported here, but they are not read as substantive change without the caveat. They function as hypotheses about possible broadening, to be believed only where a coding-independent signal corroborates them.

[25] **Source:** Authors' own named-tool extraction based on the AI-ImpactSK respondent-only transcript corpus

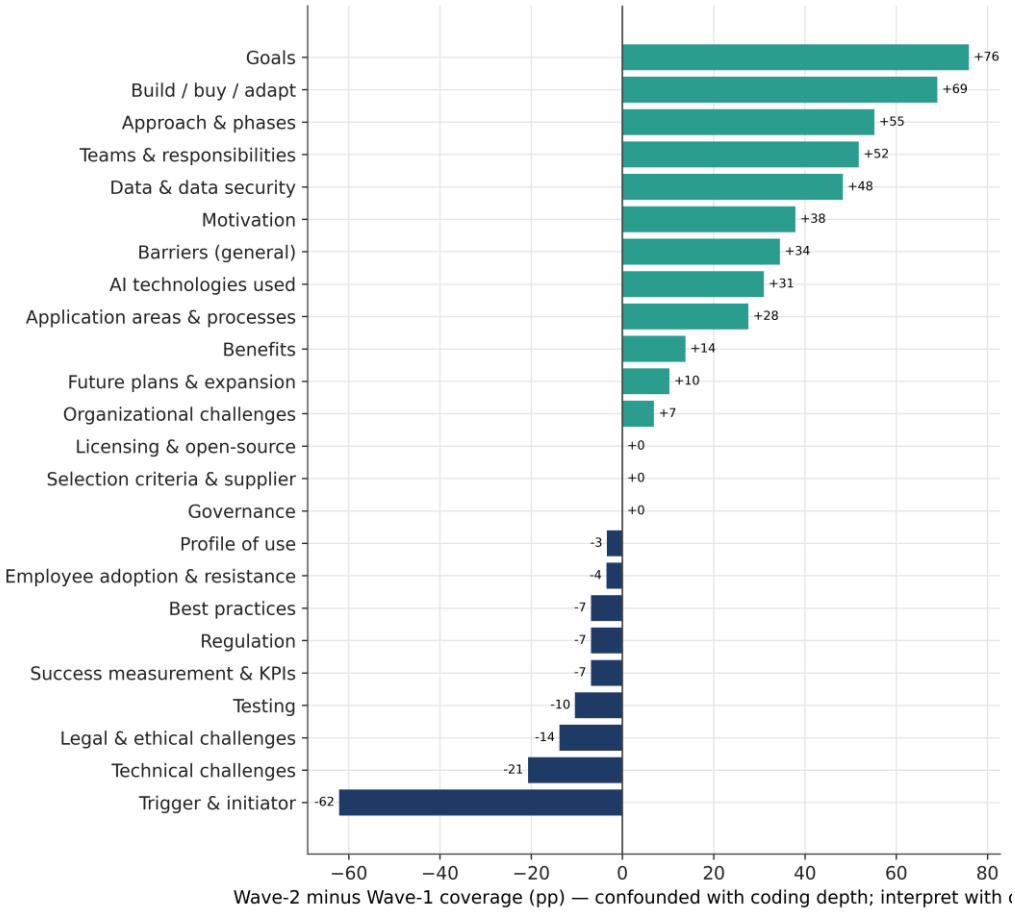


Figure 9.5 Difference in theme coverage between waves (W2–W1, percentage points). Broad positive shifts partly reflect denser Wave-2 coding; the Trigger drop and agentic rise are notable.[26]

Note: Coverage differences are coding-dependent and should be interpreted descriptively. Broad positive shifts partly reflect denser Wave 2 coding and greater respondent verbal volume.

One feature of the coverage data is too large and too contrary to the general upward drift to be an artifact, and it deserves separate comment. The Trigger theme dropped sharply, from 86.2 percent coverage in Wave 1 to 24.1 percent in Wave 2. This decline runs against the direction the artifact would push, since denser coding should if anything find triggers more often, not less, which makes the drop credible as a real change rather than a coding effect. It also coheres with the keyness finding, in which "triggering" was distinctive of Wave 1 and absent from the Wave 2 signature. Two independent measures point the same way. By 2026, firms in the sample had largely stopped narrating the moment of first encounter with AI. How adoption began, who initiated it, what prompted the first experiment: this had become old news, no longer the salient thing respondents reached for, displaced by talk of solutions, verification, and production. The decline of the trigger narrative is itself a marker of maturation — organizations that have moved from arrival to operation no longer dwell on their arrival.

[26] **Source:** Authors' own processing based on the AI-ImpactSK framework-coded interview dataset

9.4 What is artifact, what is real, and what it means

The discipline of this chapter rests on a single contrast between two kinds of wave signal. Some measures are coding-dependent. They scale with how densely the material was coded and how much respondents happened to say, and must be treated with caution. Others are coding-independent. They are computed on raw text, or validated against it, and can bear substantive interpretive weight. Table 9.1 sets the two side by side.

Table 9.1 Coding-dependent and coding-independent wave signals[27]

Wave signal	Status	Evidence and reading
Coding density (8.21 → 13.48 segments/interview)	Coding-dependent (caution)	~64% denser coding in Wave 2 (MWU p < 1e-7); an instrument shift, not a firm-level change.
Intensity (high share 0.43 → 0.91)	Coding-dependent (caution)	Wave-confounded; medium/low categories collapse in Wave 2. Not usable for substantive temporal claims.
Naive theme-coverage rises (Figure 16)	Coding-dependent (caution)	Partly mechanical, driven by coding depth and 1.6× respondent volume. Hypothesis-generating only.
Implementation formality, cross-wave	Coding-dependent (caution)	Correlates with segments/interview at r = 0.56; apparent maturation partly reflects coding resolution.
Respondent volume (696 → 1,123 words)	Coding-independent (robust)	Measured on raw text (MWU p = .0007, rank-biserial 0.52); real, and a source of the artifact in coded measures.
Agentic discourse (0.72 → 3.14 mentions; 37.9% → 69.0% of interviews)	Coding-independent (robust)	Text-based (MWU p = .0022, rank-biserial 0.44); manual code validates against text (r = 0.55, p < .0001).
Keyness shift (Wave 1: generative, machine, triggering → Wave 2: solutions, verification, production, agents)	Coding-independent (robust)	Log-likelihood on raw speech; immune to coding depth.
Trigger-theme drop (86.2% → 24.1%)	Coding-independent in import (robust)	Runs against the artifact's direction; corroborated by keyness. A real de-emphasis of the arrival narrative.
Tool-portfolio breadth (1.24 → 1.97 distinct tools)	Coding-independent (suggestive)	Text-based but only marginal (MWU p = .077); a tendency, not an established shift.
Governance / measurement, cross-wave	Coding-independent in import (robust)	Flat to slightly lower despite denser coding (24.1% → 17.2% each); the gaps are not artifacts.

Note: Coding-dependent signals are affected by wave differences in coding density and respondent verbal volume. Coding-independent signals are computed on raw respondent text or validated against it. Coverage-based changes are treated as hypothesis-generating unless corroborated by coding-independent evidence.

The asymmetry in this table is the chapter's central methodological lesson. The measures that would most readily support a triumphant story of rapid maturation, namely coverage, formality, and intensity, are exactly the measures most contaminated by the coding artifact. They cannot carry that story on their own. The measures that survive scrutiny tell a narrower but more

[27] **Source:** Authors' own processing based on the AI-ImpactSK framework-coded interview dataset and respondent-only transcript corpus

defensible tale. Respondents spoke more. They spoke increasingly in the vocabulary of agents and autonomy. Their distinctive lexicon moved from novelty toward operation, and they stopped recounting how it all began. The governance and measurement deficits did not close, which matters greatly for the interpretation below.

Taken together, the robust signals describe a coherent and short maturation. Over roughly seven months, the discourse of the firms in this sample moved along a recognizable trajectory: from the language of trigger, novelty, generative content, and machine learning toward the language of practical solutions, verification, production, and — above all — agents. This is the path from arrival to operation. In 2025 the conversation was about encountering a new general-purpose capability. In 2026 it was increasingly about running that capability inside working processes and checking its outputs. The agentic turn is the leading edge of the movement, the place where the technology itself advanced from conversational assistance toward tool-augmented, partially autonomous action, and where the firms' vocabulary advanced with it.

This shift in collective vocabulary is intelligible as a change in shared technological frames, the assumptions, expectations, and knowledge that organizational actors hold about a technology and that shape how they make sense of it (Orlikowski & Gash, 1994). A rising tide of agentic discourse, measurable in the keyness signature and the mention counts, is the trace of a frame in motion: the category through which respondents apprehend AI is shifting from "a generative tool I prompt" to "an agent that acts." The conceptual literature gives this frame its content. Agentic AI denotes LLM-based systems organized around profiling, memory, planning, and action that can pursue goals with reduced step-by-step human direction (Wang et al., 2024), and the enterprise interest these interviews register tracks the broader movement of the field toward foundation-agent architectures and multi-agent orchestration (Liu et al., 2025). The bridge from the generative present to the agentic near-future runs through the software scaffolding that surrounds the model. Exposure rises when LLMs are embedded in additional software (Eloundou et al., 2023), and the step from chat to agent is the step from a model one queries to a system one delegates to. That is precisely the step the keyness analysis catches, in the co-occurrence of "agents," "solutions," and "production" in the Wave 2 signature.

The maturation claim must be held to its proper scale. This is a discursive maturation observed in a small, ICT-skewed, two-wave sample, not a demonstrated transformation of organizational practice across the Slovak economy. The agentic rise is a change in how these firms talk and, for some ICT firms, in what they deploy. It is not evidence that agents had become widespread across sectors, nor that the firms talking about agents had built the governance to oversee them. The narrowness is the point. The robust signals establish that something real moved. They do not license the inference that everything moved, and the flat governance and measurement series are a standing reminder that some things conspicuously did not.

9.5 The agentic turn and the governance tension

The maturation narrative would be reassuring if it were the whole story. It is not, and the reason is the collision between the one signal that rose most robustly and the two that did not. Agentic discourse climbed from a minority idiom to a majority one. Codified governance and the measurement of benefits stayed flat at roughly one interview in six. The firms most fluent in the language of autonomous agents were not, on the evidence of this corpus, the firms that had built the oversight such agents demand. This is the governance tension that the temporal shift creates, and it is where Chapter 9 rejoins the monograph's argument that adoption races ahead of governance and measurement.

The tension acquires its force from the structure of autonomy risk. Mitchell et al. (2025) argue, on the basis of an autonomy-level taxonomy, that the risks an AI system poses to people increase as users cede more control to it, so that safety and other human-protective values are progressively traded away as agents are granted greater latitude to act without human checkpoints. Agentic AI is, by definition, a movement up that autonomy gradient. An agent that "writes code directly" (F2.20) or whose "core functionality consists of agents" (F2.06) is a system to which more decision latitude has been delegated than to a chatbot that drafts text for a human to approve (Štetka et al., 2026). On the Mitchell et al. (2025) analysis, the agentic turn is therefore not a neutral capability upgrade. It is a rightward shift on a risk curve, one that raises the stakes of the very governance the firms in this sample have not codified.

Placed against the findings of Chapter 6, the implication is uncomfortable. Governance in these firms is not absent in spirit. It surfaces as diffuse, informal vigilance, in the habit of checking outputs and guarding data ("I always check the outputs because they are not one hundred percent reliable," F1.06; "We are careful that our data is not used for model training," F1.06; Štetka et al., 2026). But it is not codified into policy, audit, or oversight structures, and it did not become more codified as the agentic vocabulary rose. Informal vigilance is a defensible posture toward a generative tool whose outputs a human reviews before use; the human-in-the-loop is the control. The distinctive promise of agentic AI is precisely to remove or thin that loop, to let the agent act, chain steps, and execute without a checkpoint at each one. An informal-vigilance regime calibrated to a generation-and-review workflow does not automatically transfer to an action-and-execution one, and the autonomy-risk argument implies that it should not be assumed to. As firms move up the autonomy gradient while leaving their governance at the level of individual carefulness, the gap between what the technology can do without supervision and what the organization is structured to supervise widens. The denser the agentic discourse, the more salient that widening becomes.

This is the maturation narrative's sting in the tail. The temporal signal that is most robust — the agentic rise — is also the one that most sharpens the monograph's governance concern. It moves firms toward a class of system for which informal vigilance is least adequate, without any matching movement toward formal oversight. The flat governance and measurement series, read in 2025 as a static deficiency, read in light of the 2026 agentic turn as a deficiency that is becoming more consequential even as it fails to close. A firm that does not measure the benefits of a chatbot foregoes useful information. A firm that does not govern an autonomous agent foregoes control. The 2025-to-2026 shift did not relieve the adoption-ahead-of-governance pattern that organizes this book. It raised the price of leaving it unaddressed.

The answer to RQ4 has two halves, and they must be held together. Substantively, the discourse of the firms in this sample matured in a real but circumscribed way between 2025 and 2026. Respondents spoke more, their vocabulary moved from novelty and trigger toward solutions, verification, and production, the agentic register rose from a minority to a majority idiom on triangulated evidence, and the narrative of first encounter faded. Methodologically, this answer could only be reached by first quarantining the coding-depth artifact: the roughly 64 percent denser coding of Wave 2, the 1.6-fold rise in respondent verbal volume, the wave-confounded intensity scores, and the coverage and formality measures that scale with all three. None of it can be read as substantive change without corroboration from a coding-independent signal.

Modeling that separation is itself a contribution of the chapter, and of the monograph's methodological program (Štetka et al., 2026). The honest temporal claim is the narrow one that survives the artifact, not the broad one the naive numbers invite. The narrow claim is not a

deflationary anticlimax. It lands on the study's most pointed concern. The agentic turn is the one robust temporal signal; agentic systems sit higher on the autonomy-risk gradient (Mitchell et al., 2025); the governance and measurement structures that would answer that risk did not move at all. The firms of the AI-ImpactSK corpus entered 2026 talking more like operators of autonomous systems and no more like governors of them. Real maturation in capability-talk, stasis in oversight: that asymmetry is the temporal expression of the thesis that runs through every chapter of this book.

10 Discussion

Generative AI sits in 89.7 percent of the firms in this sample. AI agents reach 44.8 percent, with a mean of 2.5 distinct AI modalities per firm (Štetka et al.

, 2026). Adoption, on these numbers, is not the problem. What firms say about governing and evaluating that adoption is a different matter — and the gap between the two is the subject of this chapter.

Chapters 6, 7, and 8 were written as if they described three separate problems. Chapter 6 documented a governance gap, Chapter 7 a measurement paradox, and Chapter 8 a set of adoption configurations. Each built its own evidentiary case and reached its own verdict. Treating them separately was necessary for analytic clarity, but it understates what the evidence shows when read as a whole. The three findings are better understood as three faces of one regime. Adoption is broad, generative-led, and overwhelmingly informal. Governance rarely surfaces in firms' own accounts. Evaluation surfaces a little more often but is rarer still as a practice. And the shape adoption takes is not one story but several, conditional on the kind of firm doing the adopting. We characterize this as a *thin-governance, thin-measurement, configuration-dependent* regime, and its defining feature is decoupling: adoption runs ahead of the machinery that would govern it and the machinery that would tell anyone whether it is working.

