International Journal of Knowledge Processing Studies
(KPS)
Homepage: http://kps.artahub.ir/

ORIGINAL RESEARCH ARTICLE

A Conceptual Review of Agentic Banking and the Linkage of Autonomy, Architecture, and Governance

Reza Eyvazi Fesghandisi[1], Mohammad Hasanzadeh[2*], Zahra Alipour Darvishi[3], Maryam Shoar[4]

[1] *Ph.D. candidate of Information Technology Management (Business Intelligence), Department of Management, NT.C., Islamic Azad University, Tehran, Iran. reza.eyvazi@iau.ac.ir, 0009-0003-5336-7172.*
[2*] *Professor, Dept. of Knowledge and Information Science, Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran (corresponding author). hasanzadeh@modares.ac.ir, 0000-0002-6175-0855.*
[3] *Associate Professor, Department of Management, NT.C., Islamic Azad University, Tehran, Iran. z_alipour109@iau.ac.ir, 0000-0001-5546-7159.*
[4] *Assistant Professor, Department of Management, NT.C., Islamic Azad University, Tehran, Iran. m.shoar@iau.ac.ir, 0000-0002-7811-3924.*

ARTICLE INFO

Article History:
Received: 2026-03-01
Revised: 2026-04-24
Accepted: 2026-05-27
Published Online: 2026-06-01

Keywords:
Agentic Banking; Autonomous AI Agent; AI Governance; Human Oversight; Autonomy Levels.

Number of Reference: 46
Number of Figures: 5
Number of Tables: 7

DOI: 10.22034/kps.2026.605571.1317

CC BY

ABSTRACT

Autonomous AI agents have given rise to agentic banking, yet the literature remains fragmented, and its distinction from conventional smart banking, its architecture and its governance requirements lack a coherent formulation. This conceptual integrative review develops an integrated framework of agentic banking from 45 sources (regulatory documents, peer-reviewed studies, working papers, preprints and industry reports) published mainly between 2024 and 2026. Sources were identified through institutional websites and scholarly search platforms; a documented OpenAlex search (four Boolean strings, 1,988 records, two screening stages) and citation tracking then verified and extended the corpus, and sources were coded with an extraction form. Agentic banking differs from traditional automation and from analytics-driven smart banking: an intelligent agent can interpret a goal, plan, use tools and act within a defined scope of authority. Industry reports indicate wide piloting of such agents, whereas peer-reviewed disclosure data place adoption at an early stage, and a survey of finance leaders suggests that controls lag behind ambitions. No dedicated Persian-language source on agentic banking was found. The proposed framework treats the risk exposure of an agent's action, defined by the likelihood, severity, scope and reversibility of an adverse outcome, as the joint outcome of agentic capabilities, architectural enablers and autonomy level, and proposes that governance intensity be proportionate to that exposure rather than to any single dimension. The review derives three testable propositions, suggests measures and methods for testing them, and concludes that agentic banking requires coordinated redesign of technical architecture, governance structures and the regulatory framework. ©**authors**.

►**Citation:** Eyvazi Fesghandisi, R., Hasanzadeh, M., Alipour Darvishi, Z., & Shoar, M. (2026). A Conceptual Review of Agentic Banking and the Linkage of Autonomy, Architecture, and Governance. *International Journal of Knowledge Processing Studies (KPS)*, 6(2): 53-73. Doi: 10.22034/kps.2026.605571.1317

1. Introduction

Digital transformation in the banking industry has in recent years moved beyond channel digitalization and data analytics to a stage in which artificial intelligence (AI) is no longer only a tool for analysis or prediction. It increasingly appears as agents that can interpret a goal, plan how to reach it, and advance part of a bank's processes with full or partial autonomy (Aldasoro et al., 2024; Rizinski & Trajanov, 2026). This phenomenon, known in the emerging literature as agentic banking, raises new questions about technical architecture, governance patterns, and the boundaries of responsibility within the banking system.

Two developments make these questions pressing. First, adoption is advancing faster than the evidence about it. An industry survey reports wide piloting of AI agents in banks (The Financial Brand, 2025), whereas peer-reviewed disclosure data place adoption at an early stage (Ul Mustafa & Aysan, 2026). Second, supervisors have begun to address AI agents explicitly (Infocomm Media Development Authority [IMDA], 2026; Monetary Authority of Singapore [MAS], 2025). Banks therefore have to decide how much autonomy to grant an agent, and how much control to place around it, before a settled body of evidence exists.

The literature that could guide this decision is divided by topic: definitions and taxonomies (Bandi et al., 2025; Sapkota et al., 2026), levels of autonomy (Feng et al., 2025), trust and acceptance (Pathak & Bansal, 2024), financial agents and their architecture (Huang et al., 2025; Rizinski & Trajanov, 2026) and governance requirements (IMDA, 2026; Kurshan et al., 2026). Each strand treats one dimension. In the sources reviewed, no formulation states how an agent's capabilities, its architectural access and its autonomy together determine the risk of its actions, and how governance should respond. No dedicated Persian-language study was found either (Section 2.5).

This paper aims to review and develop an integrated conceptual picture of this phenomenon: What is agentic banking, and how does it differ from traditional automation and conventional smart banking? What architecture and levels of autonomy are observed in the literature and in practice? What risks and governance requirements have been raised? And what is the current state of the domestic (Persian-language) body of knowledge in this area?

Contribution of this study. This paper makes three contributions. First, it synthesizes a fragmented literature into a single conceptual framework (Section 5) that connects five elements with distinct roles. Three are determinants: agentic capabilities (goal interpretation, planning, tool use), architectural enablers (data, APIs, orchestration, identity), and autonomy level. The fourth is the risk exposure of an action, defined by four dimensions (likelihood, severity, scope and reversibility), as the intermediate outcome, and the fifth is governance intensity, as the response proportionate to that exposure. Second, it sets the basis for calibrating governance at the joint risk exposure produced by the three determinants and not at any single one of them, a relationship that the sources reviewed do not formulate explicitly. Third, it translates this formulation into three testable propositions, with suggested measures and methods (Appendix C), and into practical implications for banks and supervisors.

The study is a conceptual integrative review (Snyder, 2019) of 45 sources, focused on retail banking at autonomy levels 2 and 3; its procedures are reported in Section 3. Section 2 reviews the literature, Section 4 presents the synthesis, Section 5 the framework and its propositions, and Section 6 the discussion and limitations.

2. Literature Review

2.1 Evolution of banking technology and definition of agentic banking

Industry analysis published by the banking technology firm Backbase (Rutten, 2026) identifies four generations of banking technology: digital channels, engagement platforms, AI-powered platforms, and, most recently, an AI-native banking operating system that coordinates employees, operations, and intelligent agents on a single, governed foundation. This industry narrative aligns with a more precise academic

formulation. Gong (2026) describes a shift in financial automation from isolated prediction tasks to integrated decision systems and defines agentic finance as financial market environments in which autonomous or semi-autonomous AI systems participate in information processing, decision support, monitoring, and execution workflows. The convergence of these narratives suggests that agentic banking is the latest stage of a longer evolution rather than an isolated technological change.

