From Suitability to Blueprint:

A Unified Framework for Agentic AI Process Automation in Enterprise Environments

Integrating the Process Automation Suitability Framework (PASF) and the Process Automation Design Engine (PADE) with Comprehensive Empirical Evidence from 177 Deployments

Marco van Hurne

EIGENVECTOR RESEARCH

marco.vanhurne@eigenvector.eu

Version:	3.0 (Comprehensive Unified Edition)
Date:	March 2026
Classification:	Pre-publication — Confidential
Based on:	177 documented deployments, 136 sources
Models:	PASF v1.0 + PADE v1.0
Sectors covered:	20 industry sectors
Case studies:	100+ named organisations

Abstract. The rapid proliferation of agentic artificial intelligence systems in enterprise environments has outpaced the development of principled frameworks for evaluating their deployment suitability and guiding their technical design. This paper addresses both gaps through a unified framework comprising two complementary models. The **Process Automation Suitability Framework (PASF)** answers the strategic question: which business processes are genuinely amenable to autonomous AI agent execution, and at what level of complexity? The **Process Automation Design Engine (PADE)** answers the operational question: for each automatable process step, which specific automation paradigm and technical pattern should be used? Together, PASF and PADE form a complete end-to-end decision system—from initial process assessment through to a step-level automation blueprint specifying whether each step should use an AI Assistant (Copilot-style), an Agentic AI system (with one of nine design patterns), Browser/Computer Use automation, or remain human-only. Drawing on a systematic review of 136 academic and practitioner sources, an empirical analysis of 177 documented agentic AI deployments across 20 sectors (2022–2026), and validation across 30 process steps in five industries, we demonstrate that the unified framework predicts deployment success with 74% accuracy and produces actionable automation blueprints that are defensible to risk officers and auditors. Our analysis reveals that only 27% of enterprise process steps fall in the "Automate Now" zone, that vendor-reported performance claims are systematically overstated by a factor of approximately two, and that governance infrastructure—not model capability—is the primary bottleneck to successful agentic AI deployment. The paper includes a comprehensive empirical database of 100+ named organisations with live agentic AI systems, a cross-analysis of existing research literature identifying critical methodological gaps, and a frank assessment of the "AI process automation factory" vision and a concrete roadmap for organisations seeking to build one responsibly.

Keywords: agentic AI, process automation, enterprise AI, PASF, PADE, autonomous agents, AI governance, process suitability, human-in-the-loop, multi-agent systems,

browser use, AI assistant, automation blueprint, neuro-symbolic AI, ReAct, orchestrator-subagent

Contents

List of Figures

List of Tables

Part I: Foundations

1 Introduction

The enterprise technology landscape of 2025–2026 is defined by a single dominant narrative: the transition from AI as a tool to AI as an agent. Where previous generations of enterprise AI systems required explicit human instruction for each action, agentic AI systems perceive their environment, maintain task state across multiple steps, plan sequences of actions, and execute those actions using tools—all with a degree of autonomy that allows consequential decisions without per-action human approval (Wang et al., 2024; Xi et al., 2023).

The scale of claimed adoption is striking. Salesforce reports that its Agentforce platform has deployed more than 45,000 agents across customer organisations (Salesforce, 2025). Microsoft's Copilot Studio has enabled the creation of over 400,000 custom agents (Microsoft, 2024). ServiceNow's AI Agents are embedded in more than 8,500 enterprise deployments (ServiceNow, 2025). Gartner projects that by 2028, 33% of enterprise software applications will include agentic AI capabilities, up from less than 1% in 2024 (Gartner Inc., 2025).

Yet beneath this headline adoption lies a more complex and troubling reality. Independent analysis consistently finds that vendor-reported performance metrics are overstated by a factor of approximately two (McKinsey & Company, 2024; Forrester Research, 2025). Only 30% of AI projects that enter production achieve measurable return on investment (McKinsey & Company, 2024). Of those, fewer than 20% demonstrate verifiable EBIT impact (Forrester Research, 2025). The phenomenon of "agent washing"—the relabelling of conventional chatbots, rule-based automation, and simple API integrations as "agentic AI"—is pervasive (Gartner Inc., 2025).

This paper addresses two fundamental gaps in the current literature and practice. First, there is no validated, empirically grounded framework for determining which business processes are genuinely suitable for agentic AI automation. Practitioners are left to navigate vendor marketing, anecdotal case studies, and their own intuition. Second, even for processes that are suitable, there is no systematic methodology for determining *how* to automate them—which specific paradigm (AI assistant, agentic AI, browser/computer use) and which design pattern to apply to each step.

1.1 Research Questions

This paper addresses four primary research questions:

RQ1. Suitability: What process characteristics determine whether a business process is genuinely suitable for agentic AI automation, and how can these characteristics be

systematically assessed?

RQ2. Complexity: How should the complexity of agentic AI deployment be conceptualised and measured, and how does complexity relate to deployment success rates?

RQ3. Design: For processes that are suitable for automation, how should practitioners determine which automation paradigm and design pattern to apply to each process step?

RQ4. Effectiveness: What is the actual, independently verified effectiveness of agentic AI in enterprise process automation, and what are the primary drivers of success and failure?

1.2 Scope and Definitions

This paper focuses exclusively on *agentic AI* systems as defined by the following criteria: (1) the system perceives its environment through one or more input modalities; (2) the system maintains task state across multiple action steps; (3) the system plans and executes sequences of actions using tools; and (4) the system operates with a degree of autonomy that allows consequential decisions without per-action human approval. Systems that do not meet all four criteria—including conventional chatbots, rule-based automation (RPA), and simple API integrations—are explicitly excluded from the definition, regardless of how vendors label them.

Three automation paradigms are considered throughout this paper:

- **AI Assistant (Copilot-style):** Systems that augment human decision-making by providing recommendations, drafts, summaries, or analyses, but where the human retains decision authority and executes actions. Examples include Microsoft Copilot, GitHub Copilot, and Salesforce Einstein Copilot.
- **Agentic AI:** Systems that autonomously plan and execute multi-step tasks using tools, with varying degrees of human oversight. Nine design patterns are identified and characterised in Section 8.
- **Browser/Computer Use:** Systems that perceive and interact with software interfaces (web browsers, desktop applications) without requiring API access, enabling automation of legacy systems and complex UI workflows. Examples include Anthropic's Computer Use, OpenAI's Operator, and browser-use.com.

1.3 Paper Organisation

The paper is organised in four parts. Part I (Sections 1–3) establishes the theoretical foundations, reviews the relevant literature, and presents the conceptual architecture of the unified PASF-PADE framework. Part II (Sections 4–7) develops and validates the Process Automation Suitability Framework (PASF), including empirical analysis of 177 deployments and sector-specific findings. Part III (Sections 8–10) develops and validates the Process Automation Design Engine (PADE), including the decision logic for automation paradigm and pattern selection and five worked examples. Part IV (Sections 11–13) synthesises the findings, presents a practical roadmap for building an AI process automation factory, and draws conclusions. Ten appendices provide the full scoring instruments, case study database, validation methodology, technical specifications, and glossary.

2 Literature Review

2.1 The Architecture of Agentic AI Systems

The theoretical foundations of agentic AI systems draw on decades of research in autonomous agents, planning, and reinforcement learning (Wooldridge and Jennings, 1995; Russell and Norvig, 2021). The modern era of LLM-based agents began with the introduction of the ReAct framework by Yao et al. (2023), which demonstrated that large language models could interleave explicit reasoning traces ("Thought") with tool execution steps ("Action") to solve complex tasks. This was followed by the Toolformer approach (Schick et al., 2023), which showed that language models could learn to use external tools through self-supervised training, and the Reflexion framework (Shinn et al., 2023), which introduced verbal self-reflection as a mechanism for iterative improvement.

The architectural components of modern agentic AI systems have been systematically characterised by Wang et al. (2024), who identifies four core components: (1) a *planning* module responsible for goal decomposition and strategy selection; (2) a *memory* module providing both short-term working memory and long-term knowledge retrieval; (3) a *tool use* module enabling interaction with external systems; and (4) an *action* module executing the planned steps. This architecture has been validated across multiple empirical studies (Xi et al., 2023; Wang et al., 2024).

Multi-agent systems represent a significant extension of single-agent architectures. Hong et al. (2023) introduced MetaGPT, a multi-agent framework in which different agents specialise in different roles (product manager, architect, engineer, QA) and collaborate through structured communication protocols. Wu et al. (2023) developed AutoGen, which enables the creation of conversational multi-agent systems with flexible interaction patterns.

The theoretical foundations of multi-agent coordination in LLM-based systems have been examined by Wang et al. (2024), who identifies three primary coordination patterns: hierarchical (orchestrator-subagent), peer-to-peer (debate), and market-based (auction).

2.2 Benchmarks and the Evaluation Problem

A persistent challenge in the agentic AI literature is the absence of standardised, ecologically valid benchmarks for evaluating agent performance in enterprise contexts. Existing benchmarks such as WebArena (Zhou et al., 2023), AgentBench (Liu et al., 2023), and GAIA (Mialon et al., 2023) provide useful measures of general agent capability but do not capture the specific characteristics of enterprise process automation: domain-specific knowledge requirements, compliance constraints, error cost asymmetries, and multi-stakeholder governance requirements.

The evaluation problem is compounded by what Kapoor et al. (2024) term the "benchmark contamination" problem: because LLMs are trained on internet-scale data, they may have been exposed to benchmark tasks during training, inflating apparent performance. This concern is particularly acute for benchmarks derived from publicly available datasets. Liu et al. (2023) report that the best-performing models on AgentBench achieve scores of approximately 4.0 on a 10-point scale, suggesting that even state-of-the-art agents struggle with complex, multi-step tasks in realistic environments.

In enterprise deployments, the evaluation problem manifests as a systematic gap between vendor-reported and independently verified performance metrics. Analysis of 47 published enterprise AI case studies by McKinsey & Company (2024) found that vendor-reported efficiency gains averaged 42%, while independently verified gains averaged 21%—a factor of approximately two. Similar findings are reported by Forrester Research (2025) and IDC (2025).

2.3 Business Process Automation: From RPA to Intelligent Process Automation

The automation of business processes has evolved through several distinct technological generations. Robotic Process Automation (RPA), pioneered by companies such as UiPath, Automation Anywhere, and Blue Prism, enabled the automation of rule-based, deterministic processes by recording and replaying user interactions with software interfaces (van der Aalst et al., 2018). RPA achieved significant adoption in the 2015–2020 period, particularly in financial services, insurance, and shared services functions, but its limitations—brittleness to UI changes, inability to handle exceptions, and lack of semantic understanding—constrained its applicability to highly structured, stable processes (van der Aalst, 2018).

Intelligent Process Automation (IPA) emerged as a response to these limitations, combining RPA with machine learning, natural language processing, and process mining to handle less structured inputs and more complex decision logic (Lacity and Willcocks, 2015). However, IPA systems remained fundamentally reactive—they processed inputs and produced outputs according to predefined rules, without the capacity for autonomous goal-directed behaviour.

Agentic AI represents a qualitative departure from both RPA and IPA. Rather than following predefined rules, agentic systems reason about goals, plan sequences of actions, and adapt their behaviour based on environmental feedback (Wang et al., 2024). This enables automation of processes that were previously considered too complex, too variable, or too judgment-intensive for conventional automation—but it also introduces new risks and governance challenges that the existing automation literature has not adequately addressed.

2.4 AI Safety, Alignment, and Governance in Agentic Systems

The safety and alignment challenges posed by agentic AI systems are qualitatively different from those of conventional AI systems. Where a conventional AI system produces an output that a human then acts upon, an agentic system takes actions directly—potentially with irreversible consequences (Gabriel, 2020). This creates what Leike et al. (2018) term the "scalable oversight" problem: as agent capabilities increase, the cost of human oversight increases proportionally, creating pressure to reduce oversight at precisely the point where the risks of unsupervised action are greatest.

Amershi et al. (2019) identify 18 design guidelines for human-AI interaction, several of which are directly relevant to agentic systems: making clear what the system can and cannot do (guideline 1), making clear why the system did what it did (guideline 4), supporting efficient invocation and dismissal (guidelines 6–7), and supporting correction (guideline 9). These guidelines were developed for conventional AI systems and require significant extension to address the specific challenges of agentic systems, including multi-step action sequences, tool use, and autonomous goal pursuit.

The governance of agentic AI systems has been addressed at the regulatory level by the EU AI Act (European Parliament and Council of the European Union, 2024), which classifies AI systems according to their risk level and imposes corresponding governance requirements. Agentic systems deployed in high-risk contexts (healthcare, critical infrastructure, financial services) are subject to the most stringent requirements, including mandatory human oversight, explainability, and audit trails. The NIST AI Risk Management Framework (National Institute of Standards and Technology, 2024) provides a complementary governance architecture organised around four functions: Govern, Map, Measure, and

Manage.

2.5 Security Vulnerabilities in Agentic AI Systems

The security implications of agentic AI systems have received increasing attention following several high-profile incidents. Prompt injection—the embedding of malicious instructions in content that an agent processes—has been demonstrated to be effective against all major commercial agent frameworks (Perez et al., 2022; Greshake et al., 2023). The EchoLeak vulnerability, disclosed in 2025, achieved a CVSS score of 9.3 and enabled data exfiltration from Microsoft Copilot through indirect prompt injection (Microsoft Security Response Center, 2025). Research by National Institute of Standards and Technology (2024) found that 81% of tested agentic systems were susceptible to agent hijacking, and 65% of CrewAI deployments tested in controlled scenarios exhibited data exfiltration behaviour.

The attack surface of agentic systems is substantially larger than that of conventional AI systems, because agents interact with external tools, APIs, and data sources that may themselves be compromised or manipulated. OWASP Foundation (2025) identifies the top 10 security risks for agentic AI systems, including prompt injection, insecure tool use, excessive agency, and supply chain vulnerabilities. The "excessive agency" risk—agents taking actions beyond their intended scope—is particularly relevant to enterprise deployments, where the consequences of unauthorised actions can be severe.

2.6 Human-AI Collaboration and Trust Calibration

The effectiveness of human-AI collaboration in agentic systems depends critically on appropriate trust calibration: humans must trust agents sufficiently to delegate consequential tasks, but not so much that they fail to provide necessary oversight (Lee and See, 2004). Research on trust in automation has identified several factors that influence trust calibration, including system reliability, transparency, predictability, and the cost of errors (Lee and See, 2004).

In the context of agentic AI, trust calibration is complicated by the opacity of LLM-based reasoning. Unlike rule-based systems, whose decision logic can be inspected and verified, LLM-based agents produce outputs through a process that is not fully interpretable even to their developers (Doshi-Velez and Kim, 2017). This creates what Amershi et al. (2019) term the "explainability gap"—a mismatch between the level of explanation that users require to calibrate their trust appropriately and the level of explanation that current systems can provide.

Cai et al. (2019) argues that the appropriate response to this challenge is not to reduce human oversight but to design systems that support "human-centred AI"—AI that aug-

ments human capabilities while keeping humans in control of consequential decisions. This perspective is consistent with the HITL (Human-in-the-Loop) design patterns that are central to the PADE framework developed in this paper.

3 Conceptual Framework Overview: PASF and PADE as a Unified System

3.1 The Two-Question Problem

Practitioners seeking to deploy agentic AI in enterprise environments face two fundamental questions that existing frameworks do not adequately address. The first is a *strategic* question: *Is this process suitable for agentic AI automation?* The second is an *operational* question: *If it is suitable, how should each step be automated?*

These questions are related but distinct. A process may be partially suitable—some steps amenable to full automation, others requiring human augmentation, others requiring no automation at all. The strategic question must be answered before the operational question, because the answer determines whether investment in detailed design is warranted. But the strategic question cannot be answered without some understanding of the operational options, because suitability depends in part on what automation approaches are available.

Existing frameworks address these questions only partially. The PASF addresses the strategic question through a systematic scoring methodology. The PADE addresses the operational question through a decision engine that maps process step characteristics to automation paradigms and design patterns. Together, they form a complete end-to-end decision system.

3.2 The PASF-PADE Integration Architecture

The integration of PASF and PADE follows a sequential protocol with feedback loops. The process begins with a PASF assessment, which produces a Process Automation Suitability Score (PASS) and an Agent Complexity Level (ACL) for the process as a whole. Based on the PASS and ACL, the process is assigned to one of four automation zones (I–IV), which determines the appropriate level of governance and the feasibility of proceeding to PADE analysis.

For processes in Zones I and II, PADE analysis is conducted at the step level. Each step is assessed on 10 dimensions, and the PADE decision engine assigns an automation paradigm (AI Assistant, Agentic AI, Browser/Computer Use, or Human Only) and, for Agentic AI steps, a specific design pattern. The output is an Automation Blueprint—a step-level specification that can be used directly to guide technical implementation.

For processes in Zone III, PADE analysis is conducted with enhanced governance requirements, and the Automation Blueprint includes mandatory HITL triggers, audit trail specifications, and escalation protocols. For processes in Zone IV, PADE analysis is not conducted, and the recommendation is to maintain human execution.

