DMPedia Lecture Notes in Computer Science & Engineering
Volume 1, Issue 2, pp.353-365, 2026
ICICCS-2026

# From Robotic Process Automation to Agentic AI: A Systematic Review, Taxonomy, and Capability Assessment Framework for Intelligent Automation in Enterprise Accounting

Rahul Rao Juvvadi
Principal Software Engineer, Microsoft
rahulraoj99@gmail.com

**Abstract.**

Enterprise accounting is undergoing a structural transition from rule-based robotic process automation (RPA) toward agentic artificial intelligence: systems built on large language models that plan multi-step workflows, invoke tools, and adapt their execution autonomously. Research on this transition remains fragmented across information systems, accounting, and artificial intelligence venues, and no unified conceptual structure yet characterises the emerging class of agentic accounting systems. We conduct a PRISMA 2020-compliant systematic literature review, screening 2,387 retrieved records down to 60 included studies, and synthesise the technological trajectory of intelligent automation across four generations RPA, intelligent process automation (IPA), hyperautomation, and agentic AI and across core accounting subfunctions including procure-to-pay, order-to-cash, record-to-report, reconciliation, audit, and tax. Applying Nickerson et al.'s method, we derive a six-dimensional taxonomy spanning autonomy level, learning capability, process scope, human oversight, integration depth, and accounting subfunction. We extend the taxonomy into a capability assessment framework with five maturity levels (L0–L4) per dimension and ground it in four archetypal system classes, from the classic RPA bot to the autonomous multi-agent reconciliation system. The review identifies five critical research gaps: agent governance, hallucination mitigation, benchmark scarcity, multi-agent orchestration standards, and explainability and outlines a research agenda for trustworthy agentic accounting systems.

**Keywords:** Robotic Process Automation, Agentic AI, Intelligent Agents, LLM-based Agents, Enterprise Accounting, Systematic Literature Review, Taxonomy, Capability Assessment Framework

## 1 Introduction

Financial operations teams face unrelenting pressure to compress close cycles, cut processing costs, and improve reporting accuracy under growing regulatory complexity. Enterprise accounting platforms ERP systems, subledgers, the general ledger, and document repositories generate enormous transaction flows still largely classified, matched, and reconciled by hand. Automation promises to deflect this routine work, but accounting differs from domains such as customer service: every automated action must remain auditable, controlled, and defensible to regulators and external auditors. Accounting automation must therefore be not merely efficient but trustworthy.

Over the past decade, three automation waves reached the finance function: robotic process automation (RPA), scripted routine, high-volume task sequences; ML-augmented intelligent process automation (IPA), added probabilistic classification learned from historical data; and hyperautomation, orchestrated specialised tools into end-to-end coverage. Each widened automation’s reach, yet all three remained prescribed: workflows designed in advance, decisions hard-coded or trained offline, humans controlling every consequential step.

A fourth wave is emerging. Frontier large language models with tool-use interfaces and memory enable agentic AI software that sets subgoals, plans multi-step action sequences, reasons over novel scenarios, and adapts execution at runtime [14, 15]. Given accounting data, such agents can gather audit evidence, draft journal entries with supporting narratives, resolve reconciliation variances, and flag anomalies. Yet the literature on these systems is scattered across disciplines and lacks a shared vocabulary for distinguishing agentic systems from predecessors or assessing their maturity.


Proceedings of ICICCS-2026 (Poster Papers), DMPedia Lecture Notes in Computer Science & Engineering.
ISBN: 978-81-993813-1-5

This paper addresses that gap through three contributions. **C1 Systematic review:** a PRISMA 2020-compliant review that screened 2,387 retrieved records to 60 included studies, charting the evolution from RPA to agentic AI in enterprise accounting and mapping which subfunctions each generation has reached. **C2 Taxonomy:** a six-dimensional taxonomy autonomy, learning, scope, oversight, integration, and subfunction derived from the corpus using Nickerson et al.'s method [10]. **C3 Capability framework:** a maturity model assigning levels L0–L4 to each taxonomy dimension, illustrated through four archetypal system classes observed in practice.

Four research questions structure the review. **RQ1:** How has intelligent automation in accounting evolved across the four generations, and what capabilities distinguish each? **RQ2:** Which accounting subfunctions has each generation addressed, and where is coverage thin? **RQ3:** What dimensions form a useful taxonomy of intelligent automation systems in accounting? **RQ4:** What research gaps remain particularly in governance, auditability, and evaluation of agentic systems and what research agenda follows?

Section 2 reviews the four generations and prior reviews; Section 3 documents the methodology; Section 4 synthesises findings; Section 5 derives the taxonomy; Section 6 presents the capability framework and its application; Section 7 identifies research gaps and concludes.

## 2 Background and Related Work

### 2.1 Four Generations of Intelligent Automation

**Robotic process automation (2016–2023).** RPA automates high-volume, rule-bound processes through user-interface automation [1, 2]: bots execute hard-coded IF-THEN workflows triggered by predefined events. In accounting, the first wave targeted narrow, repetitive tasks accounts payable invoice entry [6], template-driven general ledger reconciliation, and rule-based cash application. RPA delivers measurable speed and cost benefits [3], but fixed selectors and exhaustively enumerated decision paths make it brittle when interfaces change, or transactions deviate from anticipated formats [4].