Agentic banking can be defined as a form of smart banking in which artificial intelligence is no longer limited to passive analytical tools, such as intelligent assistants or simple conversational systems, and takes the form of agents that draw on tools and databases, interact with other systems, and take necessary action within a defined scope. What defines agentic banking, therefore, is not the addition of artificial intelligence to a bank but the reorganization of its processes around agents that act (Aldasoro et al., 2024; Rizinski & Trajanov, 2026).

The operational definition adopted in this paper follows from this: agentic banking refers to a system in which the intelligent agent simultaneously exhibits five properties: purposeful but bounded autonomy, context-aware perception, multi-step agency, interoperability with core banking systems, and controlled adaptivity. The emerging literature on agentic AI has not yet converged on a single set of necessary conditions; these five properties are therefore used here as an operational definition, and a system lacking any one of them is treated in this paper as conventional smart banking.

What sets agentic banking apart from traditional robotic process automation (RPA) is its degree of independence, flexibility, and decision-making capacity. A traditional automation system simply runs a fixed sequence of predefined rules and stops or fails when it meets an unforeseen situation, whereas an intelligent agent can break a problem into smaller parts, choose a suitable tool for each, and ask a human for confirmation or additional data whenever it encounters ambiguity (Okpala et al., 2025). This distinction is also reflected in supervisory practice: the Bank of Thailand's guiding principles for AI risk management explicitly include agentic AI within the scope of AI systems while excluding rule-based automation such as robotic process automation (Bank of Thailand, 2025).

From an institutional perspective, a related four-stage evolution of digital banking can be distinguished: channel digitalization; data analytics and algorithmic scoring; generative AI and natural human–machine interaction; and now the agentic stage, in which AI systems are planners and executors rather than mere generators or predictors (Aldasoro et al., 2024). Aldridge et al. (2026) formulate this distinction around three defining features that separate financial agents from traditional trading algorithms: goal-directed autonomy, continuous learning, and multi-agent coordination, and note that the literature on autonomous, multi-agent financial systems remains scattered across disciplines.

2.2 AI agents and agentic AI

To sharpen this definition, two closely related concepts that are sometimes used interchangeably should be distinguished: AI agents and agentic AI. In their conceptual taxonomy, Sapkota et al. (2026) characterize AI agents as modular systems driven by large language models and designed for narrow, task-specific automation, whose capability derives from tool integration, prompt engineering, and reasoning enhancements. Agentic AI, by contrast, is described as a paradigm shift marked by four features: multi-agent collaboration, dynamic task decomposition, persistent memory, and coordinated autonomy. In a systematic review of 143 studies, Bandi et al. (2025) likewise identify the distinctive feature of agentic AI as the combination, within a single system, of strategic planning, memory that preserves context over time, the use of external tools, and collaboration with other agents. They also show that more than ninety percent of the studies in this area were published in 2024 and 2025, which indicates a field whose definitions and frameworks have not yet stabilized.

This distinction carries specific implementation requirements. First,

persistent memory requires a data infrastructure that retains the context of prior interactions and decisions in a secure and traceable manner. Second, dynamic task decomposition and multi-agent collaboration require an orchestration layer that coordinates the role of each agent, the sequence of actions, and the boundaries of their authority. Third, tool use requires standardized application programming interfaces and access control based on each agent's identity. Fourth, Sapkota et al. (2026) list hallucination, brittleness, and emergent behavior as characteristic challenges; these turn continuous monitoring and complete logging of agents' actions from an option into a requirement.

For agentic banking, the significance of this distinction is that many banking processes are inherently multi-step and cross-unit. Assessing a credit application, for example, requires gathering data from several systems, checking compliance, analyzing risk, and preparing a report; a single task-specific agent covers only part of this chain, whereas an agentic system can divide these steps among specialized agents and coordinate them while preserving the context of the case. Multi-step banking processes therefore typically require a transition from individual agents to coordinated agentic systems. That coordination, however, also widens the risk exposure of an action and reinforces the need for governance intensity proportionate to that exposure (Section 5). Table 1 summarizes this distinction.

Table 1. *Comparison of AI agents and agentic AI and implications for agentic banking (middle columns based on Sapkota et al., 2026, and Bandi et al., 2025; last column is this paper's inference; the patterns are typical and the boundary between the two categories is not absolute)*

Dimension	AI agents	Agentic AI	Implication for agentic banking
Scope of action	A specific, bounded task within a single workflow	Multi-step goals across several workflows	Covering cross-unit processes such as credit assessment
Structure	A single, modular agent	Several specialized, collaborating agents	Need for an orchestration layer
Task decomposition	Predefined	Dynamic	Explicit boundaries of each agent's authority
Memory	Limited to the current interaction	Persistent and context-preserving	Secure and traceable data infrastructure
Autonomy	Bounded to the task	Coordinated across agents	Governance intensity proportionate to risk exposure
Characteristic challenges	Hallucination and brittleness	In addition, emergent behavior	Continuous monitoring and complete action logging

2.3 Applications and levels of autonomy

An intelligent agent in financial services connects data, analysis, decision and action. The five properties listed in Section 2.1 are visible across customer interaction, credit and lending processes, fraud and money-laundering detection, payment operations and liquidity management, and corporate back-office automation (Aldasoro & Desai, 2025; Okpala et al., 2025; Yang et al., 2025). A spectrum of specialized agents can be identified in the industry, including customer, account, credit, fraud and security-monitoring, product, marketplace, governance, and blockchain-node agents, which illustrates one common implementation pattern of agentic banking based on specialized agents (Huang et al., 2025).

Artificial intelligence in financial services may appear either as a decision aid or as a delegated agent (Pathak & Bansal, 2024), spanning four graduated levels of autonomy: informational advisor, active recommender, supervised executor, and delegated (autonomous) agent. As an agent moves toward an acting role, the need for transparency, explainability, access control, and human oversight correspondingly increases (Mitchell et al., 2025). For the near term, Gong (2026) expects bounded autonomy, with agents working as supervised co-pilots and constrained execution modules inside human decision processes.

2.4 Risks and governance

On the governance side, Kurshan et al. (2026), in a working paper, identify multi-

layered risks associated with autonomous agents, including errors of generative models, bias in algorithms, breaches of privacy, adversarial attacks and automation bias, and argue that traditional model-risk frameworks are incompatible with the continuous-learning, emergent-behavior nature of multi-agent systems. Leading regulators and standard-setters, including the European Union (AI Act, Digital Operational Resilience Act [DORA]), the Bank of Thailand, and the National Institute of Standards and Technology (NIST), have adopted or proposed differing but converging approaches emphasizing transparency, human oversight, and technical documentation (Bank of Thailand, 2025; European Parliament & Council of the European Union, 2022, 2024; NIST, 2023, 2024; Vassilev et al., 2025).