The interest is in the structure of that decoupling and the consistency with which the data display it. The firms in this corpus do not adopt AI tentatively and then neglect governance and measurement because adoption is shallow. They adopt widely, and the governing and measuring lag anyway. Formal governance was coded in only 13.8 percent of firms — explicit attention to regulation in only 6.9 percent. Benefits were articulated by 65.5 percent of firms but measured by only 20.7 percent. No subgroup difference in governance salience was detected by firm size or sector, though that comparison is underpowered, so the data cannot establish that the pattern is uniform across firm types. The picture is not one of some firms catching up on accountability while others fall behind. Across the sample, adoption has moved well ahead of measurement and explicit governance, and it has done so along several distinct paths at once.

A clarification before the argument proceeds. The governance figures need careful handling. The interview protocol contained no standalone governance or policy question. Governance and regulation entered the record only when a respondent raised them, under the prompt on legal and ethical challenges or the forward-looking question about expected regulation. The 13.8 percent and 6.9 percent figures therefore measure how often governance was foregrounded in firms' own accounts of adoption, not whether governance practices exist. Measurement is on firmer ground: it was directly probed, through a goal-measurability question and an explicit question about how firms measure success and which figures they track. The measurement paradox is well grounded for that reason, and this chapter treats it as the stronger of the two findings.

The remainder of the chapter proceeds in four moves. It names the triad and shows how the descriptive structure of the corpus encodes it. It then sets out what the regime contributes theoretically, with the contribution stated modestly. It places the findings in dialogue with the literatures on SME and Central and Eastern European (CEE) adoption, on generative and agentic AI, and on governance. Finally it reads the maturation narrative and the boundary conditions that configuration-dependence imposes. The limits of the evidence stay in view throughout.

Fifty-eight purposive interviews, an ICT-skewed sample, and a coding process whose depth differed across the two waves describe a hypothesis about a population, not a measured fact about it.

10.1 The adoption–governance–evaluation triad

The organizing device is a triad. Adoption, governance, and evaluation are three activities an organization deploying a consequential technology might be expected to undertake together. Adoption is the act of putting AI to work. Governance sets rules for how it is put to work: who may use it, on what data, with what oversight, against what regulatory constraint. Evaluation finds out whether the work is worth doing, by measuring benefits and attaching numbers to claims. In a mature technology regime these three move in rough proportion. An organization that deploys a tool at scale builds the policies that contain its risks and the metrics that justify its cost. The AI-ImpactSK data describe a regime in which they do not move together. Adoption is the dense core of the corpus — governance and evaluation sit at its sparse edge.

Consider the proportions in which the two flanks lag the center. Governance lags more sharply, with one caveat carried forward from the chapter opening: the governance figures index how often the topic was spontaneously raised, not a census of practice. Category G, governance and regulation, accounts for just 1.9 percent of all coded segments in the corpus (Štetka et al., 2026). Formal governance appears in 8 of 58 firms (13.8 percent, 95 percent Wilson score interval 7.2 to 24.9) and explicit regulatory awareness in 4 of 58 (6.9 percent, 95 percent Wilson score interval 2.7 to 16.4). Chapter 6 showed this is not a coding artifact. The denser Wave 2 coding inflated many coverage measures but did not raise the governance signal, which held essentially flat across waves. Risk is not simply absent from firms' minds. A coding-independent governance lexicon, with words such as policy, compliance, GDPR, audit, and oversight, does appear in the transcripts. It does not align with the formal governance code (point-biserial $r = -0.20$, not significant). What firms exhibit is not codified governance but a diffuse, informal vigilance — a habit of verifying outputs and protecting data that lives in individual practice rather than organizational rule. The informal-vigilance posture has a name in the behavioral literature: an aversion to delegating judgment to a system whose errors the user has seen (Dietvorst, Simmons, & Massey, 2015). Evaluation lags less dramatically but is, if anything, the more analytically surprising, because here firms demonstrably possess the thing that measurement would quantify. Benefits are articulated by 38 of 58 firms (65.5 percent) and measured by only 12 (20.7 percent). The probability that a firm measures a benefit given that it claims one is 4 in 38, or 10.5 percent. The dissociation between claiming and measuring is statistically robust (McNemar exact statistic 8.0, $p = 6.9 \times 10^{-5}$, surviving false-discovery-rate correction). More than half the sample, 58.6 percent, asserts a benefit it never measures.

These are not three independent shortfalls that happen to co-occur. They are plausibly connected rather than merely co-occurring, and the connection is a shared informality in how AI enters the firm. The retrospective accounts gathered here read as describing a sequence in which the doing came first and the governing and measuring came later or not at all. We infer that ordering from how respondents narrate their adoption, not from any sequence the study observed directly. A tool that visibly saves time does not need a business case to survive. An employee who finds ChatGPT useful does not wait for a policy before using it. The informality that Chapter 8 documented as the default mode of adoption, bottom-up, individual, voluntary, frequently without a defined project, pilot, or set of key performance indicators, is the soil in which the governance and measurement gaps grow. Where adoption enters through the front door of a sanctioned project, governance and measurement can be attached as conditions of

approval. Where it enters through the side door of individual experimentation, there is no approval gate to attach them to. A shared-mechanism reading is one the data are consistent with — not a causal relationship the study can establish. The same informality that makes adoption fast and broad is associated with governance and evaluation being slow and rare, and that association holds across firm types rather than concentrating in any one.

The retrospective accounts also caution against reading too much intention into the lag. They do not suggest firms are reckless or that they will never govern or measure. Several point the other way, toward an expectation that measurement will come. "At the moment it is more of a gut feeling, but I believe we will gradually get to measurement," one ICT respondent said, and the temporal posture in that sentence is characteristic: measurement is deferred, not denied (F1.03; Štetka et al., 2026). The pattern is one of ordering and asymmetry of effort. Adoption is treated as urgent and self-evidently worthwhile. Governance and measurement are treated as deferrable refinements, to be added once the value is obvious. Whether the deferral reflects disposition or simply the early stage of adoption is a question Section 10.5 takes up directly, because the consequences differ sharply depending on the answer.

The thesis encoded in the co-occurrence structure

The triad is not only an inference from base rates. It is legible in the shape of the corpus. Figure 5.5, the radial co-occurrence map of the twenty-four themes, renders the regime as a topology. At its center sits a dense, tightly linked core of adoption-narrative themes: application areas (B1, n = 48), future plans (H1, n = 47), AI technologies used (C1, n = 45), and benefits (F1, n = 38), bound together by strong participant-level Jaccard associations (B1 with H1 at J = .73, B1 with benefits at J = .72, B1 with C1 at J = .69). These themes co-occur because they are how firms talk about adopting AI: what they use it for, what they plan next, which technologies they have, what they believe they are gaining. Around this core, governance, regulation, testing, and KPIs sit at or near the periphery, near-isolated, with few or no strong links to the adoption core. A firm narrating its application areas and future plans is, statistically, narrating a connected story. A firm's governance, when it appears at all, appears as an island. The decoupling that the base rates establish numerically, the co-occurrence structure displays spatially. The things that might be tied to adoption — its governance and its measurement — are the things that float free of it.

10.2 Theoretical contributions

The empirical regime does more than fill a gap in the descriptive record. It lets the monograph apply four theoretical lenses and generate propositions for later testing. Two of those applications sharpen the underlying theory enough to count as contributions in their own right — the TOE framework shapes the form of adoption more than its fact, and the measurement deficit is self-confirming. The other two applications are read here as productive uses of existing theory, offered as hypotheses for representative work, not as extensions the present design can secure.

The technology–organization–environment framework: context shapes form, not fact

The technology–organization–environment (TOE) framework holds that a firm's adoption of an innovation is shaped by three interacting contexts, the technological, the organizational, and the environmental (Tornatzky & Fleischer, 1990). In classical use, the dependent variable is adoption itself: whether, and how much, a firm takes up a technology. The AI-ImpactSK evidence invites a reframing. Where the technology is a general-purpose tool available to anyone with a

browser and a subscription, the fact of adoption is nearly invariant across the contexts TOE specifies. Generative AI is in 89.7 percent of firms regardless of their organizational sophistication or environmental pressure. What varies is not whether firms adopt but how — the formality, the configuration, the depth of implementation. The most defensible sector finding illustrates this. ICT firms do not adopt AI more than non-ICT firms in any meaningful sense. They tend to adopt it more formally, with a mean implementation-formality index of 2.50 against 1.78 for non-ICT firms (Cliff's $\delta = 0.36$). That tendency is modest and directionally consistent, suggestive rather than established: the one confound-clean test, a within-wave stratified permutation, returns $p = .059$, and the within-wave δ falls to 0.21 in Wave 2. So the contrast is best read as a lead worth pursuing, not a settled effect. Read with that caution, the organizational and environmental contexts appear to bear on the form of adoption more than on its presence. This is the first genuine sharpening: it relocates TOE's explanatory interest from the adoption decision, which the spread of general-purpose AI has largely settled, toward the implementation architecture, which it has not. The contexts still matter. They appear to matter for the shape of practice.

An institutional reading: legitimacy and decoupling

If TOE bears on the variation in form, institutional theory offers a reading of the uniformity in fact. DiMaggio and Powell's (1983) account of institutional isomorphism holds that organizations in a field grow alike through coercive, mimetic, and normative pressures, and that they often adopt practices to secure legitimacy rather than because they have established that the practices improve efficiency. The Slovak pattern is consistent with mimetic and normative adoption running ahead of any governed, efficiency-tested process. Firms adopt generative AI under high uncertainty about its value, the uncertainty they acknowledge when they say they cannot quantify what it does for them, which is the condition under which DiMaggio and Powell expect firms to copy what peers and competitors are visibly doing. In these terms, the rapid diffusion of ChatGPT-like tools can be read as a mimetic cascade. A technology becomes legitimate to adopt because everyone is adopting it.

The sharper construct for the adoption–governance gap is decoupling. Meyer and Rowan (1977) argued that organizations adopt legitimating structures that are loosely coupled to, or formally separated from, their actual work. Bromley and Powell (2012) extended this to means–ends decoupling, where a policy is adopted but its connection to outcomes goes unexamined. The juxtaposition at the heart of the measurement paradox fits the first idea. Why would firms invest effort in adopting a tool whose benefits they will not measure? Because where adoption is driven by legitimacy, measurement is beside the point. One adopts to be the kind of firm that has adopted. The benefit that licenses the adoption is reputational, not a measured return. The governance gap follows from the same logic. Where adoption is driven by the pursuit of legitimacy through conformity rather than an internal rational-control process, the apparatus of internal control — policies, audits, regulatory mapping — is not what is being sought and is not what gets built. Decoupling supplies the mechanism by which adoption can be enthusiastic and ungoverned at once. The enthusiasm is for the legitimacy, and legitimacy does not require governance. This reading is offered as a proposition for designs that can observe policy and practice directly. The present data, which record only how often governance was raised, are consistent with it but cannot confirm it.

Dynamic capabilities: sensing and seizing without reconfiguring

The dynamic-capabilities framework decomposes a firm's capacity to profit from change into three activities: sensing opportunities, seizing them, and reconfiguring the firm's assets and routines to capture their value (Teece, 2007). Read this way, the AI-ImpactSK firms are vivid

sensers and able seizers who appear to under-invest in reconfiguration, and the measurement paradox is the signature of that under-investment. Sensing is everywhere in the corpus. Firms have noticed AI, identified use cases, and formed clear expectations of benefit, with two-thirds articulating concrete gains. Seizing is widespread too. They have put the tools to work, often across multiple modalities. The third activity is the thin one. Reconfiguration — the rebuilding of routines, the construction of measurement systems, the embedding of the tool into governed and instrumented processes — is the work that converts episodic use into durable capability, and it is the work the data show firms doing least. The 20.7 percent measurement rate reads, in these terms, as a reconfiguration deficit. A tool whose effects are not measured cannot be deliberately improved, scaled on evidence, or defended in a budget. The same deficit appears on the governance side, where the low salience of policy and oversight is consistent with control routines that have not been rebuilt around the new technology. The useful question this lens raises is which of the three capabilities binds in non-frontier AI adoption. Not sensing: the firms sense well. Not seizing: general-purpose tools are cheap to seize. Reconfiguration is the candidate — the most organizationally demanding of the three, the one that requires building rather than buying. This is a proposition the data motivate rather than prove.

Configurational theory: equifinal paths to adoption

The fourth lens concerns plurality. Configurational theory in organization and information-systems research holds that firms are best understood not as points on a set of independent variables but as coherent gestalts, a small number of internally consistent types, and that multiple distinct configurations can be equifinal, reaching a similar outcome by different routes (Fiss, 2011; Miller, 1986). The four indicative adopter configurations of Chapter 8 instantiate this logic for AI adoption. The multi-modal technical adopters (C0, n = 16), the vision and agent specialists (C1, n = 10), the generative mainstream of broad-economy adopters (C2, n = 23), and the generative-only minimalists (C3, n = 9) are not four points on a sophistication scale. They are four qualitatively different ways of being an AI-adopting firm, distinguished by modality mix, breadth, and formality. The configurational reading matters because it suggests the thin-governance, thin-measurement regime is not produced by a single deficient adopter type. It appears to be reached by several. The vision and agent specialists are the broadest adopters in the sample (breadth 3.80, the highest of any configuration) yet measure benefits at a rate of zero percent. The generative-only minimalists are the narrowest and least formal (breadth 1.00, formality 1.33) yet measure at 33 percent — a higher rate than the technically sophisticated specialists. There is no monotonic relation here between adoption sophistication and accountability, with firms arriving at non-measurement and non-governance from opposite ends of the technical spectrum.

That reading must be held at the level of confidence its derivation warrants. The four configurations were obtained by k-modes clustering on coding-independent metadata and AI-profile variables. Internal diagnostics actually favored a two-cluster split (silhouette 0.369). The four-configuration solution is a heuristic descriptive partition with weak separation (bootstrap adjusted Rand index 0.35), which is why Chapter 8 styled the groups as indicative rather than definitive. The small per-cluster cells (zero of ten measuring in C1 against three of nine in C3, for instance) are best read as descriptive, not as an established pattern. They are a defensible reading of structure in the data, not a claim that exactly four types exist in the Slovak economy. What survives the caution is a qualitative point, offered as a proposition: more than one path appears to lead to the regime, and that point does not depend on the precise number or boundaries of the clusters.

The measurement paradox as a firm-level re-run of the productivity paradox

The contribution that most directly sharpens existing theory concerns the measurement paradox. Read it as a firm-level re-run of the productivity paradox, but a re-run with a twist the macroeconomic literature did not anticipate. The classical productivity paradox held that the value of information technology was real but did not show up in aggregate statistics, and the leading resolution was a measurement resolution: the value existed but was intangible and hard to measure, so the apparent absence of returns was an artifact of measurement rather than a fact about value (Brynjolfsson & Hitt, 2000). Brynjolfsson, Rock, and Syverson (2017) re-opened exactly this argument for AI, framing the gap between AI's promise and its measured returns as a predictable consequence of the time needed to build the intangible complements — skills, process redesign, new routines — that make value visible. On its face the Slovak measurement paradox looks like a micro-instance of this lag. Firms perceive benefits before they have built the apparatus to measure them, and a J-curve account would predict that the measurement will come as the complements accumulate (Brynjolfsson, Rock, & Syverson, 2021). The benefits-realization literature makes the same point from the management side: value from an IT investment has to be actively planned for and tracked, and is not captured by default (Ward & Daniel, 2012).