**Intelligent process automation (2018–2024).** As supervised learning matured, vendors layered predictive models onto deterministic workflows: classifiers and anomaly detectors [5] supporting invoice classification, account prediction, and embedding-based invoice-to-purchase-order matching. Field evidence documented adoption in public accounting and corporate finance [7, 8], with intelligent document processing OCR paired with learned field extraction as the signature pattern [25]. IPA remains bounded by its training distribution: models need labelled data, are retrained offline, and defer novel cases to human review.

**Hyperautomation (2020–present).** Hyperautomation integrates RPA, document processing, process mining, and analytics under a common orchestration layer, extending automation to end-to-end processes such as full procure-to-pay flows [30]. Practitioner studies describe orchestration embedded in ERP landscapes, including financial-close control frameworks [23] and agent platforms native to enterprise suites [37]. The trade-off is coordination complexity: more moving parts demand more governance and maintenance.

**Agentic AI (2023 present).** Large language models enable a qualitatively different paradigm: systems that decompose goals into multi-step plans, call tools and APIs through function-calling interfaces, retain memory across interactions, and explain intermediate reasoning [14, 15]. Reported applications include autonomous audit-evidence gathering and policy-compliance analysis [17], constraint-aware adoption frameworks for corporate finance [18], and end-to-end transaction lifecycle orchestration [28]. The defining shift is to goal-driven autonomy: the agent, not the designer, decides the execution path.

### 2.2 Positioning Against Prior Reviews

Prior systematic reviews examined RPA within business process management [1, 3, 5] and RPA or AI adoption in accounting and auditing [6, 7, 8, 9]; all predate the agentic turn and propose no classification scheme extending to LLM-based agents. General surveys of LLM-based autonomous agents [14, 15] are domain-agnostic, omitting the control, auditability, and regulatory constraints that dominate accounting deployments. This review bridges the two literatures, extending accounting-automation scholarship to the agentic generation and grounding agent-architecture concepts in the governance realities of the finance function.

## 3 Review Methodology

This systematic review followed PRISMA 2020 guidelines [11] and established procedures for evidence-based software-engineering and information-systems reviews [12, 13]. The protocol research questions, search string, and inclusion/exclusion criteria was fixed before screening began.

### 3.1 Search Protocol

The master Boolean search string paired an automation-term group ("robotic process automation" OR RPA OR "intelligent process automation" OR hyper automation OR "agentic AI" OR "LLM agent*" OR "autonomous agent*") with an accounting-term group (accounting OR audit* OR "accounts payable" OR "accounts receivable" OR reconciliation OR "financial close" OR "finance function"), applied to titles and abstracts, restricted to English-language records published 2018–2026. Searches were executed on 11 June 2026 against the Semantic Scholar Academic Graph, which aggregates Crossref, DBLP, PubMed, arXiv, and publisher feeds; because the API does not accept the full master string, it was decomposed into 24 logically equivalent sub-queries crossing each automation term with each accounting term. Copy-ready strings for Scopus, Web of Science, IEEE Xplore, and the AIS eLibrary are documented in supplementary materials for replication.

The search retrieved 2,387 records; deduplication by paper identifier, DOI, and normalised title removed 574 duplicates (24% inter-query overlap), leaving 1,813 unique records. Within these, 67% carry Crossref DOIs, 29% arXiv, 20% DBLP, and 6% PubMed identifiers (categories overlap), confirming that publisher-registered and computer-science preprint sources dominate coverage. The supplementary Scopus, Web of Science, IEEE Xplore, and AIS eLibrary searches needed chiefly to reach regional accounting journals under-indexed by the aggregator have not yet been executed; their documented strings bound the present coverage claim (Sect. 3.4).

### 3.2 Inclusion and Exclusion Criteria

Records were included when they (I1) address automation technology applied within not merely mentioning accounting or finance processes; (I2) address at least one of the four generations or a hybrid; (I3) are peer-reviewed articles, conference papers, book chapters, or preprints (flagged as grey literature); (I4) are in English; and (I5) were published 2018–2026. Records were excluded as (E1) acronym homonyms "RPA" as random-phase approximation, remotely piloted aircraft, or recombinase polymerase amplification; (E2) automation outside the finance function; (E3) editorials without research content; (E4) student theses; or (E5) non-English texts.

### 3.3 Screening and Selection

Title-and-abstract screening applied the criteria to all 1,813 unique records, logging an explicit exclusion reason per decision: 44 homonyms (E1), 52 non-finance applications (E2), 9 editorials (E3), 11 theses (E4), and 1,309 records failing inclusion, for 1,425 exclusions. The surviving 388 records entered a full-text proxy assessment scoring recency, peer-review status, empirical method, generation specificity, subfunction specificity, and abstract completeness; the 60 top-scoring studies were selected for extraction. Figure 1 depicts the PRISMA 2020 flow diagram of the four-stage process.