Three recent institutional developments clarify where this discussion stands. First, the Financial Stability Board (FSB) warns that the rapid adoption of artificial intelligence may amplify existing vulnerabilities of the financial sector, including third-party dependencies, market correlations, cyber risk, and model risk (FSB, 2024); its 2025 monitoring report reiterates the need for continuous monitoring of adoption and for cooperation among authorities (FSB, 2025). Second, IMDA has published a governance framework dedicated to agentic AI, organized around four dimensions: assessing and bounding risk before deployment, meaningful human accountability, technical and process controls, and end-user responsibility (IMDA, 2026). This framework makes the permissible level of autonomy of each agent conditional on a risk-based approach, taking the domain's tolerance for error, access to sensitive data, and the reversibility of actions as its basis. This is the same proportionality logic on which the framework proposed here rests. Third, in November 2025 MAS issued a consultation paper proposing guidelines on AI risk management for all financial institutions; the proposed guidelines cover AI agents and are to be applied in proportion to the size of each institution and the risk profile of its AI use (MAS, 2025).

An industry forecast points in the same direction. According to Gartner (2026), a single governance model applied to every agent, irrespective of its autonomy level and scope of authority, leads deployments to fail, and by 2027 forty percent of enterprises will demote or decommission their autonomous agents because of governance gaps that are identified only after incidents in production. This figure is an analyst prediction for enterprises in general, not a survey result.

Peer-reviewed studies of 2025 and 2026 specify these risks. Restrepo Amariles et al. (2026) identify liability gaps in payments initiated by AI agents and argue that payment rules designed for human-intermediated transactions fit them poorly. Akkireddy et al. (2026) report that, in retail banking, instructions injected through retrieved documents remain the main attack path and that current authorization scopes are too coarse for the bounded delegation an agent needs. Kshetri (2025) names data quality, compliance, security and human oversight as conditions for trustworthy use.

The sources differ on the basis for setting governance. Grégoire (2026) relates controls to the autonomy level; Feng et al. (2025) treat autonomy as a design decision, capped by a certificate that reflects capability and operational environment. The EU AI Act attaches obligations to the intended purpose of a system (European Parliament & Council of the European Union, 2024). Others hold that the single agent is not a sufficient unit: controls on each agent do not ensure acceptable collective outcomes (de la Chica Rodriguez et al., 2026), and the binding constraint is whether a decision can be verified afterwards, not capability (Han, 2026). Mitchell et al. (2025) reject full autonomy altogether. Section 5.3 compares these positions with the framework.

2.5 Persian-language literature and research gap

A purposive search of Persian-language scholarly sources did not surface a dedicated Persian-language work directly addressing agentic banking. The closest domestic works address digital-transformation governance and the role of artificial intelligence in

banking service transformation, without engaging the distinguishing features of agency (Garoussi et al., 2026 [1405 SH]; Khosropour et al., 2024 [1403 SH]; Oloumi Dodaran et al., 2025 [1404 SH]). This gap constitutes the point of departure for the present review. More generally, the reviewed literature supplies definitions, architectures, autonomy scales, lists of risks and governance proposals, but does not state how these elements depend on one another.

3. Methodology

3.1 Review design

This study is a conceptual integrative review, which combines sources of different kinds to develop a conceptual framework and does not claim exhaustive coverage (Snyder, 2019). The design suits a field in which regulatory documents, peer-reviewed studies, preprints and industry reports describe the same phenomenon with different levels of evidence. It is not a PRISMA-based systematic review, but its steps are reported so that they can be followed and so that the documented search and screening can be repeated (Table 2; details in Appendices A and B). The synthesis is organized around four analytical axes derived from the research questions: (1) conceptual definition and distinction, (2) capabilities and architecture, (3) autonomy and risk, and (4) governance requirements.

Table 2. *Review protocol*

Element	Description
Review type	Conceptual integrative review (Snyder, 2019), not a PRISMA-based systematic review
Scope and time window	Retail banking at autonomy levels 2 and 3; publications from 1 January 2024 to 30 September 2026, with earlier acts and frameworks only where sources of the window rely on them
Channels	(1) Websites of eight institutions; (2) Google Scholar, Web of Science, arXiv, OpenAlex and the Google search engine, with four search phrases documented as Boolean strings in OpenAlex; (3) backward and forward citation tracking; Civilica for Persian-language sources
Search dates	Topic searches up to September 2026
Inclusion criteria	Publication within the time window; direct relevance to at least one analytical axis at the level of the bank or of financial services; source credibility (peer-reviewed publication, recognized institutional body, preprint by identifiable authors in a recognized repository, or industry report from an identifiable source, the latter two reported with explicit caution)
Exclusion criteria	Another sector or function; technical system for a single task; redundancy; lack of traceable documentation (both defined in Section 3.3)
Final corpus (45 reviewed sources)	Peer-reviewed journal articles, book chapters and conference papers: 15; regulatory and standard-setting documents: 14; institutional working and research papers: 3; preprints: 7; industry reports: 6

3.2 Search strategy

Sources were identified through three channels. Institutional documents were taken from the websites of eight bodies: the Bank for International Settlements (BIS), the FSB, NIST, the European Union (EUR-Lex), MAS, IMDA, the Bank of Thailand and the Banking Industry Architecture Network (BIAN). They were chosen because they set international banking standards, issue the AI risk frameworks most cited in the literature, have published guidance that names AI agents, or maintain the reference architecture of the industry. Their publications on artificial intelligence from 2024 to 2026 were examined; earlier acts and frameworks were kept only where sources of the window rely on them. This channel does not cover all jurisdictions, and no other body was examined.

Scholarly and industry sources were sought up to September 2026 in Google Scholar and Web of Science (journal articles), arXiv (preprints), OpenAlex (an open index of journals, proceedings and preprint repositories that returns the complete record list of a query) and the Google search engine (industry reports), with four phrases: “agentic banking,” “autonomous AI agents in finance,” “AI agent autonomy levels financial services,” and “AI governance financial agents.” To give a record that can be repeated, the search was re-run in OpenAlex for this revision and used to verify and extend the corpus: the phrases were run as Boolean strings on titles and abstracts for works published from 1 January 2024 to 30 September 2026 and returned 1,988 records (strings and counts in Appendix A).

Backward and forward citation tracking in OpenAlex for five finance-specific sources gave 206 records; 32 met the title criterion, 10 of them already retrieved, and none of the other 22 qualified (Appendix A). In Civilica,

the Persian equivalent of "agentic banking" returned no record and that of "intelligent agent in banking" two records, which concern smart banking and not agentic banking.

3.3 Screening and selection

After removal of 470 duplicates, 1,518 records were screened. Stage 1 retained records whose title referred both to AI agents or agentic AI and to banking, payments or financial services (1,196 excluded). Stage 2, on title and abstract, retained records that treat agentic AI at the level of the bank or of financial services and address at least one axis; records on another sector or function and technical systems for a single task were excluded (168). The remaining 154 records formed the candidate pool. The general works on agentic AI (Bandi et al., 2025; Feng et al., 2025; Mitchell et al., 2025; Sapkota et al., 2026) fall outside the stage 1 criterion and were included because finance sources rely on them for definitions and autonomy levels.