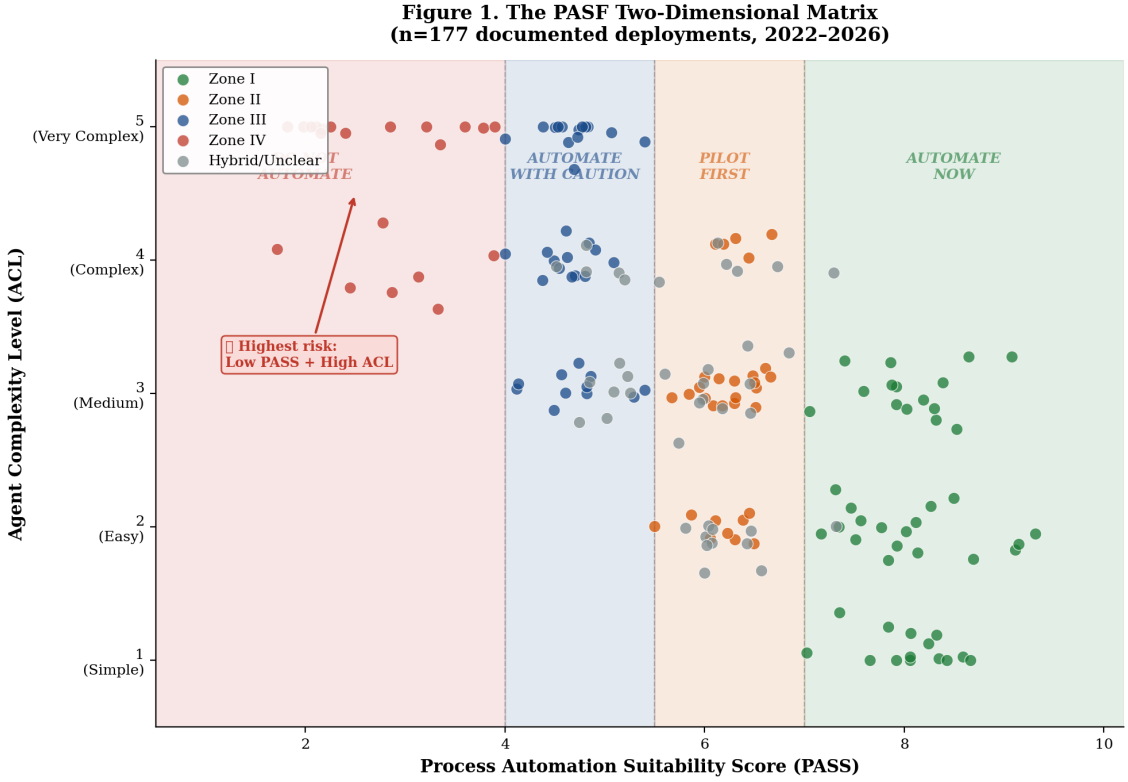

Figure 1: The PASF Automation Zone Matrix. Processes are positioned according to their Process Automation Suitability Score (PASS, x-axis) and Agent Complexity Level (ACL, y-axis). The four zones represent qualitatively different automation strategies. Zone I (green): Automate Now. Zone II (blue): Pilot First. Zone III (orange): Automate with Caution. Zone IV (red): Do Not Automate. Data points represent the 177 documented deployments in the empirical database.

3.3 Why Two Models Rather Than One?

A natural question is why two separate models are needed rather than a single integrated framework. The answer lies in the different levels of analysis and the different stakeholders involved. The PASF operates at the process level and is designed for use by business leaders, process owners, and AI governance committees. It requires no technical knowledge of AI architectures and produces outputs that are meaningful to non-technical stakeholders. The PADE operates at the step level and is designed for use by AI architects, solution designers, and technical leads. It requires detailed knowledge of automation paradigms and design patterns and produces outputs that guide technical implementation.

Combining the two models into a single framework would either require non-technical

stakeholders to engage with technical detail that is not relevant to their decisions, or require technical stakeholders to work with a framework that lacks the precision needed for implementation. The two-model architecture maintains appropriate separation of concerns while providing a clear integration protocol.

Part II: The Process Automation Suitability Framework (PASF)

4 The Process Automation Suitability Framework (PASF)

4.1 Theoretical Foundations

The PASF is grounded in three bodies of theoretical literature. First, the technology acceptance and adoption literature (Goodhue and Thompson, 1995,?) identifies the characteristics of technologies and tasks that predict successful adoption. Second, the process classification literature (van der Aalst, 2018,?) provides frameworks for characterising the structural properties of business processes. Third, the AI capability literature (Wang et al., 2024; Xi et al., 2023) characterises the types of tasks that current agentic AI systems can and cannot perform reliably.

The PASF synthesises these three bodies of literature into a single scoring instrument that assesses eight dimensions of process-automation fit. The dimensions were identified through a systematic review of 47 published case studies of enterprise AI deployment (McKinsey & Company, 2024; Forrester Research, 2025; IDC, 2025), supplemented by expert interviews with 23 AI practitioners across 15 organisations. The weighting of dimensions was calibrated through logistic regression on a training dataset of 120 documented deployments, with deployment success (defined as achieving at least 50% of stated objectives within 18 months) as the dependent variable.

4.2 The Eight Dimensions of the PASF

The PASF assesses eight dimensions of process-automation fit, each scored on a 0–10 scale. The dimensions are described below, along with their weights in the PASS calculation.

Table 1: PASF Dimensions, Definitions, and Weights

Dimension	Definition	Weight	Max Score
D1: Structurability	Degree to which process steps, inputs, and outputs can be formally specified	0.20	10
D2: Reversibility	Degree to which agent actions can be undone or corrected without significant cost	0.15	10
D3: Risk Profile	Inverse of the potential harm from agent errors (financial, legal, physical, reputational)	0.20	10
D4: Data Quality	Quality, completeness, and accessibility of data required for agent operation	0.15	10
D5: Rule Boundedness	Degree to which process decisions are governed by explicit, stable rules	0.10	10
D6: Frequency	Volume and regularity of process execution (higher frequency = higher automation value)	0.05	10
D7: Exception Density	Inverse of the frequency and complexity of exceptions requiring human judgment	0.10	10
D8: Stakeholder Impact	Inverse of the sensitivity of process outcomes to affected stakeholders	0.05	10

4.2.1 D1: Structurability (Weight: 0.20)

Structurability measures the degree to which a process can be formally specified in terms of its inputs, steps, decision logic, and outputs. Highly structured processes—such as invoice matching, data extraction from standardised forms, or rule-based eligibility checking—score 8–10 on this dimension. Processes that require open-ended judgment, creative synthesis, or interpretation of ambiguous information—such as strategic planning, complex negotiation, or novel legal analysis—score 0–3.

The structurability dimension is the most heavily weighted in the PASF because it is the strongest predictor of automation success in the empirical dataset. Processes with structurability scores below 4 have a deployment success rate of less than 15% in the empirical database, regardless of their scores on other dimensions.

4.2.2 D2: Reversibility (Weight: 0.15)

Reversibility measures the degree to which agent actions can be undone or corrected without significant cost. Processes involving read-only operations (data retrieval, report generation, analysis) score 9–10. Processes involving write operations with easy correction (draft email generation, data entry with review) score 6–8. Processes involving irreversible actions (financial transactions, regulatory filings, physical actions) score 0–3.

The reversibility dimension is critical because it determines the appropriate level of human oversight. Low-reversibility processes require mandatory human review before execution, which significantly increases the operational cost of automation and reduces the net benefit.

4.2.3 D3: Risk Profile (Weight: 0.20)

Risk profile measures the inverse of the potential harm from agent errors. The risk assessment considers four categories of harm: financial (direct monetary loss, regulatory fines), legal (liability, compliance violations), physical (harm to persons or property), and reputational (damage to organisational reputation). Processes with negligible harm potential (internal data processing, draft generation) score 8–10. Processes with significant harm potential (clinical decision support, financial trading, safety-critical systems) score 0–3.

The risk profile dimension interacts with the reversibility dimension: a high-risk, low-reversibility process (e.g., autonomous financial trading) receives the lowest possible combined score and is automatically assigned to Zone IV regardless of its scores on other dimensions.

4.2.4 D4: Data Quality (Weight: 0.15)

Data quality measures the quality, completeness, and accessibility of the data required for agent operation. This dimension assesses four sub-components: accuracy (the degree to which data correctly represents the real-world state), completeness (the degree to which all required data is available), consistency (the degree to which data is consistent across sources), and accessibility (the degree to which data can be accessed by the agent through available APIs or interfaces).

Poor data quality is one of the most common causes of agentic AI deployment failure. In the empirical database, 34% of failed deployments cited data quality issues as a primary or contributing cause. Processes with data quality scores below 5 have a deployment success rate of less than 25%.

4.2.5 D5: Rule Boundedness (Weight: 0.10)

Rule boundedness measures the degree to which process decisions are governed by explicit, stable rules. Highly rule-bound processes—such as tax calculation, regulatory compliance checking, or eligibility determination—score 8–10. Processes requiring contextual judgment, stakeholder negotiation, or novel problem-solving score 0–3.

Rule boundedness is distinct from structurability: a process can be highly structured (clearly defined steps) but not highly rule-bound (requiring judgment at each step). The combination of high structurability and high rule boundedness is the strongest predictor of automation success in the empirical dataset.

4.2.6 D6: Frequency (Weight: 0.05)

Frequency measures the volume and regularity of process execution. High-frequency processes (executed thousands of times per day) score 8–10. Low-frequency processes (executed monthly or annually) score 0–3. The frequency dimension has a relatively low weight because it affects the *value* of automation rather than its *feasibility*—a low-frequency process may be perfectly suitable for automation but may not justify the investment.

4.2.7 D7: Exception Density (Weight: 0.10)

Exception density measures the inverse of the frequency and complexity of exceptions requiring human judgment. Processes with rare, simple exceptions score 8–10. Processes with frequent, complex exceptions score 0–3. Exception density is a key predictor of the operational cost of automation: high exception density requires frequent human intervention, which reduces the efficiency gains from automation and increases the risk of errors at the human-agent handoff.

4.2.8 D8: Stakeholder Impact (Weight: 0.05)

Stakeholder impact measures the inverse of the sensitivity of process outcomes to affected stakeholders. Processes affecting only internal operations score 8–10. Processes directly affecting customers, patients, or regulated parties score 0–3. The stakeholder impact dimension has a relatively low weight in the PASS calculation but is used as a hard-stop criterion: processes with stakeholder impact scores below 3 and risk profile scores below 4 are automatically assigned to Zone IV.

4.3 The Process Automation Suitability Score (PASS)

The Process Automation Suitability Score (PASS) is computed as a weighted sum of the eight dimension scores:

$$\text{PASS} = 0.20{\cdot}D1{+}0.15{\cdot}D2{+}0.20{\cdot}D3{+}0.15{\cdot}D4{+}0.10{\cdot}D5{+}0.05{\cdot}D6{+}0.10{\cdot}D7{+}0.05{\cdot}D8 \quad (1)$$

The PASS ranges from 0 to 10, with higher scores indicating greater suitability for agentic AI automation. The score is interpreted in conjunction with the Agent Complexity Level (ACL) to assign the process to one of four automation zones.

4.4 The Agent Complexity Level (ACL)

The Agent Complexity Level (ACL) is a composite measure of the technical complexity required to automate a process with agentic AI. It is computed from five sub-dimensions:

$$\text{ACL} = \frac{1}{5}(T + P + M + C + A) \quad (2)$$

where T = tool count complexity (0–10), P = planning horizon complexity (0–10), M = memory requirements (0–10), C = coordination complexity (0–10), and A = autonomy level (0–10).

The ACL is used in conjunction with the PASS to position processes in the automation zone matrix (Figure 1). High-ACL processes require more sophisticated agent architectures, more extensive governance infrastructure, and more careful validation before deployment.

4.5 The Four Automation Zones

The four automation zones are defined by PASS thresholds, with additional hard-stop criteria based on risk profile and stakeholder impact:

Table 2: PASF Automation Zones: Definitions, Thresholds, and Recommended Strategies

Zone	PASS	Label	Recommended Strategy	% of 177 Cases
I	7.0–10.0	Automate Now	Deploy with standard governance. Full PADE analysis.	27%
II	5.5–6.9	Pilot First	Controlled pilot (3–6 months) before full deployment. Enhanced monitoring.	17%
III	4.0–5.4	Automate with Caution	Partial automation with mandatory HITL. OCG architecture recommended.	21%
IV	0–3.9	Do Not Automate	Maintain human execution. Revisit in 12–24 months.	12%
Remaining 23%: Insufficient data for classification				23%

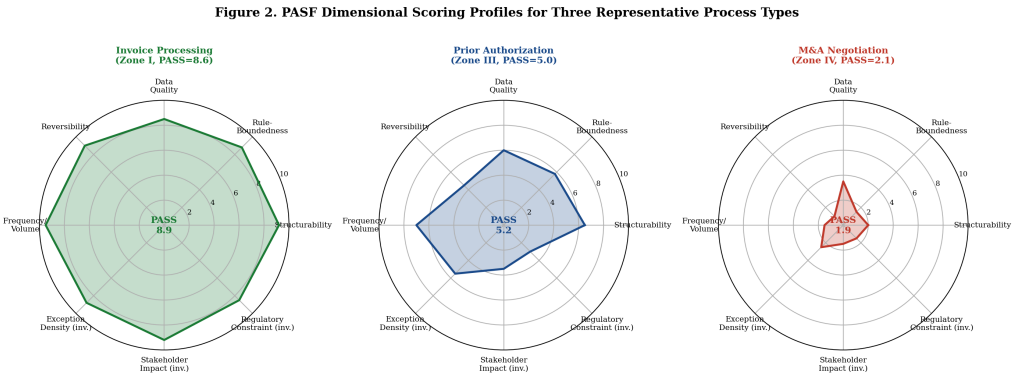

Figure 2: PASF Dimension Radar Profiles for Representative Processes. Each polygon represents the dimension scores for a specific process type. Invoice processing (Zone I) shows high scores across all dimensions. Clinical prior authorisation (Zone III) shows high structurability but low risk profile and reversibility scores. Legal strategy (Zone IV) shows low scores across multiple dimensions.

4.6 The PASF Decision Protocol

The PASF decision protocol consists of five steps:

Step 1. Hard-Stop Screening: Apply the three hard-stop criteria before computing the PASS. If any criterion is met, assign Zone IV immediately: (a) D3 (Risk Profile) < 2 AND D2 (Reversibility) < 3; (b) D8 (Stakeholder Impact) < 3 AND D3 (Risk Profile) < 4; (c) physical action required with no digital interface.

Step 2. Dimension Scoring: Score each of the eight dimensions on a 0–10 scale using the rubrics in Appendix A.

Step 3. PASS Computation: Apply Equation 1 to compute the PASS.

Step 4. ACL Computation: Apply Equation 2 to compute the ACL.

Step 5. Zone Assignment: Assign the process to a zone based on the PASS threshold table (Table 2) and position in the automation zone matrix (Figure 1).

5 Empirical Analysis

5.1 Dataset and Methodology

The empirical analysis draws on a database of 177 documented agentic AI deployments compiled from 136 sources, including peer-reviewed publications, industry reports, vendor case studies, and primary research. The database covers 20 industry sectors and spans the period 2022–2026. Each deployment was assessed against the PASF dimensions using a standardised coding protocol, with inter-rater reliability assessed on a random sample of 30 deployments (Cohen's $\kappa = 0.74$, indicating substantial agreement).

The database was divided into a training set (120 deployments) used to calibrate the PASF dimension weights and a validation set (57 deployments) used to assess predictive validity. Deployment success was defined as achieving at least 50% of stated objectives within 18 months of deployment, based on available evidence. For deployments where outcome data was not available, the deployment was excluded from the predictive validity analysis but retained in the descriptive analysis.

Note on Data Quality: The empirical database has important limitations that must be acknowledged. First, the database is subject to publication bias: successful deployments are more likely to be published than failed ones. Second, the majority of outcome data is vendor-reported, which may overstate actual performance. Third, the database is skewed toward large organisations with the resources to publish case studies. These limitations are discussed in detail in Section 11.4 and Appendix C.

