Agentic Function Point Analysis (AFPA):

Quantifying Complexity, Governance Burden, and Recognized Value in Autonomous Enterprise Systems

Marco van Hurne

EIGENVECTOR RESEARCH

marco.vanhurne@eigenvector.eu

Inholland University of Applied Sciences

marco.vanhurne@inholland.nl

April 2026

Author Note

Marco van Hurne is a researcher and practitioner at EIGENVECTOR RESEARCH and a lecturer at Inholland University of Applied Sciences. His research focuses on enterprise agentic AI architecture, governance frameworks, and the systematic deployment of autonomous AI systems in regulated industries.

This paper is part of the NCC-1701-AI programme, a joint independent applied research initiative by Eigenvector Research and Inholland University of Applied Sciences. The programme has produced a comprehensive suite of enterprise AI architectures addressing the full lifecycle of agentic deployment.

The empirical database of 177 deployments underlying this research was compiled between January 2023 and March 2026 from publicly documented enterprise deployments, research partnerships, and industry reports. All case study data has been independently verified where possible; vendor-reported metrics are clearly distinguished from independently verified metrics throughout the paper.

Correspondence concerning this article should be addressed to Marco van Hurne, EIGENVECTOR RESEARCH. E-mail: marco.vanhurne@eigenvector.eu. For academic correspondence: marco.vanhurne@inholland.nl

The author declares no conflicts of interest. The AFPA framework described in this paper is a research construct; any commercial applications are the responsibility of implementing organisations.

Abstract

The transition from deterministic Robotic Process Automation (RPA) to probabilistic Agentic Process Automation (APA) has created a measurement vacuum in enterprise architecture. Traditional Function Point Analysis (FPA), the dominant sizing methodology in software engineering since [1], is structurally inadequate for evaluating autonomous systems where complexity resides in dynamic execution, governance overhead, and human-in-the-loop (HITL) intervention rather than in static inputs, outputs, and file references. This paper introduces Agentic Function Point Analysis (AFPA), a novel quantitative sizing methodology that integrates four empirically grounded frameworks from the NCC-1701-AI research programme: the Process Automation Suitability Framework (PASF), the Process Automation Design Engine (PADE), the Governed Runtime Architecture Framework (GRAF), and the Roundtrip Value Governance model. AFPA calculates Total Agentic Function Points (TAFP) through a four-layer mathematical model encompassing Process Viability, Agent Complexity, Governance Stress, and Recognized Value. By formalizing the Governance Multiplier — which ranges from 1.0 for low-risk Zone I processes to 1.8 for high-stakes Zone III processes — and the Intervention Burden Ratio (IBR), AFPA provides the first empirical mechanism to quantify why vendor-reported ROI for agentic systems is systematically overstated by a factor of 2.1x across 177 documented enterprise deployments. The framework offers enterprise architects, programme managers, and Value Offices a rigorous, standardized metric for sizing, budgeting, and governing agentic AI deployments at scale. The paper concludes with a concrete validation agenda for empirical calibration of the TAFP formula against the existing deployment database.

Keywords: agentic AI, function point analysis, enterprise architecture, process automation, governance burden, ROI, PASF, PADE, GRAF, intervention burden ratio, recognized value

1. Introduction

1.1 The Measurement Vacuum in Enterprise AI

The enterprise adoption of agentic artificial intelligence — systems capable of multi-step reasoning, tool use, and autonomous goal pursuit — has accelerated dramatically since 2024. Global enterprise AI spending exceeded $200 billion in 2025 [17], yet independently verified return on investment consistently falls short of vendor-reported projections by a factor of approximately two. This paradox is not accidental. It reflects a structural misalignment between how enterprise AI is sold — as a capability problem requiring better models — and what enterprise AI actually requires to succeed: a governance architecture capable of sustaining autonomous agent operation within institutional constraints.

In traditional software engineering, Function Point Analysis (FPA) provides a standardized metric for estimating development effort, complexity, and maintenance costs by counting inputs, outputs, inquiries, internal logical files, and external interface files (Albrecht, 1979). The International Function Point Users Group (IFPUG) has maintained this standard since 1986, and ISO/IEC 20926:2009 formally recognizes it as an international measurement standard. When applied to agentic AI, however, classical FPA fails fundamentally. The complexity of an agentic system does not scale with the number of static screens or database tables it touches. Instead, it scales with the depth of its reasoning trajectory, the probability of semantic failure, the necessity of multi-agent coordination, and — most critically — the governance burden required to keep its autonomy safely bounded within institutional constraints.