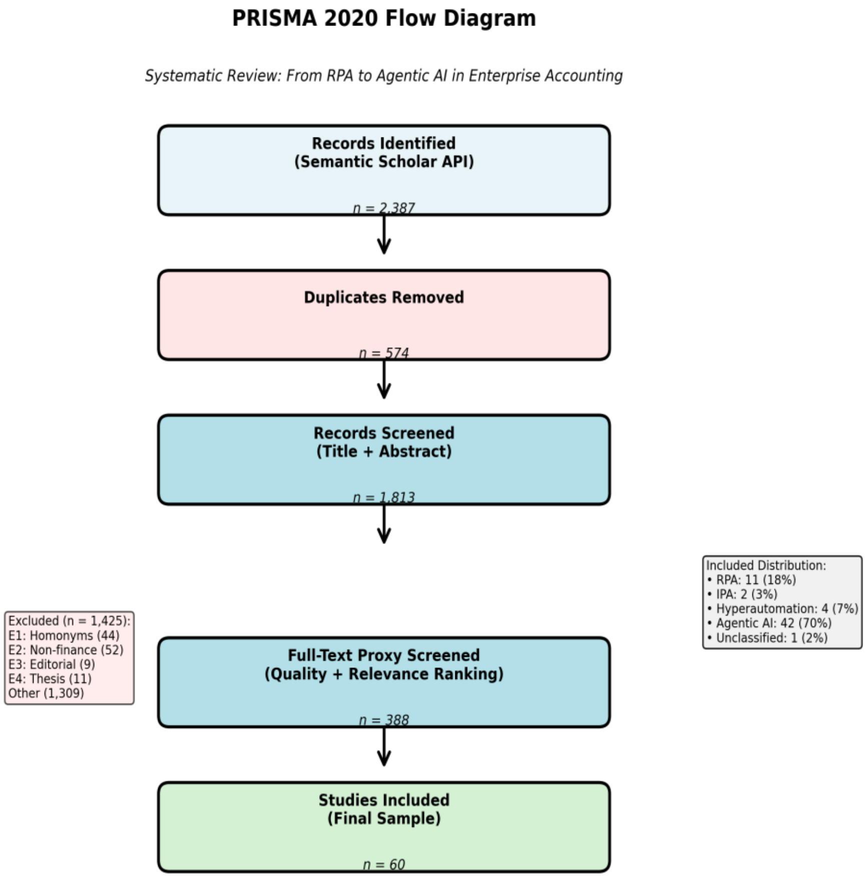


**Fig. 1.** PRISMA 2020 flow diagram of the four-stage screening process.

For each included study, we extracted year, venue, generation, accounting subfunction(s), evaluation method, autonomy level, and key findings relevant to RQ1–RQ4.

***Screening reliability:*** To quantify screening reliability, a second reviewer an independent LLM-based screener applying the written criteria, blinded to first-pass decisions re-screened a random subset of 100 records (5.5% of the screening pool). Raw agreement was 84% (Cohen's $\kappa = 0.43$; prevalence-adjusted PABAK = 0.68). All 16 disagreements fell into two borderline classes: domain-generic agent studies that used audit or accounting vocabulary in non-financial senses, and records lacking abstracts, for which the criterion-based first pass was conservative. The author adjudicated all 16 disagreements against the written criteria; adjudication upheld the original decisions, clarifying that criterion I1 admits agent-governance studies whose methods bear directly on accounting deployment contexts four such records appear among the final 60. The I1 domain boundary nonetheless remains the least reliable screening judgment (Sect. 3.4).

### 3.4 Threats to Validity

*Construct validity.* The four-generation framing was imposed ex ante; to mitigate anchoring, the taxonomy in Section 5 was derived inductively from the corpus rather than from the framing, thereby improving internal *validity.* Primary screening was single-pass; pre-specified criteria and logged exclusion reasons provide auditability, and the blinded second-pass subset yielded 84% raw agreement ($\kappa = 0.43$), with all disagreements resolved by author adjudication (Sect. 3.3) conclusions about agent governance should be read knowing the corpus deliberately admits domain-adjacent trustworthiness studies. The taxonomy dimensions and capability framework have not yet been validated by an independent expert panel or piloted with practitioners; both are planned (Sect. 7.2). *External validity.* Coverage rests on a single if broad aggregator whose identifier profile (Sect. 3.1) skews toward publisher-registered and computer-science sources; regional accounting journals are likely under-indexed, and the documented Scopus, Web of Science, IEEE Xplore, and AIS eLibrary searches remain to be executed. *Conclusion validity.* All 60 included studies date from 2025–2026: the recency component of the quality score, citation-velocity weighting in the source database, and the genuine post-2024 surge in agentic research jointly crowded out 2018–2024 empirical work. The synthesis is therefore weighted toward the agentic frontier, and generational claims about RPA and IPA rest on the foundational 2016–2024 literature [1–9] rather than on systematically selected studies from those years; re-running the selection with a relaxed recency weight to admit foundational empirical studies is a planned robustness check.

## 4 Findings: The Evolution of Intelligent Automation in Accounting

### 4.1 Generational Overview

Table 1 compares the four generations across eight defining attributes. The trajectory is one of steadily increasing autonomy, learning capability, process scope, and integration depth: from deterministic task scripts operating on user interfaces, through supervised models embedded in workflows, to orchestrated multi-tool platforms, and finally to goal-driven agents reasoning over enterprise data.