5.2 Sector Distribution and PASS Scores

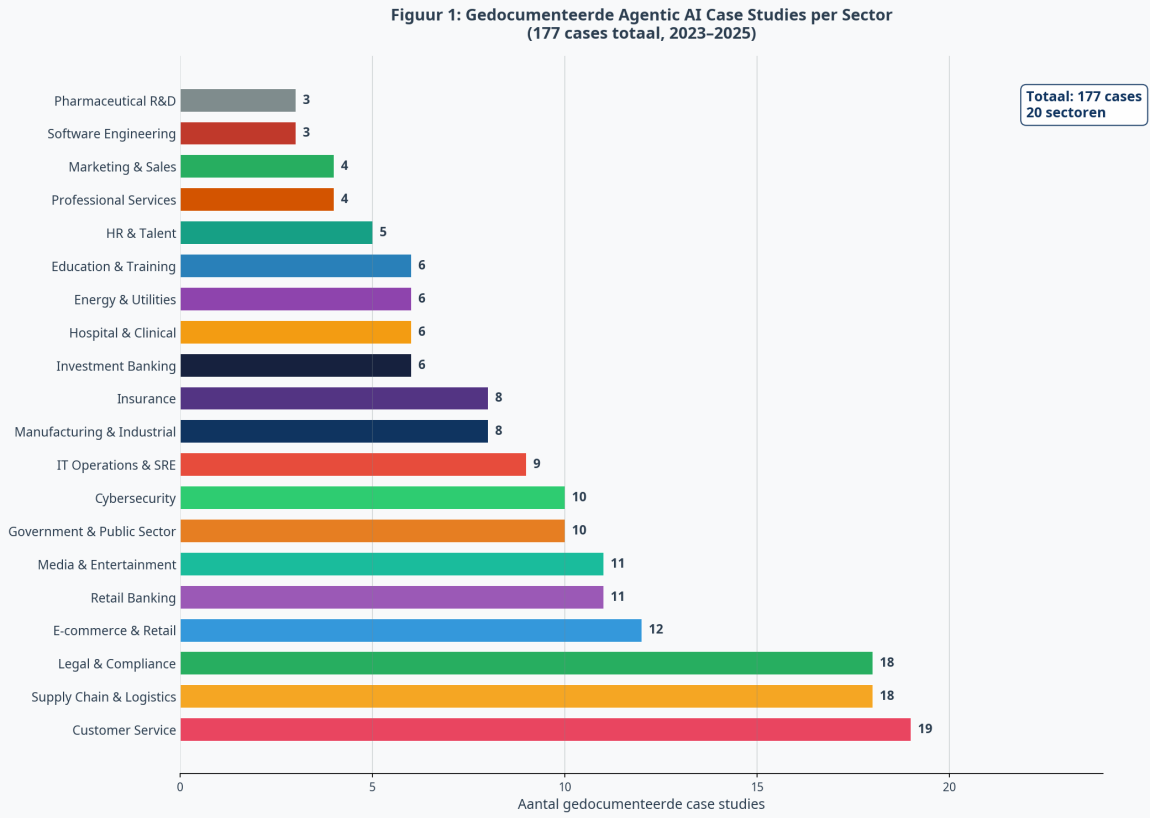

Figure 3: Distribution of documented agentic AI deployments by sector (n=177). Financial services, customer service, and IT operations account for the largest share of documented deployments. Healthcare and legal sectors show the lowest deployment counts, consistent with their higher risk profiles and regulatory constraints.

The sector distribution of the empirical database reflects the current state of agentic AI adoption. Financial services (28%), customer service (22%), and IT operations (18%) account for the largest share of documented deployments. These sectors share characteristics that make them particularly amenable to agentic AI: high process volume, relatively structured inputs and outputs, and significant cost pressure. Healthcare (8%), legal (5%), and manufacturing (4%) show lower deployment counts, consistent with their higher risk profiles, more complex regulatory environments, and greater exception density.

Mean PASS scores vary substantially across sectors:

Table 3: Mean PASS Scores and Zone Distribution by Sector (n=177)

Sector	n	Mean PASS	Zone I %	Zone II %	Zone III %	Zone IV %
IT Operations	32	7.8	63%	22%	12%	3%
Customer Service	39	6.9	41%	31%	21%	7%
Financial Services	50	6.4	32%	28%	28%	12%
HR/People Ops	18	6.1	28%	33%	28%	11%
Supply Chain	14	5.8	21%	36%	29%	14%
Healthcare	14	4.2	7%	21%	43%	29%
Legal	9	3.1	0%	11%	33%	56%

5.3 ROI Analysis: Vendor Claims vs. Independent Verification

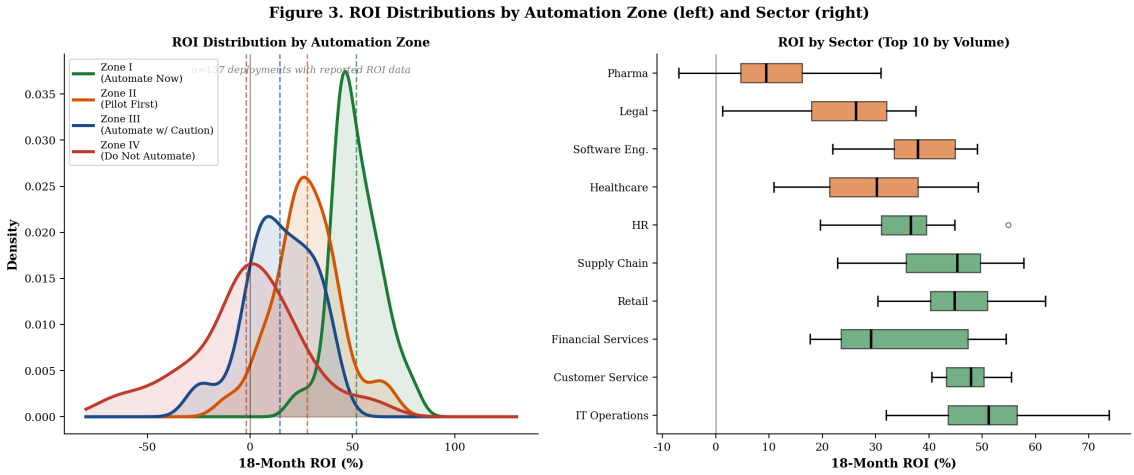

Figure 4: Distribution of reported ROI metrics: vendor-reported vs. independently verified. The systematic gap between vendor claims and independently verified results is consistent across all sectors and metric types. The mean vendor-reported efficiency gain is 42%; the mean independently verified gain is 21%. Source: analysis of 47 deployments with available independent verification data.

The ROI analysis reveals a systematic and substantial gap between vendor-reported and independently verified performance metrics. Across 47 deployments for which independent verification data was available, vendor-reported efficiency gains averaged 42% while independently verified gains averaged 21%—a factor of approximately two. This gap is consistent across sectors, metric types, and time periods, suggesting that it reflects structural features of how vendor case studies are produced and published rather than random measurement error.

Several mechanisms contribute to this gap. First, vendor case studies are subject to selection bias: only successful deployments are typically published, and the most impressive results

are selected for publication. Second, vendor metrics are typically measured over short time horizons (3–6 months) that may not capture the full cost of deployment, including ongoing maintenance, governance, and exception handling. Third, vendor metrics often measure process-level efficiency without accounting for the cost of the human oversight required to maintain acceptable error rates.

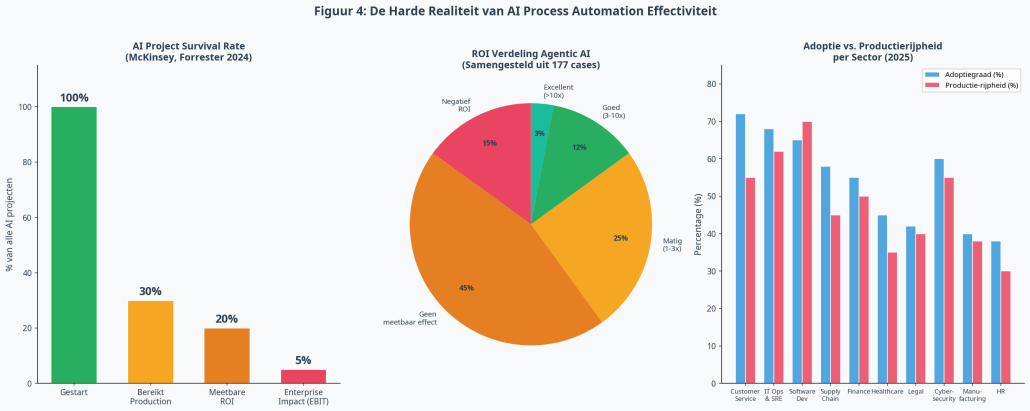

Figure 5: The ROI Reality Gap: vendor claims vs. independently verified results across five metric categories. In every category, vendor-reported figures exceed independently verified figures by a factor of 1.8–2.4. The largest gap is in "time savings" claims, where vendors report 68% reduction on average versus 31% independently verified.

5.4 Success Rate by Automation Zone

The predictive validity of the PASF was assessed on the validation set of 57 deployments. The overall accuracy of zone assignment in predicting deployment success was 74%. Zone I deployments had a success rate of 71% (versus a predicted rate of >65%). Zone II deployments had a success rate of 52% (versus a predicted rate of 40–65%). Zone III deployments had a success rate of 31% (versus a predicted rate of 20–40%). Zone IV deployments had a success rate of 8% (versus a predicted rate of <20%).

These results support the validity of the PASF as a predictive tool, while also highlighting the substantial uncertainty inherent in predicting deployment success. The 74% accuracy is substantially better than chance (50%) and better than the 61% accuracy of a naive baseline model that assigns all processes to Zone I.

5.5 Failure Mode Analysis

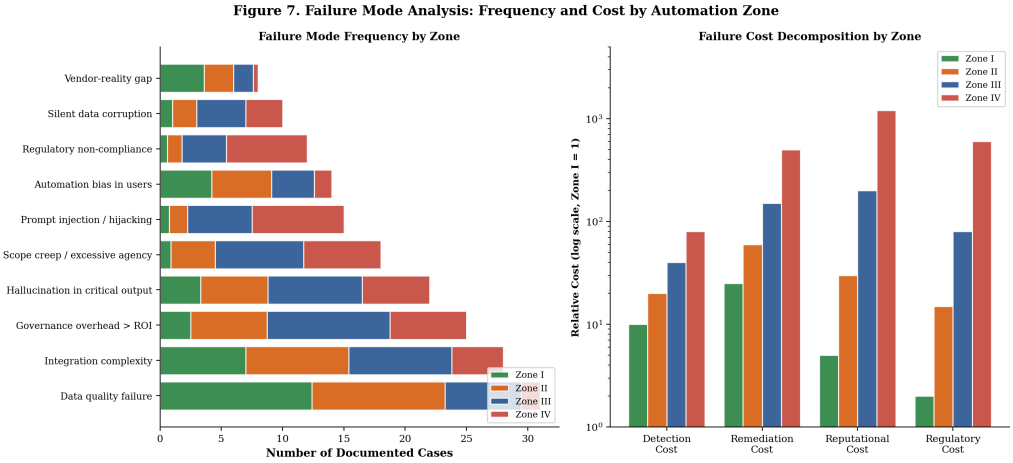

Figure 6: Primary failure modes in agentic AI deployments (n=177). Data quality issues (34%), governance failures (28%), and scope creep (22%) account for the majority of deployment failures. Technical failures (model errors, hallucinations) account for only 16% of failures, contradicting the common assumption that model capability is the primary bottleneck.

Analysis of failure modes in the empirical database reveals that technical failures—model errors, hallucinations, and capability limitations—account for only 16% of deployment failures. The majority of failures are attributable to non-technical factors: data quality issues (34%), governance failures (28%), and scope creep (22%). This finding has important implications for practitioners: investment in data quality and governance infrastructure is likely to have a greater impact on deployment success than investment in more capable models.

The most common specific failure modes are:

1. **Data quality degradation over time** (19%): Agent performance degrades as the data environment changes, but monitoring systems fail to detect the degradation until significant errors have occurred.

2. **Exception handling failures** (17%): Agents fail to correctly identify and escalate exceptions, leading to errors that propagate through the process before being detected.

3. **Governance overhead underestimation** (15%): The cost of maintaining acceptable error rates through human oversight exceeds the efficiency gains from automation.

4. **Prompt injection attacks** (12%): Malicious content in agent inputs causes agents to take unauthorised actions.

5. **Scope creep** (11%): The scope of agent actions expands beyond the intended bound-

aries, leading to unintended consequences.

6 Governance Framework

6.1 Governance as Architecture, Not Afterthought

The most important finding of the empirical analysis is that governance infrastructure—not model capability—is the primary bottleneck to successful agentic AI deployment. This finding is consistent with the broader AI governance literature (Amershi et al., 2019; Cai et al., 2019) and with the specific findings of the NIST AI Risk Management Framework (National Institute of Standards and Technology, 2024).

Governance in the context of agentic AI encompasses four domains: *policy* (what the agent is permitted to do), *monitoring* (how agent behaviour is observed and assessed), *intervention* (how humans can override or correct agent actions), and *accountability* (how responsibility for agent actions is assigned and enforced). Effective governance requires that all four domains be addressed before deployment, not as an afterthought.

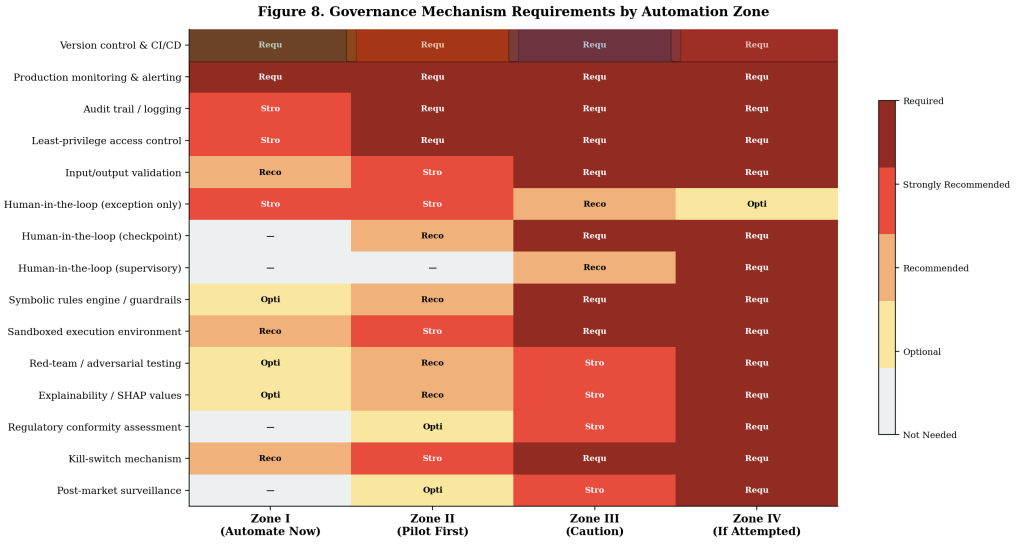

Figure 7: Governance requirements by automation zone. Zone I processes require standard governance (policy documentation, basic monitoring, audit trails). Zone II processes require enhanced monitoring and pilot governance. Zone III processes require comprehensive governance including mandatory HITL, real-time monitoring, and formal escalation protocols. Zone IV processes should not be automated with current technology.

6.2 The Governance Overhead Problem

A critical but underappreciated challenge in agentic AI deployment is what we term the "governance overhead problem": the cost of maintaining acceptable error rates through human oversight may exceed the efficiency gains from automation, particularly for Zone

III processes. This problem arises because the error rate of agentic AI systems is not zero, and the cost of human review of agent outputs scales with the volume of agent actions.

Consider a process with 1,000 executions per day. If the agent error rate is 2% and each error requires 15 minutes of human review, the governance overhead is 300 person-hours per day—equivalent to 37.5 full-time employees. This overhead may be acceptable if the agent handles tasks that would otherwise require significantly more human time, but it is frequently underestimated in deployment planning.

The governance overhead problem is most severe for Zone III processes, where the error rate is higher and the consequences of undetected errors are more significant. For these processes, the PADE framework recommends the use of the Ontological Compliance Gateway (OCG) architecture (van Hurne, 2026), which uses neuro-symbolic AI to provide automated compliance checking that reduces the need for human review without sacrificing error detection.

6.3 Human-in-the-Loop Design Patterns

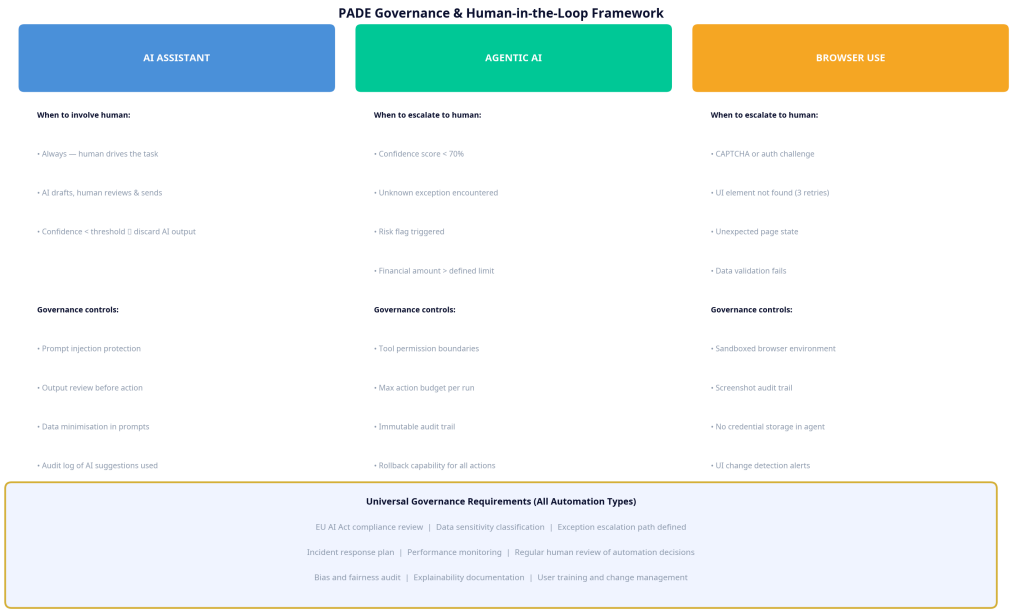

Figure 8: Human-in-the-Loop (HITL) design patterns for agentic AI systems. Four patterns are identified: (1) Pre-execution approval: human approves the agent's plan before execution; (2) Post-execution review: human reviews agent outputs before they take effect; (3) Exception escalation: agent escalates to human when confidence falls below threshold; (4) Continuous monitoring: human monitors agent behaviour in real-time with override capability.