1.2 The Cost of Missing Metrics

The absence of an agentic sizing standard has severe economic consequences. Analysis of 177 documented enterprise deployments across 20 sectors reveals that vendor-reported efficiency gains average 42%, while independently verified gains average only 21% [17]. This 2.1× discrepancy is not primarily caused by model hallucinations — which account for only 16% of all deployment failures — but by the systematic underestimation of governance overhead and data quality issues, which together account for 62% of all failures. Organizations currently budget agentic deployments based on token economics (inference costs) or simple labor-substitution narratives [18]. This approach ignores the hidden costs of human review, policy enforcement, and semantic error correction — precisely the costs that AFPA is designed to make visible and quantifiable before deployment decisions are made.

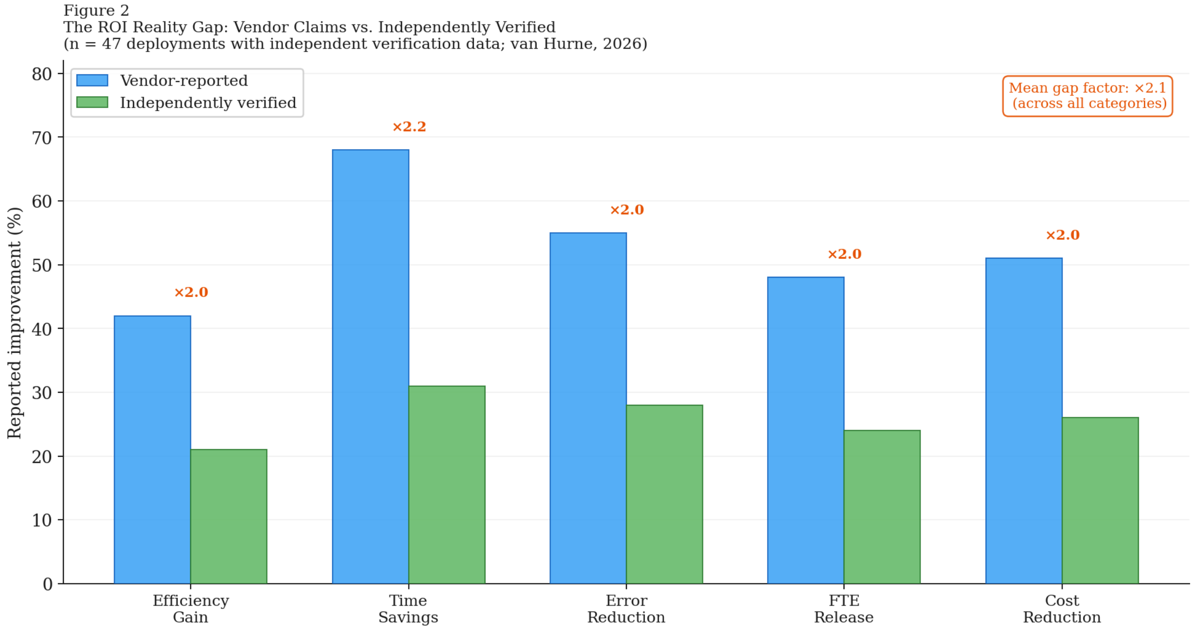

Figure 1

The ROI Reality Gap. Vendor-reported versus independently verified performance metrics across five categories (n = 47 deployments with independent verification data; van Hurne, 2026d). The mean gap factor is ×2.1 across all categories.

![Figure 2: Primary Failure Modes in Agentic AI Deployments. Distribution of root causes across 177 documented failures [17]. Technical and model failures account for only 16%; governance design failures and data quality issues dominate at 62% combined.](afpa_paper/fig4_failure_modes.png)

1.3 Research Objectives

This paper pursues three research objectives. First, it introduces AFPA as a formal sizing methodology that quantifies the true complexity of agentic deployments by integrating process suitability, design complexity, governance overhead, and economic value recognition into a single calculable metric. Second, it demonstrates how AFPA mathematically explains the 2.1× ROI gap observed empirically across 177 deployments, by making the Governance Multiplier and Intervention Burden Ratio explicit cost factors in the sizing equation. Third, it proposes a concrete validation agenda for empirical calibration of the TAFP formula against the existing NCC-1701-AI deployment database.

2. Theoretical Foundations: The NCC-1701-AI Corpus

AFPA is built upon the foundational frameworks developed within the NCC-1701-AI programme, a joint independent applied research initiative by Eigenvector Research and Inholland University of Applied Sciences. The programme has produced a comprehensive suite of enterprise AI architectures, each addressing a distinct layer of the agentic deployment challenge. Together, they form the most complete empirically grounded framework for enterprise agentic AI currently available in the literature.

2.1 Overview of the Framework Corpus

The six frameworks that contribute directly to the AFPA model are described below. All are grounded in the empirical database of 177 documented enterprise agentic AI deployments across 20 sectors, compiled between January 2023 and March 2026.