**Table 1.** Four generations of intelligent automation in accounting.

| Attribute | RPA | IPA | Hyperautomation | Agentic AI |
|---|---|---|---|---|
| **Era** | 2016–2023 | 2018–2024 | 2020–present | 2023–present |
| **Trigger** | Rule-based, event-driven | Rules + ML filters | Orchestration + rules + ML | Goal-driven; the agent decides |
| **Decision logic** | Hard-coded IF-THEN rules | Rules + ML predictions | Composite (rules, ML, heuristics) | LLM reasoning; chain-of-thought |
| **Learning** | None; deterministic | Supervised ML (labelled data) | Limited (offline training) | In-context; multi-step; adaptive |
| **Autonomy** | Scripted | Assisted | Supervised-agentic | Autonomous-agentic |
| **Process scope** | Task-level | Process-level | End-to-end | End-to-end (adaptive) |
| **Integration** | UI automation | API | Deep API + orchestration | Native/embedded |
| **Typical tasks** | Invoice entry, GL posting | Invoice coding, AR matching | Full P2P, full O2C, close | Autonomous audit, reconciliation |

Within the 60-study sample, agentic AI dominates with 42 studies (70%), followed by RPA (11; 18%), hyperautomation (4; 7%), IPA (2; 3%), and one unclassifiable study (2%)—reflecting both the acceleration of agent research after frontier models reached accounting-task competence in 2023–2024 and the recency weighting of our quality score. The thin IPA representation suggests the label has been absorbed: its techniques persist inside hyperautomation stacks and agent toolchains rather than as a standalone research category.

### 4.2 Subfunction Coverage and Gaps

**Strong coverage.** Audit and assurance dominate (52 studies, 87%), spanning AI-driven financial-reporting accuracy [16], algorithmic-auditor applications [21], zero-touch compliance auditing through LLM-based agents [20], and field evidence on audit quality in multinational firms [34]. Audit's regulatory mandate makes automation both visible and governance-critical, and researchers have gravitated toward the auditability and explainability challenges agentic systems raise.

**Moderate coverage.** Procure-to-pay and order-to-cash each appear in 12 studies (20%), consistent with their document-intensive character invoice processing pipelines [25], payment-card reconciliation automation [29], and AI-supported revenue assurance [22]. Reconciliation appears in 9 studies (15%), with a discernible progression from rule-based matching toward multi-agent variance resolution [40].

**Thin coverage.** Record-to-report and financial close (6 studies, 10%), tax (4; 7%), and FP&A (3; 5%) remain under-represented despite their prominence in finance-transformation budgets. Existing close studies concentrate on ERP-embedded control frameworks [23], centralised close optimisation [32], and AI-driven period-end closing in cloud ERP [38]; tax and FP&A combine jurisdictional variability with forward-looking judgment properties resisting both hard-coded rules and offline-trained models. Industry analyses confirm adoption pressure across the profession [19, 24, 26, 35], sharpening the contrast with thin academic coverage; we return to these gaps in Section 7.

### 4.3 Temporal and Venue Distribution

The sample skews sharply recently: 24 studies from 2025 and 36 from 2026 (partial year). In the 2018–2024 study, only those that survived the quality-weighted selection, a pattern attributable to citation-velocity weighting in the source database and to the post-2024 explosion of agentic research. 57 of the 60 studies (95%) are peer-reviewed journal or conference publications; the remaining 3 are preprints. Venues span IEEE and ACM proceedings, specialist accounting and information-systems journals, and interdisciplinary outlets evidence of the fragmentation this review consolidates. The 2025–2026 concentration is a deliberate design consequence with an interpretive cost: longitudinal claims, such as the absorption of IPA into later stacks, rest on the foundational 2016–2024 literature [1–9] rather than on corpus-internal evidence, and should be read accordingly (Sect. 3.4).

## 5 A Taxonomy of Intelligent Automation Systems

### 5.1 Development Method

We followed Nickerson et al.'s iterative method [10], chosen because the corpus provides empirical grounding for inductive dimension discovery, while the nascent agentic literature offers no framework for deducing dimensions. Iteration 1 (empirical-to-conceptual) coded studies against provisional dimensions, from which six stable dimensions emerged; Iteration 2 (conceptual-to-empirical) applied the refined schema to all 60 studies. Ending conditions were verified: every study classified, exactly one characteristic per dimension, no new dimensions in the second iteration, and meaningful discrimination across the population. Related efforts notably a taxonomy of trustworthy multi-agent LLM coordination [19] address security properties of agent systems, not the accounting deployment context targeted here.

### 5.2 Dimensions and Characteristics

**D1 Autonomy level.** *Scripted* systems execute fully predetermined paths. *Assisted* systems recommend actions that humans validate before execution. *Supervised-agentic* systems make high-confidence decisions autonomously while escalating exceptions. *Autonomous-agentic* systems self-direct end-to-end, with humans setting guardrails and reviewing outcomes post hoc.

**D2 Intelligence and learning.** *Deterministic* systems embody fixed rule logic. *ML-augmented* systems learn from labelled data and are retrained offline. *LLM-reasoning* systems perform in-context, multi-step reasoning and generalise to scenarios absent from any training set.

**D3 Process scope.** Coverage ranges from a single *task* (e.g., invoice entry), through a multi-step *process* within one subfunction (e.g., procure-to-pay from purchase order to payment), to an *end-to-end function* such as the complete financial close.

**D4 Human oversight.** *In-the-loop* configurations require human validation of every action; *on-the-loop* configurations have humans monitoring autonomous execution and intervening on anomalies; *out-of-loop* configurations confine humans to post-hoc review above defined escalation thresholds. Oversight design interacts directly with internal-control requirements, since control frameworks presume identifiable approval points [31].

**D5 Integration depth.** *UI/surface* integration scrapes screens and is brittle to interface changes; API integration connects systems through stable service interfaces; native/embedded integration embeds automation within the platform, providing direct access to data and workflow state.