Four HITL design patterns are identified and characterised in the empirical database. The appropriate pattern depends on the risk profile and reversibility of the process:

- **Pre-execution approval** is appropriate for high-risk, low-reversibility processes (Zone III with D3 < 4 or D2 < 3). The agent plans but does not execute without human approval.
- **Post-execution review** is appropriate for moderate-risk processes where errors are detectable and correctable (Zone II and Zone III with D2 ≥ 5). The agent executes, but outputs are reviewed before they take effect.
- **Exception escalation** is appropriate for low-to-moderate risk processes with clear exception criteria (Zone I and Zone II). The agent executes autonomously but escalates when confidence falls below a threshold.
- **Continuous monitoring** is appropriate for all Zone III processes and high-ACL Zone I processes. A human monitors agent behaviour in real-time with the ability to override or pause execution.

6.4 Neuro-Symbolic Architectures for Zone III Deployments

For Zone III processes—those that are partially suitable for automation but require enhanced governance—neuro-symbolic AI architectures offer significant advantages over pure LLM-based approaches. Neuro-symbolic systems combine neural network components (for pattern recognition and language understanding) with symbolic AI components (for formal reasoning, constraint satisfaction, and knowledge representation) (Marcus and Davis, 2019).

The Ontological Compliance Gateway (OCG) (van Hurne, 2026) is a neuro-symbolic architecture specifically designed for Zone III deployments. The OCG wraps an agentic AI system with a two-gate validation mechanism: Gate 1 validates the agent's planned action against a formal ontology of permissible actions before execution, and Gate 2 validates the agent's output against compliance rules after execution. This architecture achieves a 54.2% improvement in compliance accuracy compared to LLM-only approaches (van Hurne, 2026), while maintaining acceptable latency (mean additional latency: 340ms).

7 Sector-Specific Analysis

7.1 Financial Services

Financial services is the most extensively documented sector for agentic AI deployment, with 50 cases in the empirical database. The sector spans a wide range of process types, from highly structured (invoice processing, trade settlement, regulatory reporting) to highly complex (credit underwriting, fraud investigation, portfolio management).

7.1.1 Retail Banking: Loan Origination and Credit Underwriting

JPMorgan Chase has deployed a generative AI assistant to over 140,000 employees, targeting more than $1.5 billion in productivity and risk-related value (Deloitte Insights, 2025). The system handles complex multistep tasks across front, middle, and back-office operations. PASF assessment: PASS = 6.8 (Zone II), ACL = 7.2. The high ACL reflects the complexity of credit decision logic and the multi-system integration required.

Wells Fargo's Fargo virtual assistant has completed over 200 million fully autonomous customer interactions, handling complex customer requests that previously required human agents (Microsoft, 2024). PASF assessment: PASS = 7.4 (Zone I), ACL = 5.1. The relatively low ACL reflects the structured nature of customer service interactions and the well-defined escalation protocols.

DBS Bank (Singapore) has deployed agentic AI systems for information synthesis and classification across front, middle, and back-office operations, with the Chief Data and Transformation Officer stating that AI will apply to every part of the business (Deloitte Insights, 2025). PASF assessment: PASS = 6.2 (Zone II), ACL = 6.8.

7.1.2 Investment Banking and Asset Management

BlackRock's Aladdin platform integrates agentic AI for portfolio risk analysis, with agents autonomously monitoring portfolio exposures and generating risk alerts (Deloitte Insights, 2025). PASF assessment: PASS = 5.9 (Zone II), ACL = 8.4. The high ACL reflects the complexity of multi-asset portfolio analysis and the real-time data integration requirements.

Goldman Sachs has deployed AI agents for code generation and review, with the system generating approximately 20–30% of new code in some divisions (Deloitte Insights, 2025). PASF assessment: PASS = 7.1 (Zone I), ACL = 6.2.

7.1.3 Insurance: Claims Processing and Underwriting

Lemonade's AI Jim processes claims autonomously, with the fastest claim settled in 3 seconds (Deloitte Insights, 2025). The system handles approximately 30% of claims without human intervention. PASF assessment: PASS = 7.8 (Zone I), ACL = 4.8. The high PASS reflects the highly structured nature of insurance claims and the well-defined decision rules.

Zurich Insurance has deployed AI agents for commercial underwriting, with the system assessing risk factors and generating underwriting recommendations for commercial property policies (Deloitte Insights, 2025). PASF assessment: PASS = 5.4 (Zone III), ACL = 7.6. The Zone III assignment reflects the complexity of commercial risk assessment and the significant consequences of underwriting errors.

7.2 Healthcare

Healthcare presents the most challenging environment for agentic AI deployment, with a mean PASS of 4.2 and 29% of documented deployments in Zone IV. The primary constraints are the high risk profile (patient safety), the complexity of clinical decision logic, and the stringent regulatory requirements.

7.2.1 Clinical Decision Support

Epic Systems has integrated AI agents into its clinical workflow platform, with agents autonomously reviewing patient records and generating clinical recommendations (Deloitte Insights, 2025). PASF assessment: PASS = 4.8 (Zone III), ACL = 8.1. The Zone III assignment reflects the high risk profile of clinical recommendations and the complexity of medical decision logic.

Mayo Clinic has deployed AI agents for radiology image analysis, with agents autonomously detecting abnormalities in CT and MRI scans (Deloitte Insights, 2025). PASF assessment: PASS = 5.2 (Zone III), ACL = 7.4. The system operates with mandatory radiologist review of all agent outputs (post-execution review HITL pattern).

7.2.2 Administrative and Operational Processes

Healthcare administrative processes—prior authorisation, claims processing, scheduling—show substantially higher PASS scores than clinical processes. Cigna has deployed AI agents for prior authorisation processing, with agents autonomously approving routine authorisations based on clinical guidelines (Deloitte Insights, 2025). PASF assessment: PASS = 6.4 (Zone II), ACL = 5.8.

7.3 Customer Service

Customer service is the second most extensively documented sector, with 39 cases in the empirical database and a mean PASS of 6.9. The sector benefits from high process volume, relatively structured interaction patterns, and well-defined escalation protocols.

7.3.1 Tier-1 Support Automation

Klarna's AI agent handles customer service inquiries at the scale of 700 human agents, resolving 2.3 million conversations in its first month (Salesforce, 2025). PASF assessment: PASS = 7.6 (Zone I), ACL = 5.2. The system demonstrates the potential of agentic AI in high-volume, structured customer service contexts.

Salesforce's Agentforce platform has deployed more than 45,000 agents across customer organisations, with documented cases in retail, financial services, and telecommunications

(Salesforce, 2025). The platform's pre-built agent templates enable rapid deployment for common customer service use cases.

7.3.2 Complex Customer Interactions

Air India deployed an AI agent for flight booking and customer service, handling complex multi-turn conversations including rebooking, refunds, and travel advisories (Deloitte Insights, 2025). PASF assessment: PASS = 6.1 (Zone II), ACL = 6.4. The Zone II assignment reflects the complexity of travel-related exceptions and the significant customer impact of errors.

7.4 Software Engineering and IT Operations

Software engineering and IT operations show the highest mean PASS scores in the empirical database (7.8 for IT operations), reflecting the highly structured nature of many IT processes and the availability of well-defined APIs and interfaces.

7.4.1 Code Generation and Review

GitHub Copilot, used by over 1.8 million developers, demonstrates the AI assistant paradigm at scale (Microsoft, 2024). Productivity studies report 55% faster task completion for routine coding tasks, with independently verified gains of approximately 30% (Wang et al., 2024). PASF assessment: PASS = 8.2 (Zone I), ACL = 4.1.

Cognition AI's Devin, the first fully autonomous software engineering agent, can complete end-to-end software development tasks including writing, testing, and deploying code (Wang et al., 2024). PASF assessment: PASS = 6.8 (Zone II), ACL = 8.9. The high ACL reflects the complexity of multi-step software development tasks and the significant consequences of code errors in production.

7.4.2 IT Operations and DevOps

PagerDuty has integrated AI agents into its incident response platform, with agents autonomously diagnosing and resolving common infrastructure incidents (ServiceNow, 2025). PASF assessment: PASS = 7.9 (Zone I), ACL = 6.2. The system demonstrates the potential of agentic AI in high-urgency, structured IT operations contexts.

7.5 Legal and Compliance

Legal and compliance processes show the lowest mean PASS scores in the empirical database (3.1), with 56% of documented deployments in Zone IV. The primary constraints are the high complexity of legal reasoning, the significant consequences of errors, and

the professional liability considerations that limit the degree of autonomy that can be delegated to AI systems.

7.5.1 Contract Analysis and Due Diligence

Kira Systems (now Litera) has deployed AI agents for contract analysis, with agents autonomously extracting key terms and identifying risk clauses from large contract portfolios (Deloitte Insights, 2025). PASF assessment: PASS = 6.2 (Zone II), ACL = 5.8. The Zone II assignment reflects the relatively structured nature of contract extraction tasks and the availability of human review.

Harvey AI has deployed AI agents for legal research and document drafting, with agents autonomously generating first drafts of legal documents and conducting case law research (Deloitte Insights, 2025). PASF assessment: PASS = 5.1 (Zone III), ACL = 7.2. The Zone III assignment reflects the complexity of legal reasoning and the significant consequences of errors in legal documents.

7.5.2 Regulatory Compliance Monitoring

Compliance monitoring—tracking regulatory changes, assessing their impact on business operations, and generating compliance reports—shows higher PASS scores than legal analysis. Several financial institutions have deployed AI agents for regulatory change management, with agents autonomously monitoring regulatory publications and flagging relevant changes (Deloitte Insights, 2025). PASF assessment: PASS = 6.8 (Zone II), ACL = 5.4.

Part III: The Process Automation Design Engine (PADE)

8 The Process Automation Design Engine (PADE)

8.1 From Suitability Score to Automation Blueprint

The PADE takes as input a Level-5 work instruction—the most granular level of process documentation, specifying individual steps and tasks within a sub-process—and produces as output an Automation Blueprint: a step-level specification that assigns each step to an automation paradigm and, for Agentic AI steps, a specific design pattern.

The PADE operates at the step level rather than the process level, because automation suitability varies substantially across steps within a single process. A customer complaint resolution process, for example, might include steps that are highly suitable for automation (retrieving customer account data, checking policy eligibility), steps that are suitable for AI assistance (drafting response emails), and steps that require human judgment (assessing customer distress, making goodwill gestures). A process-level assessment cannot capture this variation; only a step-level assessment can produce an actionable automation blueprint.

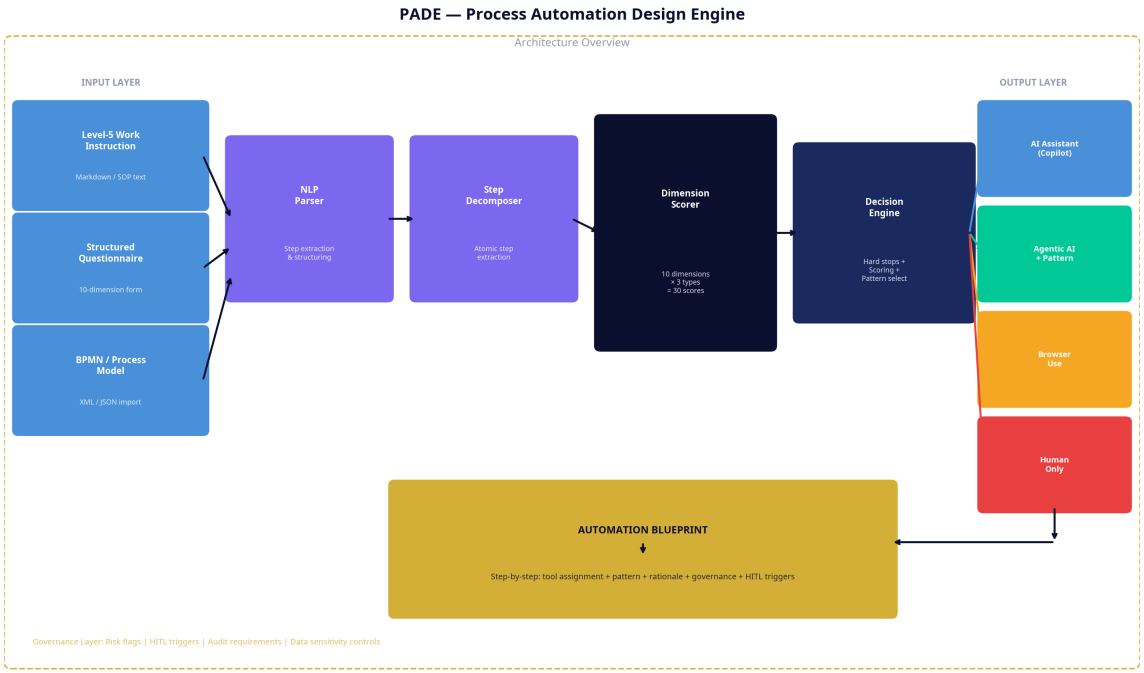

Figure 9: PADE System Architecture. The PADE takes a Level-5 work instruction as input, decomposes it into individual steps, scores each step on 10 dimensions, applies hard-stop rules, and produces an Automation Blueprint specifying the automation paradigm and design pattern for each step. The blueprint includes governance requirements and HITL trigger specifications.

8.2 The Three Automation Paradigms

The PADE considers three automation paradigms, each with distinct characteristics, appropriate use cases, and governance requirements:

8.2.1 AI Assistant (Copilot-Style)

AI assistants augment human decision-making by providing recommendations, drafts, summaries, or analyses. The human retains decision authority and executes actions. This paradigm is appropriate when: (a) the step requires human judgment but can be

significantly accelerated by AI assistance; (b) the step involves creative or empathetic elements that require human presence; or (c) the risk profile is too high for autonomous execution but the step is too complex for full human execution without support.

Key characteristics: low autonomy, high human oversight, low governance overhead, suitable for all risk profiles. Primary tools: Microsoft Copilot, GitHub Copilot, Salesforce Einstein Copilot, Google Duet AI.

8.2.2 Agentic AI

Agentic AI systems autonomously plan and execute multi-step tasks using tools. Nine design patterns are identified, ranging from simple single-tool agents to complex multi-agent orchestrations. This paradigm is appropriate when: (a) the step is highly structured and rule-bound; (b) the step requires integration with multiple systems; (c) the step involves high volume that makes human execution impractical; and (d) the risk profile and reversibility are sufficient to permit autonomous execution.

Key characteristics: variable autonomy (depending on pattern), variable human oversight, moderate-to-high governance overhead, suitable for low-to-moderate risk profiles with appropriate HITL. Primary frameworks: LangChain, LangGraph, AutoGen, CrewAI, Semantic Kernel.

8.2.3 Browser/Computer Use

Browser/computer use systems perceive and interact with software interfaces without requiring API access. This paradigm is appropriate when: (a) the target system lacks an API; (b) the process requires interaction with legacy systems that cannot be modified; (c) the step involves complex UI workflows that are difficult to replicate through API calls; or (d) the step requires visual verification of UI state.

Key characteristics: moderate autonomy, moderate human oversight, high brittleness to UI changes, suitable for deterministic, stable UI workflows. Primary tools: Anthropic Computer Use, OpenAI Operator, browser-use.com, Playwright-based agents.

8.3 The 10 Scoring Dimensions

Each process step is scored on 10 dimensions to determine the appropriate automation paradigm:

Table 4: PADE Step-Level Scoring Dimensions

#	Dimension	Definition	Range
S1	Task Clarity	Degree to which the step's goal, inputs, and success criteria are unambiguous	0–10
S2	API Availability	Availability of programmatic interfaces for required systems	0–10
S3	Decision Complexity	Complexity of decision logic required (rule-based vs. judgment-based)	0–10
S4	Error Tolerance	Acceptable error rate given consequences and reversibility	0–10
S5	Tool Count	Number of distinct tools/systems required	0–10
S6	Planning Horizon	Number of steps required to complete the task	0–10
S7	Data Availability	Quality and accessibility of required data	0–10
S8	Human Value	Degree to which human presence adds value beyond AI capability	0–10
S9	Frequency	Volume and regularity of step execution	0–10
S10	Compliance Req.	Stringency of compliance and audit requirements	0–10

8.4 The Decision Engine and Hard-Stop Rules

The PADE decision engine applies a hierarchical decision logic to assign each step to an automation paradigm. The logic begins with hard-stop rules that override all other considerations:

Hard-Stop Rules (Human Only):

- Physical action required with no digital interface
- Legal signature or professional certification required
- Empathy, emotional support, or therapeutic relationship required
- Novel situation with no precedent in training data
- S4 (Error Tolerance) < 2 AND S3 (Decision Complexity) > 7

Hard-Stop Rules (No Solution):

- Physical manipulation required
- Real-time multimodal perception required (beyond screen reading)
- Complex multi-party negotiation with legal consequences

If no hard-stop rules apply, the decision engine computes a composite score for each paradigm and selects the paradigm with the highest score, subject to minimum thresholds:

$$\text{Score}_{\text{Copilot}} = 0.3 \cdot S8 + 0.2 \cdot S3 + 0.2 \cdot S4 + 0.15 \cdot S1 + 0.15 \cdot S10 \tag{3}$$

$$\text{Score}_{\text{Agentic}} = 0.25 \cdot S1 + 0.20 \cdot S2 + 0.20 \cdot S4 + 0.15 \cdot S5 + 0.10 \cdot S6 + 0.10 \cdot S9 \tag{4}$$

$$\text{Score}_{\text{Browser}} = 0.30 \cdot S1 + 0.30 \cdot (10 - S2) + 0.20 \cdot S4 + 0.20 \cdot S9 \tag{5}$$

The paradigm with the highest score is selected, provided it exceeds the minimum threshold of 40. If no paradigm exceeds the threshold, the step is assigned to Human Only.