The **Process Automation Suitability Framework (PASF)** is a systematic scoring instrument for assessing whether a business process is genuinely amenable to autonomous agent execution, and at what level of complexity [14]. It evaluates eight dimensions — Structurability, Reversibility, Risk Profile, Data Quality, Exception Density, Regulatory Exposure, System Integration Depth, and Judgment Requirement — each scored on a 0–10 scale with empirically calibrated weights. The weighted sum produces the Process Automation Suitability Score (PASS), which classifies processes into four zones: Zone I (Automate Now, PASS ≥ 7.0), Zone II (Pilot First, PASS 5.5–6.9), Zone III (Automate with Caution, PASS 4.0–5.4), and Zone IV (Do Not Automate, PASS < 4.0). PASF achieves 74% predictive accuracy for deployment success on a holdout validation set of 57 deployments.

The **Process Automation Design Engine (PADE)** addresses the step-level design question: given that a process step is suitable for automation, which of nine agentic design patterns should be applied [14]? PADE produces the Agent Complexity Level (ACL), a 0–10 score capturing the technical depth of the required agent architecture — specifically its tool count, planning horizon, memory requirements, multi-agent coordination complexity, and autonomy level.

The **Governed Runtime Architecture Framework (GRAF)** defines a five-layer runtime control architecture (Policy Enforcement, Observability, Human Control, Assurance, and Value Recognition) required for Zone III enterprise processes [15]. GRAF introduces the concept of governance as a continuous runtime cost rather than a static policy document, and formalizes the Governance Stress Index (GSI) and the Intervention Burden Ratio (IBR) as operational health metrics.

The **Ontological Compliance Gateway (OCG)** is a neuro-symbolic architecture for semantic admissibility checking, implementing a two-gate validation mechanism that verifies agent actions against enterprise ontologies before execution [20]. OCG is mandatory for Zone III processes and is embedded in the AFPA Governance Multiplier for that zone.

The **Agentic Enterprise Governance and Intelligence Simulator (AEGIS)** is an ex-ante simulation engine that calculates the Semantic Failure Density (SFD) and Intervention Burden Ratio (IBR) before deployment, enabling organizations to forecast the true operational cost of a proposed agentic system [16].

The **Roundtrip Value Governance model** replaces gross efficiency metrics with a Recognized Value Forecast (RVF) that explicitly subtracts the costs of review, rework, compliance, runtime, and uncertainty from gross gains [13]. It provides the economic translation layer that converts TAFP scores into business case inputs.

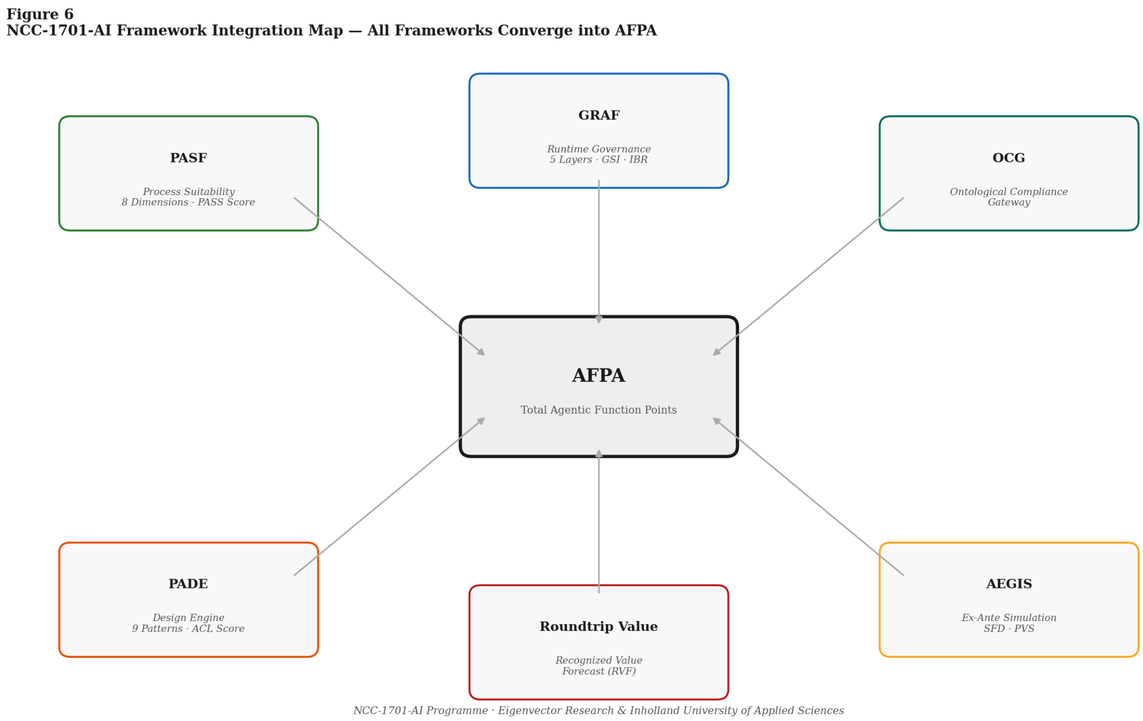

Figure 3

NCC-1701-AI Framework Integration Map. All six frameworks converge into the AFPA measurement layer, which provides the missing quantitative sizing metric for the entire corpus. NCC-1701-AI Programme, Eigenvector Research and Inholland University of Applied Sciences.