The AI-ImpactSK evidence supports a sharper and partly different reading, and this is the second genuine sharpening. In the macroeconomic version, the payoff is unmeasured primarily because measurement lags, with value accumulating as intangible capital that has not yet surfaced. In the Slovak firm-level version, the payoff is unmeasured for reasons the lag account does not fully cover. Many firms are not building measurement systems that will eventually catch up with the value. They are not measuring, and they say so. "I cannot quantify it because we did not track it"; "Completely honestly, I did not quantify it"; "That is also why we are not measuring things dramatically at this stage" (F1.12, F2.02, F2.23; Štetka et al., 2026). Whether this reflects disposition, limited capability, or simply work not yet undertaken cannot be determined here, and Section 10.5 weighs the early-stage possibility carefully. What the data do show is that, for now, the apparatus to detect the benefit is largely absent.

This is where the second strand of the value literature becomes decisive, because it turns the absence into something self-perpetuating. Kohli and Devaraj (2003) established that the detection of IT payoff is endogenous to measurement design: whether value is seen depends on how, and whether, it is measured. Applied at the firm level, this is the engine of the paradox. A firm that does not measure does not see returns. A firm that does not see returns has no evidential basis on which to justify the investment in measurement that would let it see returns. Non-measurement reproduces itself. The benefit remains a "gut feeling" not because it is unreal but because the firm has never built the instrument that would make it real to its own decision processes. The contrast with the controlled generative-AI experiments shows this is a measurement gap and not a value gap. Where measurement is imposed, the benefits are large and unambiguous: a 15 percent productivity gain for customer-support agents, a 40 percent reduction in task time for professional writers, a 55.8 percent speed-up for developers (Brynjolfsson, Li, & Raymond, 2025; Noy & Zhang, 2023; Peng, Kalliamvakou, Cihon, & Demirer, 2023). The Slovak firms operate in the same technological reality as those experiments. What separates them is the presence of the apparatus to detect the benefit. The monograph's contribution is to specify the firm-level mechanism the macro-paradox literature left implicit. The modern productivity paradox, reproduced inside the firm, is sustained not only by lag but by an endogenous measurement deficit that, once it takes hold, is self-confirming. The jagged-frontier evidence sharpens the stakes further. Because AI's benefits are real but uneven, and

because on tasks beyond its competence AI can degrade rather than improve performance (Dell'Acqua et al., 2026), a firm that does not measure cannot know which side of the frontier it is on. The unmeasured benefit and the unmeasured harm are the same blind spot.

10.3 Dialogue with the literature and the maturation trajectory

The regime documented here both confirms and extends several adjacent literatures, and locating it against them is part of establishing what is new in the Slovak evidence.

The closest comparator is Schwaeke, Peters, Kanbach, Kraus, and Jones (2024), whose synthesis of SME AI adoption concluded that adoption has become normalized but remains predominantly informal, tool-level, and experimental, with structured, strategic implementation the exception. The AI-ImpactSK corpus confirms this informal-normalization thesis, and in a context the original synthesis did not cover: a single, SME-heavy CEE economy observed across two closely spaced waves. The informality is the organizing default, visible in the testimony that adoption "happens organically," that "each person uses it independently," and that "some we may be using, but it is not formally defined" (F1.01, F2.04, F2.01; Štetka et al., 2026). The monograph extends Schwaeke et al. in two ways the synthesis could not. First, it attaches base rates to a pattern the prior literature described qualitatively: 13.8 percent, 6.9 percent, 20.7 percent, with the governance figures read as salience rates rather than census counts. Second, it shows that informal normalization is not a static endpoint but a state with internal dynamics. Across the two waves the corpus reveals a maturation toward agentic AI and toward a more production-oriented way of talking, which suggests informal normalization is a phase firms move through rather than a terminus at which they rest. The Slovak evidence turns a cross-sectional generalization into a process.

That maturation connects the findings to the agentic-AI literature, and here the contribution is one of location. The discussion of agentic AI has been conducted largely in and about the frontier, the firms, the labs, and the economies at the technological edge (Mitchell, Ghosh, Luccioni, & Pistilli, 2025; Wang et al., 2024). The AI-ImpactSK corpus locates the agentic turn in a non-frontier economy, and shows that the turn is not confined to the frontier. The coding-independent rise in agentic discourse documented in Chapter 9 — respondent mentions of agentic terms rising from a mean of 0.72 to 3.14 per interview (Mann–Whitney $p = .0022$, rank-biserial 0.44), and the breadth of agentic talk rising from 37.9 percent to 69.0 percent of interviews — establishes that the ICT firms in this sample were absorbing the agentic vocabulary, and in some cases the agentic practice, within months of its arrival at the frontier. This matters for the governance discussion because the agentic turn raises the stakes of the governance gap. Mitchell et al. (2025) argue that the risks of AI agents rise with their autonomy, as users cede more control to systems that act rather than advise. A regime that has not foregrounded governance for advisory generative AI is less prepared still for agents that write code directly and operate with reduced human mediation. The Slovak data show firms moving toward such agents while governance discourse remains at the 1.9-percent-of-segments level it occupied for the less autonomous tools. The agentic turn, arriving into a thin-governance regime, widens the distance between the autonomy of what is being adopted and the oversight discussed around it.

On governance proper, the findings speak to two literatures. The first is the principles-to-practice critique, of which Mittelstadt (2019) is the sharpest statement: high-level ethical principles cannot by themselves produce ethical AI, because they lack the professional structures, proven methods, and accountability mechanisms that would translate values into practice. The Slovak

evidence is the empirical shadow of that critique. What the firms exhibit is values without binding structures, a diffuse, informal vigilance about verifying outputs and protecting data that rarely crystallizes into codified governance, at least not in how firms narrate their adoption. The non-alignment between the firms' governance lexicon and their formal governance code ($r = -0.20$, not significant) is the principles-to-practice gap made measurable. The principles are present in the talk; the practice is not foregrounded in the structure. The second governance literature is the work on shadow IT and its generative-AI successor, shadow AI (Silić, Silić, & Kind-Trüller, 2025). The Slovak pattern of bottom-up, individual, voluntary adoption that proceeds without sanction or policy is shadow AI in organizational form, and the corpus supplies independent, SME-heavy, two-wave field evidence of a phenomenon the shadow-AI literature has largely theorized. One firm's words are telling here, and they cut against a flat reading. The respondent says "we do not have a precise policy on how to use AI," yet the firm holds ISO 27001 certification and is pursuing ISO 45001 (F2.02; Štetka et al., 2026). The accurate point is not that the firm lacks governance, but that it has no AI-specific policy despite formal security certification. Set beside "I always check the outputs because they are not one hundred percent reliable," it captures the defining tension of shadow AI — unsanctioned use coexisting with a private, uncodified risk-awareness that the organization has not converted into rule (F1.06; Štetka et al., 2026).

The temporal evidence of Chapter 9, handled with the caution the coding-depth artifact demands, supports a reading of the regime as moving rather than static, along two legible dimensions. The first is a shift in modality, from generative toward agentic. The second is a shift in posture, from novelty toward production. The keyness analysis traces both. The vocabulary distinctive of Wave 1, insurance, assessment, learning, triggering, generative, machine, is the language of arrival and appraisal. The vocabulary distinctive of Wave 2, solutions, verification, screening, practical, context, production, claude, agents, is the language of operation and consolidation (Štetka et al., 2026). Firms are moving from talking about AI as a new and broad category to be assessed toward talking about specific instruments deployed into working systems and routinely checked. This rests on coding-independent measures that survive the artifact.

The implication for the trajectory is double-edged, and the chapter resists resolving it optimistically. On the encouraging side, the rise of "verification" and "screening" in the Wave 2 lexicon suggests the informal vigilance documented on the governance side is becoming a more routinized part of practice, with firms building, in everyday use, some of the checking behavior that codified governance would formalize. On the discouraging side, the maturation is occurring along the adoption and modality dimensions far more visibly than along the governance and measurement ones. The robustness check of Chapter 9 is the constraint. The denser Wave 2 coding did not raise the governance or measurement signals, which held flat or edged downward even as agentic discourse surged. The regime is maturing on the axis it was already strong on, and maturing slowly if at all on the axes it was weak on. Should the trajectory continue on its current bearing, the firms in this sample would become more sophisticated adopters of more autonomous AI without becoming correspondingly better governors or evaluators of it. That would widen the decoupling rather than close it. The maturation narrative is not, on present evidence, a story of the gaps closing.

10.4 The early-adoption-stage rival explanation

Before drawing boundary conditions, one rival explanation deserves a direct hearing, because it bears on the two findings the chapter rests most weight on. The design is an eight-month, two-

wave cross-section. It cannot separate "not measured, and low in governance salience, by disposition" from "not yet, because adoption is early." Those are different claims with different implications, and the data underdetermine the choice between them.

Take measurement first. A dispositional reading says firms are content to run on gut feeling and will keep doing so. A stage reading says the firms are early enough that measurement systems have not been stood up yet, and would be built as use stabilizes and budgets start to demand justification. The same forward-looking quotes that the chapter has used elsewhere fit the stage account at least as well as any other. "I believe we will gradually get to measurement" is, on its face, a statement about sequence and timing (F1.03). A firm eight months into serious generative-AI use has not had long to build evaluation routines that, in the IT literature, often lag deployment by years (Ward & Daniel, 2012). The self-confirming dynamic of Section 10.3.5 does not settle the question. It explains why non-measurement, once entrenched, persists. It does not establish that what we observe is entrenchment rather than an early and reversible state. Both readings are live, and a cross-section taken this close to the technology's arrival cannot adjudicate between them.

The same applies, with more force, to governance. Here the elicitation point compounds the stage point. Governance was not directly probed, so its low salience may reflect a topic not yet front of mind rather than an absence of practice, and either of those, in turn, may be a function of how early the firms are. Codified AI governance tends to follow scale, incidents, and external pressure, none of which a young adoption has had much time to generate. The forward-looking regulatory question drew thin responses, which a dispositional reading treats as indifference and a stage reading treats as prematurity: regulation that is still taking shape is hard to foreground. The honest position is that an eight-month window cannot tell a firm that will codify governance later from one that will not. The base rates fix where these firms were in 2025 and 2026 — not where they are heading, and the stage explanation remains a serious competitor to the dispositional one for both findings. Longitudinal follow-up is the only way to choose between them, which is why the next chapters call for it.

The third term in the thesis, configuration-dependent, is also the principal boundary condition on what the chapter has argued. Taking it seriously means resisting the temptation to tell a single adoption story. The regime is thin-governance and thin-measurement across the sample, but it is not uniform in adoption architecture, and the configurational evidence forbids flattening the four adopter groups into one. There is no single Slovak AI adopter. There is a sophisticated multi-modal technical builder who measures about a third of the time. There is a vision-and-agent specialist who is the broadest adopter and measures not at all. There is a broad-economy generative-mainstream adopter who is the largest group and the most likely to govern. And there is a generative-only minimalist who is the lightest user yet measures more often than the specialist. Any claim that begins "Slovak firms adopt AI by…" is incomplete unless it specifies which configuration it describes. The thesis locates the uniformity and the plurality at different levels. The decoupling of adoption from governance and measurement is the sample-wide regularity. The form of adoption — its modality, breadth, formality, and even its measurement rate — is configuration-specific. Configuration-dependence is a constitutive part of the thesis, not a hedge appended to it.

Several further boundary conditions discipline the reach of these claims, and they have been in view throughout rather than relegated to a closing apology. The sample is fifty-eight firms, purposively and conveniently recruited rather than drawn to represent the Slovak economy, so the base rates here are characterizations of this corpus and hypotheses about the population,

not measured population parameters. The articulated and coded rates are lower bounds — a practice not mentioned in a roughly 33-minute interview may still exist, which matters most for governance, where there was no standalone question. The sample is skewed toward the ICT sector, which comprises 44.8 percent of the firms. Because ICT firms are early and formal adopters, this skew most plausibly biases the picture toward more formality, measurement, and sophistication than the broader Slovak economy would display, so the true population-level gaps may be wider than these already-stark figures suggest rather than narrower. The two waves were coded at different depths, Wave 2 roughly 64 percent more densely, so every cross-wave comparison built on coverage has been treated as hypothesis-generating and anchored, where it bears weight, in coding-independent measures. The maturation narrative of Section 10.4 rests on those measures for that reason. The configurations carry their own humility (adjusted Rand index 0.35) and are read as indicative. The inferential statistics that survive correction, the McNemar test on the measurement paradox most prominently, establish that the dissociations are not chance features of these fifty-eight firms, but they do not license claims about a population the sampling design was never built to represent. And the early-stage confound of Section 10.6 means the disposition-versus-timing question stays open. These limits do not dissolve the regime. The decoupling is too consistent across base rates, co-occurrence structure, triangulated text measures, and qualitative testimony to be an artifact of any one of them. But they fix the epistemic status of the claim. The thin-governance, thin-measurement, configuration-dependent regime is the best available description of how a small, SME-heavy, ICT-inflected sample of Slovak firms was adopting AI in 2025 and 2026, offered as a foundation for the representative and longitudinal work the next chapters call for.

10.5 Synthesis

The three findings of the monograph compose a single regime, intelligible as a coherent organizational phenomenon rather than a coincidence of deficits. The firms in this sample adopt AI broadly, informally, and increasingly along an agentic frontier. They foreground governance rarely and measure benefits rarely. They do this along several distinct configurational paths that share an accountability gap while differing in much else. The triad of adoption, governance, and evaluation comes apart in a characteristic way, with the doing ahead of the governing and the measuring, and that single pattern — rather than three separate ones — is what the governance gap, the measurement paradox, and the configurations of adoption each show from a different angle. The theoretical payoff is stated modestly. Two sharpenings are offered as contributions: TOE bears on the form of adoption more than its fact, and the measurement deficit is self-confirming once it sets in. The institutional reading, framed through decoupling, the dynamic-capabilities reading, and the configurational reading are offered as propositions for representative and longitudinal work to test. Whether the gaps reflect disposition or the early stage of adoption is a question this cross-section cannot answer. What it can do is map where these firms stood — and hand the next study a sharper set of questions than it started with.

11 Implications for Practice, Policy, and Method

Fifty-eight firms took up artificial intelligence widely; far fewer wrote down how they govern or evaluate it. That descriptive pattern, traced across the preceding chapters, is what this chapter turns into guidance for three audiences: managers and practitioners, policymakers in the European and Slovak arenas, and the methodological community of qualitative and mixed-methods researchers. The recommendations are deliberately proportionate. They rest on a purposive, ICT-skewed sample of fifty-eight interviews across two waves (Štetka et al., 2026), so they follow from associations and descriptive patterns rather than from established causes, and each is hedged accordingly. The pattern in this sample is that codified governance and routine measurement rarely surfaced alongside enthusiastic adoption. The most useful interventions, on that reading, are the ones that let firms keep the speed of bottom-up adoption while gradually supplying structure, evidence, and oversight where the accounts suggest these are thin.