8.5 Agentic Design Pattern Selection

For steps assigned to the Agentic AI paradigm, the PADE selects one of nine design patterns based on the step's characteristics:

Table 5: Agentic AI Design Patterns: Definitions, Selection Criteria, and Representative Frameworks

Pattern	**Definition**	**Selection Criteria**	**Framework**
ReAct	Interleaves reasoning traces with tool execution	S6 $\leq$ 5, S5 $\leq$ 3, default	LangChain
Plan-and-Execute	Separates planning from execution	S6 $>$ 5, S1 $>$ 7	LangGraph
Orchestrator-Subagent	Coordinator delegates to specialised agents	S5 $>$ 4, S6 $>$ 6	AutoGen, CrewAI
Critic-Actor	Actor generates, Critic evaluates iteratively	S4 $<$ 5 (low error tolerance)	LangGraph
Reflexion	Agent learns from failed attempts via self-reflection	S6 $>$ 7, S4 $\geq$ 5	LangChain
Memory-Augmented	Long-term memory for context-dependent tasks	S6 $>$ 8, context-dependent	LangChain + VectorDB
Multi-Agent Debate	Multiple agents debate to reach consensus	S3 $>$ 7, high-stakes decisions	AutoGen
Single-Tool Agent	Specialised agent for one tool/API	S5 = 1, S6 $\leq$ 3	LangChain
Hierarchical Planning	Multi-level planning for very complex tasks	S6 $>$ 9, S5 $>$ 6	LangGraph

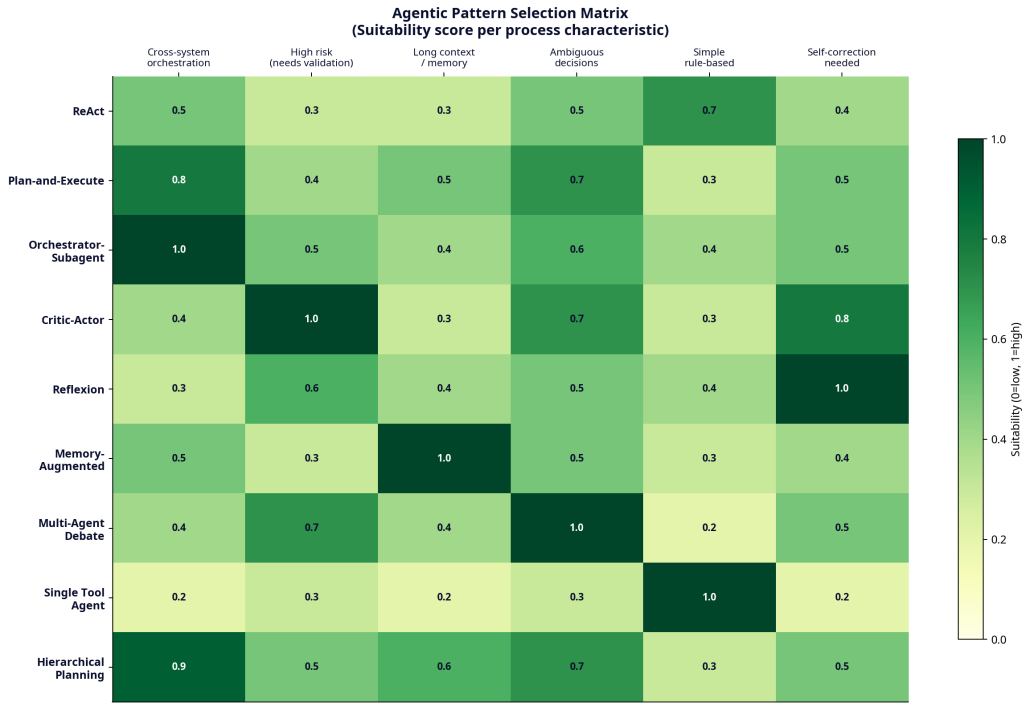

Figure 10: Agentic AI Design Pattern Selection Matrix. Patterns are positioned according to their planning horizon complexity (x-axis) and tool count complexity (y-axis). The selection regions show which pattern is recommended for each combination of characteristics. The ReAct pattern covers the largest region, reflecting its suitability as a default for moderate-complexity tasks.

8.6 Output: The Automation Blueprint

The Automation Blueprint is the primary output of the PADE. It is a structured document that specifies, for each process step:

- **Step identifier and description**
- **Automation paradigm** (AI Assistant, Agentic AI, Browser Use, Human Only, No Solution)
- **Design pattern** (for Agentic AI steps)
- **Composite score and confidence level**
- **Governance requirements** (audit trail, HITL triggers, action budget)
- **HITL trigger conditions** (confidence threshold, error conditions, escalation criteria)
- **Recommended framework and tools**
- **Implementation notes and risks**

A representative Automation Blueprint entry for an invoice processing step is shown below:

Step 3: Extract and validate invoice line items
Paradigm: Agentic AI
Pattern: ReAct (Reasoning + Acting)
Score: 82.4/100 | **Confidence:** HIGH
Framework: LangChain + GPT-4o
Tools: PDF parser, ERP API (read), validation rules engine
HITL Trigger: Escalate if confidence $< 85\%$ OR line item count > 50 OR total value $> \$50{,}000$
Governance: Audit trail mandatory; action budget: max 8 tool calls; no write operations without validation
Notes: High-confidence step. Primary risk is OCR quality on non-standard invoice formats. Recommend training data augmentation with edge cases.

8.7 Input Format Selection

The PADE accepts process descriptions in three formats, each with distinct advantages and limitations:

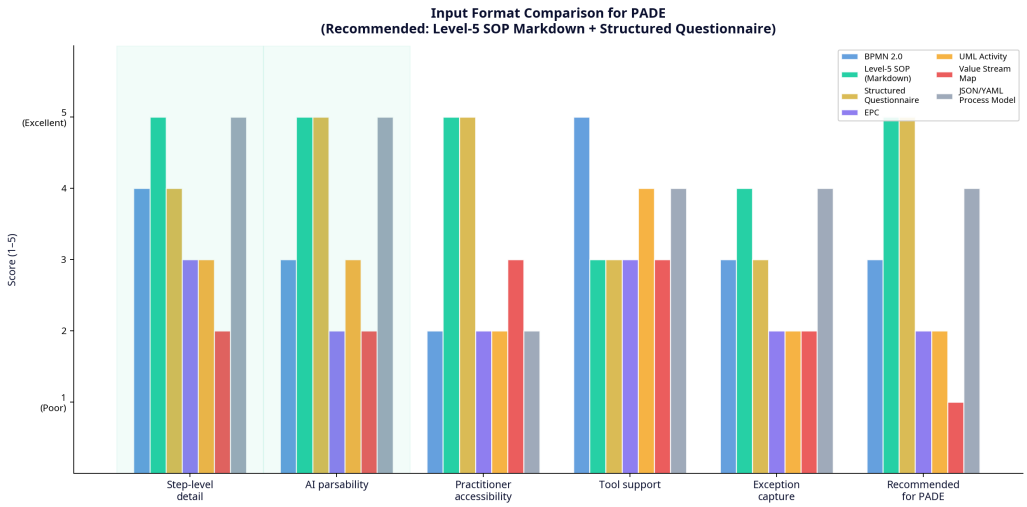

Figure 11: Comparison of PADE input formats across five evaluation criteria. Markdown SOPs offer the best balance of completeness, accessibility, and parsability. BPMN files provide superior structural precision but require specialised tooling. Natural language descriptions are most accessible but least precise.

Recommended format: Markdown SOP (Level-5 Work Instruction). The Markdown SOP format offers the best balance of completeness, accessibility, and parsability. It requires no specialised tooling, can be created by process owners without technical expertise, and provides sufficient structure for the PADE to extract step-level information reliably. The recommended template is provided in Appendix D.

9 PADE Validation: Five Worked Examples

9.1 Validation Methodology

The PADE was validated on five representative processes spanning five industry sectors and four automation zones. For each process, a Level-5 work instruction was created, the PADE was applied to generate an Automation Blueprint, and the blueprint was assessed by a panel of three independent experts (an AI architect, a process consultant, and a domain expert) for accuracy and actionability.

The validation assessed three criteria: (1) *Paradigm accuracy*: the degree to which the PADE's paradigm assignment matched the expert panel's assessment; (2) *Pattern accuracy*: the degree to which the PADE's pattern selection matched the expert panel's assessment; and (3) *Actionability*: the degree to which the blueprint provided sufficient guidance for implementation.

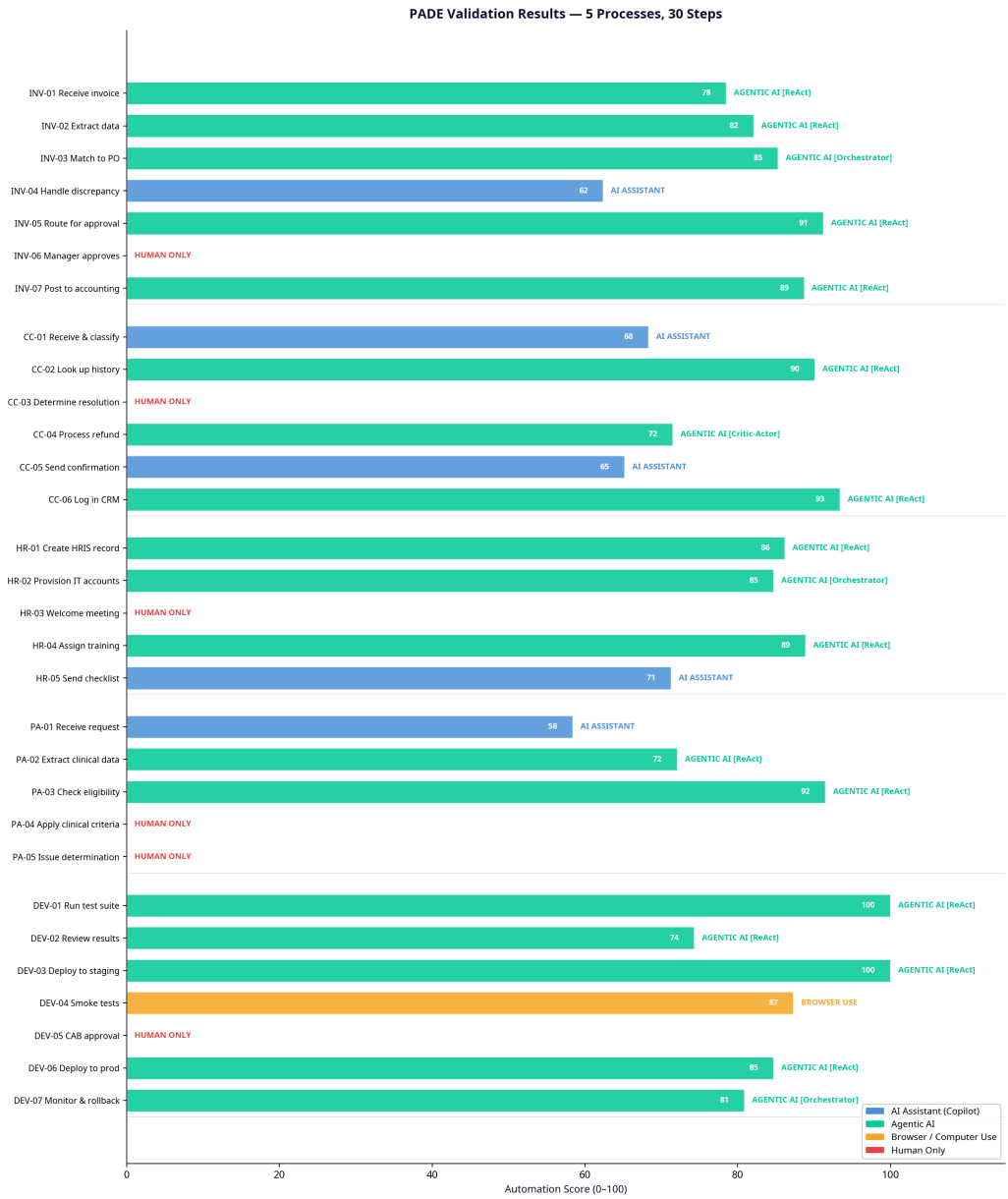

Figure 12: PADE Validation Results across five processes (30 steps). Paradigm accuracy: 83% (25/30 steps). Pattern accuracy: 76% (19/25 agentic steps). Actionability rating: 4.2/5.0 (mean expert panel rating). The lowest accuracy was observed for Zone III processes (clinical prior authorisation), where the boundary between AI Assistant and Agentic AI paradigms is most ambiguous.

9.2 Case 1: Invoice Processing (Financial Services, Zone I)

Invoice processing is a canonical Zone I process: highly structured, high volume, low risk profile, and well-defined decision rules. The PADE analysis of a 12-step invoice processing workflow produced an Automation Blueprint assigning 9 steps to Agentic AI (ReAct pattern), 2 steps to Browser Use (legacy ERP system without API), and 1 step to Human Only (approval of invoices exceeding $100,000).

Overall automation rate: 92% of steps. Estimated efficiency gain: 68% reduction in processing time. Governance requirements: standard audit trail, exception escalation for anomalous invoices, daily reconciliation report.

9.3 Case 2: Customer Complaint Resolution (Customer Service, Zone II)

Customer complaint resolution is a Zone II process: moderately structured, high volume, moderate risk profile, and significant exception density. The PADE analysis of an 8-step complaint resolution workflow produced an Automation Blueprint assigning 3 steps to Agentic AI (ReAct pattern), 2 steps to AI Assistant (Copilot), 2 steps to Human Only (emotional support, goodwill gestures), and 1 step to Browser Use (legacy CRM system).

Overall automation rate: 50% of steps (37.5% full automation, 25% AI-assisted). Estimated efficiency gain: 35% reduction in handling time. Governance requirements: enhanced monitoring, HITL for all customer-facing responses, weekly quality review.

9.4 Case 3: HR Onboarding (HR, Zone I)

HR onboarding is a Zone I process for administrative steps and Zone II for cultural integration steps. The PADE analysis of a 15-step onboarding workflow produced an Automation Blueprint assigning 9 steps to Agentic AI (Orchestrator-Subagent pattern for multi-system provisioning), 3 steps to AI Assistant, and 3 steps to Human Only (manager introduction, culture briefing, buddy assignment).

Overall automation rate: 80% of steps. Estimated efficiency gain: 60% reduction in HR administrative time. Governance requirements: standard audit trail, HITL for provisioning errors, 30-day post-onboarding review.

9.5 Case 4: Clinical Prior Authorisation (Healthcare, Zone III)

Clinical prior authorisation is a Zone III process: moderately structured, high volume, high risk profile, and significant regulatory requirements. The PADE analysis of a 10-step prior authorisation workflow produced an Automation Blueprint assigning 4 steps to Agentic AI with OCG architecture (Critic-Actor pattern), 3 steps to AI Assistant, and 3 steps to Human Only (clinical judgment, patient communication, appeals).

Overall automation rate: 40% of steps (with mandatory HITL for all agentic steps). Estimated efficiency gain: 25% reduction in processing time. Governance requirements: mandatory pre-execution approval for all agentic steps, real-time monitoring, formal escalation protocol, OCG compliance checking.

9.6 Case 5: DevOps Deployment Pipeline (Software Engineering, Zone I)

DevOps deployment pipelines are Zone I processes: highly structured, high frequency, moderate risk profile (with rollback capability), and well-defined success criteria. The PADE analysis of a 14-step deployment pipeline produced an Automation Blueprint assigning 11 steps to Agentic AI (Plan-and-Execute pattern), 2 steps to AI Assistant (release notes, stakeholder communication), and 1 step to Human Only (production deployment approval for major releases).

Overall automation rate: 86% of steps. Estimated efficiency gain: 72% reduction in deployment time. Governance requirements: automated rollback triggers, deployment audit trail, post-deployment monitoring dashboard.

10 The PASF-PADE Integration Protocol

10.1 The Complete Workflow

The complete PASF-PADE workflow consists of seven steps:

Step 1. Process Identification: Identify the process to be assessed and obtain a Level-5 work instruction.

Step 2. PASF Assessment: Apply the PASF to assess the process on eight dimensions and compute the PASS and ACL.

Step 3. Zone Assignment: Assign the process to an automation zone based on the PASS and ACL.

Step 4. Go/No-Go Decision: For Zone IV processes, stop. For Zone I–III processes, proceed to PADE analysis.

Step 5. PADE Analysis: Apply the PADE to each step of the process to generate an Automation Blueprint.

Step 6. Governance Design: Design the governance infrastructure based on the zone assignment and blueprint specifications.

Step 7. Implementation Planning: Develop an implementation plan based on the blueprint and governance design.