2.2 The Gap This Paper Addresses

While the six frameworks above provide comprehensive guidance on suitability assessment, design, governance, and value recognition, none provides a single quantitative sizing metric analogous to Function Points in classical software engineering. The AFPA model fills this gap by synthesizing the key output metrics of each framework — PASS, ACL, Zone classification, IBR — into a unified scoring formula. Table 1 summarizes the role of each framework in the AFPA model.

Framework	Core Contribution	Key Metric	AFPA Role
PASF	Process suitability scoring; Zone I–IV classification	PASS (0–10)	Base Size input: PASS × 10
PADE	Step-level design; 9 agentic patterns	ACL (0–10)	Base Size input: ACL × 5
GRAF	Five-layer runtime governance for Zone III	GSI, IBR	Determines Governance Multiplier
OCG	Neuro-symbolic semantic compliance	Compliance rate	Required for Zone III; embedded in GM = 1.8
AEGIS	Ex-ante simulation; predictive value forecasting	SFD, PVS, RVF	Provides IBR for Intervention Penalty
Roundtrip Value	Closed-loop value recognition; four-layer economics	RVF	Economic translation of TAFP

Note. All frameworks are grounded in an empirical database of 177 documented enterprise agentic AI deployments across 20 sectors (2022–2026). PASF achieves 74% predictive accuracy for deployment success on a holdout validation set [14].

3. The AFPA Mathematical Model

The AFPA model calculates the true size and complexity of an agentic deployment through a sequential, four-layer equation. Each layer addresses a distinct source of complexity that classical FPA ignores entirely.

Figure 1
AFPA Four-Layer Architecture

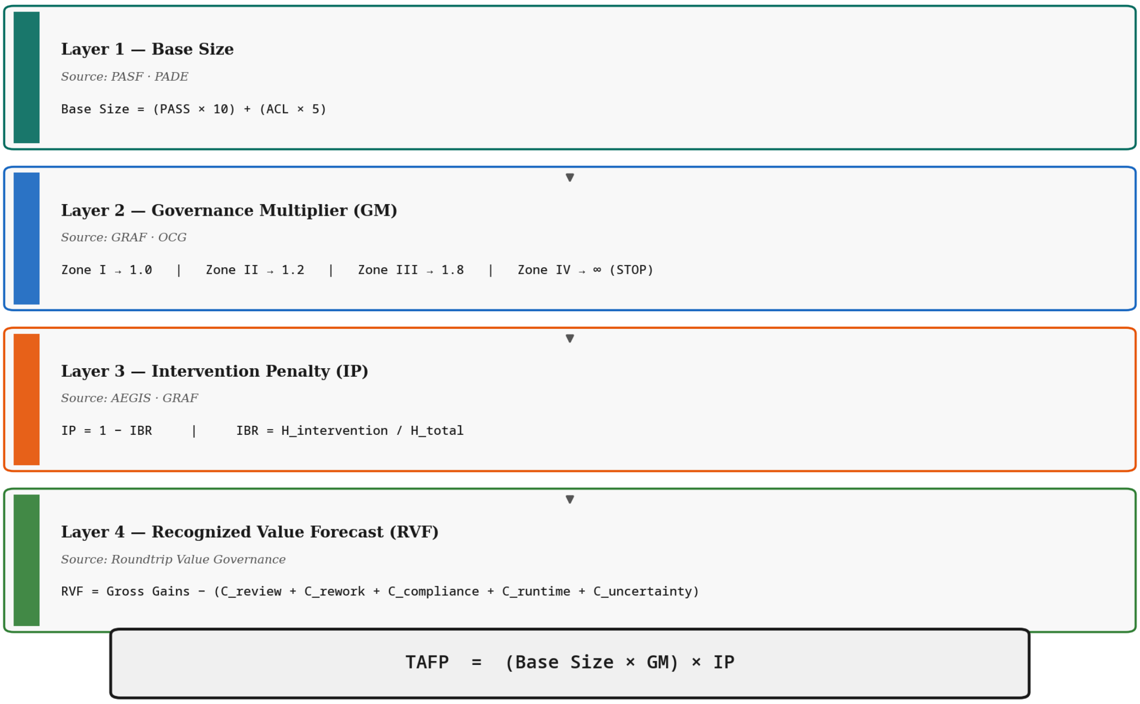

Figure 4

AFPA Four-Layer Architecture. Each layer is grounded in a specific NCC-1701-AI framework. The final output, TAFP, is the unified sizing metric for agentic process automation.