11.1 Implications for managers and practitioners

The managerial implications follow from the central asymmetry documented in the study: firms are already experimenting with AI, but the routines that would govern, evaluate, and scale that experimentation remain thin. The task for managers is therefore not to slow adoption down or to replace bottom-up initiative with heavy bureaucracy. It is to convert existing informal vigilance into lightweight, reviewable practices: simple policies, tool registries, verification rules, and a small number of task-level metrics. The recommendations below are deliberately proportionate to the SME-heavy and ICT-skewed character of the corpus. They are offered as practical starting points rather than as a universal maturity model.

Lightweight governance built on existing vigilance

The governance pattern documented in Chapter 6 is the most actionable finding for managers. Explicit governance was raised spontaneously in only 8 of 58 interviews (13.8%, 95% Wilson score CI 7.2–24.9) and regulation in only 4 (6.9%, CI 2.7–16.4), with Category G accounting for a mere 1.9% of all coded segments. These rates measure how often governance was foregrounded in firms' own accounts of adoption, not whether governance practices exist. The interview protocol carried no standalone governance question, so the topic entered the record mainly when a respondent volunteered it. Read that way, the finding is not negligence. The triangulation evidence shows that governance talk is diffuse: a coding-independent governance lexicon (verification, data protection, oversight) appears across interviews but does not align with the formal governance code (point-biserial $r = -0.20$, ns) and is flat across waves. Firms appear to practise informal vigilance without codifying it, the familiar principles-to-practice gap (Mittelstadt, 2019; Morley et al., 2020). The managerial opportunity is therefore formalization rather than cultural change.

The most leverageable asset is the verification habit already embedded in practice. When a respondent says, "I always check the outputs because they are not one hundred percent reliable" [F1.06, W1, ICT, SME], that is a human-in-the-loop control awaiting documentation. Lightweight governance means writing such habits down. A one-page acceptable-use policy can specify which tools may be used for which data; a simple AI tool registry can record what is in use and by whom; an explicit human-in-the-loop verification standard can name the tasks for which outputs must be reviewed before they leave the firm. These instruments are proportionate to SME capacity. They require no internal audit function of the kind designed for AI-developing

organizations (Raji et al., 2020), yet they convert tacit caution into reviewable practice. They also map onto established definitions of codified organizational AI governance (Mäntymäki et al., 2022), onto the "Govern" function of the NIST AI Risk Management Framework (Tabassi, 2023), and onto the externalizable, modular "ethics-as-a-service" support that resource-poor firms can realistically draw on (Morley, Elhalal, et al., 2021). Whether the low salience of governance is uniform across firm types cannot be settled here: governance-any did not differ significantly by size (Large 33.3% vs SME 14.3%; Fisher $p = .327$, ns) or by ICT status (OR 0.87, $p = 1.0$), but the size comparison is underpowered (Large $n = 9$), so no subgroup difference was detected rather than uniformity established. The recommendation is offered broadly across the firms in the sample for that reason, not because a systemic gap has been demonstrated.

A particular risk attends the generative tools at the centre of this informal usage. The NIST Generative AI Profile (National Institute of Standards and Technology, 2024) enumerates confabulation, data leakage, and information-integrity harms as distinctive to these systems, and respondents independently named the same hazards: "The biggest problem is so-called hallucination" [F1.03, W1, ICT] and "We are careful that our data is not used for model training" [F1.06]. Documenting a verification standard and a data-handling rule is, on this evidence, a direct response to risks the firms themselves identify.

Measurement by design

The measurement paradox of Chapter 7 is the flagship managerial problem, and unlike governance it was directly probed. Benefits were articulated in 65.5% of interviews (95% Wilson score CI 52.7–76.4) but measured in only 20.7% (CI 12.3–32.8), and the conditional probability of measuring given a benefit claim was just 4 of 38, or 10.5%. The asymmetry is strong and statistically robust (McNemar exact $p = 6.9 \times 10^{-5}$; surviving BH-FDR at $q = .0007$), and the dissociation between claiming and measuring spans 34 of 58 firms (58.6%, CI 45.8–70.4). The interpretive literature reads this not as carelessness but as the early stage of a productivity J-curve in which intangible benefits are genuinely "disproportionately difficult to measure" (Brynjolfsson & Hitt, 2000; Brynjolfsson et al., 2017, 2021). Firms sense value before they hold the complementary capabilities to quantify it, and payoff detection is endogenous to the act of measurement (Kohli & Devaraj, 2003).

The practical remedy is measurement by design: defining a small number of outcome metrics before adoption rather than attempting reconstruction afterward, when, as one respondent put it, "I cannot quantify it because we did not track it" [F1.12, W1, finance, SME]. Two or three task-level indicators suffice. Time saved on a recurring task, error or rework rates, throughput per worker — any of these can turn a "gut feeling" [F1.03] into evidence, and this benefits-realization discipline is well established outside the AI setting (Ward & Daniel, 2012). External benchmarks indicate the stakes: controlled studies report 15% support-productivity gains (Brynjolfsson, Li, & Raymond, 2025), 40% time reductions on writing tasks (Noy & Zhang, 2023), and large but jagged effects that can turn negative on ill-suited tasks (Dell'Acqua et al., 2026). That jaggedness is itself an argument for measurement. Where returns are uneven and occasionally negative, unmeasured adoption can quietly mask quality losses, so instrumentation works as a safeguard. There is also a plausible payoff to building the habit, since in this sample measurement was associated with implementation formality (formality mean 2.83 among measurers vs 1.91 among non-measurers; Cliff's $\delta = 0.41$, $p = .026$; FDR $q = .053$, borderline) and not with firm size ($V = .03$) or ICT status ($V = .14$, ns). The pattern is consistent with small, non-technical firms becoming measurers by formalizing a few practices rather than by growing or hiring specialists.

Configuration-aware playbooks

Adoption in this sample looks configuration-dependent rather than uniform, so generic advice will tend to misfire. Chapter 8 identified four indicative groupings (offered with humility, bootstrap ARI = 0.35), and each suggests a different next step. The C0 multi-modal technical adopters (n = 16; breadth 3.56; measurement 31%; governance 19%) are sophisticated builders whose binding constraint looks like oversight rather than tooling or breadth; the pattern points to codified governance and lifecycle documentation commensurate with the autonomy of the agentic systems they deploy (Mitchell et al., 2025). The C1 vision/agent specialists (n = 10) are the starkest case, technically the broadest grouping (breadth 3.80) yet measuring at 0%, for whom a minimal measurement scaffold would be the highest-return move. The C2 generative mainstream (n = 23, the largest and most ordinary-economy group; governance highest at 26% but measurement only 17%) would benefit most from lightweight templates that fit non-technical firms across industry and finance. The C3 generative-only minimalists (n = 9; breadth 1.00; formality lowest at 1.33) are light single-tool users, often in public sector and "other" settings, for whom the appropriate next step is modest: a single usage policy and one outcome metric, not an enterprise framework that would exceed both their needs and their slack. Reading these groupings as gestalts (Miller, 1986; Fiss, 2011; Park et al., 2017) rather than as rungs on a maturity ladder guards against prescribing large-firm scaffolding to firms that differ in kind, not merely in degree (Mittal et al., 2018). The cell counts behind each grouping are small, so these are descriptive starting points, not validated segments.

Converting shadow use into governed capability

These three strands converge on a single dynamic-capabilities prescription. The default adoption mode in the sample is informal and bottom-up: "We started using the technology individually, each person in their own way" [F1.01, W1, retail]; "Some we may be using, but it is not formally defined" [F2.01, W2, ICT, SME]. That is the shadow-AI pattern recent work reframes not as a threat to be suppressed but as latent innovation to be surfaced and governed (Silić et al., 2025; Waters-Lynch et al., 2025). In Teece's (2007) terms, firms in this sample appear to have accomplished sensing and seizing while stalling at reconfiguring. They have detected AI's promise and put tools to work, but the routines that turn improvised use into a durable capability — registries, policies, metrics, verification standards — are largely not yet evidenced. The managerial agenda here is a reconfiguring agenda: channel the energy of shadow adoption into governed capability without extinguishing the experimentation that produced the value.

11.2 Implications for policymakers (European and Slovak)

The policy implications follow from the same asymmetry, but at field level. If firms adopt AI faster than they codify governance or measure benefits, the policy task is not simply to demand compliance after the fact. It is to make the first steps of governance and evaluation cheap, understandable, and reachable, especially for SMEs and for firms outside the ICT core. The evidence points toward an enabling posture: proportionate guidance under the EU AI Act, reusable templates, shared advisory capacity, proactive outreach to hard-to-reach firms, and simple measurement toolkits attached to public support. These recommendations are framed as capacity-building measures, not as claims that the observed sample represents the Slovak economy as a whole.

Proportionality and support under the EU AI Act

The headline message for policymakers is one of posture. Codified AI governance rarely surfaced in these firms' accounts, and the values of responsible AI appear diffusely while seldom being instantiated as written policy. That is consistent with a readiness gap rather than a collective lapse of diligence, which argues for an enabling rather than a punitive stance. Governance did not vary significantly by firm size or sector in this sample (Fisher tests ns), though those comparisons are underpowered and cannot establish that the condition is uniform. The EU AI Act (Regulation (EU) 2024/1689) already anticipates an enabling posture through SME-oriented measures such as regulatory sandboxes and proportionate documentation, and the scholarly critique of the Act's design warns that its self-assessment-heavy, standards-dependent obligations bear especially hard on smaller actors (Veale & Zuiderveen Borgesius, 2021). The finding that regulation surfaced in only 6.9% of interviews is better read as low awareness to be closed with support than as non-compliance to be sanctioned. Concretely, the pattern in this sample favours ready-made templates (model acceptable-use policies, tool registries, data-handling checklists), accessible sandbox provision, and plain-language guidance bridging the Act and the overlapping GDPR regime these firms already nominally operate under (Schwartmann et al., 2024). The aim is to lower the codification threshold to a level SMEs can actually meet, so that proportionality is realized in practice and not only promised in the text.

Recovery-and-Resilience Capacity Building

Slovakia's policy instruments are well placed to deliver that support. The AI-ImpactSK project itself is funded through the EU's NextGenerationEU and the Recovery and Resilience Plan, and the same architecture can underwrite the capacity-building the evidence points to. The dominant SME-AI barriers in the wider literature are organizational and managerial — skills, strategy, governance — rather than technological availability (Oldemeyer et al., 2024; Ghobakhloo et al., 2022), a pattern echoed in neighbouring V4 economies where leadership engagement, not technology, is the binding factor (Jaciow et al., 2026). Recovery-and-Resilience funding directed at shared governance services, template libraries, and practitioner training would address that constraint directly. The "ethics-as-a-service" model fits here (Morley, Elhalal, et al., 2021). Few SMEs of this size can sustain in-house governance expertise, so publicly supported, externalizable governance capacity — shared sandboxes, sectoral advisory services, reusable compliance modules — is a more realistic route to codified practice than expecting each firm to build it alone.

Reaching the hard-to-reach long tail

The conduct of the fieldwork carries its own lesson. The response rate on dialled numbers fell sharply between waves, from 22.2% to 6.4%, even as the dialling frame expanded from 135 to 435 numbers, and the achieved sample skewed heavily toward ICT and software firms (26 of 58, 44.8%). These are signals about who can be reached. The collapse is consistent with acute survey fatigue and with a hard-to-reach long tail — smaller, less digital, outside the ICT core — that is least likely to answer a research call and, by extension, least likely to be reached by voluntary governance guidance or to self-identify for support. The ICT skew implies that even an engaged sample over-represents early adopters, so the governance and measurement patterns documented here may be conservative relative to the broader economy. The practical implication is that outreach cannot rely on firms coming forward. Reaching the long tail will require proactive, sector-specific channels — chambers of commerce, sectoral associations, accountants and IT vendors as intermediaries — rather than open calls the most digitally marginal firms will never hear.

Measurement-capacity programs

The measurement paradox has a policy dimension that mirrors its managerial one. If firms cannot evidence the value of AI, neither can the public programs that subsidize its adoption, and the case for continued investment weakens for want of data. The dissociation between asserting and measuring benefits is a field-level problem long recognized in information-systems research (Schryen, 2013), not a peculiarly Slovak failing, and controlled evidence shows the gains are real and measurable with modest apparatus (Noy & Zhang, 2023; Peng et al., 2023). Public measurement-capacity programs would help on both sides: simple, standardized outcome-metric toolkits distributed with adoption funding; light-touch reporting that builds evaluation into subsidized projects; benchmarking services that let firms compare against credible external estimates. Such programs would let firms evidence AI value while generating, in aggregate, the data the policy system needs to steer. They would also seed the complementary measurement capability whose absence the J-curve literature predicts at this stage of diffusion (Brynjolfsson et al., 2021).

11.3 Implications for method and researchers

The methodological implications follow from the study's own research design and from the artifact it uncovered. AI-assisted coding made it possible to analyze a two-wave qualitative corpus systematically, but the study also shows why such assistance must remain human-reviewed and independently triangulated. The central lesson for researchers is therefore double. First, AI-assisted framework coding can be useful when it is anchored in an explicit codebook, reviewed by researchers, and checked against coding-independent text signals. Second, any cross-wave or cross-site qualitative comparison must report coding depth and respondent volume, because apparent thematic change may otherwise reflect the changing resolution of the analytic instrument rather than change in the field.

AI-Assisted Coding with Computational Triangulation

The study's design is itself a contribution, and one other researchers can adopt. The analysis combined AI-assisted framework thematic analysis (Gale et al., 2013; Braun & Clarke, 2006) — transcription, translation, and deductive coding against an explicit eight-category, twenty-four-theme codebook, with the model guided by the codebook and every output reviewed by the authors — with computational text analysis used as an independent check on the manual codes. This human-in-the-loop division of labour is consistent with the emerging consensus that large language models can assist, but not replace, codebook-based qualitative coding (Xiao et al., 2023; Chew et al., 2023), and that their outputs require expert validation and show real limits in interpretive depth (De Paoli, 2023; Mathis et al., 2024; Prescott et al., 2024). The distinctive move is triangulation: each substantive coded construct was tested against a coding-independent lexical signal drawn from the respondents' own words. The measurement code aligned with a KPI/measurement lexicon (mean hits 2.58 when coded vs 1.35 when not; point-biserial $r = 0.27$, $p = .041$), and the manual "AI agents" code validated strongly against agentic text ($r = 0.55$, $p < .0001$), whereas the formal governance code did not align with governance talk ($r = -0.20$, ns). That divergence was itself informative, consistent with governance as diffuse vigilance rather than codified practice. Reported transparently and cross-checked against the original-language data wherever possible (Squires, 2009; van Nes et al., 2010), this template offers a reproducible pattern for qualitative research at a scale that manual coding alone would render costly.

A coding-depth reporting standard

The strongest methodological recommendation arises from this study's own artifact, and it is offered as a candid contribution rather than a confession. The two waves were coded at markedly different densities. Wave 2 averaged 13.48 segments per interview against Wave 1's 8.21 (Mann–Whitney $p < 1 \times 10^{-7}$), Wave 2 respondents simply spoke more (1,123 vs 696 words per interview; $p = .0007$), and intensity coding was wave-confounded to the point of being unusable for cross-wave claims (the "high"-intensity share was 0.43 in Wave 1 against 0.91 in Wave 2). Because the formality index correlates with segments per interview ($r = 0.56$), naive cross-wave coverage increases partly reflect how densely each wave was coded rather than real change in the firms. The authors caught this only because they computed coding-independent signals. Without that check, the artifact could have passed for substantive temporal change.