Selection from the pool followed three rules. (1) Evidential rank: regulatory documents, peer-reviewed publications, institutional working papers, preprints by identifiable authors and industry reports, in that order; a lower-ranked source is not preferred to a higher-ranked source that supports the same point. (2) Contribution: a record is selected when it supplies a definition, component, autonomy scale, risk category or governance mechanism not yet coded, or evidence or a position that differs from the sources already selected. (3) Redundancy: a record is not selected when all its content relevant to the axes is already coded from a source of equal or higher rank; a record that offers an alternative explanation, contradictory evidence or an extension is by definition not redundant. Lack of traceable documentation means no identifiable author or issuing body, no date, no DOI or stable URL, or figures without a stated source. Five records of the pool were already in the corpus, eight were added (six peer-reviewed publications and two preprints whose positions differ from the framework) and 141 were not selected: eight had no listed source, and 133, judged on title and abstract, offered nothing under rule 2 beyond the sources already coded (rule 3); full reading could change this for individual records.

3.4 Data extraction and synthesis

Each source was recorded in an extraction form: source type and evidential rank, channel, analytical axis, content relevant to the axis, the basis, sample and date of any figure, and statements that fit no axis (summarized in Appendix B). Synthesis had four steps: grouping of statements under the axes, which were fixed in advance; merging of statements with the same content into categories (Figure 1, Tables 3 to 5); separate examination of statements that fit no axis, through which user trust and acceptance was retained as a theme (Figure 3); and comparison of the categories of axes 2 to 4 to form the framework. To limit the influence of the predetermined axes and of the authors' position, rule 2 retains contradicting sources (de la Chica Rodriguez et al., 2026; Han, 2026; Mitchell et al., 2025), and the framework is compared with alternative bases for governance and with the closest studies (Table 7). Figures from industry reports, forecasts and preprints are attributed with their basis and are not the sole support for any construct.

3.5 Corpus and saturation

The corpus contains 45 sources (Table 2; distribution by type and axis in Appendix B); the methodological reference is not part of it. Saturation is claimed only for the constructs of the framework. Every axis is covered by at least five regulatory or peer-reviewed sources; the twelve sources that entered last could be coded with the existing categories, according to the first author's coding; and citation tracking added no source. Two of the twelve raise considerations that the framework does not model (collective outcomes of agent populations and verifiability); these are stated as boundaries (Section 6.5), not as saturated. The corpus is not a census: about three quarters of the unique records retrieved were published in 2026.

4. Findings

This section reports the categories obtained under axes 2 to 4 (axis 1 is treated in Section 2), followed by the evidence on adoption. Within the scope of this review, namely retail banking at autonomy levels 2 and 3, a subset of specialized agents (customer agent, credit agent, account/deposit agent, product agent) is placed between a governed, agile data core and a layer of human oversight (Figure 1).

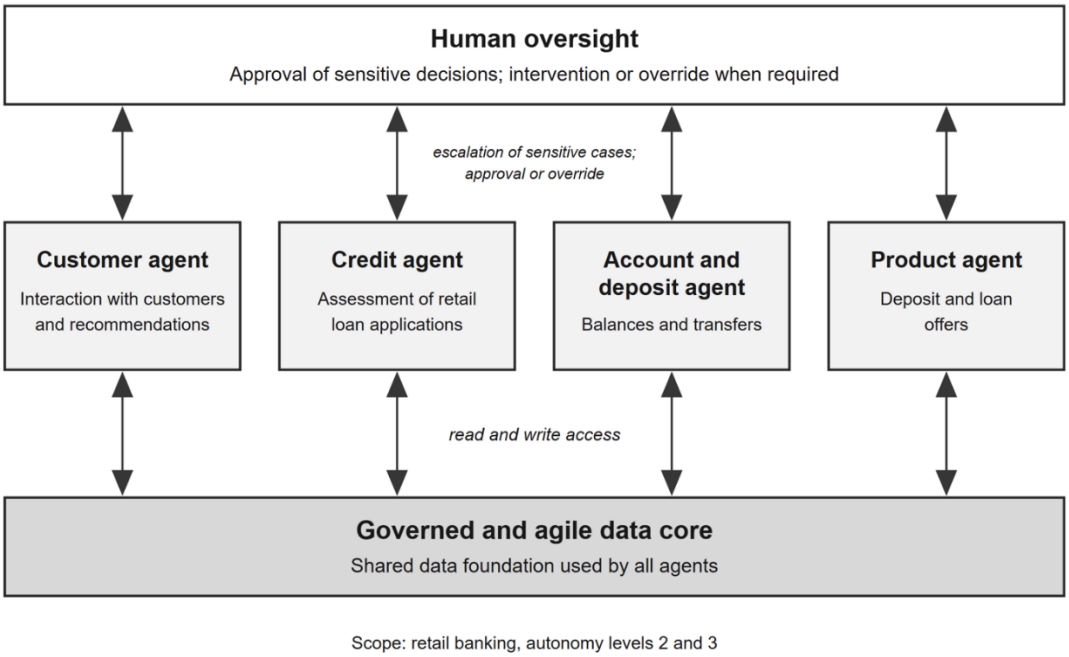


Figure 1. *Architecture of agentic banking in retail banking: four specialized agents placed between human oversight and a governed data core (autonomy levels 2 and 3). Sources of the concept: Huang et al. (2025), Feng et al. (2025) and BIAN (2023).*

Each specialized agent in Figure 1 can be related to one or more BIAN service domains; the customer agent, for instance, aligns with the Customer Domain and the credit agent with the Credit Domain (BIAN, 2023).