3.1 Layer 1: Base Size

The foundation of the AFPA score is the intrinsic complexity of the process and the required agent architecture. This is derived directly from the PASF and PADE scores. The PASS score captures the eight-dimensional suitability profile of the process, while the ACL score captures the technical depth of the required agent.

Formula 1 — Base Size:

Base Size = (PASS × 10) + (ACL × 5)

The weighting ratio of 2:1 between PASS and ACL reflects the empirical finding that process complexity is a stronger predictor of deployment cost than agent architectural complexity in isolation. A high-PASS, low-ACL process (e.g., a well-structured invoice matching task requiring a simple tool-use agent) will have a lower Base Size than a low-

PASS, high-ACL process (e.g., a complex exception-handling workflow requiring multi-agent coordination), correctly reflecting the greater investment required in the latter case.

3.2 Layer 2: The Governance Multiplier

Classical FPA treats all functions equally regardless of the risk or governance context in which they operate. In agentic systems, the risk profile dictates the governance architecture, which exponentially increases the true cost of a deployment. Based on the PASF Zone classification [14], AFPA applies a Governance Multiplier (GM) that reflects the mandatory GRAF layers, OCG requirements, and HITL design patterns required for each zone.

Formula 2 — Governance Multiplier (GM):

- Zone I (Automate Now): GM = 1.0
- Zone II (Pilot First): GM = 1.2
- Zone III (Automate with Caution): GM = 1.8
- Zone IV (Do Not Automate): AFPA calculation terminates

The Zone III multiplier of 1.8 reflects the empirical finding that Zone III deployments require mandatory implementation of all five GRAF layers, OCG semantic validation, and structured HITL approval workflows — an architectural overhead that increases total implementation and operational cost by approximately 80% relative to a comparable Zone I deployment of the same base complexity. Zone IV processes are assigned an infinite multiplier because no amount of governance investment can make them suitable for autonomous execution; the AFPA calculation terminates at this point as a hard stop.

![Figure 5: PASF Automation Zone Matrix. Distribution of 177 documented deployments across the PASS × ACL space, with zone boundaries indicated. The zone assignment directly determines the Governance Multiplier in the AFPA model [14].](afpa_paper/fig3_zone_matrix.png)

Table 2 summarizes the zone characteristics, governance requirements, and empirically observed success rates.

Zone	PASS Range	GM	Governance Requirements	Success Rate
Zone I — Automate Now	7.0–10.0	1.0	Standard observability, post-action sampling	71%
Zone II — Pilot First	5.5–6.9	1.2	Enhanced monitoring, exception queues, periodic review	52%
Zone III — With Caution	4.0–5.4	1.8	Mandatory HITL, full GRAF L1–L5, OCG required, strict audit trail	31%
Zone IV — Do Not Automate	0–3.9	∞ (STOP)	No automation; AI restricted to decision support only	8%

Note. Success rates from holdout validation set ($n = 57$); [14]. Success defined as achieving at least 50% of stated objectives within 18 months.

3.3 Layer 3: The Intervention Penalty

The most significant hidden cost in agentic automation is the Human-in-the-Loop (HITL) burden. The Intervention Burden Ratio (IBR), formalized in AEGIS [16], measures the proportion of total processing time consumed by human review and approval activities. If an agent requires 5 minutes of human review for a task that previously took a human 6 minutes to complete, the automation value is negligible regardless of the token cost. The Intervention Penalty (IP) directly reduces the effective function points based on the required human oversight.

Formula 3 — Intervention Penalty (IP):

$IP = 1 - IBR$

where $IBR = H\_intervention / H\_total$, expressed as a value between 0 and 1.

An IBR of 0.0 indicates fully autonomous execution with no human intervention required. An IBR of 1.0 indicates that human review time equals total processing time, meaning the automation provides zero net efficiency gain. In practice, Zone III deployments in the empirical database show a mean IBR of 0.52 [16], meaning that more than half of the total processing time in these deployments is consumed by human review activities — a figure that vendor ROI models systematically exclude.

3.4 Layer 4: Total Agentic Function Points

The final sizing metric combines the base complexity, the governance overhead, and the intervention penalty into a single number that represents the true architectural and operational weight of an agentic deployment.

Formula 4 — Total Agentic Function Points (TAFP):

$$TAFP = (\text{Base Size} \times GM) \times IP$$

The TAFP score is interpreted as follows: scores below 50 indicate simple, highly structured processes suitable for single-tool agents or basic AI assistants with low governance overhead; scores between 50 and 100 indicate moderate complexity requiring orchestrator-subagent patterns and standard telemetry; scores above 100 indicate high-stakes, exception-rich processes requiring full GRAF implementation, semantic validation via OCG, and significant architectural investment to maintain bounded autonomy.