That experience motivates a concrete proposal: a coding-depth reporting standard for AI-assisted and team-based qualitative research. At minimum, studies that compare across waves, sites, or coders should report segments per interview (with dispersion), speaker-volume distributions, and intensity distributions, broken down by wave and by coder, so readers can see whether apparent differences in coverage track differences in coding effort. More strongly, any cross-wave or cross-site substantive claim should be triangulated against a coding-independent signal, as was done here for the agentic turn, where the rise survived the test (respondent agentic mentions rising from 0.72 to 3.14; Mann–Whitney $p = .0022$; mentioned by 37.9% → 69.0% of interviews) and could be reported as genuine maturation rather than artifact. The robustness anchor illustrates the payoff. Denser Wave 2 coding did not inflate the headline patterns — measurement and governance were, if anything, slightly lower in Wave 2 (each 17.2% vs 24.1% in Wave 1) — which is why those patterns can be reported with reasonable confidence while raw coverage trends are flagged as hypothesis-generating. As AI-assisted coding makes deeper, faster coding routine, coding depth becomes a variable in its own right, and the field will need shared conventions for disclosing it. A standard of this kind would let the efficiency gains of computational coding be realized without quietly importing depth artifacts into substantive conclusions.

11.4 Synthesis

Read together, the three strands describe one program. Managers can keep the speed of bottom-up adoption while supplying its missing scaffolding — registries, usage policies, verification standards, and a handful of outcome metrics defined before adoption — calibrated to the grouping they actually occupy. Policymakers can make that scaffolding cheap and reachable through proportionate EU AI Act implementation, Recovery-and-Resilience-funded capacity building, proactive outreach to the hard-to-reach long tail, and measurement-capacity programs that let firms and the public purse alike see what AI is worth. Researchers can adopt the study's AI-assisted, triangulated design while observing a coding-depth reporting standard that this study's own experience shows to be necessary. Every recommendation is bounded by the limits restated throughout: a purposive, ICT-skewed sample of fifty-eight firms across two closely spaced waves (Štetka et al., 2026), with minor independence caveats and unknown sizes for a third of the sample. These are transferable, hedged suggestions, not a representative mandate. Within those bounds the task ahead is less to accelerate adoption, which is already underway, than to help firms govern and measure it — to close, deliberately and proportionately, the distance the pattern in this sample reveals between adopting and knowing.

12 Limitations and Future Research

Every percentage in this book describes 58 firms, not the Slovak economy. That single fact governs how far any of the preceding claims can travel, and it is the first of several boundaries worth stating plainly rather than burying. The findings cohere into an important picture: an adoption regime in which Slovak enterprises reach for artificial intelligence enthusiastically while governing and measuring it thinly, and in which the discourse of practice is already turning toward agentic systems. But the confidence a reader can place in any one of these claims is bounded by how the evidence was produced. This chapter does two things. It enumerates the study's limitations one by one, naming for each the specific threat to validity and the mitigation, if any, that the design allowed. It then converts those limitations, together with the literature gaps catalogued earlier (see the research-gap synthesis underlying Chapters 1 and 3), into an agenda for future research. The governing principle is the one that has organized the whole monograph: claims are hedged to the strength of their warrant, effect sizes accompany every association, and the purposive, non-probability character of the sample (N = 58) is treated as a standing caveat.

12.1 Limitations

The limitations fall into four broad groups. The first concerns sampling and external validity: the study is purposive, ICT-skewed, partly affected by unknown firm size, and shaped by a steep response-rate decline between waves. The second concerns measurement and elicitation: governance was not directly probed, measurement was unevenly administered, and all coded rates should be read as lower bounds rather than censuses. The third concerns method: AI assistance, coding-depth differences, and intensity artifacts condition how the evidence can be interpreted. The fourth concerns design: the study is single-country, cross-sectional within each wave, and reliant on one informant per organization. Taken together, these limits do not invalidate the findings, but they fix their epistemic status: the study offers a disciplined description of one corpus and a set of hypotheses for broader and longitudinal testing.

Sampling strategy and non-representativeness

The most consequential limitation is the sampling design. The AI-ImpactSK corpus comprises 58 semi-structured computer-assisted telephone interviews recruited through a purposive and, in practice, convenience procedure across two waves (Štetka et al., 2026). Participation required willingness to discuss organizational AI use by telephone, which means the sample was never drawn to mirror the population of Slovak enterprises and cannot support prevalence or point-estimate claims about that population. The threat is straightforward. External validity is limited, and every percentage reported in this book — whether the 65.5% of firms that articulate benefits or the 13.8% that evidence formal governance — describes this sample rather than the Slovak economy. The mitigation is partly procedural and partly rhetorical. Procedurally, the two waves and the breadth of sectors and roles were intended to capture variety rather than a representative cross-section, an aim well suited to the interpretive, theory-building purpose of the study. The sample comfortably exceeds conventional thematic saturation thresholds, beyond which the core range of codes is typically recovered (Guest et al., 2006), and the two-wave design was used to move past code saturation toward meaning saturation (Hennink et al., 2017). Rhetorically, we have consistently framed quantitative results as descriptive of the corpus and as hypothesis-generating for the population, never as estimates of national incidence. The

reader should nonetheless treat the prevalence-like figures as characterizations of a deliberately assembled qualitative sample.

Sectoral skew toward ICT and over-representation of directly involved respondents

Within that sample two compositional biases pull in the same, adoption-positive, direction. First, information and communication technology and software firms account for 26 of 58 cases (44.8%), a pronounced over-representation of an early-adopter sector relative to its weight in the economy. Because ICT firms are precisely those most likely to build, extend, and reflect articulately on AI, their prominence inflates the apparent sophistication and breadth of adoption and risks making the corpus read as more AI-mature than the broader enterprise base. Second, the respondents were overwhelmingly people close to the technology: 43 of 58 had directly implemented the AI in question, with only a handful of pure users, project managers, or merely informed observers. Self-selection of the directly involved compounds the sectoral skew, since those who implement AI are disposed to perceive and recount its benefits. The combined threat is a selection bias toward adoption-positive accounts, which would tend to overstate benefit articulation and understate indifference or failure. Two considerations mitigate, without removing, this concern. The ICT skew was confronted analytically rather than ignored. ICT versus non-ICT contrasts were tested directly, and the headline governance and measurement gaps proved broad rather than ICT-specific (for governance, the size contrast yielded a non-significant Fisher test; measurement was not associated with the ICT sector, Cramér's $V = 0.14$, ns). The most important deficits are not artifacts of who happened to answer the telephone. First-hand involvement, while biasing affect, also improves the factual fidelity of process accounts, since implementers can describe what was actually done rather than what was nominally planned.

Unknown company sizes and the size–wave confound

Firm size is incompletely recorded: 21 of 58 organizations could not be assigned a size band, leaving 28 small and medium-sized enterprises, 9 large firms, and a substantial unknown stratum. This missingness threatens any analysis that conditions on size, both by shrinking cell counts to the point where subgroup tests have little power and by raising the possibility that the unknowns differ systematically from the classified cases. More seriously, size is confounded with wave — the recruitment of Wave 2 drew disproportionately on smaller firms, so that size and wave cannot be disentangled in this design. The mitigation was disciplinary restraint. Where size and wave were entangled, most importantly in the relationship between firm size and implementation formality, we declined to interpret the apparent effect and said so explicitly, rather than reporting a size effect that might in fact be a wave (and hence coding-depth) effect. Size contrasts that we did report, such as the governance comparison between large firms and SMEs (odds ratio 2.93, Fisher $p = .327$, ns), were presented as descriptive and underpowered. This is a real limitation on the study's ability to adjudicate the classic firm-size argument familiar from the CEE technology-adoption literature (Horváth & Szabó, 2019). It is not, we judge, a threat to the conclusions we actually draw, because we drew none that rest on the size variable.

Response-rate collapse and non-response bias

The fieldwork exhibited a steep decline in cooperation between waves. On numbers dialed, the response rate fell from 22.2% in Wave 1 to 6.4% in Wave 2, the latter obtained against a much larger dialing frame (435 versus 135 numbers). A response rate of this magnitude raises the spectre of non-response bias in acute form. If the small fraction who agreed to participate differs systematically from the great majority who declined, the resulting picture may be skewed in unknown directions, and self-selection toward the AI-interested is the most plausible such

difference. The threat is to the representativeness already discussed and to the comparability of the two waves, since the determinants of cooperation may themselves have shifted over the eight months separating the fieldwork periods. Mitigation here is limited and consists largely of transparency. The dialing, connection, and completion figures are reported in full so that readers can weigh the selection context for themselves, and the cross-wave analyses were designed to lean on coding-independent textual signals rather than on naive comparisons that non-response could contaminate. We also note the minor independence caveats in the corpus, one participant appearing in both waves and one interview covering two companies, which, while affecting only 2 of 58 records, are disclosed for completeness. We cannot rule out self-selection toward AI-positive firms, and the response-rate collapse makes this caution especially pertinent for Wave 2.

AI-assisted transcription, translation, and first-pass coding

The analytical pipeline used artificial intelligence at three junctures, transcription, translation, and a first pass of deductive coding, each of which introduces its own validity threats. The translation interface is the first concern. Interviews were conducted and analyzed in Slovak, which remains the source of record, with English renderings produced by AI assistance and reviewed by the authors. Cross-language qualitative research nonetheless introduces systematic threats at precisely this interface, where conceptual equivalence can quietly erode (Squires, 2009), and subtle meaning is most at risk when analysis migrates away from the original language (van Nes et al., 2010). We mitigated this by treating the Slovak as authoritative, conducting interpretation as far as possible against the Slovak text, and reserving English for presentation, in line with the principle of delaying and limiting translation. The coding interface is the second concern. First-pass coding against the codebook was AI-assisted, a workflow with growing but still cautionary precedent: large language models guided by an expert codebook can achieve fair-to-substantial agreement with human coders yet cannot replace them (Xiao et al., 2023; Chew et al., 2023), and open-source models can produce broadly comparable thematic output while still requiring human interpretation (Mathis et al., 2024). The specific risks are familiar from this literature. They include hallucinated or spurious code assignments, and a tendency toward smoothing whereby the model normalizes idiosyncratic or ambiguous passages into tidy, expected categories, flattening exactly the nuance that interpretive analysis prizes (De Paoli, 2023; Prescott et al., 2024). Two mitigations were applied. First, every AI-suggested code was subject to author review, consistent with a human-in-the-loop stance in which the analyst retains interpretive authority and the model serves efficiency and scale; this posture is congruent with reflexive thematic analysis, in which rigor derives from theoretically informed researcher judgment rather than from inter-coder agreement as such (Braun & Clarke, 2019). Second, and more distinctively, the substantive conclusions were not allowed to rest on the manual codes alone but were triangulated against coding-independent computational text analysis, lexical incidence, keyness, and the like, so that a claim survived only if it appeared in both the human-reviewed codes and in measures untouched by the coding process. The governance, measurement, and agentic-turn findings each passed this dual test. The governance lexicon, for instance, was analyzed entirely independently of the formal governance code. AI assistance shaped the corpus, and we do not pretend otherwise, but the load-bearing claims were constructed to be robust to the particular failure modes the method invites.

Coding-depth and intensity artifacts

A methodological finding internal to the study constitutes a further limitation on cross-wave comparison and is disclosed here as such. Wave 2 was coded considerably more densely than

Wave 1: segments per interview rose from a mean of 8.21 to 13.48 (Mann–Whitney, $p < 1e\text{-}7$), respondents in Wave 2 spoke roughly 1.6 times as many words (means of 696 versus 1123; $U = 203$, $p = .0007$, rank-biserial 0.52), and the formality index correlates with segments per interview at $r = 0.56$. The consequence is that naive cross-wave increases in thematic coverage partly reflect deeper coding rather than real organizational change, and the intensity coding is wave-confounded to the point of being unusable for substantive comparison (the "high"-intensity share is 0.43 in Wave 1 against 0.91 in Wave 2). The threat is to the internal validity of any temporal claim: an apparent rise might be an artifact of how the later wave was processed. The mitigations were strict and were applied consistently throughout the book. Intensity was never used for cross-wave substantive claims; cross-wave coverage shifts were reported only with the coding-depth caveat attached; and temporal conclusions were anchored to coding-independent signals (for example, the rise in agentic discourse, $U = 234$, $p = .0022$, rank-biserial 0.44, validated against text at point-biserial $r = 0.55$). One robustness anchor deserves repetition. The denser Wave 2 coding did not inflate the deficits that matter, since both measurement and any-governance coverage were marginally lower in Wave 2 than in Wave 1 (17.2% versus 24.1% in each case). The governance and measurement gaps are therefore not coding-depth artifacts. The limitation is real for temporal interpretation and was managed by refusing to read coverage changes as substantive without independent corroboration.

The governance gap is an elicitation limit, not a census

The most important correction to make about the governance finding concerns how it was elicited. The interview protocol contained no standalone governance or policy question. Governance and regulation entered the record only when a respondent volunteered them, under the broad legal-and-ethical-challenges prompt or the forward-looking question about whether AI regulation was expected. The headline figures, formal governance in 13.8% of firms and explicit regulatory awareness in 6.9%, therefore measure how often governance was spontaneously raised, not whether governance practices exist. The threat is a specific and serious one — spontaneous salience and actual practice are not the same quantity, and reading the former as the latter would overstate the case. A firm that maintains an informal review habit, or a security policy it did not think to mention in a thirty-odd-minute call about AI, would not register on a measure built from unprompted mentions. We have accordingly reframed the governance result throughout the book as a discourse-level finding. Governance rarely surfaced spontaneously; it was a low-salience theme in firms' own accounts of adoption. That is a defensible and, we think, meaningful claim. It is not the claim that firms "do not govern," and the difference bounds how the governance gap should be read. It indexes the place AI governance occupies in managerial attention, not a verified absence of practice. This asymmetry matters because measurement, the study's other central deficit, is on entirely different footing. Measurement was probed directly, through a goal-measurability question and an explicit prompt about how firms gauge success and which figures they track. The measurement paradox is well grounded for exactly that reason and is not softened the way the governance finding is.

Uneven administration of the measurement probe

The measurement probe, though direct, was not administered uniformly across the corpus. Most of the gaps in its administration fall in Wave 1, where the prompt about concrete figures and KPIs reached fewer respondents than in Wave 2. The consequence runs in a known direction. The 20.7% articulated measurement rate understates measurement somewhat and is best read as a near-lower bound rather than a point estimate. The threat is modest and bounded. It affects the marginal rate — the share of all firms coded as measuring — but it does not touch

the paired within-firm contrast at the heart of the paradox. The McNemar test compares, firm by firm, those that articulate a benefit without measuring it against those that measure without articulating, and that comparison is unaffected by uneven coverage of the marginal question. So the exact figure may drift upward under more uniform probing, but the dissociation itself, and the McNemar result that certifies it, does not depend on where 20.7% lands.