**D6 Accounting subfunction.** The primary process addressed: procure-to-pay, order-to-cash, record-to-report, reconciliation, audit, tax, or FP&A, as characterized in Section 4.

### 5.3 Corpus Distribution

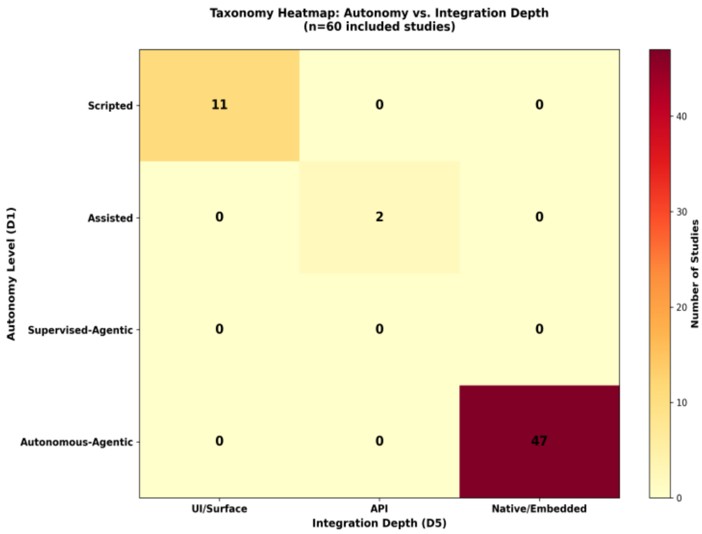


**Fig. 2.** Taxonomy heatmap: distribution of the 60 included studies across autonomy level (D1) and integration depth (D5).

Figure 2 maps the 60 studies across autonomy (D1) and integration depth (D5). The distribution concentrates in the autonomous-agentic × native/embedded cell (47 studies, 78%): current research targets deeply integrated, highly autonomous designs rather than incremental extensions of surface-level bots. Near-empty intermediate cells suggest the field is leapfrogging the assisted/API quadrant, raising the governance questions of Section 7 precisely. The full classification of all 60 studies is in the supplementary materials .

## 6 Capability Assessment Framework

### 6.1 Maturity Levels

To make the taxonomy actionable, we extend dimensions D1–D5 into five-level maturity ladders (L0–L4). For autonomy (D1): L0 (no automation), L1 (scripted), L2 (assisted), L3 (supervised-agentic), L4 (autonomous-agentic). Analogous ladders govern learning (none → deterministic → ML-augmented → LLM-reasoning → continuous learning), scope (none → isolated task → multi-step task → multi-subfunction → end-to-end function), oversight (none → in-the-loop rule → in-the-loop exception → on-the-loop → out-of-loop), and integration (none → UI/surface → API → deep API → native/embedded). Assigning one level per dimension yields a five-axis capability signature for benchmarking, gap analysis, and roadmap design.

### 6.2 Four Archetypes

Four archetypal system classes, recurrent across the corpus, anchor the framework.

**Archetype 1 Classic RPA bot (L1).** A rule-based bot automating one task over the user interface, such as accounts payable invoice data entry [6, 29]. It fails due to formatting deviations and interface changes, leading to manual exception queues.

**Archetype 2 ML-enhanced IDP pipeline (L2).** An intelligent document-processing pipeline combining OCR, document classification, and learned field extraction, with humans validating predictions before posting [25]. Technically sophisticated yet bounded by its training data and by mandatory human validation.

**Archetype 3 LLM-assisted financial close (L3).** An LLM-based assistant supporting the close with journal-entry suggestions, accrual narratives, and variance explanations; the close team retains escalation authority [23, 38]. Oversight is in-the-loop for critical entries, on-the-loop for routine ones.

**Archetype 4 Autonomous multi-agent reconciliation (L4).** A multi-agent system that reconciles general-ledger accounts against subledgers, explains variances, and auto-posts routine reconciling entries, with agents coordinated through a central orchestrator and humans reviewing only high-variance items [40]. Emerging agent platforms embedded in major ERP suites indicate this archetype is approaching production viability [37].

Figure 3 visualises the archetypes as radar plots across the five maturity dimensions. The widening profile from Archetype 1 to 4 makes the capability frontier legible at a glance. In contrast, asymmetric profiles expose targeted improvement opportunities—an Archetype 2 deployment peaked at integration but was flat at autonomy, signalling that the constraint is trust, not technology.