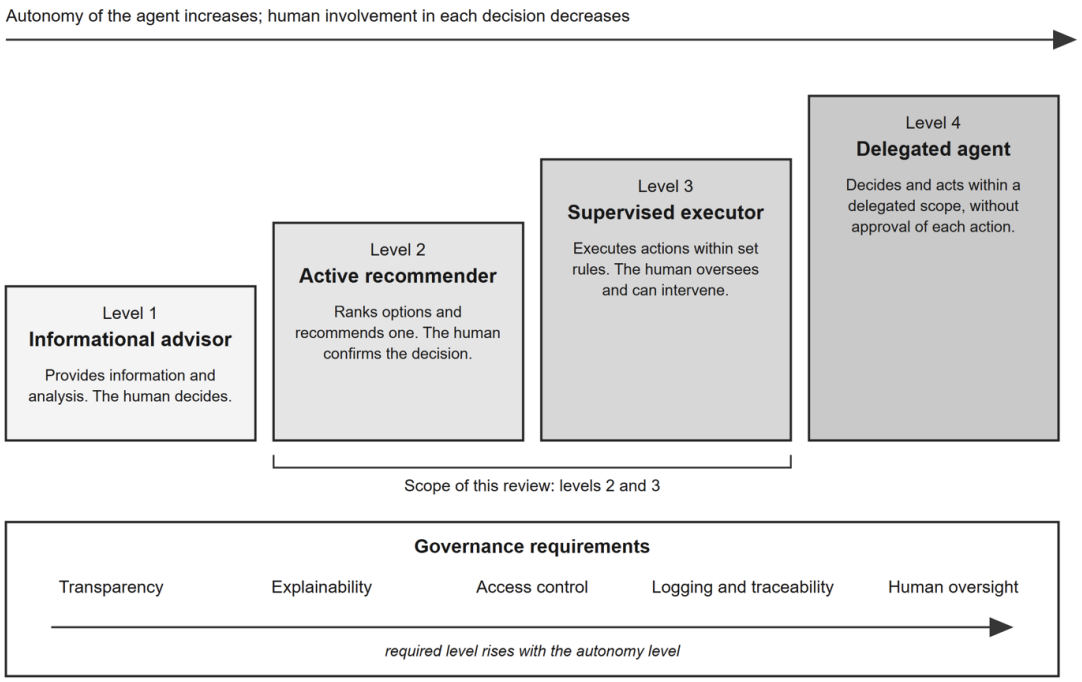


Figure 2. *Four levels of agent autonomy in financial services and the governance requirements that rise with them. Sources of the concept: Pathak and Bansal (2024) and NIST (2023, 2024).*

The transition across autonomy levels is a technical decision, but it may also be influenced by mechanisms of user trust and acceptance. The empirical findings of Pathak and Bansal (2024) indicate that this acceptance rests on three dimensions of trust (cognitive, affective, and social) and results in two distinct patterns: accepting the agent as a decision aid, or accepting it as a delegated agent. Figure 3 depicts this model and its acceptance pathways.

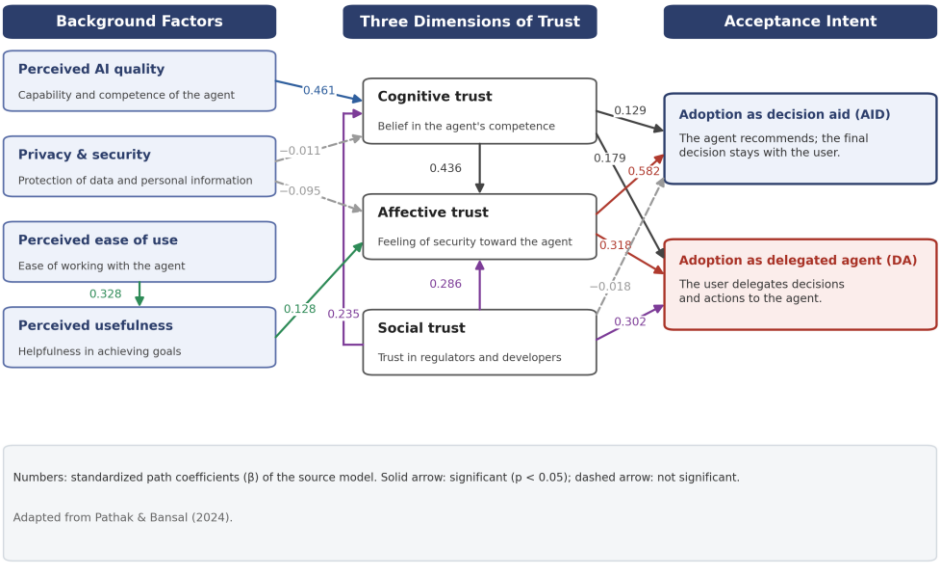


Figure 3. *Three-dimensional trust model and agent-acceptance pathways (conceptual source: Pathak & Bansal, 2024).*

This trust model is linked to the framework but is not one of its constructs (Figure 5). Trust was not one of the four predetermined axes; it was retained because the acceptance of an agent as decision aid or as delegated agent rests on it (Pathak & Bansal, 2024). Conceptually, increasing an agent's autonomy may increase perceived risk, which may call for more structured human oversight; the resulting oversight may in turn influence the user's cognitive and affective trust and, ultimately, the degree of agent acceptance. This is an inference from the reviewed sources and has not been tested in this study.

Table 3 summarizes the four autonomy levels of this spectrum. The level descriptions are adapted from Feng et al. (2025) and Mitchell et al. (2025), and the labels draw on Pathak and Bansal (2024); the four-level scale is proposed here and does not reproduce the levels of these sources one to one. Levels one and four are included only to mark the boundaries of the spectrum. The applied focus of this paper and its framework is on levels two and three.

Table 3. Proposed spectrum of four AI agent autonomy levels in financial services (adapted from Feng et al., 2025)

Level	Description	Corresponding Label in Figure 2
1	The agent supplies information and analysis; the human takes the decision	Informational advisor
2	The agent recommends a decision; the human confirms it before execution	Active recommender
3	The agent executes within set rules; the human supervises and can intervene	Supervised executor
4	The agent decides and acts within a delegated scope, without human approval of each action	Delegated (autonomous) agent

Industry survey figures are consistent with this spectrum: according to The Financial Brand (2025), which reports a 2025 survey of 250 banking executives by MIT Technology Review Insights in association with EY, most banking executives believe current systems can advise and assist, but far fewer believe present-day technology is ready for full digital autonomy. Infrastructure requirements can be formulated as ten interconnected components spanning data, processing, identity, security, model governance, logging, human oversight, compliance, testing, and organizational capacity (Table 4). The table is

the authors' synthesis; its last column names the main sources of each requirement.

Table 4. *Ten infrastructure requirements of agentic banking*

Category	Requirement	Focus	Main sources
A. Technical Enablers	1. Unified data infrastructure	Accurate, timely, and governable data	Aldasoro et al. (2024); BIS (2024)
A. Technical Enablers	2. Processing infrastructure and agent-ready architecture	Orchestration layer for planning, memory, and API invocation	Yang et al. (2025)
B. Control & Security	3. Authorization, identity, and access control	Principle of least privilege; traceable agent identity	Akkireddy et al. (2026); IMDA (2026)
B. Control & Security	4. Cybersecurity and operational resilience	Sandboxing, adversarial testing, emergency stop	NIST (2024); Vassilev et al. (2025)
C. Governance & Assurance	5. Model governance and risk management	Prompt logging, model versioning, drift monitoring	Okpala et al. (2025)
C. Governance & Assurance	6. Logging, auditing, and explainability	Complete logging, scenario replay, chain of evidence	BIS (2024); NIST (2023)
C. Governance & Assurance	7. Evaluation, testing, and phased deployment	Low-risk processes first, gradual extension	IMDA (2026); Lee and Yang (2026)
D. Organizational & Regulatory	8. Meaningful human oversight	Monitoring dashboards, prioritized alerts	NIST (2024)
D. Organizational & Regulatory	9. Regulatory and legal compliance	Embedded from the design stage	BIS (2024)
D. Organizational & Regulatory	10. Organizational infrastructure and human capital	Data specialists, AI architects, risk and compliance managers	Aldasoro et al. (2024)

As Table 4 shows, these ten requirements are organized into four interlocking categories: (A) technical enablers, (B) control and security, (C) governance and assurance, and (D) organizational-regulatory. As an agent's autonomy level and scope of action increase, the intensity of requirements in all four categories, particularly categories C and D, increases accordingly.