Coded rates are lower bounds, not censuses

A caveat that applies across the study, and with most force to governance and measurement, is that theme-presence prevalence are coded rates rather than censuses. They count the firms for which a practice surfaced clearly enough in a roughly thirty-three-minute interview to be coded. A practice that did not come up may still exist. This is the elementary but easily forgotten point that absence of evidence is not evidence of absence — an interview is a sample of what a respondent chose to say in a bounded conversation, not an inventory of everything the firm does. Every prevalence figure in the book is therefore a lower bound on actual practice. The threat is a tendency to read silence as proof of non-practice, which would inflate the apparent deficits. The two findings most exposed are precisely the two this chapter dwells on. Governance, which was never directly asked, is the most exposed of all, and measurement, with its uneven probe, is the next. We have framed both as articulated or coded rates throughout, and the reader should carry that framing into every number. The deficits are real as features of the discourse and of what firms foreground; they are floors, not ceilings, on what firms actually do.

The early-adoption-stage confound

An eight-month, two-wave, cross-sectional design cannot separate two very different explanations for the low governance and measurement signals. One is dispositional: these firms tend not to govern or measure their AI. The other is developmental — they have not done so yet, because adoption is early and these are activities that mature after deployment rather than alongside it. The data are consistent with both, and the design cannot adjudicate between them. The threat this poses is to interpretation rather than to the figures. A finding read as a stable trait of Slovak AI adoption may instead be a snapshot of an early moment that later thickens into governance and measurement as firms accumulate experience. The same forward-looking testimony that appears elsewhere in the book, the recurrent framing of measurement as a later-stage aspiration ("we will gradually get to measurement"), is exactly what the stage account predicts, and it should caution against reading the deficits as settled dispositions. We weigh this rival explanation explicitly in the discussion and flag it here as a standing limit. Whether the thin governance and thin measurement observed in 2025 and 2026 reflect how these firms are or merely where they are in an adoption trajectory cannot be determined from a cross-section, and only a design that follows firms over time can settle it.

Instability and heuristic status of the adopter configurations

The four adopter configurations developed in Chapter 8 are a heuristic descriptive partition, not a validated typology, and two diagnostics make that status concrete. The internal-validity criteria did not point to four groups at all: a two-cluster split scored best on the silhouette criterion (0.369), so the four-way solution is a choice made for descriptive richness rather than a structure the data demanded. And that solution separates only weakly. A bootstrap adjusted Rand index of 0.35 means the cluster assignments shift under resampling and the boundaries between groups are soft. The threat is over-reification: readers might treat the four configurations as a settled taxonomy when the data warrant a far more tentative reading. Any claim resting on a

small per-cluster cell, such as the contrast between C1's zero-of-ten and C3's three-of-nine on measurement, is descriptive rather than established. We mitigated this by labeling the configurations as indicative at every appearance, deriving them by k-modes clustering on coding-independent metadata and AI-profile variables so they are not circular with the outcomes they organize, and grounding them in configurational and typology-development methodology that itself counsels humility about boundary cases and equifinality (Fiss, 2011; Park et al., 2017; Nickerson et al., 2013). The configurations are offered as a structured hypothesis about how AI adoption differs in form across firms, to be tested and refined, not as a validated instrument.

Single-country context

The study is set entirely within Slovakia. This bounds transferability. Slovakia is a small, SME-heavy, manufacturing-and-ICT economy operating under a particular configuration of EU regulation, national digital maturity, and Recovery and Resilience funding, and the patterns observed here may not hold in larger or differently structured economies, even neighboring ones. The threat is to generalization beyond the national case, and it cannot be neutralized within a single-country design. What can be offered is, first, candor about the boundary, and second, the observation that the central findings resonate with adjacent evidence: the normalization of informal SME adoption documented in broader European samples (Schwaeke et al., 2024) and the partial, function-specific AI use reported in the Slovak SME segment elsewhere (Dvorský, 2025). These suggest the phenomena are not idiosyncratically Slovak even though their precise magnitudes are local. Transferability remains, in Lincoln and Guba's (1986) sense, a judgment for the reader equipped with sufficient contextual detail, which Chapters 2 and 4 supply.

Cross-sectional design within each wave

Although the study spans two waves, each wave is cross-sectional. Every firm was interviewed once within a wave (with the single noted exception), and the design captures perceptions and reported practices at a moment rather than tracking change within organizations over time. This precludes causal inference. We can describe associations, that measurement co-occurs with higher implementation formality (Cliff's $\delta = 0.41$, $p = .026$), or that ICT firms display greater formality ($\delta = 0.36$, $q = .036$), but we cannot establish that formality causes measurement, that measurement causes formality, or that some third factor drives both, and we have been careful never to phrase such associations causally. The threat is the standard one of mistaking covariation for causation, and the mitigation was simply to honor the design's limits in every interpretive sentence. Even the temporal comparison between waves is a comparison of two independent cross-sections rather than a within-firm trajectory, so the maturation narrative is an inference about the discourse of the population of adopters, not a demonstrated path traveled by particular firms. It is closely related to the early-stage confound discussed above, and the two limits compound: the design can neither establish causal direction nor distinguish disposition from timing.

Social-desirability and single-informant bias

Two response-level biases attend the interview method. Because the topic is one on which firms may wish to appear competent and responsible, social-desirability pressure could inflate accounts of careful practice, of checking outputs, protecting data, or governing use, relative to what occurs. This bias would tend to overstate governance and measurement, which makes the thinness we actually observe the more striking. The gaps are unlikely to be artifacts of

respondents understating their virtue. The second bias is reliance on a single informant per organization (again with the one exception). One person's account, however well placed, is a partial view of a firm's practice and cannot be cross-validated against other organizational members or against documentary evidence, and the risk is that idiosyncratic perception is mistaken for organizational reality. The directly involved character of most respondents mitigates the factual side of this concern, since implementers describe practices they personally enacted, but it does not remove the structural limitation that each firm speaks through one voice. Triangulation across informants was beyond the resources of this design and is flagged below as a priority for future work.

12.2 Future research

The limitations above are not merely confessions. Each points toward research that would advance the field. We organize the agenda around the gaps identified in the literature and the openings created by the present study's design constraints.

Establishing prevalence: a representative Slovak and CEE survey

The most direct response to the non-representativeness limitation is a probability-based survey of Slovak and, ideally, Central and Eastern European enterprises designed to estimate the prevalence of the patterns this study could only characterize qualitatively. Such a survey would convert the indicative rates reported here — benefit articulation, the measurement gap, the governance deficit — into population estimates with proper confidence intervals, and would allow the firm-size, sector, and capability correlates to be tested with adequate power and without the size–wave confound that hampered the present analysis. Validated multi-item constructs for AI capability of the kind already developed in the literature (Mikalef & Gupta, 2021) could anchor such an instrument. A representative design would also resolve whether the ICT skew and adoption-positive selection of the interview corpus materially inflated the picture, by situating the qualitative findings against a population baseline. Crucially, a survey could ask about governance directly rather than waiting for it to be volunteered, converting the discourse-level governance finding of this study into a measured estimate of practice and closing the elicitation gap that bounds the present reading. This is the natural complement to the present work: depth here, breadth there.

Tracking the agentic turn: a longitudinal panel

The cross-sectional-within-wave limitation, the coding-depth confound, and the early-adoption-stage confound all argue for a genuine longitudinal panel that follows the same firms over time. Where this study inferred maturation from two independent cross-sections, a panel could observe trajectories within organizations: whether the agentic discourse documented here (mentioned by 37.9% of interviews in Wave 1 and 69.0% in Wave 2) translates into sustained agentic deployment, and, critically, whether governance and measurement eventually catch up with adoption or whether the regime of thin oversight persists as autonomy increases. A panel is also the only design that can settle the disposition-versus-timing question this study had to leave open: it would show directly whether low governance and measurement are stable traits or early-stage features that thicken with experience. Given that the risks of agentic systems rise with the autonomy ceded to them (Mitchell et al., 2025), the question of whether oversight matures in step with capability is not academic. A panel would additionally retire the coding-depth artifact as a confound, provided coding protocols were held constant across waves, and would let the field distinguish real organizational learning from measurement-side change.

Cross-country comparison across CEE

The single-country boundary invites explicit cross-country comparison. A coordinated multi-country study across several CEE economies, pairing Slovakia with V4 neighbors and beyond, would establish which findings are regional and which are national, and would test whether the informal-adoption pattern and the governance and measurement gaps generalize across distinct institutional, linguistic, and market contexts within a shared EU regulatory frame. Recent V4 evidence on leadership-driven AI maturity in Polish SMEs (Jaciow et al., 2026) and risk-aware adoption in Slovakia and Poland (Snieška et al., 2020) suggests both common threads and national particularities worth disentangling. Comparison would also illuminate whether the EU AI Act's phased obligations (Regulation (EU) 2024/1689) diffuse evenly across member states or unevenly by national capacity, a question of direct policy relevance.

Governance and measurement intervention studies

Two of the book's central findings — the governance gap and the measurement paradox — pose interventional questions that observational designs cannot answer. Do lightweight governance instruments actually change outcomes? A field experiment or quasi-experiment that introduced, say, an AI tool registry, a minimal use policy, or KPI-by-design templates to some firms and not others could test whether such proportionate, SME-feasible scaffolds, of the kind that surface-and-govern approaches to shadow AI envisage (Waters-Lynch et al., 2025), and that frameworks such as the NIST AI Risk Management Framework's "Govern" and "Measure" functions formalize (Tabassi, 2023), improve safety, value capture, or both, relative to the diffuse informal vigilance this study observed. Because detection of value is itself endogenous to whether and how firms measure (Kohli & Devaraj, 2003), an intervention that simply instituted measurement might reveal benefits currently sensed but invisible, testing directly whether the measurement paradox is a capability gap rather than an absence of value. Such studies would move the field from documenting deficits to learning what relieves them.

Validated instruments for AI value measurement

The measurement paradox also exposes a tooling gap. Firms that wish to measure AI value frequently lack usable, validated instruments scaled to their resources. Methodological research that developed and validated lightweight AI-value measurement instruments, capable of capturing the intangible, process-mediated benefits that the IT-value literature has long identified as disproportionately hard to quantify (Brynjolfsson & Hitt, 2000; Melville et al., 2004; Schryen, 2013), would address a genuine practitioner need and a research one. The controlled-experiment literature demonstrates that GenAI gains are real and measurable with modest apparatus, from customer support to writing to software development (Brynjolfsson et al., 2025; Noy & Zhang, 2023; Peng et al., 2023), yet also that those gains are task-contingent and occasionally negative across the jagged frontier (Dell'Acqua et al., 2026). Instruments that let ordinary firms detect both the gains and the losses, rather than relying on impression, would make adoption decisions evidence-based and would supply the dependent variables that the intervention studies above require.

Deeper inter-coder reliability and human-versus-LLM coding studies

The AI-assisted coding limitation opens a methodological research front of its own. Although reflexive thematic analysis does not treat inter-coder agreement as the arbiter of rigor (Braun & Clarke, 2019, 2021), the rise of LLM-assisted coding makes systematic human-versus-model comparison newly valuable, not to police agreement for its own sake, but to map where models

reliably assist and where they hallucinate or smooth. Studies that ran independent human coders alongside LLM coders on the same interview corpora, characterized the divergences qualitatively, and probed the smoothing tendency directly would sharpen the human-in-the-loop protocols on which credible AI-assisted qualitative work depends, building on the feasibility-and-limits evidence already accumulating (De Paoli, 2023; Prescott et al., 2024; Mathis et al., 2024). The computational-triangulation strategy used in this study, admitting a claim only when human-reviewed codes and coding-independent text analysis concur, is itself a candidate template for replication and refinement, and the coding-depth reporting introduced here could become a routine transparency standard for two-wave qualitative designs.

Sectoral deep-dives beyond ICT

The ICT skew suggests that the configurations and gaps documented here should be examined within the sectors this corpus under-sampled. Deep-dives into industry and construction, finance and professional services, healthcare and pharma, retail, and the public sector would test whether the adoption-without-measurement pattern and the informal-governance posture take sector-specific forms, and whether the "generative mainstream" of broad-economy adopters that dominated the present sample fractures into distinct sectoral logics on closer inspection. Sector-bounded studies could also probe sourcing strategies, the build/buy/adapt continuum and the retrieval-augmented adaptation path that underlies it (Lewis et al., 2020; Klotz, 2026), where they are most consequential, for instance in regulated finance and healthcare settings whose risk profiles differ sharply from those of software firms. Such work would populate the indicative configurations of Chapter 8 with the sectoral detail that a single ICT-heavy corpus could only gesture toward.

Taken together, these directions trace a path from the present study's deliberate but limited qualitative depth toward the breadth, longitudinal traction, comparative reach, and interventional leverage that the questions raised here ultimately demand. The contribution of this monograph is to have characterized, with as much rigor and candor as a purposive two-wave design allows, an adoption regime that races ahead of its own governance and measurement. The contribution of the research it invites will be to measure that regime at population scale, to watch it move, and to discover what helps it govern and measure what it has so eagerly adopted.

Conclusion

By the spring of 2026, a fair number of the firms in this study were delegating real work to systems that plan, retrieve, and act on their own. Almost none could say what those systems were worth to them, and almost none could point to a written rule for using them. That combination is the subject this book has tried to pin down.

The aim was modest in scope and, we hope, exact in execution. Fifty-eight Slovak organizations were interviewed by telephone across two waves, the first in the late summer and autumn of 2025, the second the following spring, with the express purpose of catching enterprise AI mid-flight rather than after the fact (Štetka et al., 2026). The reading of those transcripts was qualitative in its primary method and quantitative in its checks: a framework analysis of eight categories and twenty-four themes across 629 coded segments, with every substantive claim held against a coding-independent reading of the same words. Slovakia earned its place here not as a stand-in for anywhere else but as a setting the literature has largely passed over, where a small open economy adopts a frontier technology under the same European rules as its larger neighbours and with far less of the surrounding apparatus.

12.3 Three findings, in brief

Three results carry the argument, and they are easiest to state apart before they are drawn together.

Governance was rarely something these firms brought up on their own. Where formal AI governance appeared at all, a policy, a named owner, a codified rule of use, it surfaced in 13.8% of organizations, and a respondent volunteered any awareness of regulation in only 6.9%; the whole governance-and-regulation category drew 1.9% of the coded material, the slimmest share of any category by a wide margin (Štetka et al., 2026). A caveat governs how those figures should be read, and it matters. The interview carried no standalone governance question. Governance and regulation entered the record only when someone raised them unbidden, so the numbers track how often the topic was foregrounded, not whether a control exists somewhere in the firm. What did recur, when the subject came up, was a habit rather than a structure: people checked the output, and they were careful about what data they fed a model. Chapter 6 reads this through the decoupling tradition, in which adopted forms and everyday practice drift apart (Meyer & Rowan, 1977; Bromley & Powell, 2012), and it is the principles-to-practice gap brought down to the level of what firms actually describe themselves doing (Mittelstadt, 2019).