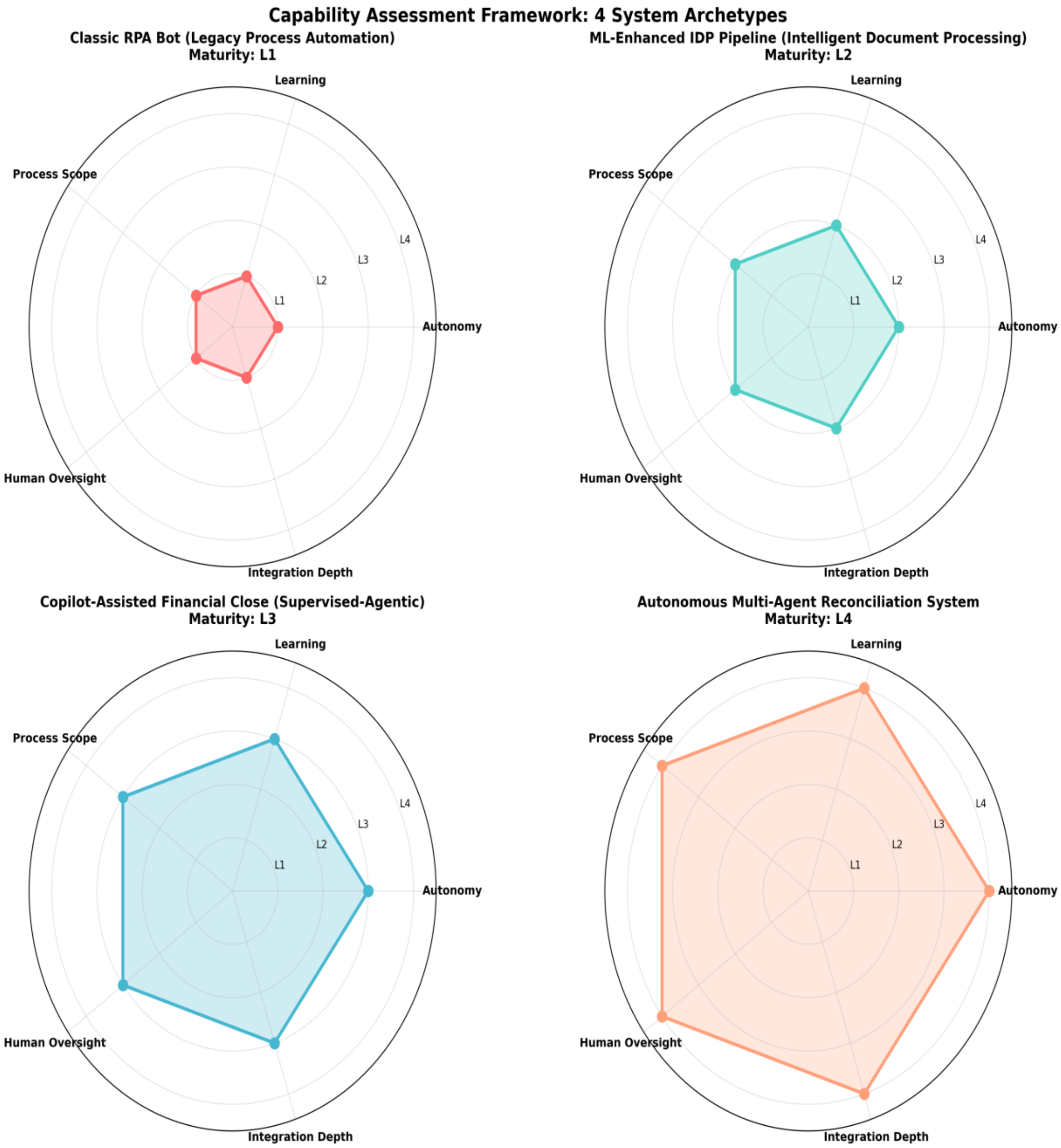


**Fig. 3.** Capability maturity radar profiles of the four archetypes across dimensions D1–D5.

### 6.3 Applying the Framework

An assessment proceeds in five steps: (1) scope a single deployment serving a defined subfunction (D6) as the unit of assessment; (2) gather evidence—design documents, control descriptions, and run logs—covering triggers, decision logic, learning mechanism, escalation paths, and integration interfaces; (3) assign each dimension D1–D5 the highest maturity level whose defining condition the evidence fully satisfies, defaulting to the lower level when uncertain; (4) compare the resulting five-level signature against the nearest archetype, investigating any dimension that deviates by two or more levels—in particular, an autonomy level exceeding the oversight level signals governance debt requiring immediate control review; and (5) set a target signature and sequence investments, treating integration and oversight maturity as preconditions for autonomy increases. Table 2 condenses the procedure into a practitioner checklist. The framework's decision rules derive from the corpus synthesis;

structured validation—expert panel review of the dimensions and a multi-site practitioner pilot—remains future work (Sect. 7.2).

**Table 2.** Practitioner capability-assessment checklist.

| Step | Key questions and decision rule |
|---|---|
| 1. Scope (D6) | Which subfunction does the deployment serve? One deployment = one assessment unit. |
| 2. Evidence | Are design documents, control descriptions, and run logs available for triggers, decision logic, learning, escalation, and integration? |
| 3. Profile D1–D5 | Assign the highest level (L0–L4) for each dimension with full evidence; if uncertain, assign the lower level. |
| 4. Compare & govern | Match the signature to the nearest archetype; investigate deviations ≥2 levels. If autonomy (D1) exceeds oversight (D4), freeze autonomy increases pending control review. |
| 5. Roadmap | Define the target signature; sequence investments with integration and oversight maturity as preconditions for autonomy upgrades; reassess each control cycle. |

## 7 Research Gaps, Future Directions, and Conclusion

### 7.1 Five Critical Gaps (RQ4)

**Gap 1 Governance and auditability of autonomous agents.** Out-of-loop autonomy collides with control frameworks that presume pre-authorisation of material postings [31]. Early work explores agentic compliance in adjacent domains financial crime [33] and principal-agent liability [39] but no included study proposes a control architecture under which an autonomous agent's postings would satisfy internal-control attestation requirements. Hybrid designs (agents draft, humans commit) and cryptographically anchored decision logs are promising directions.

**Gap 2 Hallucination and fabrication risk.** LLM-based agents can generate plausible but false outputs; in accounting, fabricated figures or audit evidence are categorically unacceptable. Surveyed mitigations remain generic [19]; accounting-specific guardrails such as deterministic verification of every numeric claim against systems of record are essentially unstudied.