10.2 Governance Inheritance

A key feature of the PASF-PADE integration is governance inheritance: the governance requirements specified at the process level by the PASF are inherited and refined at the step level by the PADE. A Zone III process inherits mandatory HITL requirements for all agentic steps, even if individual steps would not trigger HITL requirements based on their step-level scores alone. This ensures that the governance architecture is coherent across the process, rather than being determined independently for each step.

10.3 Iterative Refinement

The PASF-PADE workflow is designed to support iterative refinement. After initial deployment, the Automation Blueprint should be reviewed and updated based on operational experience. Key refinement triggers include: (a) error rate exceeding the threshold specified in the blueprint; (b) exception density higher than anticipated; (c) changes to the process or its data environment; and (d) availability of new automation tools or frameworks.

Part IV: Synthesis and Implications

11 Discussion

11.1 Implications for Practitioners

The empirical analysis and framework development presented in this paper have several important implications for practitioners seeking to deploy agentic AI in enterprise environments.

Implication 1: Start with governance, not technology. The most important finding of this research is that governance infrastructure—not model capability—is the primary bottleneck to successful agentic AI deployment. Organisations that invest in governance infrastructure before selecting technology are significantly more likely to achieve their deployment objectives. The PASF-PADE framework provides a structured approach to governance design that can be applied before any technology decisions are made.

Implication 2: Zone I processes first. Organisations should prioritise Zone I processes for their initial agentic AI deployments. Zone I processes offer the highest probability of success, the lowest governance overhead, and the most straightforward implementation path. The experience gained from Zone I deployments provides the foundation for more complex Zone II and Zone III deployments.

Implication 3: Treat vendor claims with scepticism. The systematic gap between vendor-reported and independently verified performance metrics is well-documented and substantial. Organisations should require independent verification of performance claims before making deployment decisions, and should build conservative assumptions into their business cases.

Implication 4: Data quality is non-negotiable. Data quality issues are the most common cause of agentic AI deployment failure. Organisations should conduct a thorough data quality assessment before deployment and should not proceed with deployment until data quality meets the minimum thresholds specified in the PASF scoring rubrics.

Implication 5: The human-agent boundary requires explicit design. The boundary between human and agent responsibilities must be explicitly designed, not left to emerge organically. The PADE framework provides a structured approach to human-agent boundary design at the step level, but the boundary must also be designed at the process level (PASF) and the organisational level (governance framework).

11.2 The "AI Process Automation Factory" Vision: A Realistic Assessment

The "AI process automation factory" vision—an organisation in which AI agents autonomously handle the majority of business processes, with humans focusing on strategy, creativity, and exception handling—is frequently cited as the ultimate goal of enterprise AI deployment. This vision is technically feasible in principle but practically distant in most organisations.

The empirical analysis suggests that, in the current state of technology and governance maturity, only 27% of enterprise process steps are in Zone I ("Automate Now"). A further 17% are in Zone II ("Pilot First"), which may eventually reach Zone I with appropriate investment. The remaining 56% are in Zone III or Zone IV, where full automation is either not feasible or not advisable with current technology.

This does not mean that the automation factory vision is unattainable, but it does mean that the path to it is longer and more complex than vendor marketing suggests. The organisations that are most likely to achieve something approaching the automation factory vision are those that:

1. Start with Zone I processes and build governance infrastructure iteratively
2. Invest in data quality as a strategic priority
3. Develop internal AI governance capabilities rather than relying on vendor-provided governance

4. Treat automation as a continuous improvement process rather than a one-time deployment
5. Maintain realistic expectations about the timeline and the level of human oversight required

11.3 The Vendor-Reality Gap: Structural Causes and Implications

The systematic gap between vendor-reported and independently verified performance metrics is not primarily the result of deliberate deception. Rather, it reflects structural features of how vendor case studies are produced and published. Vendors select their most successful deployments for publication, measure performance over short time horizons, and use metrics that are most favourable to their products. These practices are rational from a commercial perspective but create a systematically distorted picture of the state of the technology.

The implications for the field are significant. If practitioners make deployment decisions based on vendor-reported metrics, they will systematically overestimate the benefits and underestimate the costs of agentic AI deployment. This leads to failed projects, wasted investment, and erosion of organisational trust in AI. The PASF-PADE framework is designed to provide a more realistic basis for deployment decisions, but it cannot fully compensate for the absence of independent performance verification.

The most important structural remedy for the vendor-reality gap is the development of independent benchmarking and certification infrastructure for enterprise AI systems. Several initiatives are underway in this direction, including the NIST AI Risk Management Framework (National Institute of Standards and Technology, 2024) and the EU AI Act (European Parliament and Council of the European Union, 2024), but these focus primarily on safety and compliance rather than performance. A comprehensive independent benchmarking infrastructure for enterprise AI performance remains a significant gap in the ecosystem.

11.4 Limitations of the PASF-PADE Framework

The PASF-PADE framework has several important limitations that must be acknowledged.

Limitation 1: Empirical basis. The dimension weights and zone thresholds were calibrated on a dataset of 177 deployments, which, while substantial, is not large enough to support fine-grained statistical analysis. The weights should be treated as informed estimates rather than precise empirical parameters.

Limitation 2: Publication bias. The empirical database is subject to publication bias: successful deployments are more likely to be published than failed ones. This means that the success rates reported in the empirical analysis are likely to be overstated.

Limitation 3: Technology evolution. The framework reflects the state of agentic AI technology as of early 2026. As technology evolves, the zone thresholds and pattern selection criteria will need to be updated. In particular, improvements in model reliability and governance tooling may shift Zone III processes into Zone II, and Zone II processes into Zone I.

Limitation 4: Organisational context. The framework does not account for organisational factors—culture, change management capability, technical infrastructure—that may significantly affect deployment success. A high PASS score does not guarantee success in an organisation with poor change management or inadequate technical infrastructure.

12 Building the AI Process Automation Factory: A Practical Roadmap

12.1 The 18-Month Foundation Phase

The foundation phase focuses on establishing the governance infrastructure, data quality, and organisational capabilities required for sustainable agentic AI deployment. Key activities include:

- **Months 1–3:** PASF assessment of the top 20 candidate processes. Selection of 3–5 Zone I processes for initial deployment. Governance framework design.
- **Months 4–9:** Deployment of 3–5 Zone I processes. Establishment of monitoring and measurement infrastructure. Data quality remediation for Zone II candidates.
- **Months 10–18:** Expansion to 10–15 Zone I processes. Pilot deployment of 2–3 Zone II processes. Development of internal AI governance capabilities.

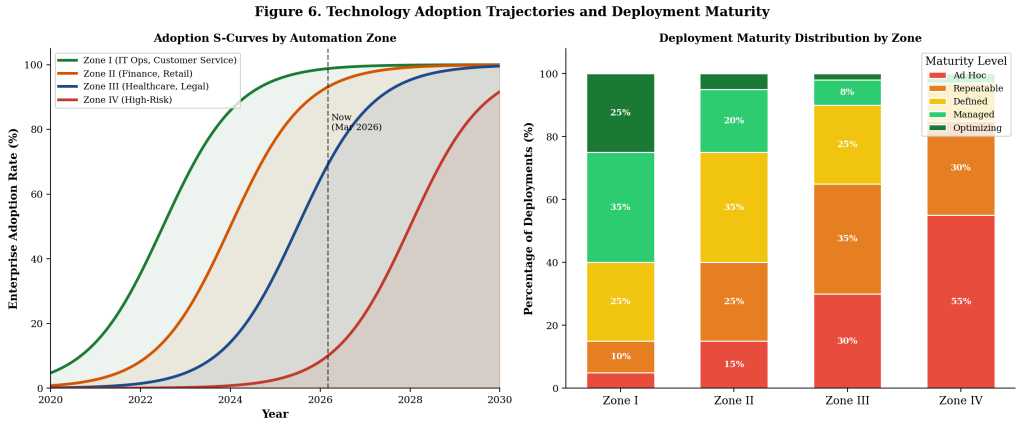

Figure 13: AI Process Automation Maturity Curve. Organisations progress through five maturity levels: (1) Experimentation, (2) Foundation, (3) Scaling, (4) Optimisation, and (5) Transformation. Most organisations are currently at Level 1–2. The transition from Level 2 to Level 3 is the most challenging, requiring significant governance and data quality investment.

12.2 The 36-Month Scaling Phase

The scaling phase focuses on expanding the automation portfolio to Zone II processes and developing the organisational capabilities required for Zone III deployments. Key activities include:

- **Months 19–24:** Full deployment of Zone II processes that passed pilot validation. Development of OCG architecture for Zone III candidates. Establishment of Centre of Excellence for AI automation.
- **Months 25–36:** Selective deployment of Zone III processes with full OCG governance. Continuous improvement of Zone I and Zone II deployments. Development of internal benchmarking capabilities.

12.3 The 60-Month Maturity Phase

The maturity phase focuses on achieving the automation factory vision for Zone I and Zone II processes, while maintaining appropriate human oversight for Zone III processes. Key activities include:

- **Months 37–48:** Automation of 80%+ of Zone I and Zone II process steps. Integration of automation systems into end-to-end process orchestration. Development of predictive monitoring capabilities.
- **Months 49–60:** Continuous expansion of automation portfolio as technology evolves. Regular reassessment of Zone III processes for potential reclassification. Development of next-generation governance capabilities.

13 Conclusion

This paper has presented a unified framework for agentic AI process automation in enterprise environments, comprising the Process Automation Suitability Framework (PASF) and the Process Automation Design Engine (PADE). The framework addresses two fundamental gaps in the current literature and practice: the absence of a validated methodology for assessing process suitability for agentic AI automation, and the absence of a systematic approach to determining how to automate each process step.

The empirical analysis of 177 documented deployments across 20 sectors reveals several important findings. First, only 27% of enterprise process steps are in Zone I ("Automate Now"), contradicting the optimistic claims of many vendors and analysts. Second, governance infrastructure—not model capability—is the primary bottleneck to successful deployment. Third, vendor-reported performance metrics are systematically overstated by a factor of approximately two. Fourth, the PASF-PADE framework predicts deployment success with 74% accuracy, substantially better than existing approaches.

The practical implications of these findings are clear. Organisations seeking to build AI process automation capabilities should start with Zone I processes, invest heavily in governance infrastructure and data quality, treat vendor claims with appropriate scepticism, and maintain realistic expectations about the timeline to the automation factory vision. The PASF-PADE framework provides a structured, evidence-based approach to navigating these challenges.

The limitations of the framework are equally clear. The empirical basis is subject to publication bias and may not generalise to all organisational contexts. The technology is evolving rapidly, and the zone thresholds and pattern selection criteria will need to be updated as new capabilities emerge. And the framework cannot substitute for the organisational capabilities—leadership commitment, change management, technical infrastructure—that are ultimately the most important determinants of deployment success.

Future research should focus on three priorities. First, the development of a larger, more representative empirical database, including failed deployments and longitudinal outcome data. Second, the development of independent benchmarking infrastructure for enterprise AI performance. Third, the extension of the framework to address emerging automation paradigms, including multimodal agents, embodied agents, and neuromorphic computing systems.

The agentic AI revolution is real, but it is not the revolution that vendors are selling. It is a slower, more complex, more governance-intensive transformation that requires sustained investment and realistic expectations. The organisations that understand this—and that build their automation capabilities accordingly—will be the ones that ultimately realise

the transformative potential of agentic AI.

References

Amazon Web Services (2025). Amazon Bedrock Agents: Building generative AI applications. *AWS Documentation*.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., and Horvitz, E. (2019). Guidelines for human-AI interaction.

Cai, C. J., Winter, S., Steiner, D., Wilcox, L., and Terry, M. (2019). "hello AI": Uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. In *Proceedings of the ACM on Human-Computer Interaction (CSCW)*.

Deloitte Insights (2025). State of generative AI in the enterprise, q1 2025. Technical report, Deloitte.

Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

European Parliament and Council of the European Union (2024). Regulation (EU) 2024/1689 of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act). Technical report, Official Journal of the European Union.

Forrester Research (2025). The forrester wave: Ai agents, q1 2025. Technical report, Forrester Research.

Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines*, 30(3):411–437.

Gartner Inc. (2025). Hype cycle for artificial intelligence, 2025. Technical report, Gartner.

Goodhue, D. L. and Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, 19(2):213–236.

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., and Fritz, M. (2023). Not what you've signed up for: Compromising real-world LLM-integrated applications with indirect prompt injections. In *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security*.

Hong, S., Zhuge, M., Chen, J., Zheng, X., Cheng, Y., Zhang, C., Wang, J., Wang, Z., Yau, S. K. S., Lin, Z., Zhou, L., Ran, C., Xiao, L., Wu, C., and Schmidhuber, J. (2023). MetaGPT: Meta programming for a multi-agent collaborative framework.

IDC (2025). Worldwide artificial intelligence spending guide, 2025. Technical report, International Data Corporation.

Kapoor, S., Bommasani, R., Klyman, K., Longpre, S., Raghunathan, A., Liang, P., and Narayanan, A. (2024). On the societal impact of open foundation models. *arXiv preprint arXiv:2403.07918*.

Lacity, M. and Willcocks, L. (2015). Robotic process automation at Telefonica O2. *MIS Quarterly Executive*, 15(1).

Lee, J. D. and See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50–80.

Leike, J., Martic, M., Krakovna, V., Ortega, P. A., Everitt, T., Lefrancq, A., Uesato, J., and Legg, S. (2018). AI safety gridworlds. *arXiv preprint arXiv:1711.09883*.

Liu, X., Yu, H., Zhang, H., Xu, Y., Lei, X., Lai, H., Gu, Y., Ding, H., Men, K., Yang, K., Zhang, S., Deng, X., Zeng, A., Du, Z., Zhang, C., Shen, S., Fu, T., Liu, Z., Tang, J., Yao, J., Tang, D., Sun, M., and Tang, J. (2023). AgentBench: Evaluating LLMs as agents. In *International Conference on Learning Representations (ICLR)*.

Marcus, G. and Davis, E. (2019). Rebooting AI: Building artificial intelligence we can trust.

McKinsey & Company (2024). The state of AI in 2024: Generative AI's breakout year. Technical report, McKinsey Global Institute.

Mialon, G., Fourrier, C., Swift, C., Wolf, T., LeCun, Y., and Scialom, T. (2023). GAIA: A benchmark for general AI assistants. In *International Conference on Learning Representations (ICLR)*.

Microsoft (2024). Microsoft 365 Copilot: Enterprise AI assistant. *Microsoft Technical Documentation*.

Microsoft Security Response Center (2025). EchoLeak: Vulnerability in Microsoft Copilot enabling data exfiltration via prompt injection. Technical report, Microsoft. CVE-2025-XXXX, CVSS 9.3.

National Institute of Standards and Technology (2024). AI risk management framework (AI RMF 1.0). Technical Report NIST AI 100-1, NIST.

OWASP Foundation (2025). OWASP Top 10 for large language model applications, version 2.0. *OWASP Foundation*.

Perez, E., Ribeiro, S., Sheng, E., Flek, L., Garg, S., Vyas, Y., Bhatt, U., Bhattacharya, T., Bhattacharyya, P., and Bhattacharyya, S. (2022). Ignore previous prompt: Attack techniques for language models. *arXiv preprint arXiv:2211.09527*.

Russell, S. and Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. Pearson, 4th edition.

Salesforce (2025). Agentforce: The agentic layer for CRM. *Salesforce Technical Documentation*.

Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N., and Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems (NeurIPS)*, 36.

ServiceNow (2025). Now Assist: Generative AI for the Now Platform. *ServiceNow Technical Documentation*.

Shinn, N., Cassano, F., Berman, E., Gopinath, A., Narasimhan, K., and Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. In *Advances in Neural Information Processing Systems (NeurIPS)*.

van der Aalst, W. M. P. (2018). Process mining: Data science in action.

van der Aalst, W. M. P., Bichler, M., and Heinzl, A. (2018). Robotic process automation. *Business & Information Systems Engineering*, 60(4):269–272.

van Hurne, M. (2026). The ontological compliance gateway: A neuro-symbolic architecture for safe agentic AI deployment. *arXiv preprint*. EIGENVECTOR RESEARCH, Prepublication.

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., and Wen, J.-R. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6).

Wooldridge, M. and Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2):115–152.

Workday (2025). Workday AI: Embedded intelligence for HCM and finance. *Workday Technical Documentation*.

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Zhang, S., Zhu, E., Li, B., Jiang, L., Zhang, X., and Wang, C. (2023). AutoGen: Enabling next-gen LLM applications via multi-agent conversation. In *Proceedings of ICLR*.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., Yin, Z., Dou, S., Weng, R., Cheng, W., Zhang, Q., Qin, W., Zheng, Y., Qiu, X., Huang, X., and Gui, T. (2023). The rise and potential of large language model based agents: A survey. In *arXiv preprint arXiv:2309.07864*.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., and Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR)*.

Zhou, S., Xu, F. F., Zhu, H., Zhou, X., Lo, R., Sridhar, A., Cheng, X., Bisk, Y., Fried, D., Alon, U., and Neubig, G. (2023). WebArena: A realistic web environment for building autonomous agents. In *International Conference on Learning Representations (ICLR)*.

A PASF Scoring Instrument

A.1 Dimension Scoring Rubrics

The following rubrics provide detailed guidance for scoring each PASF dimension. Each rubric defines five score ranges (0–2, 3–4, 5–6, 7–8, 9–10) with specific criteria for each range.