Measurement is a sharper finding, and a firmer one, because it was asked about directly. Respondents were pressed on whether their goals were measurable and on how, concretely, they judged success and which figures they tracked. Two-thirds, 65.5%, named a concrete benefit from AI. One in five, 20.7%, could attach any measurement to it. Among the firms claiming a benefit, the chance of having measured it was 10.5%, and the split between asserting value and showing it held up under an exact McNemar test ($p < .001$) after correction for multiple comparisons (Štetka et al., 2026). This is the information-technology productivity paradox in a new instrument (Brynjolfsson & Hitt, 2000; Kohli & Devaraj, 2003), and it is the benefits-realization problem that the IT-value literature has long described, in which value asserted is not value managed (Ward & Daniel, 2012). The articulated rate is a floor, not a census: a metric never raised in a thirty-odd-minute call may still sit in a spreadsheet somewhere. Even read as a floor, the distance between believing in the technology and counting it is hard to miss.

The third result is that adoption is not one thing. Informality is the baseline, much of it individual, voluntary, and bottom-up, run without a defined project or a set of indicators. Against that baseline, implementation was somewhat more formal among ICT firms than among the rest, a tendency that was directionally consistent but suggestive rather than settled once the cleanest within-wave test is applied. The firms also sorted, indicatively, into four adopter groupings, from multi-modal builders through agent-minded specialists and the generative mainstream to generative-only minimalists, that differ in how broadly they use AI and how formally they run it. These groupings are descriptive partitions with weak separation, offered to show that the governance and measurement deficits do not fall evenly, not to certify a typology.

12.4 What the method contributes

The way these findings were produced is itself part of the book's case. Transcription, translation, and coding were assisted by large language models and reviewed by the authors, and every claim was cross-checked against computational text analysis of the respondents' own language, so a manual code and an independent lexical signal had to agree before a finding was trusted (Štetka et al., 2026). That is a workable pattern for qualitative research at a pace manual coding strains to match, and a small piece of evidence that machine assistance and rigour can coexist when a person stays in the loop and the machine's reading is checked against another.

What keeps the workflow honest is a disclosure most studies could make and few do. The second wave was coded about 64% more densely than the first, and those respondents also simply said more, so a naive comparison across waves measures the analysis as much as the world. The book reports this plainly and treats no cross-wave movement as real without independent corroboration. One temporal claim clears that bar. Talk of agents, autonomy, and retrieval-augmented generation moved from a minority to a majority of interviews, rising from 37.9% to 69.0% on triangulated evidence that survives the density caveat. Tellingly, the same density that lifted other measures left governance and measurement flat, which is the strongest sign that those two deficits belong to the firms and not to the coding.

12.5 The pattern, and its stakes

Set side by side, the findings describe one posture. We characterize it as adoption that has run ahead of both governing and measuring: firms take up a general-purpose technology with conviction, while the means to say what it is worth and the means to say how it may be used both lag well behind. The reading is not that anyone here is careless. It is more plausibly legitimacy-driven and capability-thin. Firms reach for AI in part because their environment now expects it, and they reach for it informally because the in-house expertise, the slack, and the absorptive capacity that formal measurement and governance demand are exactly what resource-constrained firms off the frontier tend to lack. For a tool whose output a person still reads before acting, the habit of checking is a reasonable response, and for an ordinary chatbot it may be enough.

Two things complicate that consolation. The first is endogeneity. A firm that does not measure cannot tell the uses that pay from the ones that merely please, and because detecting a benefit depends on looking for it, the firm that does not look will tend not to see what it is getting, for good or ill (Kohli & Devaraj, 2003). The second is autonomy. The agentic shift these interviews caught in motion is precisely the development that makes a habit of glancing at the output least sufficient, since the value of an agent lies in the steps a person no longer supervises (Mitchell et al., 2025). Both run against the firms in this sample, not for them.

There is also a policy edge, and it is uncomfortable. These are EU firms, the study itself is EU-funded, and the European Union's Artificial Intelligence Act was phasing in across the very months of the fieldwork (European Parliament & Council, 2024). That a respondent volunteered any awareness of regulation in fewer than one firm in fourteen says less about the law's text than about its reach, and the gap is one of salience in firms' own accounts as much as a documented gap in practice. Whether this low profile is a matter of disposition or of timing, of firms uninclined to govern and measure, or simply of firms early enough that they have not yet, cannot be resolved by an eight-month, two-wave, cross-sectional design. The forward-looking remarks point both ways: a respondent who says the firm will get to measurement in time fits the early-stage account at least as well as the dispositional one.

Proportion is due. The evidence comes from fifty-eight purposively recruited firms in one country, tilted toward ICT and toward early adopters, and the adopter groupings in particular are indicative. The book has argued for patterns, not for population estimates, and these conclusions inherit that limit. What the firms in this sample show is that the distance between adopting and accounting for AI is wide, that the measurement finding in particular is robust and broad, and that the agentic turn is making the distance costlier. Whether the same holds across Central and Eastern Europe and the wider field of small firms is a question this work sharpens rather than answers, and the chapters around this one set out how it might be tested.

12.6 From use to capability

What would it take to narrow the distance? Nothing in these findings makes the narrowing happen by itself. The productivity-paradox literature offers an encouraging story, that early adopters create value not yet visible in any metric they keep, and that measurement and governance will follow as complementary capabilities accumulate (Brynjolfsson & Hitt, 2000). The way these firms' language matured over seven months, from novelty toward solutions, verification, and production, is faintly consistent with it. The flat governance and measurement series are the warning against leaning on it. Across the same seven months, the talk advanced and the oversight did not move. Speaking more like an operator of autonomous systems is not the same as becoming a governor of them.

So the work is construction, not patience, and construction sized to the firms that have to do it. On measurement, that means lightweight indicators attached at the moment of adoption, so that a benefit a firm is already willing to assert becomes one it is equipped to show. On governance, it means something codified but bearable, a short policy, an owner, a rule of use a small firm can actually keep, with shared, outsourced, or platform-provided scaffolding for those that cannot stand up oversight alone. And it means meeting adopters where they sit, asking different things of a multi-modal builder than of a generative-only minimalist. None of this is exotic, and all of it is within reach of firms that decide the distance is worth closing. The firms in this study entered 2026 sounding more like users of autonomous systems and no more like stewards of them. Closing that distance, before the technology's autonomy and the law now rising to meet it make the gap expensive, is the task this book has tried to make both visible and urgent.

References

Bagozzi, R. P. (2007). The legacy of the technology acceptance model and a proposal for a paradigm shift. Journal of the Association for Information Systems, 8(4), 244–254. https://doi.org/10.17705/1jais.00122

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. Electronic Markets, 33(1), Article 63. https://doi.org/10.1007/s12525-023-00680-1

Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. Journal of Organizational Behavior, 45(2), 159–182. https://doi.org/10.1002/job.2735

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health, 11(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qualitative Research in Psychology, 18(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Bromley, P., & Powell, W. W. (2012). From smoke and mirrors to walking the talk: Decoupling in the contemporary world. Academy of Management Annals, 6(1), 483–530. https://doi.org/10.5465/19416520.2012.684462

Brynjolfsson, E. (1993). The productivity paradox of information technology. Communications of the ACM, 36(12), 66–77. https://doi.org/10.1145/163298.163309

Brynjolfsson, E., & Hitt, L. (1996). Paradox lost? Firm-level evidence on the returns to information systems spending. Management Science, 42(4), 541–558. https://doi.org/10.1287/mnsc.42.4.541

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. (2023). Generative AI at work (NBER Working Paper No. 31161). National Bureau of Economic Research. https://doi.org/10.3386/w31161

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. The Quarterly Journal of Economics, 140(2), 889–942. https://doi.org/10.1093/qje/qjae044

Brynjolfsson, E., Rock, D., & Syverson, C. (2017). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics (NBER Working Paper No. 24001). National Bureau of Economic Research. https://doi.org/10.3386/w24001

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

Chatterjee, S., Rana, N. P., Dwivedi, Y. K., & Baabdullah, A. M. (2021). Understanding AI adoption in manufacturing and production firms using an integrated TAM-TOE model. Technological Forecasting and Social Change, 170, 120880. https://doi.org/10.1016/j.techfore.2021.120880

Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). LLM-assisted content analysis: Using large language models to support deductive coding (arXiv:2306.14924). arXiv. https://doi.org/10.48550/arXiv.2306.14924

Cihon, P., Schuett, J., & Baum, S. D. (2021). Corporate governance of artificial intelligence in the public interest. Information, 12(7), 275. https://doi.org/10.3390/info12070275

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

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

De Paoli, S. (2023). Performing an inductive thematic analysis of semi-structured interviews with a large language model: An exploration and provocation on the limits of the approach. Social Science Computer Review, 42(4), 997–1019. https://doi.org/10.1177/08944393231220483

Dell'Acqua, F., McFowland, E., III, Mollick, E., Lifshitz-Assaf, H., Kellogg, K., 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

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review, 48(2), 147–160. https://doi.org/10.2307/2095101

Dvorský, J. (2025). Impact of artificial intelligence on enterprise risk management: A case study from the Slovak SME segment. Journal of Business Sectors, 3(1), 96–103. https://doi.org/10.62222/caja0666

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy artificial intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion, 99, 101896. https://doi.org/10.1016/j.inffus.2023.101896

El Sawy, O. A., Malhotra, A., Park, Y., & Pavlou, P. A. (2010). Research commentary—Seeking the configurations of digital ecodynamics: It takes three to tango. Information Systems Research, 21(4), 835–848. https://doi.org/10.1287/isre.1100.0326

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models [Preprint]. arXiv. https://arxiv.org/abs/2303.10130

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. Information Systems Frontiers, 24(5), 1709–1734. https://doi.org/10.1007/s10796-021-10186-w

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689, 12.7.2024. http://data.europa.eu/eli/reg/2024/1689/oj

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

Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. Academy of Management Journal, 54(2), 393–420. https://doi.org/10.5465/amj.2011.60263120

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People — An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. International Journal of Production Economics, 210, 15–26. https://doi.org/10.1016/j.ijpe.2019.01.004

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, 117. https://doi.org/10.1186/1471-2288-13-117

Ghobakhloo, M., Iranmanesh, M., Vilkas, M., Grybauskas, A., & Amran, A. (2022). Drivers and barriers of Industry 4.0 technology adoption among manufacturing SMEs: A systematic review and transformation roadmap. Journal of Manufacturing Technology Management, 33(6), 1029–1058. https://doi.org/10.1108/jmtm-12-2021-0505

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. Field Methods, 18(1), 59–82. https://doi.org/10.1177/1525822X05279903

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough? Qualitative Health Research, 27(4), 591–608. https://doi.org/10.1177/1049732316665344

Hickman, E., & Petrin, M. (2021). Trustworthy AI and corporate governance: The EU's Ethics Guidelines for Trustworthy Artificial Intelligence from a company law perspective. European Business Organization Law Review, 22(4), 593–625. https://doi.org/10.1007/s40804-021-00224-0

Horváth, D., & Szabó, R. Z. (2019). Driving forces and barriers of Industry 4.0: Do multinational and small and medium-sized companies have equal opportunities? Technological Forecasting and Social Change, 146, 119–132. https://doi.org/10.1016/j.techfore.2019.05.021

Jaciow, M., Adamczyk, A., Bartuś, K., Bratnicka-Myśliwiec, K., Hoffmann-Burdzińska, K., Skórska, A., Strzelecki, A., Szojda, G., & Wolny, R. (2026). Perceived benefits, leadership engagement and AI maturity in Polish SMEs: A socio-technical perspective on sustainable digital transformation under competitive pressure. Sustainability, 18(10), 4807. https://doi.org/10.3390/su18104807

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kanbach, D. K., Heiduk, L., Blueher, G., Schreiter, M., & Lahmann, A. (2024). The GenAI is out of the bottle: Generative artificial intelligence from a business model innovation perspective. Review of Managerial Science, 18(4), 1189–1220. https://doi.org/10.1007/s11846-023-00696-z

Kljajić Borštnar, M., & Pucihar, A. (2021). Multi-attribute assessment of digital maturity of SMEs. Electronics, 10(8), 885. https://doi.org/10.3390/electronics10080885

Klotz, D. (2026). The buy-or-build decision, revisited: How agentic AI changes the economics of enterprise software [Preprint]. arXiv. https://arxiv.org/abs/2604.26482

Kohli, R., & Devaraj, S. (2003). Measuring information technology payoff: A meta-analysis of structural variables in firm-level empirical research. Information Systems Research, 14(2), 127–145. https://doi.org/10.1287/isre.14.2.127.16019

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks [Preprint]. arXiv. https://arxiv.org/abs/2005.11401

Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. New Directions for Program Evaluation, 1986(30), 73–84. https://doi.org/10.1002/ev.1427

Liu, B., Li, X., Zhang, J., Wang, J., He, T., Hong, S., Liu, H., Zhang, S., Song, K., Zhu, K., … Wu, C. (2025). Advances and challenges in foundation agents: From brain-inspired intelligence to evolutionary, collaborative, and safe systems [Preprint]. arXiv. https://arxiv.org/abs/2504.01990

Marcineková, K., Sujová, A., & Ďurica, R. (2025). Implementing AI chatbots in customer service optimization — A case study in micro-enterprise. Information, 16(12), 1078. https://doi.org/10.3390/info16121078

Mariani, M. M., & Dwivedi, Y. K. (2024). Generative artificial intelligence in innovation management: A preview of future research developments. Journal of Business Research, 175, Article 114542. https://doi.org/10.1016/j.jbusres.2024.114542

Markus, M. L., & Silver, M. S. (2008). A foundation for the study of IT effects: A new look at DeSanctis and Poole's concepts of structural features and spirit. Journal of the Association for Information Systems, 9(10), 609–632. https://doi.org/10.17705/1jais.00176

Mathis, W. S., Zhao, S., Pratt, N., Weleff, J., & De Paoli, S. (2024). Inductive thematic analysis of healthcare qualitative interviews using open-source large language models: How does it compare to traditional methods? Computer Methods and Programs in Biomedicine, 255, 108356. https://doi.org/10.1016/j.cmpb.2024.108356

Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Information technology and organizational performance: An integrative model of IT business value. MIS Quarterly, 28(2), 283–322. https://doi.org/10.2307/25148636

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. American Journal of Sociology, 83(2), 340–363. https://doi.org/10.1086/226550

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

Miller, D. (1986). Configurations of strategy and structure: Towards a synthesis. Strategic Management Journal, 7(3), 233–249. https://doi.org/10.1002/smj.4250070305

Mitchell, M., Ghosh, A., Luccioni, A. S., & Pistilli, G. (2025). Fully autonomous AI agents should not be developed [Preprint]. arXiv. https://arxiv.org/abs/2502.02649

Mittal, S., Khan, M. A., Romero, D., & Wuest, T. (2018). A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs). Journal of Manufacturing Systems, 49, 194–214. https://doi.org/10.1016/j.jmsy.2018.10.005

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

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2). https://doi.org/10.1177/2053951716679679

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., & Mökander, J. (2021). Ethics as a service: A pragmatic operationalisation of AI ethics. Minds and Machines, 31(2), 239–256. https://doi.org/10.1007/s11023-021-09563-w

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Science and Engineering Ethics, 26(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2021). Operationalising AI ethics: Barriers, enablers and next steps. AI & Society, 38(1), 411–423. https://doi.org/10.1007/s00146-021-01308-8

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

National Institute of Standards and Technology. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.600-1

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. European Journal of Information Systems, 22(3), 336–359. https://doi.org/10.1057/ejis.2012.26