These ten requirements are not independent components of the proposed framework. They are supporting evidence drawn from the literature and used to derive two of its constructs, architectural enablers and governance intensity. Requirements concerning data, application programming interfaces, orchestration, and processing infrastructure are aggregated into architectural enablers, requirements concerning model risk management, human oversight, access limits, and traceability into governance intensity, and requirements concerning human capital and organizational readiness are treated as preconditions for both. The identity infrastructure on which access limits depend belongs to the architectural enablers.

Institutional governance approaches vary but converge on transparency, human oversight, and documentation (Table 5).

Table 5. *Comparison of institutional approaches to agentic AI governance in the financial system*

Institution/Framework	Focus of Requirement	Key Mechanism
EU AI Act	Risk classification (credit scoring of individuals = high-risk; fraud detection excluded)	Transparency, human oversight, documentation prior to deployment
DORA (EU, since 2025)	Digital operational resilience	Continuous monitoring of IT systems
Bank of Thailand (2025)	Strategic banking functions	Risk-proportionate human oversight (e.g., loan approval, account opening)
NIST (USA)	Technical standardization	Transparency, accountability, fairness, safety, oversight
Kurshan et al. (2026)	Risk of learning-based, multi-agent models	Four-layer governance: self-regulation to third-party audit

According to an industry report by The Financial Brand (2025), summarized in Figure 4, roughly two-thirds of the banking executives surveyed report use of agentic AI through existing deployments (16 percent) or pilot projects (52 percent). The underlying survey of 250 banking executives does not state the respondents' job titles, bank sizes or countries, so the proportion should be read as an indication of direction rather than an industry-wide estimate.

According to McKinsey's Global Banking Annual Review (2025), fifty of the world's largest banks announced more than 160 AI use cases, a count based on public announcements compiled by a third-party

tracker and not on a survey. McKinsey's scenario analysis estimates that gradual adoption of this technology could reduce banking costs by fifteen to twenty percent; this figure is a forward-looking estimate, not a realized outcome (McKinsey & Company, 2025).

Peer-reviewed evidence is more cautious. In a panel of 2,500 annual filings of 500 U.S. finance firms for 2021 to 2025, references to agentic AI were absent through 2023 and appeared in 0.4 percent of firm-years in 2024 and 1.6 percent in 2025; evidence of autonomy was rare and concentrated where control language was denser (Ul Mustafa & Aysan, 2026). Industry reports record announcements, pilots and intentions, whereas disclosures record what firms report as material; together they indicate wide experimentation and limited deployment with real autonomy.

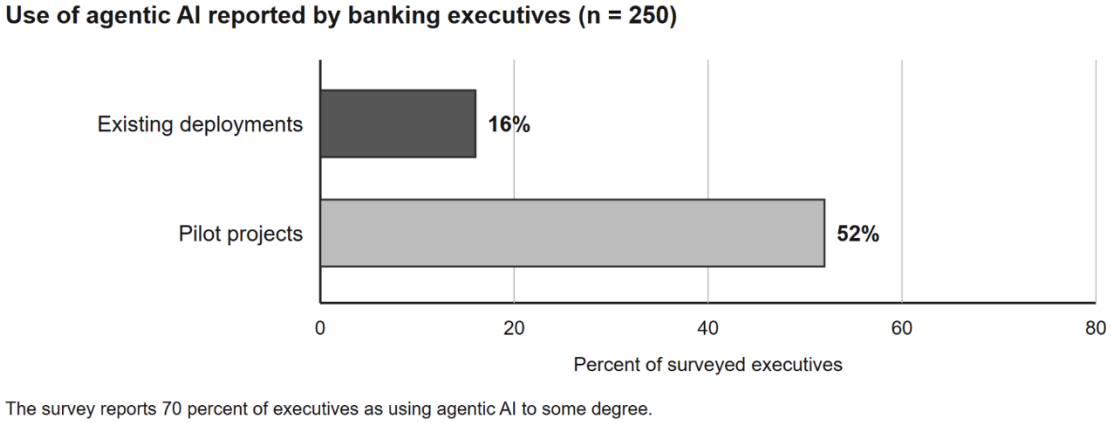


Figure 4. *Industry evidence on agentic banking deployment (2025). Source: The Financial Brand (2025), reporting a survey of 250 banking executives by MIT Technology Review Insights and EY.*

4. Integrated Conceptual Framework

4.1 Constructs and relations

Section 3 treated architecture (Figure 1), the autonomy spectrum (Figure 2), the trust model (Figure 3), infrastructure requirements (Table 4) and governance approaches (Table 5) one at a time. The contribution of this paper lies in their integration. Figure 5 formulates this integration as a unified framework in which the elements play distinct roles.

Agentic capabilities, architectural enablers, and autonomy level act as conceptual determinants that jointly shape an agent's risk exposure. Governance intensity is conceptually proposed to be proportionate to that exposure and is implemented through human oversight and controls, with safe and reliable banking action as the intended outcome. The constructs derive from the synthesis: agentic capabilities from Table 1, architectural enablers and governance intensity from Tables 4 and 5, autonomy level from Table 3, and risk exposure from NIST (2023) and IMDA (2026).

The novelty of the present framework does not lie in introducing new concepts, since agentic capabilities, architectural enablers, autonomy levels, risk, and governance have each been examined in the existing literature. Nor is proportionality new: IMDA (2026) and MAS (2025) apply it, and Grégoire (2026) proposes governance in proportion to autonomy. What this paper adds is the formulation of the dependent, integrated relationship among these elements: the risk exposure of an action is treated as the joint outcome of three determinants, and governance intensity is conceptually proposed to be proportionate to that exposure rather than to any single dimension such as autonomy level. The contribution is therefore one of integration and formulation rather than the discovery of a new relationship.

The framework rests on two relationships. Regarding the first, risk exposure is determined by the three constructs jointly, not by any one of them alone. An agent with multi-step planning capability but without access to executable tools has a limited risk exposure; the same capability, once combined with transactional application programming interfaces and a higher autonomy level, alters the risk exposure qualitatively rather than merely increasing it quantitatively. This is a

conceptual argument; whether the determinants combine additively or interact is an empirical question, addressed by Proposition 2. Regarding the second, governance is not a binary variable; it has an intensity that must be commensurate with the risk exposure. Governance below the required level leaves the system vulnerable to irreversible errors, while governance heavier than the risk exposure reduces the amount of usable autonomy without a corresponding gain in safety.