**Gap 3 Evaluation and benchmarking.** No standardised benchmark exists for accounting agents, unlike in mainstream NLP [15]. A domain benchmark covering matching, journal generation, and variance analysis with gold-standard solutions and audit-trail quality metrics would let researchers and vendors compare systems on common ground.

**Gap 4 Multi-agent orchestration standards.** As deployments scale to agent fleets, coordination, conflict resolution, and shared-context management become first-order problems. Orchestration research advances in other domains [27, 36], but accounting-specific protocols who posts, who approves, how agents share evidential context are absent.

**Gap 5 Explainability and reasoning transparency.** Auditors must verify why an agent posted an entry, what evidence it consulted, and what assumptions it made. Current applications expose conclusions, not reasoning chains [20, 21]; auditor-verifiable explanation formats remain an open problem.

### 7.2 Subfunction Agenda

The thin coverage of record-to-report, tax, and FP&A defines a complementary agenda: close automation requires judgment under audit scrutiny; tax must track shifting, jurisdiction-specific rules, favouring retrieval-grounded reasoning; FP&A requires probabilistic, forward-looking reasoning; rule-based systems cannot supply each property aligning naturally with agentic capabilities. A methodological agenda targets this review's own instruments: executing the documented supplementary database searches to capture regional journal coverage; conducting a Delphi-style expert panel review of the taxonomy dimensions; and conducting a multi-site practitioner pilot of the Section 6.3 assessment procedure with per-dimension reliability reporting.

### 7.3 Conclusion

This review synthesised 60 studies tracing the evolution of intelligent automation in enterprise accounting across four generations. The trajectory is one of rising autonomy, learning, scope, and integration depth, with agentic systems dominating the research frontier (RQ1); coverage concentrates in audit while record-to-report, tax, and FP&A remain underexplored (RQ2); a six-dimensional taxonomy captures the design space and, extended into an L0–L4 maturity framework with four archetypes and an assessment procedure, supports practical capability assessment (RQ3); and five gaps governance, hallucination mitigation, benchmarking, orchestration, and explainability constitute the agenda for trustworthy agentic accounting (RQ4). As finance functions move agents from pilot to production, closing these gaps will determine whether agentic autonomy can coexist with the fiduciary and regulatory accountability on which financial reporting depends.

### References


[1] van der Aalst, W.M.P., Bichler, M., Heinzl, A.: Robotic process automation. Business & Information Systems Engineering 60(4), 269–272 (2018). https://doi.org/10.1007/s12599-018-0542-4

[2] Lacity, M., Willcocks, L.: A new approach to automating services. MIT Sloan Management Review 58(1), 41–49 (2016)

[3] Syed, R., et al.: Robotic process automation: contemporary themes and challenges. Computers in Industry 115, 103162 (2020). https://doi.org/10.1016/j.compind.2019.103162

[4] Hofmann, P., Samp, C., Urbach, N.: Robotic process automation. Electronic Markets 30, 99–106 (2020). https://doi.org/10.1007/s12525-019-00365-8

[5] Ivančić, L., Suša Vugec, D., Bosilj Vukšić, V.: Robotic process automation: systematic literature review. In: Business Process Management Workshops, LNBIP. Springer (2019)

[6] Moffitt, K.C., Rozario, A.M., Vasarhelyi, M.A.: Robotic process automation for auditing. Journal of Emerging Technologies in Accounting 15(1), 1–10 (2018). https://doi.org/10.2308/jeta-10589

[7] Kokina, J., Blanchette, S.: Early evidence of digital labour in accounting: innovation with robotic process automation. International Journal of Accounting Information Systems 35, 100431 (2019). https://doi.org/10.1016/j.accinf.2019.100431

[8] Cooper, L.A., et al.: Robotic process automation in public accounting. Accounting Horizons 33(4), 15–35 (2019)

[9] Tiron-Tudor, A., et al.: Perspectives on how robotic process automation is transforming accounting and auditing services. Accounting Perspectives 23(2) (2024)