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

Oldemeyer, L., Jede, A., & Teuteberg, F. (2024). Investigation of artificial intelligence in SMEs: A systematic review of the state of the art and the main implementation challenges. Management Review Quarterly, 75(2), 1185–1227. https://doi.org/10.1007/s11301-024-00405-4

Orlikowski, W. J., & Gash, D. C. (1994). Technological frames: Making sense of information technology in organizations. ACM Transactions on Information Systems, 12(2), 174–207. https://doi.org/10.1145/196734.196745

Park, Y., El Sawy, O. A., & Fiss, P. C. (2017). The role of business intelligence and communication technologies in organizational agility: A configurational approach. Journal of the Association for Information Systems, 18(9), 648–686. https://doi.org/10.17705/1jais.00467

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

Pirola, F., Cimini, C., & Pinto, R. (2019). Digital readiness assessment of Italian SMEs: A case-study research. Journal of Manufacturing Technology Management, 31(5), 1045–1083. https://doi.org/10.1108/jmtm-09-2018-0305

Prescott, M. R., Yeager, S., Ham, L., Rivera Saldana, C. D., Serrano, V., Narez, J., Paltin, D., Delgado, J., Moore, D. J., & Montoya, J. L. (2024). Comparing the efficacy and efficiency of human and generative AI: Qualitative thematic analyses. JMIR AI, 3, e54482. https://doi.org/10.2196/54482

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT '20), 33–44. https://doi.org/10.1145/3351095.3372873

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. MIS Quarterly, 43(3), 695–719. https://doi.org/10.25300/misq/2019/13747

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2017). Saturation in qualitative research: Exploring its conceptualization and operationalization. Quality & Quantity, 52(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8

Schryen, G. (2013). Revisiting IS business value research: What we already know, what we still need to know, and how we can get there. European Journal of Information Systems, 22(2), 139–169. https://doi.org/10.1057/ejis.2012.45

Schwaeke, J., Peters, A., Kanbach, D. K., Kraus, S., & Jones, P. (2024). The new normal: The status quo of AI adoption in SMEs. Journal of Small Business Management, 63(3), 1297–1331. https://doi.org/10.1080/00472778.2024.2379999

Schwartmann, R., Keber, T., Zenner, K., & Kurth, S. (2024). Data protection aspects of the use of artificial intelligence — Initial overview of the intersection between GDPR and AI Act. Computer Law Review International, 25(5), 145–150. https://doi.org/10.9785/cri-2024-250503

Short, J. C., Payne, G. T., & Ketchen, D. J. (2008). Research on organizational configurations: Past accomplishments and future challenges. Journal of Management, 34(6), 1053–1079. https://doi.org/10.1177/0149206308324324

Silić, M., & Back, A. (2014). Shadow IT — A view from behind the curtain. Computers & Security, 45, 274–283. https://doi.org/10.1016/j.cose.2014.06.007

Silić, M., Silić, D., & Kind-Trüller, K. (2025). From shadow IT to shadow AI — Threats, risks and opportunities for organizations. Strategic Change. Advance online publication. https://doi.org/10.1002/jsc.2682

Snieška, V., Navickas, V., Havierniková, K., Okręglicka, M., & Gajda, W. (2020). Technical, information and innovation risks of Industry 4.0 in small and medium-sized enterprises – case of Slovakia and Poland. Journal of Business Economics and Management, 21(5), 1269–1284. https://doi.org/10.3846/jbem.2020.12279

Squires, A. (2009). Methodological challenges in cross-language qualitative research: A research review. International Journal of Nursing Studies, 46(2), 277–287. https://doi.org/10.1016/j.ijnurstu.2008.08.006

Strong, D. M., Volkoff, O., Johnson, S. A., Pelletier, L. R., Tulu, B., Bar-On, I., Trudel, J., & Garber, L. (2014). A theory of organization-EHR affordance actualization. Journal of the Association for Information Systems, 15(2), 53–85. https://doi.org/10.17705/1jais.00353

Sánchez, E., Calderón, R., & Herrera, F. (2025). Artificial intelligence adoption in SMEs: Survey based on TOE–DOI framework, primary methodology and challenges. Applied Sciences, 15(12), 6465. https://doi.org/10.3390/app15126465

Štetka, P., Grisáková, N.& Baziuk, A. (2026). Interview Dataset: Adoption and Implementation of Artificial Intelligence in Slovak Enterprises, Bilingual Qualitative Dataset (Two Waves) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.21891565

Tabassi, E. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. Strategic Management Journal, 28(13), 1319–1350. https://doi.org/10.1002/smj.640

Tiago, F., & Almeida, A. (2026). Environmental, organizational, and individual determinants of AI adoption: A multilevel knowledge and analysis. Journal of Innovation & Knowledge, 13, 100934. https://doi.org/10.1016/j.jik.2025.100934

Tick, A. (2023). Evaluation of Industry 4.0 familiarity at SMEs in Central-Eastern Europe using machine learning algorithms. In 2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 643–648). IEEE. https://doi.org/10.1109/saci58269.2023.10158645

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. International Journal for Quality in Health Care, 19(6), 349–357. https://doi.org/10.1093/intqhc/mzm042

Tornatzky, L. G., & Fleischer, M. (1990). The processes of technological innovation. Lexington Books.

Türkeş, M. C., Oncioiu, I., Aslam, H. D., Marin-Pantelescu, A., Topor, D. I., & Căpușneanu, S. (2019). Drivers and barriers in using Industry 4.0: A perspective of SMEs in Romania. Processes, 7(3), 153. https://doi.org/10.3390/pr7030153

van Nes, F., Abma, T., Jonsson, H., & Deeg, D. (2010). Language differences in qualitative research: Is meaning lost in translation? European Journal of Ageing, 7(4), 313–316. https://doi.org/10.1007/s10433-010-0168-y

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International, 22(4), 97–112. https://doi.org/10.9785/cri-2021-220402

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Quarterly, 36(1), 157–178. https://doi.org/10.2307/41410412

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. Journal of Business Research, 70, 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J.-R. (2024). A survey on large language model based autonomous agents. Frontiers of Computer Science, 18(6), Article 186345. https://doi.org/10.1007/s11704-024-40231-1

Ward, J., & Daniel, E. (2012). Benefits management: How to increase the business value of your IT projects (2nd ed.). Wiley.

Waters-Lynch, J., Allen, D. W. E., Potts, J., & Berg, C. (2025). Shadow user innovation: Governing covert generative-AI use for dynamic-capability renewal. Innovation: Organization & Management. Advance online publication. https://doi.org/10.1080/14479338.2025.2519546

Wei, R., & Pardo, C. (2022). Artificial intelligence and SMEs: How can B2B SMEs leverage AI platforms to integrate AI technologies? Industrial Marketing Management, 107, 466–483. https://doi.org/10.1016/j.indmarman.2022.10.008

Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P.-Y. (2023). Supporting qualitative analysis with large language models: Combining codebook with GPT-3 for deductive coding. In Companion Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23 Companion) (pp. 75–78). ACM. https://doi.org/10.1145/3581754.3584136

Yeung, K. (2020). Recommendation of the Council on Artificial Intelligence (OECD). International Legal Materials, 59(1), 27–34. https://doi.org/10.1017/ilm.2020.5

Zhu, K., & Kraemer, K. L. (2005). Post-adoption variations in usage and value of e-business by organizations: Cross-country evidence from the retail industry. Information Systems Research, 16(1), 61–84. https://doi.org/10.1287/isre.1050.0045

Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. Government Information Quarterly, 38(3), 101577. https://doi.org/10.1016/j.giq.2021.101577

Appendix A. Thematic Codebook (8 categories, 24 themes)

Code	Theme	Category	Coverage (n, %)	Operational definition
A1	Trigger & initiator	Context & motivation	32 (55%)	What set AI adoption in motion and who initiated it (individual, management, CTO/HR, external trend).
A2	Motivation	Context & motivation	39 (67%)	Reasons/drivers for adopting AI: market and competitive pressure, internal need, efficiency, survival.
A3	Goals	Context & motivation	30 (52%)	Objectives set for AI adoption, formal or informal, including KPIs or their explicit absence.
B1	Application areas & processes	Use areas & data	48 (83%)	Concrete processes, functions or areas where AI is used.
B2	Profile of use	Use areas & data	13 (22%)	Whether AI use is individual/ad hoc or systematic/organization-wide, including adoption rate.
B3	Data & data security	Use areas & data	28 (48%)	Data quality, availability and protection (GDPR, on-premise, no training on company data).
C1	AI technologies used	Technology & sourcing	45 (78%)	Types of AI technologies/tools deployed (generative AI, LLMs, agents, predictive analytics, NLP/OCR, computer vision).
C2	Build / buy / adapt	Technology & sourcing	28 (48%)	Whether AI is built in-house, bought off-the-shelf, or adapted; proprietary AI products.
C3	Selection criteria & supplier	Technology & sourcing	4 (7%)	Criteria and suppliers/partners for choosing AI solutions.
C4	Licensing & open-source	Technology & sourcing	2 (3%)	Licensing model: paid vs free, open-source, open vs closed platforms.
D1	Approach & phases	Implementation process	40 (69%)	How implementation proceeded — piloting, phasing, idea-to-operation, proof of concept.
D2	Teams & responsibilities	Implementation process	29 (50%)	Who implemented — internal teams, new roles, external partners.
D3	Testing	Implementation process	5 (9%)	Testing and validation of AI before/during deployment.
E1	Technical challenges	Challenges	24 (41%)	Technical difficulties: hallucination, accuracy, data quality, scaling, integration.
E2	Organizational challenges	Challenges	10 (17%)	Organizational obstacles: scale and slow coordination, managing expectations, missing institutional support.
E3	Legal & ethical challenges	Challenges	16 (28%)	Legal/ethical issues: regulation, liability, responsible use, ethics.
E4	Barriers (general)	Challenges	20 (34%)	General or resource barriers: capital, time, scale economics, rapid obsolescence.
E5	Employee adoption & resistance	Challenges	27 (47%)	How users/employees received AI — resistance, fear of replacement, skepticism, enthusiasm, skills gaps.
F1	Benefits	Results & practices	38 (66%)	Achieved benefits — time, cost, quality, capacity and productivity gains.
F2	Success measurement & KPIs	Results & practices	12 (21%)	How AI success is measured, or the explicit absence of measurement.
F3	Best practices	Results & practices	22 (38%)	Practices and recommendations that proved effective.

Code	Theme	Category	Coverage (n, %)	Operational definition
G1	Governance	Governance & regulation	8 (14%)	Internal governance of AI use — control, oversight, internal models, transparency.
G2	Regulation	Governance & regulation	4 (7%)	External regulation and state-level rules (AI Act, sectoral supervision, expected regulation).
H1	Future plans & expansion	Future	47 (81%)	Plans to expand or evolve AI use (new processes, technologies, scaling, vision).

Appendix B. Confirmatory Tests, Effect Sizes, and FDR

Primary confirmatory family with Benjamini–Hochberg FDR across non-caveated tests (N = 58, purposive). Effect sizes: Cliff's δ, Cramér's V, odds ratio, Spearman ρ, or rank-biserial as noted.

Family	Test	Statistic	p (raw)	Effect	Effect type	q (BH)	Sig. (.05)
RQ2_measurement	McNemar benefits vs measurement	8.000	0.0001	26.000	discordant diff	0.001	yes
RQ2_measurement	measurement x size	0.057	0.9720	0.031	CramersV	1.000	no
RQ2_measurement	measurement x ict	1.116	0.2908	0.139	CramersV	0.409	no
RQ2_measurement	formality by measurement(MWU)	388.000	0.0265	0.406	Cliffs delta	0.053	no
RQ3_governance	gov_any: Large vs SME (Fisher)		0.3270	2.932	OddsRatio	0.409	no
RQ3_governance	gov_any: ICT vs nonICT (Fisher)		1.0000	0.870	OddsRatio	1.000	no
RQ1_config	formality ICT vs nonICT (MWU)	567.500	0.0144	0.364	Cliffs delta	0.036	yes
RQ1_config	formality Large vs SME (MWU)	56.500	0.0100	-0.552	Cliffs delta	0.036	yes
RQ1_config	Altypes~formality (Spearman)	0.197	0.1392	0.197	Spearman rho	0.232	no
RQ1_config	AI breadth ICT vs nonICT (MWU)	568.500	0.0137	0.367	Cliffs delta	0.036	yes
WAVE_caveated	formality W2 vs W1 (MWU, CONFOUNDED)	117.500	0.0000	0.721	Cliffs delta		
WAVE_caveated	segments W2 vs W1 (MWU)	1.000	0.0000	0.998	Cliffs delta		

Appendix C. List of Figures

Figure 3.1 Conceptual framework of AI adoption, governance, measurement, and configuration 35

Figure 5.1 Composition of the AI-ImpactSK sample (N=58): (a) wave, (b) company size, (c) sector group, (d) respondent role 48

Figure 5.2 Prevalence of AI technology types across 58 firms. 49

Figure 5.3 Coverage of the 24 themes across 58 interviews, coloured by codebook category. Coverage = share of interviews in which a theme appears at least once 52

Figure 5.4 Eight codebook categories by interview coverage and share of the 629 coded segments. Governance & regulation (G) is marginal on both measures. 53

Figure 5.5 Radial co-occurrence map of the 24 themes (chords where participant-level Jaccard ≥ 0.30; node size ∝ coverage). A densely connected adoption-narrative core (B1, H1, C1, A2, D1, F1) contrasts with near-isolated governance, regulation, testing and KPI themes. 54

Figure 6.1 Coverage of formal governance (G1), regulation (G2), legal/ethical (E3), and data-security (B3) themes with 95% Wilson score intervals. Formal governance and regulation rarely surfaced; concern was voiced mostly as data security 59

Figure 6.2 Validation of manual coding against independent lexical signals. Measurement and agentic codes are text-valid; governance talk is diffuse and weakly tied to the formal code 62

Figure 7.1 The measurement paradox. (a) 65.5% of firms articulate benefits but only 20.7% measure them (95% Wilson score CIs). (b) Only 4 firms both articulate and measure benefits; McNemar p<.001. 69

Figure 7.2 The benefit–measurement dissociation across size classes. The gap is broad, not confined to small firms (size class is confounded with wave; read descriptively) 71

Figure 8.1 Implementation formality index for ICT vs non-ICT firms within each wave, isolating the firm effect from the coding-depth/wave confound 79

Figure 8.2 Four coding-independent adopter configurations (k-modes on AI-profile and firm metadata) by share using each AI type. 80

Figure 9.1 Wave 2 interviews were coded ~64% more densely and respondents spoke ~1.6x more, confounding cross-wave coverage comparisons (a methodological caution) 86

Figure 9.2 Coding-independent text signals: agentic-AI mentions rose sharply (p=.002) and named-tool breadth widened (p=.077) between waves — a real maturation robust to coding depth 88

Figure 9.3 Keyness (log-likelihood) of respondent vocabulary by wave. Wave 1 foregrounds the generative/ML novelty and trigger; Wave 2 foregrounds practical solutions, verification, and production. 89

Figure 9.4 Most-mentioned AI tools in respondent speech (GPT/ChatGPT dominate; Claude, Copilot, Gemini follow). 90

Figure 9.5 Difference in theme coverage between waves (W2–W1, percentage points). Broad positive shifts partly reflect denser Wave-2 coding; the Trigger drop and agentic rise are notable. 91