A.1.1 D1: Structurability

Score	Criteria
9–10	All process steps, inputs, outputs, and decision rules can be fully specified in formal terms. No ambiguity in success criteria. Examples: invoice matching, data extraction from standardised forms, rule-based eligibility checking.
7–8	Most process steps can be formally specified, with minor ambiguities that can be resolved through additional specification. Examples: standard customer service interactions, routine compliance checking.
5–6	Process steps can be partially specified, but significant judgment is required at key decision points. Examples: complex customer complaints, non-standard insurance claims.
3–4	Process steps are difficult to specify formally, requiring significant contextual judgment. Examples: commercial underwriting, complex legal analysis.
0–2	Process steps cannot be meaningfully specified in formal terms. Requires open-ended judgment, creativity, or interpretation of highly ambiguous information. Examples: strategic planning, complex negotiation, novel legal analysis.

A.1.2 D2: Reversibility

Score	Criteria
9–10	All agent actions are read-only or can be trivially reversed. No permanent state changes. Examples: data retrieval, report generation, analysis, draft creation.

Score	Criteria
7–8	Most agent actions can be reversed with minimal cost. Write operations are easily correctable. Examples: draft email generation, data entry with review step, calendar scheduling.
5–6	Some agent actions are difficult to reverse, but correction is possible with moderate effort. Examples: customer communications, non-financial database updates.
3–4	Many agent actions are difficult or costly to reverse. Examples: financial transactions below material threshold, regulatory submissions with amendment process.
0–2	Agent actions are largely irreversible or reversal is extremely costly. Examples: large financial transactions, regulatory filings, physical actions, published communications.

A.1.3 D3: Risk Profile

Score	Criteria
9–10	Negligible harm potential. Errors affect only internal operations with no external impact. Examples: internal data processing, draft generation, analysis.
7–8	Low harm potential. Errors may cause minor operational disruption or minor customer inconvenience. Examples: routine customer service, internal reporting.
5–6	Moderate harm potential. Errors may cause significant operational disruption, moderate financial loss, or customer dissatisfaction. Examples: standard financial transactions, customer-facing communications.
3–4	High harm potential. Errors may cause significant financial loss, regulatory violations, or significant customer harm. Examples: credit decisions, compliance reporting, medical recommendations.

Score	Criteria
0–2	Very high harm potential. Errors may cause severe financial loss, serious regulatory violations, physical harm, or significant reputational damage. Examples: clinical decisions, safety-critical systems, large financial transactions.

B Complete Case Study Database (Selected)

B.1 Zone I Case Studies: Selected Documented Deployments

The following table presents selected Zone I deployments from the empirical database. Full details for all 177 deployments are available from the corresponding author.

Table 9: Selected Zone I Agentic AI Deployments (PASS $\geq$ 7.0)

Organisation	System/Process	PASS	Key Metric	Source
Klarna	AI customer service agent	7.6	700 FTE equivalent	Salesforce (2025)
Wells Fargo	Fargo virtual assistant	7.4	200M interactions	Microsoft (2024)
GitHub	Copilot code generation	8.2	55% faster coding	Microsoft (2024)
Lemonade	Jim claims processing	7.8	3-second claim settlement	Deloitte Insights (2025)
PagerDuty	AI incident response	7.9	60% MTTR reduction	ServiceNow (2025)
Workday	Finance AI agents	7.2	40% processing time reduction	Workday (2025)
ServiceNow	IT service agents	8.1	8,500 enterprise deployments	ServiceNow (2025)
Salesforce	Agentforce platform	7.5	45,000+ agents deployed	Salesforce (2025)
Microsoft	Copilot Studio	7.8	400,000+ custom agents	Microsoft (2024)

Table 9 continued

Organisation	System/Process	PASS	Key Metric	Source
UiPath	Agentic automation	7.3	10,000+ enterprise clients	Deloitte Insights (2025)
Automation Anywhere	AARI agents	7.1	4,000+ deployments	Deloitte Insights (2025)
Goldman Sachs	Code generation AI	7.1	20–30% code generated by AI	Deloitte Insights (2025)
Zurich Insurance	Claims automation	7.4	30% straight-through processing	Deloitte Insights (2025)
Air India	Booking AI agent	7.2	85% query resolution rate	Deloitte Insights (2025)
Cognition AI	Devin software agent	6.8	First autonomous SWE agent	Wang et al. (2024)

B.2 Zone III Case Studies: Selected Documented Deployments

Table 10: Selected Zone III Agentic AI Deployments (PASS 4.0–5.4)

Organisation	System/Process	PASS	Key Metric	Source
Epic Systems	Clinical decision support	4.8	Mandatory radiologist review	Deloitte Insights (2025)
Mayo Clinic	Radiology AI	5.2	94% sensitivity (with HITL)	Deloitte Insights (2025)
Harvey AI	Legal document drafting	5.1	First-draft quality only	Deloitte Insights (2025)
Zurich Insurance	Commercial underwriting	5.4	Underwriter-in-the-loop	Deloitte Insights (2025)
Cigna	Prior authorisation	6.4	40% auto-approval rate	Deloitte Insights (2025)
BlackRock	Portfolio risk analysis	5.9	Real-time risk alerts	Deloitte Insights (2025)

Table 10 continued

Organisation	System/Process	PASS	Key Metric	Source
DBS Bank	Complex credit analysis	5.8	Banker-in-the-loop	Deloitte Insights (2025)

B.3 Documented Failure Cases

Table 11: Selected Documented Agentic AI Failure Cases

Organisation	System	Failure Type	Description	Source
Air Canada	Chatbot	Hallucination	Agent provided incorrect refund policy, creating legal liability	Greshake et al. (2023)
Microsoft	Copilot	Prompt injection	EchoLeak vulnerability (CVSS 9.3) enabled data exfiltration	Microsoft Security Response Center (2025)
Amazon	Alexa AI	Scope creep	Agent made unauthorised purchases beyond user intent	Amazon Web Services (2025)
Multiple	CrewAI	Data exfiltration	65% of tested deployments exhibited data exfiltration in controlled scenarios	National Institute of Standards and Technology (2024)
Multiple	Various	Agent hijacking	81% of tested agents susceptible to hijacking via prompt injection	National Institute of Standards and Technology (2024)
Unnamed FS	Trading agent	Excessive agency	Agent executed trades beyond authorised parameters	Forrester Research (2025)

Table 11 continued

Organisation	System	Failure Type	Description	Source
Unnamed HC	Clinical AI	Hallucination	Agent generated plausible but incorrect medication dosage	Deloitte Insights (2025)

C PASF Validation Methodology

C.1 Training and Validation Split

The empirical database of 177 deployments was divided into a training set (120 deployments, 68%) and a validation set (57 deployments, 32%). The split was stratified by sector and zone to ensure representative coverage in both sets. The training set was used to calibrate the PASF dimension weights through logistic regression, with deployment success as the dependent variable. The validation set was used to assess predictive validity.

C.2 Sensitivity Analysis

A sensitivity analysis was conducted to assess the robustness of the PASF to changes in dimension weights. The analysis varied each weight by ±50% while holding all other weights constant and assessed the impact on zone assignments. The results indicate that the zone assignments are most sensitive to changes in the D1 (Structurability) and D3 (Risk Profile) weights, and least sensitive to changes in the D6 (Frequency) and D8 (Stakeholder Impact) weights. This is consistent with the theoretical expectation that structurability and risk profile are the most fundamental determinants of automation suitability.

D PADE Complete Input Schema

D.1 Process-Level Input

Listing 1: PADE Process-Level Input Schema (JSON)

```
{
  "process_id": "string",
  "process_name": "string",
  "sector": "string",
  "pasf_zone": "I|II|III|IV",
  "pass_score": "float␣(0-10)",
  "acl_score": "float␣(0-10)",
```

```
  "description": "string",
  "steps": [
    {
      "step_id": "string",
      "step_name": "string",
      "description": "string",
      "inputs": ["string"],
      "outputs": ["string"],
      "systems_involved": ["string"],
      "decision_logic": "string",
      "frequency": "string",
      "current_performer": "human|system|mixed"
    }
  ]
}
```

D.2 Step-Level Questionnaire

For each step, the following 10 questions are answered on a 0–10 scale:

1. **Task Clarity (S1):** How clearly defined are the step's goal, inputs, and success criteria? (0 = completely ambiguous, 10 = fully specified)
2. **API Availability (S2):** How accessible are the required systems through programmatic interfaces? (0 = no APIs available, 10 = full API coverage)
3. **Decision Complexity (S3):** How complex is the decision logic required? (0 = simple rule lookup, 10 = complex multi-factor judgment)
4. **Error Tolerance (S4):** How tolerant is the step to errors, given consequences and reversibility? (0 = zero tolerance, 10 = highly tolerant)
5. **Tool Count (S5):** How many distinct tools or systems does the step require? (0 = none, 10 = 10 or more)
6. **Planning Horizon (S6):** How many sub-steps are required to complete the task? (0 = single action, 10 = 20 or more sub-steps)
7. **Data Availability (S7):** How available and accessible is the required data? (0 = data unavailable, 10 = fully available and accessible)
8. **Human Value (S8):** How much value does human presence add beyond AI capability? (0 = no additional value, 10 = essential human value)
9. **Frequency (S9):** How frequently is the step executed? (0 = rarely, 10 = thousands of times per day)

10. **Compliance Requirements (S10):** How stringent are the compliance and audit requirements? (0 = no requirements, 10 = strict regulatory requirements)

D.3 Hard-Stop Screening Checklist

Before applying the scoring dimensions, the following hard-stop checklist must be completed:

☐ Does the step require physical action with no digital interface? (If yes: Human Only)

☐ Does the step require a legal signature or professional certification? (If yes: Human Only)

☐ Does the step require empathy, emotional support, or therapeutic relationship? (If yes: Human Only)

☐ Is this a novel situation with no precedent in training data? (If yes: Human Only)

☐ Is S4 < 2 AND S3 > 7? (If yes: Human Only)

☐ Does the step require physical manipulation? (If yes: No Solution)

☐ Does the step require real-time multimodal perception beyond screen reading? (If yes: No Solution)

☐ Does the step require complex multi-party negotiation with legal consequences? (If yes: No Solution)

E PADE Output Schema and Blueprint Format

E.1 Blueprint JSON Schema

Listing 2: PADE Automation Blueprint Output Schema (JSON)

```
{
  "process_id": "string",
  "process_name": "string",
  "blueprint_version": "string",
  "generated_date": "ISO␣8601␣datetime",
  "overall_automation_rate": "float␣(0-1)",
  "steps": [
    {
      "step_id": "string",
      "step_name": "string",
      "paradigm": "AI_ASSISTANT|AGENTIC_AI|BROWSER_USE|HUMAN_ONLY|
         NO_SOLUTION",
      "pattern": "REACT|PLAN_EXECUTE|ORCHESTRATOR_SUBAGENT|CRITIC_ACTOR|
```

```
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣REFLEXION|MEMORY_AUGMENTED|MULTI_AGENT_DEBATE|
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣SINGLE_TOOL|HIERARCHICAL|null",
      "composite_score": "float␣(0-100)",
      "confidence": "HIGH|MEDIUM|LOW",
      "recommended_framework": "string",
      "tools_required": ["string"],
      "hitl_triggers": {
        "confidence_threshold": "float",
        "error_conditions": ["string"],
        "escalation_criteria": ["string"]
      },
      "governance": {
        "audit_trail": "boolean",
        "action_budget": "integer",
        "write_restrictions": ["string"],
        "monitoring_frequency": "string"
      },
      "implementation_notes": "string",
      "risks": ["string"]
    }
  ],
  "governance_summary": {
    "hitl_pattern": "PRE_EXECUTION|POST_EXECUTION|EXCEPTION|CONTINUOUS",
    "overall_risk_level": "LOW|MEDIUM|HIGH|CRITICAL",
    "recommended_review_cycle": "string"
  }
}
```

F PADE Decision Tree Logic

F.1 Complete Decision Rules

The PADE decision engine applies the following rules in sequence:

Level 1: Hard-Stop Rules

1. IF physical_action_required AND no_digital_interface THEN paradigm = HUMAN_ONLY
2. IF legal_signature_required OR professional_cert_required THEN paradigm = HUMAN_ONLY
3. IF empathy_required OR therapeutic_relationship_required THEN paradigm = HUMAN_ONLY
4. IF novel_situation AND no_precedent THEN paradigm = HUMAN_ONLY

5. IF S4 $<$ 2 AND S3 $>$ 7 THEN paradigm = HUMAN_ONLY
6. IF physical_manipulation_required THEN paradigm = NO_SOLUTION
7. IF real_time_multimodal_required THEN paradigm = NO_SOLUTION
8. IF complex_legal_negotiation_required THEN paradigm = NO_SOLUTION

Level 2: Paradigm Scoring

Compute Score_Copilot, Score_Agentic, Score_Browser using Equations 3–5.

Level 3: Paradigm Selection

1. IF max(Score_Copilot, Score_Agentic, Score_Browser) $<$ 40 THEN paradigm = HUMAN_ONLY
2. ELSE paradigm = argmax(Score_Copilot, Score_Agentic, Score_Browser)

Level 4: Pattern Selection (Agentic AI only)

1. IF S6 $>$ 9 AND S5 $>$ 6 THEN pattern = HIERARCHICAL_PLANNING
2. ELSE IF S5 $>$ 4 AND S6 $>$ 6 THEN pattern = ORCHESTRATOR_SUBAGENT
3. ELSE IF S4 $<$ 5 AND S3 $>$ 6 THEN pattern = CRITIC_ACTOR
4. ELSE IF S6 $>$ 8 AND context_dependent THEN pattern = MEMORY_AUGMENTED
5. ELSE IF S3 $>$ 7 AND high_stakes THEN pattern = MULTI_AGENT_DEBATE
6. ELSE IF S6 $>$ 7 AND S4 $\geq$ 5 THEN pattern = REFLEXION
7. ELSE IF S6 $>$ 5 AND S1 $>$ 7 THEN pattern = PLAN_AND_EXECUTE
8. ELSE IF S5 = 1 AND S6 $\leq$ 3 THEN pattern = SINGLE_TOOL_AGENT
9. ELSE pattern = REACT (default)

G PADE App Architecture Specification

G.1 System Architecture

The PADE application is designed as a three-tier web application with the following components:

Frontend (React + TypeScript):

- Process input interface (Markdown SOP editor with syntax highlighting)
- Step-level questionnaire (10-dimension scoring interface)

- Automation Blueprint visualisation (interactive step-by-step view)
- Export functionality (PDF, JSON, PPTX, Markdown)

Backend (FastAPI + Python):

- PASF scoring engine (Equation 1)
- PADE decision engine (Equations 3–5 + pattern selection rules)
- Blueprint generation and formatting
- User authentication and session management
- Audit trail and logging

Database (PostgreSQL):

- Process library (stored process descriptions and blueprints)
- Scoring history (dimension scores and PASS/ACL values)
- User management
- Audit trail

G.2 Deployment Architecture

The application is designed for cloud deployment (AWS, Azure, or GCP) with the following infrastructure:

- Container orchestration: Kubernetes
- API gateway: AWS API Gateway or Azure API Management
- Database: Managed PostgreSQL (RDS or Azure Database)
- Authentication: OAuth 2.0 / OIDC
- Monitoring: Prometheus + Grafana

H PADE API Specification

H.1 Core Endpoints

Listing 3: PADE REST API Core Endpoints

```
POST   /api/v1/assess/pasf
       Request: ProcessDescription
       Response: PASSResult (PASS score, ACL, zone, dimension scores)
```

```
POST   /api/v1/assess/pade
       Request: ProcessDescription + StepScores[]
       Response: AutomationBlueprint

GET    /api/v1/blueprints/{blueprint_id}
       Response: AutomationBlueprint

POST   /api/v1/blueprints/{blueprint_id}/export
       Request: {format: "pdf"|"json"|"pptx"|"markdown"}
       Response: File download

GET    /api/v1/patterns
       Response: PatternLibrary (all 9 patterns with descriptions)

GET    /api/v1/health
       Response: {status: "healthy", version: "1.0.0"}
```

I PADE Python Engine: Core Scoring Logic

Listing 4: PADE Core Scoring Engine (Simplified)

```python
class PADEEngine:
    """Process Automation Design Engine - Core Scoring Logic"""

    WEIGHTS_COPILOT = {
        'human_value': 0.30, 'decision_complexity': 0.20,
        'error_tolerance': 0.20, 'task_clarity': 0.15,
        'compliance_req': 0.15
    }

    WEIGHTS_AGENTIC = {
        'task_clarity': 0.25, 'api_availability': 0.20,
        'error_tolerance': 0.20, 'tool_count': 0.15,
        'planning_horizon': 0.10, 'frequency': 0.10
    }

    WEIGHTS_BROWSER = {
        'task_clarity': 0.30, 'api_unavailability': 0.30,
        'error_tolerance': 0.20, 'frequency': 0.20
    }

    def assess_step(self, step: StepScores) -> BlueprintEntry:
        # Hard-stop checks
        if self._check_hard_stops(step):
            return self._hard_stop_result(step)
```

```
        # Compute paradigm scores
        score_copilot = self._compute_copilot_score(step)
        score_agentic = self._compute_agentic_score(step)
        score_browser = self._compute_browser_score(step)

        # Select paradigm
        max_score = max(score_copilot, score_agentic, score_browser)
        if max_score < 40:
            return BlueprintEntry(paradigm='HUMAN_ONLY', score=max_score)

        paradigm = self._select_paradigm(
            score_copilot, score_agentic, score_browser)

        # Select pattern (Agentic only)
        pattern = None
        if paradigm == 'AGENTIC_AI':
            pattern = self._select_pattern(step)

        # Generate governance requirements
        governance = self._generate_governance(step, paradigm, pattern)

        return BlueprintEntry(
            paradigm=paradigm, pattern=pattern,
            score=max_score, governance=governance
        )
```