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

[11] Page, M.J., et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372, n71 (2021). https://doi.org/10.1136/bmj.n71
[12] Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report (2007)
[13] Webster, J., Watson, R.T.: Analysing the past to prepare for the future: writing a literature review. MIS Quarterly 26(2), xiii–xxiii (2002)
[14] Wang, L., et al.: A survey on large language model-based autonomous agents. Frontiers of Computer Science 18(6), 186201 (2024)
[15] Xi, Z., et al.: The rise and potential of large language model-based agents: a survey. Science China Information Sciences (2025)
[16] Mathur, A.: Impact of artificial intelligence on financial reporting accuracy and efficiency. International Journal of Innovations & Research Analysis (2026). https://doi.org/10.62823/ijira/06.1(ii).8793
[17] Anirudh, K.: Agentic artificial intelligence and its implementation in transforming audit, policy and environmental accounting. Indian Journal of Computer Science and Technology (2026). https://doi.org/10.59256/indjcst.20260501045
[18] Jalaluddin, K.A.: The transition from automation to augmentation: a constraint-aware framework for agentic AI adoption in corporate finance functions. Veredas do Direito (2026). https://doi.org/10.18623/rvd.v23.5632
[19] Vangalapat, T., Shaikh, S.I.: Trustworthy agentic AI: a survey and taxonomy of secure coordination and hallucination mitigation in multi-agent large language model systems. International Journal of Innovative Science and Research Technology (2026). https://doi.org/10.38124/ijisrt/26feb1090
[20] Ameur, M., Brik, B., Ksentini, A.: EUROCOMPLY: enabling zero-touch AI compliance auditing via LLM-based agentic AI. IEEE Communications Magazine (2026). https://doi.org/10.1109/mcom.001.2500366
[21] Melaragni, F.: The algorithmic auditor: seven strategic applications of artificial intelligence in financial reporting and analysis. Journal of Business and Management Studies (2026). https://doi.org/10.32996/jbms.2026.8.5.5
[22] Zameer, S.A., Jagtap, S.: A study on the impact of artificial intelligence on revenue assurance. International Journal of Creative and Open Research in Engineering and Management (2026). https://doi.org/10.55041/ijcope.v2i5.459
[23] Ghory, F.H.: Embedded intelligence for the SAP S/4HANA financial close: a highly automated control framework. International Journal of Intelligent Systems and Data Science (2026). https://doi.org/10.67231/s0h5b367
[24] Sridhar, J.: The role of artificial intelligence in modern accounting systems: impact on accuracy, efficiency, and financial decision-making. International Journal of Science, Strategic Management and Technology (2026). https://doi.org/10.55041/ijsmt.v2i5.266
[25] Dameruppula, S.: Cognitive automation of invoice processing in aviation: an SAP BTP and intelligent document processing framework. Journal of Information Systems Engineering & Management (2026)
[26] Guo, C.: Research on optimisation paths for corporate financial management under digital transformation. International Business & Economics Studies (2026). https://doi.org/10.22158/ibes.v8n3p1
[27] Niu, K., Zhang, Y., Wang, X., et al.: RASO: role-aware shared reflection for multi-agent orchestration in e-commerce long-horizon planning. In: Proceedings of the International Conference on Automated Planning and Scheduling (2026). https://doi.org/10.1609/icaps.v36i1.42873

[28] Chandana, P.K.: Intelligent trade lifecycle orchestration using agentic AI for end-to-end trade automation. Journal of Information Systems Engineering & Management (2025). https://doi.org/10.52783/jisem.v10i62s.13721

[29] Alla, P.B.: Automating P-card reconciliation with UiPath and Microsoft Power Automate: a transformational approach to streamlining financial operations. In: 2025 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA) (2025). https://doi.org/10.1109/ICAIQSA67794.2025.11440607

[30] Gupta, V.: Architecting intelligent SAP workflows: a blueprint for hyperautomation with BRIM and RAR. Quantum Horizons Review Journal (2025). https://doi.org/10.63665/qhrj.v01i01.1

[31] Rusyn-Hrynyk, R., Stoianovskyi, A., Borniak, O.: The internal control system under the conditions of digital transformation of the enterprise. Business Navigator (2025). https://doi.org/10.32782/business-navigator.83-42

[32] Sharma, A.: Optimising financial closing processes with SAP S/4HANA Central Finance: a case study. International Journal of Computer Science and Information Technology Research (2025). https://doi.org/10.63530/ijcsitr_2025_06_04_003

[33] Axelsen, H., Licht, V., Damsgaard, J.: Agentic AI for financial crime compliance. In: Hawaii International Conference on System Sciences (2025). https://doi.org/10.48550/arXiv.2509.13137

[34] Olabisi, D., Olawaye, O.: Artificial intelligence and audit quality in multinational audit firms in Nigeria. International Journal of Social Science, Humanity & Management Research (2025). https://doi.org/10.58806/ijsshmr.2025.v4i10n09

[35] Cheng, X., Liu, Y., Lyu, S., et al.: The opportunities and challenges brought by AI tools to the accounting industry. SHS Web of Conferences (2025). https://doi.org/10.1051/shsconf/202521804002

[36] Chandrasekaran, M.: AURELIUS: agentic uncertainty-reasoning and learning with a hierarchical Bayesian multi-agent unified system for handling non-determinism. In: International Conference on Computing, Management and Telecommunications (2025). https://doi.org/10.1109/ComManTel68363.2025.11368546

[37] Prakashkumar, P.K.R.: Analytical study of adapting Oracle AI Agent Studio into Oracle ERP. International Journal of Entrepreneurship, Innovation, and Business Strategies (2025). https://doi.org/10.58425/ijeib.v3i1.452

[38] Shrivastava, G., Peng, S. L., Bansal, H., Sharma, K., & Sharma, M. (Eds.). (2020). New age analytics: Transforming the internet through machine learning, IoT, and trust modeling. CRC Press.

[39] Gabison, G., Xian, R., et al.: Inherent and emergent liability issues in LLM-based agentic systems: a principal-agent perspective. In: Proceedings of the 1st Workshop for Research on Agent Language Models (REALM) (2025). https://doi.org/10.18653/v1/2025.realm-1.9

[40] Mohan, B., Dayal, S., Khandelwal, A., et al.: From batch to real-time: multi-agent AI framework for autonomous financial reconciliation in wealth management. In: Educational Sciences International Conference (2026). https://doi.org/10.1109/ESIC68176.2026.11496226

[41] Juvvadi, R.R.: Robotic process automation (RPA) in accounting: measuring ROI and workforce displacement. International Journal of Intelligent Systems and Applications in Engineering 6(4), 378–385 (2018)