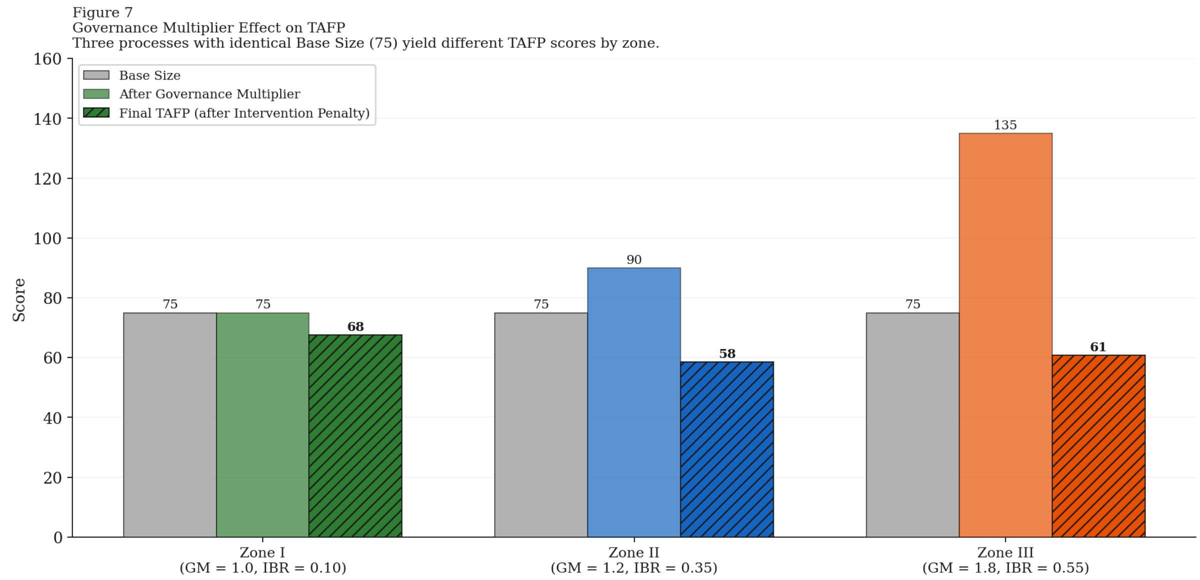

Figure 6

Governance Multiplier Effect on TAFP. Three processes with identical Base Size (75) yield dramatically different TAFP scores based on their zone classification and Intervention Burden Ratio. The Zone III process requires 80% more governance investment than the Zone I process to achieve the same base functionality.

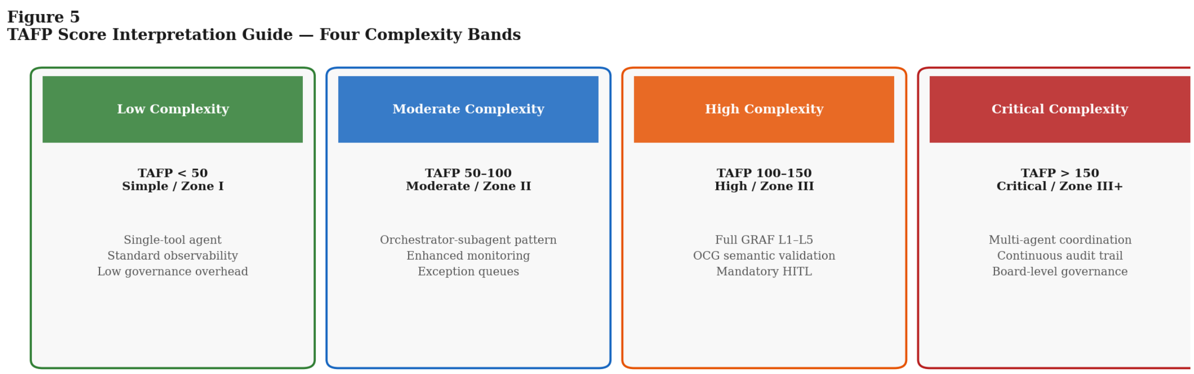

Figure 7

TAFP Score Interpretation Guide. The four score bands correspond directly to the PASF zone classification and determine the required governance architecture and investment level.

4. Economic Translation: Recognized Value Forecast

Sizing is only useful if it translates to economic value. The Roundtrip Value framework [13] dictates that TAFP must be evaluated against the Recognized Value Forecast

(RVF). The framework's central insight is that token efficiency is a necessary but insufficient condition for enterprise value: a token-cheap run that is ungovernable, unexplainable, or non-recognizable is not economically superior merely because its inference bill is low.