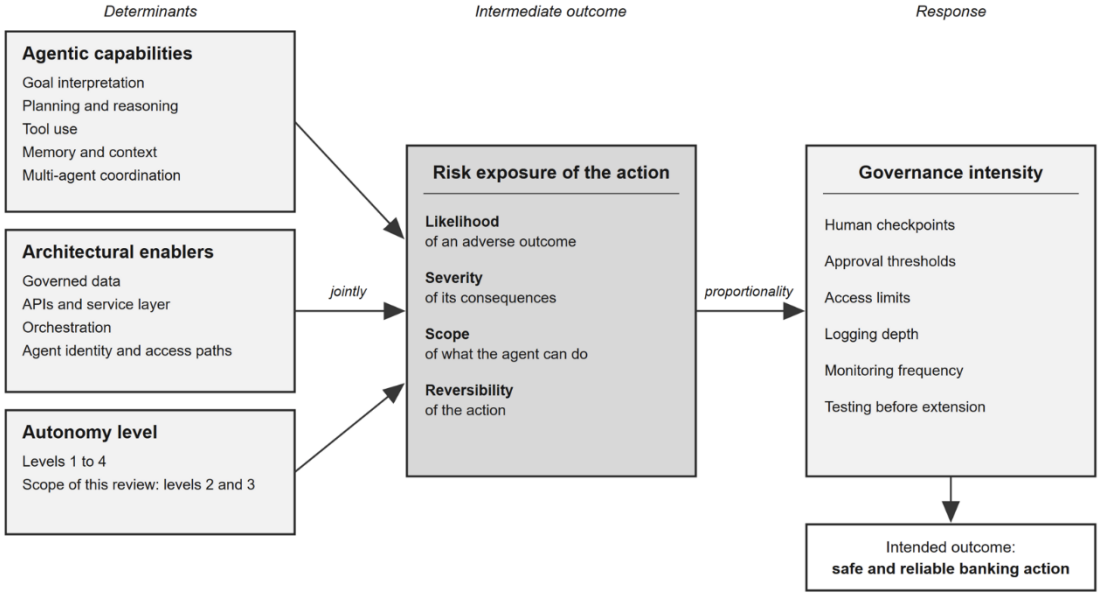


Figure 5. *Integrated conceptual framework of agentic banking, linking agentic capabilities, architectural enablers, and autonomy level to the four dimensions of risk exposure and to required governance intensity.*

4.2 Definition and dimensions of risk exposure

Risk exposure needs a precise definition. NIST (2023) defines risk as a composite of the probability of an event and the magnitude of its consequences, and IMDA (2026) adds, for acting agents, access to sensitive data and the reversibility of actions. Risk exposure is defined here as the potential for harm that a given action of an agent carries for the customer, the bank or third parties, described by four dimensions: likelihood of an adverse outcome; severity of that outcome (financial loss, harm to the customer, breach of a legal duty); scope, the range of actions and resources the agent can reach, including actions outside its mandate; and reversibility, the extent to which the outcome can be detected and undone. Exposure is a profile across these dimensions and not a single score, because they can change independently.

The determinants act on different dimensions (Table 6). Capabilities act mainly on likelihood, since longer plans, tool selection and coordination among agents add points of error. Architectural enablers act mainly on severity and scope, through the access they give: access to a payment interface raises the possible loss of an erroneous action without making the error more probable. The autonomy level acts on likelihood and on reversibility: confirmation before execution allows an erroneous proposal to be stopped, and supervision after execution allows an error to be detected and undone. An agent with high autonomy and narrow access (one that only reads balances) is therefore low on severity and scope, whereas an agent with low autonomy and wide access (one that prepares transfers across all accounts for approval) is high on both, while human approval lowers likelihood and preserves reversibility. A rating based on autonomy alone cannot separate these cases. Exposure is assessed before governance controls are added; access limits then answer severity and scope, human checkpoints likelihood, and logging depth reversibility.

Table 6. *Dimensions of risk exposure and the determinants that mainly affect them*

Dimension	Definition	Main determinant
Likelihood	Probability that the action produces an adverse outcome	Agentic capabilities (planning depth, tool use, coordination among agents); autonomy level (human confirmation before execution)
Severity	Magnitude of the outcome: financial loss, harm to the customer, breach of a legal duty	Architectural enablers (transactional interfaces, sensitivity of the data reached)
Scope	Range of actions and resources the agent can reach, including actions outside its mandate	Architectural enablers (identity and access paths); agentic capabilities (tool use)
Reversibility	Extent to which the outcome can be detected and undone before lasting harm	Autonomy level (human supervision and intervention after execution)

4.3 Comparison with related studies and alternative bases for governance

Table 7 sets the focus of each source beside its relation to the present framework, including the sources that set governance on a different basis (Section 2.4).

Table 7. *Comparison of the present framework with the closest related studies and with alternative bases for governance*

Source	Focus of the source	Relation to the present framework
Pathak and Bansal (2024)	AI in financial services as a decision aid or a delegated agent; acceptance based on cognitive, affective, and social trust	Places the two roles on a four-level autonomy spectrum and links them to governance requirements
Feng et al. (2025)	Five levels of autonomy by the user's role; autonomy as a design decision, capped by a certificate that reflects capability and operational environment	Uses the same three elements as joint determinants of risk exposure, to which governance is calibrated, in retail banking
Aldridge et al. (2026)	Defining features of financial agents (goal-directed autonomy, continuous learning, and multi-agent coordination)	Focuses on retail banking and on calibrating governance intensity to each agent's risk exposure
Huang et al. (2025)	Types of AI agents and their architecture in banking	Connects agent architecture to autonomy levels and to proportionate governance in a single framework
IMDA (2026)	Permissible autonomy depends on error tolerance, access to sensitive data and reversibility	Same proportionality logic; adds capabilities as a determinant and four dimensions of exposure
Grégoire (2026)	Autonomy-proportional governance as one of four controls	Calibrates to joint exposure, not to autonomy alone (Proposition 2)
EU AI Act (2024)	Obligations attached to the intended purpose (use case)	Use case as legal minimum; within it, exposure varies with capabilities, access and autonomy
De la Chica Rodriguez et al. (2026); Han (2026)	Governance of agent populations; verifiability as the binding constraint	Per-action exposure is necessary but not sufficient; stated as boundaries (Section 6.5)

4.4 Propositions and their operationalization

To make this formulation testable in future research, three research propositions are derived from it. Proposition 1: at the same autonomy level and with the same capabilities, an agent with access to transactional tools has a larger risk exposure on the severity dimension (its scope being wider by definition) than an agent without such access. Proposition 2: estimating risk on the basis of autonomy level alone underestimates the risk exposure of agents with broad capabilities or architectural access, and overestimates the exposure of highly autonomous agents with narrow access. Proposition 3: banks that calibrate governance intensity to the risk exposure of each application experience fewer incidents arising from insufficient control and less delay arising from heavy control than banks that apply uniform governance to all agents.