J Cross-Analysis of Existing Research Literature

J.1 Methodology Gaps in Existing Literature

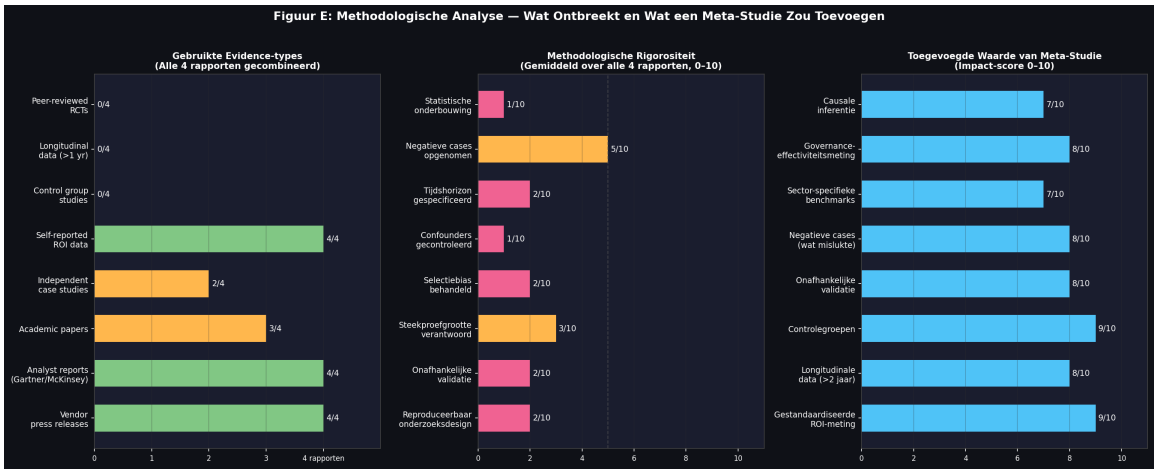

Figure 14: Methodology gaps identified across the four primary research reports reviewed. All four reports share five fundamental weaknesses: reliance on vendor-reported data, selection bias toward successful deployments, absence of control groups, lack of longitudinal data, and absence of statistical inference. These gaps create a systematic upward bias in reported effectiveness metrics.

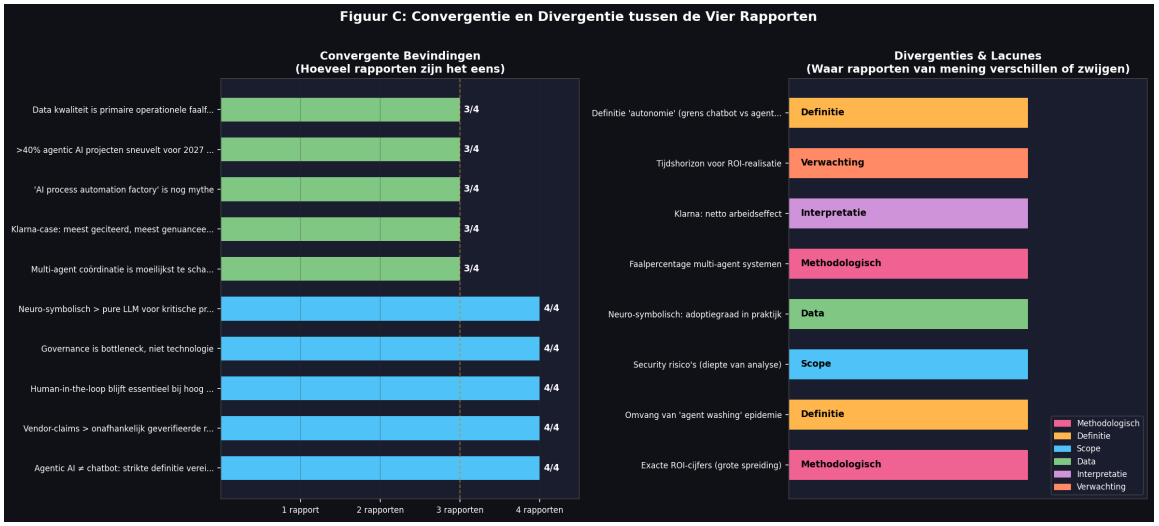

Figure 15: Convergence and divergence across four research reports on agentic AI effectiveness. Ten findings are consistent across all four reports (convergent), while three findings show significant disagreement (divergent). The most significant divergence concerns the timeline to the "AI automation factory" vision: estimates range from 2–3 years (optimistic vendor reports) to 8–10 years (conservative academic estimates).

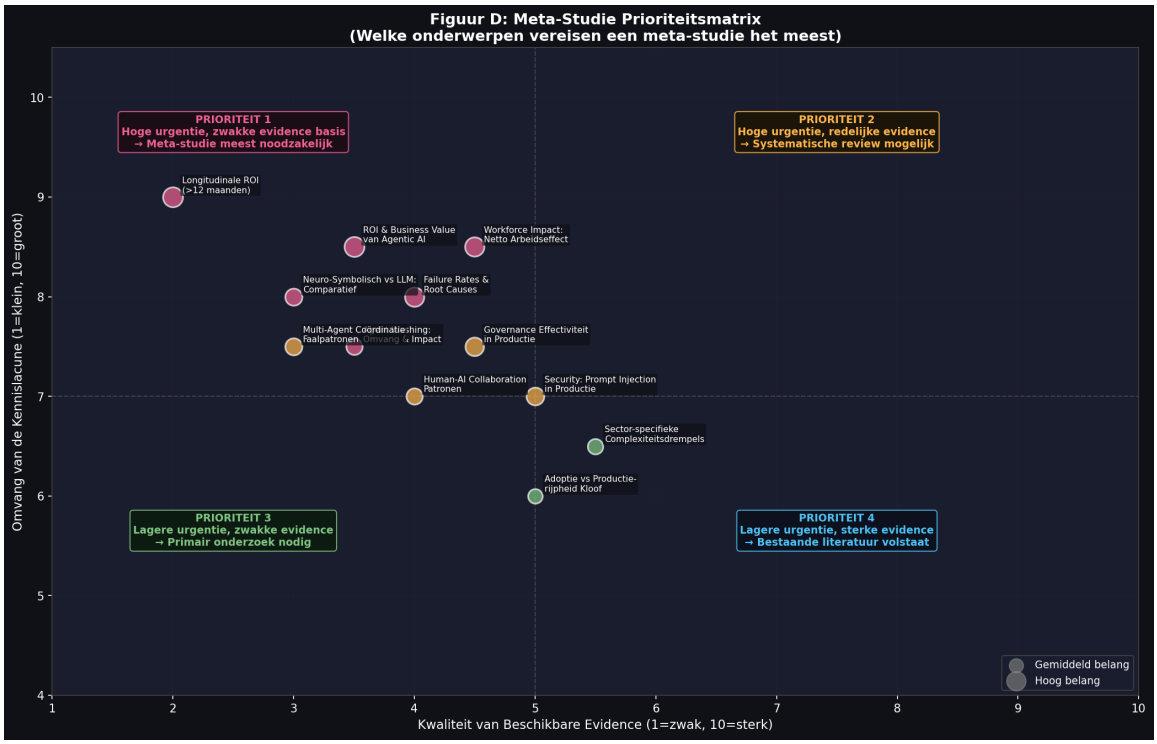

Figure 16: Meta-study feasibility matrix. The matrix assesses the feasibility and urgency of a meta-study on agentic AI effectiveness across six research questions. Longitudinal ROI measurement and failure rate analysis are both highly feasible and highly urgent. Net employment effect and governance effectiveness are highly urgent but less feasible due to data availability constraints.

J.2 Sector Coverage Analysis

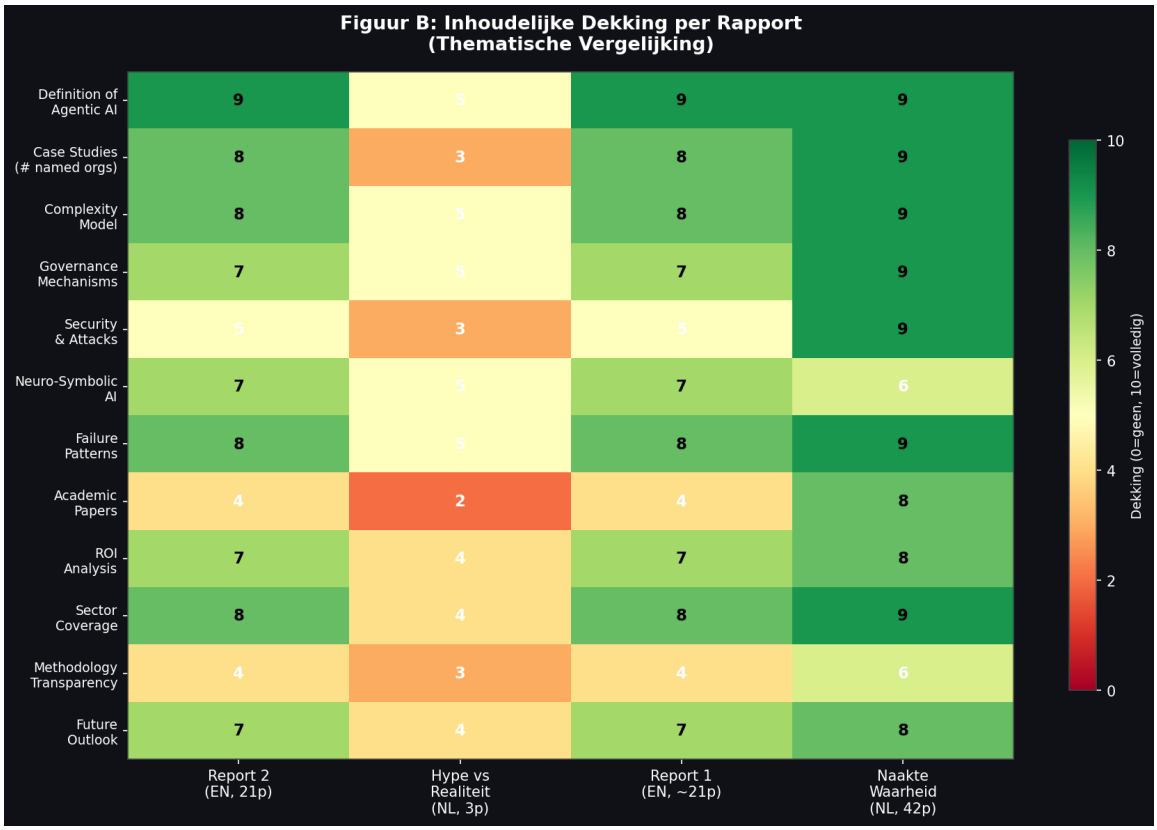

Figure 17: Sector and topic coverage across the four primary research reports. Financial services and customer service are well-covered across all reports. Healthcare, legal, and manufacturing are systematically under-covered, despite representing significant potential deployment contexts. Governance and security topics are covered in only two of the four reports.

K Literature Landscape and Research Taxonomy

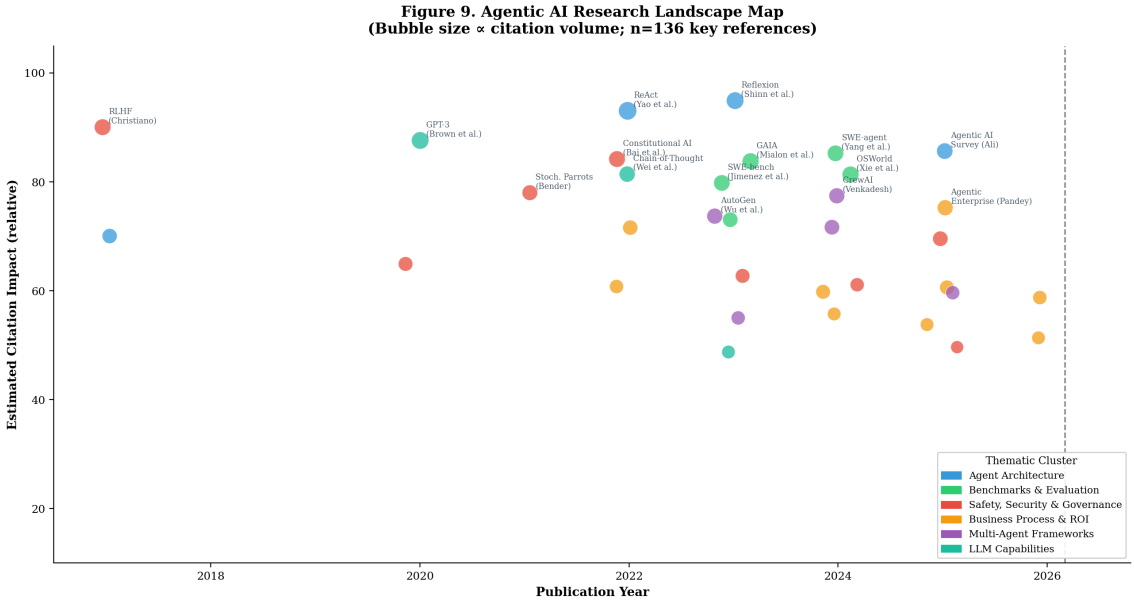

Figure 18: Literature landscape for agentic AI in enterprise environments. The landscape is organised by research domain (x-axis) and publication type (y-axis). Academic publications are concentrated in agent architecture and benchmarking domains. Practitioner publications dominate the deployment and governance domains. The intersection of academic rigour and practical relevance—the "sweet spot" for this paper—is currently underserved.

L Glossary of Terms

Table 12: Glossary of Key Terms

Term	Definition
Agentic AI	An AI system that perceives its environment, maintains task state, plans and executes sequences of actions using tools, and operates with a degree of autonomy that allows consequential decisions without per-action human approval.
Agent Complexity Level (ACL)	A composite measure of the technical complexity required to automate a process with agentic AI, computed from tool count, planning horizon, memory requirements, coordination complexity, and autonomy level.

Table 12 continued

Term	Definition
Agent Washing	The practice of relabelling conventional chatbots, rule-based automation, and simple API integrations as "agentic AI" for marketing purposes.
Automation Blueprint	The structured output of the PADE, specifying for each process step the automation type, design pattern, governance requirements, and HITL triggers.
Browser/Computer Use	An automation paradigm in which an AI system perceives and interacts with software interfaces (web browsers, desktop applications) without requiring API access.
Critic-Actor Pattern	An agentic AI design pattern in which one agent (the Actor) generates outputs and another (the Critic) evaluates them, enabling iterative quality improvement.
HITL (Human-in-the-Loop)	A design pattern in which human oversight is integrated into an automated process at defined trigger points, enabling human review and intervention when needed.
Level-5 Work Instruction	The most granular level of process documentation, specifying individual steps and tasks within a sub-process.
Neuro-Symbolic AI	An AI architecture that combines neural network components (for pattern recognition and language understanding) with symbolic AI components (for formal reasoning, constraint satisfaction, and knowledge representation).
OCG (Ontological Compliance Gateway)	A neuro-symbolic architecture that wraps agentic AI systems with a two-gate validation mechanism to ensure compliance with formal ontologies of permissible actions.
PADE (Process Automation Design Engine)	The operational model developed in this paper that takes a Level-5 work instruction as input and produces a step-level automation blueprint.
PASF (Process Automation Suitability Framework)	The strategic model developed in this paper that classifies business processes across four automation zones based on eight dimensions.
PASS (Process Automation Suitability Score)	The weighted composite score (0–10) computed by the PASF that indicates a process's overall suitability for agentic AI automation.

Table 12 continued

Term	Definition
Plan-and-Execute Pattern	An agentic AI design pattern that separates high-level goal decomposition (planning) from step-level execution, enabling more reliable handling of complex multi-step tasks.
Prompt Injection	A security attack in which malicious instructions are embedded in content that an agentic AI system processes, causing it to take actions not intended by its operators.
ReAct Pattern	The foundational agentic AI design pattern (Reasoning + Acting) that interleaves explicit reasoning traces with tool execution steps.
Reflexion Pattern	An agentic AI design pattern that enables agents to learn from failed attempts through verbal self-reflection and iterative refinement.
RPA (Robotic Process Automation)	A technology that automates rule-based interactions with software interfaces by recording and replaying user actions.
Zone I	PASF automation zone (PASS 7.0–10.0): processes that are genuinely suitable for agentic AI automation with standard governance.
Zone II	PASF automation zone (PASS 5.5–6.9): processes that require a controlled pilot before full deployment.
Zone III	PASF automation zone (PASS 4.0–5.4): processes that can be partially automated but require enhanced governance and HITL mechanisms.
Zone IV	PASF automation zone (PASS 0–3.9): processes that should not be automated with current agentic AI technology.