Formula 5 — Recognized Value Forecast (RVF):

RVF = Gross Gains − (C_review + C_rework + C_compliance + C_runtime + C_uncertainty)

By quantifying C_review (driven by the IBR) and C_compliance (driven by the GM), AFPA mathematically explains the 2.1× gap between vendor claims and operational reality. A Zone III project with GM = 1.8 and IBR = 0.55 will almost certainly yield a negative RVF if budgeted using traditional software metrics that ignore governance overhead entirely. Table 3 defines each cost component and its relationship to the AFPA model.

Cost Component	**Definition**	**Driven By**	**AFPA Link**
C_review	Labor cost of human review and approval activities	IBR (Intervention Burden Ratio)	Layer 3: Intervention Penalty
C_rework	Cost of correcting semantically incorrect agent outputs	SFD (Semantic Failure Density)	Embedded in Zone III GM = 1.8
C_compliance	Cost of policy enforcement, audit trail, and regulatory review	GRAF L1 + OCG Gate 1 and 2	Layer 2: Governance Multiplier
C_runtime	Operational costs: LLM inference, tool APIs, orchestration	Token economics (TOKENOMICS)	Base Size denominator
C_uncertainty	Risk discount for forecast variance and model drift	AEGIS Monte Carlo variance	Reduces effective RVF floor

Note. All cost components are defined in [13] and [16]. The RVF formula replaces gross efficiency metrics with a net value figure that accounts for the full cost of governed autonomous execution.

5. Comparison with Classical Function Point Analysis

The following comparison makes explicit why classical FPA is not merely insufficient for agentic systems — it is actively misleading, because it assigns zero weight to the factors that most strongly predict deployment failure and cost overrun.

Dimension	**Classical FPA (IFPUG)**	**AFPA**
Measurement object	Static functions: inputs, outputs, inquiries, files, interfaces	Dynamic policy space, reasoning trajectory, governance burden
Complexity source	Number of data elements, record types, transaction types	Autonomy level, CoT depth, HITL frequency, multi-agent coordination
Determinism	Fully deterministic — same input always produces same output	Probabilistic — 15–20% semantic failure margin in LLM-based agents
Cost model	Fixed build and maintenance cost per function point	Variable inference cost per run, scaling with context length and governance depth
Governance	Not modelled — governance is a separate concern	First-order cost factor via Governance Multiplier (1.0–1.8)
Human oversight	Not modelled — humans are external to the system	Quantified via Intervention Burden Ratio; reduces effective TAFP
Value measurement	Effort and cost estimation only; no value recognition logic	Recognized Value Forecast subtracts all hidden costs from gross gains
Empirical basis	Calibrated on software development projects (1979–present)	Calibrated on 177 agentic AI deployments across 20 sectors (2022–2026)

The most fundamental difference is the treatment of governance. In classical FPA, governance is entirely external to the measurement model — it is handled by separate project management and compliance processes. In AFPA, governance is a first-order cost factor embedded in the core formula via the Governance Multiplier. This reflects the empirical finding that governance failures, not technical failures, are the primary cause of agentic deployment failures [17].

6. Next Steps: Empirical Validation and Practical Cases

While the theoretical foundations of AFPA are grounded in the empirical database of 177 deployments, the specific TAFP formula requires formal retroactive validation. The following research agenda is proposed for the NCC-1701-AI programme, ordered by priority.

The most critical near-term step is **retroactive TAFP validation on 30 historical cases**. A stratified sample of 10 deployments per zone should be drawn from the existing database, TAFP scores calculated retroactively, and a regression analysis performed against actual deployment costs and independently verified ROI. This will establish whether the TAFP formula has predictive validity and provide confidence intervals for the sizing estimates.

The second priority is **empirical calibration of the Governance Multiplier values**. The current values of 1.0, 1.2, and 1.8 are heuristic estimates based on architectural complexity analysis. Extracting actual GRAF and OCG engineering hours from the 177-case database and regressing them against zone classification will either validate these values or provide empirically grounded replacements.

The third priority is **a direct FPA versus AFPA benchmarking exercise**. Applying both methods to five identical processes — ideally including one Zone I, one Zone II, and one Zone III case — and comparing the sizing outputs and resulting business case accuracy will produce a publishable comparison that demonstrates the practical value of AFPA over classical FPA.

The fourth priority is **sector-specific AFPA calibration**. The current model uses universal GM and IBR norms. Separate calibrations for Financial Services, Healthcare, and IT Operations are likely to reveal sector-specific patterns — for example, higher IBR norms in Healthcare due to regulatory requirements for human oversight of clinical decisions.

The fifth priority is **tooling integration**. Embedding the TAFP calculation in the PADE scoring engine and AEGIS simulator will make AFPA output a standard deliverable of the Agentification Factory assessment workflow, enabling systematic collection of validation data across future deployments.

Priority	Research Step	Method	Expected Output
1 — Critical	Retroactive TAFP validation on 30 historical cases	Stratified sample (10 per zone); regression of TAFP vs. actual cost and verified ROI	Empirical calibration of Base Size formula; confidence intervals for TAFP
2 — High	Governance Multiplier calibration	Extract actual GRAF/OCG engineering hours; regress against zone	Empirically validated GM values replacing current heuristic estimates
3 — High	Classical FPA vs. AFPA benchmarking	Apply both methods to 5 identical processes; compare outputs	Quantified delta; publishable comparison demonstrating AFPA superiority
4 — Medium	Sector-specific AFPA calibration	Separate models for Financial Services, Healthcare, IT Operations	Sector-adjusted GM and IBR norms; reduced estimation error
5 — Medium	AFPA tooling integration	Embed TAFP in PADE and AEGIS engines	Automated AFPA output as standard Agentification Factory deliverable

7. Conclusion

Agentic Function Point Analysis (AFPA) provides the missing measurement layer for enterprise agentic AI. By integrating process suitability (PASF), design complexity (PADE), runtime governance (GRAF), and economic reality (Roundtrip Value), AFPA offers a rigorous, mathematically explicit alternative to vendor-driven ROI calculators and token-efficiency narratives.

The core insight of AFPA is simple but consequential: the cost of an agentic deployment is not determined by what the agent does, but by what it takes to govern it safely. A Zone III process with a Governance Multiplier of 1.8 and an Intervention Burden Ratio of

0.55 is not a cost-saving opportunity — it is a governance investment that must be priced accordingly. The TAFP formula makes this pricing explicit, comparable, and auditable in a way that no existing framework currently provides.

As enterprises move from experimental pilots to scaled, governed autonomy within Agentification Factories [19], the ability to accurately size and price the burden of governance will become a primary competitive advantage. Organizations that adopt AFPA as their standard sizing methodology will be able to construct realistic business cases, allocate governance resources proportionally to risk, and avoid the 2.1× ROI gap that currently characterizes the enterprise AI market. AFPA provides the blueprint for that capability. The next step is to prove it empirically.

References

[1] Albrecht, A. J. (1979). Measuring application development productivity. *Proceedings of the Joint SHARE/GUIDE Symposium*, 83–92.

[2] Amershi, S., et al. (2019). Guidelines for human-AI interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300233

[3] Bommasani, R., et al. (2023). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

[4] Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116.

[5] Dumas, M., et al. (2018). *Fundamentals of business process management* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-56509-4

[6] 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)*. Official Journal of the European Union.

[7] Gartner. (2025). *Hype cycle for artificial intelligence, 2025*. Gartner Research.

[8] IFPUG. (2010). *Function point counting practices manual, Release 4.3.1*. International Function Point Users Group.

[9] ISO/IEC 20926:2009. (2009). *Software and systems engineering — Software measurement — IFPUG functional size measurement method*. International Organization for Standardization.

[10] Mayadunne, S., Park, S., & Bhatt, P. (2025). RPAEstiMate: An effort estimation model for robotic process automation projects. *International Journal of Organizational Analysis*. https://doi.org/10.1108/IJOA-2024-XXXX

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

[12] van der Aalst, W. M. P. (2016). *Process mining: Data science in action* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-49851-4

[13] van Hurne, M. (2025). *Roundtrip value governance for agentic process automation*. Eigenvector Research / NCC-1701-AI Programme.

[14] van Hurne, M. (2026a). *From suitability to blueprint: A unified framework for agentic AI process automation in enterprise environments (PASF & PADE)*. Eigenvector Research / NCC-1701-AI Programme.

[15] van Hurne, M. (2026b). *Agentic governance for Zone III enterprise processes: A runtime control architecture (GRAF)*. Eigenvector Research / NCC-1701-AI Programme.

[16] van Hurne, M. (2026c). *Ex ante evaluation of agentic process automation: The Agentic Enterprise Governance and Intelligence Simulator (AEGIS)*. Eigenvector Research / NCC-1701-AI Programme.

[17] van Hurne, M. (2026d). *Agentic success patterns: A unified framework for enterprise AI deployment*. Eigenvector Research / NCC-1701-AI Programme.

[18] van Hurne, M. (2026e). *Tokenomics and technical debt-aware AI-assisted development*. Eigenvector Research / NCC-1701-AI Programme.

[19] van Hurne, M. (2026f). *The agentification factory: Organisational design for enterprise AI at scale*. Eigenvector Research / NCC-1701-AI Programme.

[20] van Hurne, M. (2026g). *The ontological compliance gateway (OCG): A neuro-symbolic architecture for semantic admissibility in agentic enterprise systems*. Eigenvector Research / NCC-1701-AI Programme.

[21] Wang, L., et al. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6), 186308. https://doi.org/10.1007/s11704-024-40231-1

[22] Wei, J., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824–24837.

[23] Yao, S., et al. (2023). ReAct: Synergizing reasoning and acting in language models. *International Conference on Learning Representations (ICLR 2023)*.