Appendix C gives, for each proposition, the constructs, indicators, data sources and methods. Proposition 1 can be examined by a matched comparison of incident severity for agents with and without transactional access; Proposition 2 by comparing autonomy-only ratings and four-dimension ratings with observed incidents; and Proposition 3 by structural equation modeling of survey data from banks that deploy agents or by a multi-case field study of tiered and uniform governance. Propositions 1 and 2 need agent-level records of one institution, whereas Proposition 3 normally requires comparison across institutions.

5. Conclusion

5.1 The framework in relation to prior work

The reviewed sources point, indirectly, to a gap between technological readiness and governance readiness. In a Deloitte survey of 1,434 finance leaders (chief financial officers and executives one level below them, in companies of several industries with revenue of at least one billion U.S. dollars in 26 countries, spring 2026), only 14 percent were comfortable using fully autonomous agents for more critical decisions, and 42 percent reported that their ambitions exceed current capabilities because stronger controls are needed before scaling (Anderson et al., 2026). An industry report by Deloitte (Goss et al., 2026) warns that agents generate risks that existing bank risk-management frameworks may not fully address. Peer-reviewed sources are consistent with limited autonomous use: evidence of autonomy appears mainly where control language is dense (Ul Mustafa & Aysan, 2026), and fully autonomous use remains rare (Grégoire, 2026).

The framework agrees with prior work on the direction of the relation between risk and governance and differs on the basis of calibration. Grégoire (2026) scales oversight to autonomy, Feng et al. (2025) cap autonomy by capability and environment, the EU AI Act uses the use case, and IMDA (2026) error tolerance, data sensitivity and reversibility. The framework treats the use case as the legal minimum and, within it, sets governance by the exposure that capabilities, access and autonomy produce together. An analyst forecast attributes failure to a misfit between agents and controls (Gartner, 2026), but the reviewed sources contain no incident data from deployed agents

.

5.2 Liability for failures of autonomous transactions

The first unresolved issue is who bears the loss when an agent executes a wrong or unauthorized transaction. Under the second Payment Services Directive (PSD2), a payment transaction is authorized only if the payer has consented to it, and the payer's payment service provider must refund an unauthorized transaction (Articles 64 and 73; European Parliament & Council of the European Union, 2015). These rules assume that a person consents to a specific transaction. When an agent acts at the margin of a customer's mandate or outside it, it is not settled whether the transaction is authorized; Restrepo Amariles et al. (2026) count liability gaps among the legal risks of agent-initiated payments. The EU AI Act places duties of oversight, monitoring and record-keeping on the deployer of a high-risk system (Articles 14 and 26) but does not state how responsibility is divided among model provider, bank and customer when several agents contribute to one outcome. For the framework, reversibility and scope are therefore legal as well as technical dimensions: the mandate of each agent, its limits and the record of each action should be fixed before deployment. The allocation of liability itself requires legal analysis in each jurisdiction, including Iran.

5.3 Data protection across multi-agent handoffs

The second issue arises when customer data pass from one agent to another, as among the customer, credit and compliance agents of the credit example in Section 2.2. Data-protection law requires purpose limitation and data minimization and restricts decisions based solely on automated processing (General Data Protection Regulation [GDPR], Articles 5 and 22; European Parliament & Council of the European Union, 2016). Persistent memory and shared context work against these principles, because context carried forward for one task becomes available to an agent performing another. The technical means of enforcement are incomplete: authorization scopes are too coarse, so agents receive broader privileges than the task requires (Akkireddy et al., 2026), and controls verified for each agent do not guarantee an acceptable result for the agents together (de la Chica Rodriguez et al., 2026). In the framework, each handoff widens the scope dimension of exposure. The corresponding controls are purpose-bound access for each agent, minimization of the context passed at each handoff, and logging of handoffs as well as final actions.

5.4 Continuous model governance in legacy core banking environments

The third issue is operational. Model versioning, drift monitoring, complete logging and scenario replay (Table 4) presuppose that each action of an agent passes through an interface that records it and that the model stays under the bank's control. Three obstacles are reported. Periodic model validation does not fit agents that learn continuously and interact with other agents (Kurshan et al., 2026). Models are often supplied by third parties: Han (2026) reports, in a preprint, that provider releases change the actions an agentic system takes on the same inputs, so that a bank may be unable to reproduce a decision it has to explain, while DORA keeps the financial entity responsible when it uses third-party ICT services (European Parliament & Council of the European Union, 2022). Legacy systems are expected to limit integration: in the simulation of Lee and Yang (2026), which is not evidence from actual projects, production maturity goes together with low legacy-system complexity combined with data readiness and governance maturity. BIAN offers a standardized decomposition of banking activities into service domains that provides boundaries for designing distinct agents; this mapping is conceptual, and operationalization would require separate design work (BIAN, 2023). Where the logging, monitoring or replay that an agent's exposure requires cannot be provided, the proportionality principle applies in the other direction: the autonomy or access of the agent should be reduced until its exposure matches the governance the bank can deliver, which supports phased deployment starting with low-exposure processes.

5.5 Boundaries of the framework

The framework takes the single action of a single agent as its unit of analysis, which has three limits. Agents that each satisfy their local controls can together produce an unacceptable outcome (de la Chica Rodriguez et al., 2026), and systemic effects depend on how agents are coupled across institutions (Gong, 2026); a bank with many agents therefore needs monitoring at the level of the agent population as well. The framework includes logging and traceability in governance intensity but does not model verifiability as a construct; if Han (2026) is right, an agent whose decisions cannot be reproduced should be treated as highly exposed whatever its other properties. Finally, the framework is graded and does not establish that governance can make level 4 safe, which Mitchell et al. (2025) deny.

5.6 Domestic context and limitations

The absence of dedicated domestic literature, despite adjacent work on AI-enabled banking services (Garoussi et al., 2026 [1405 SH]) and smart-governance business models (Oloumi Dodaran et al., 2025 [1404 SH]), suggests that the groundwork is emerging in Iran but is not yet a dedicated body of knowledge.

This review has limitations. First, it is a conceptual integrative review and does not claim full coverage of the literature; sources were selected purposively around four analytical axes, the institutional channel is limited to the bodies listed, the documented search was run for the revision and used one index and English-language records, and screening and coding were done by the first author without a second screener. Second, because the field is emerging, part of the evidence comes from industry reports and preprints that have not undergone peer review; such evidence is reported with explicit caution. Third, the framework is conceptual and its relationships have not been tested empirically; the propositions in Section 5 are formulated for that purpose.

6. Conclusion

This review asked what agentic banking is, how it is built and governed, and what the Persian-language literature offers. In the sources reviewed, agentic banking is marked by agents that plan and act within a scope of authority, governance approaches converge on transparency, human oversight and documentation, and no dedicated Persian-language study was found. The framework applies proportionality to the joint risk exposure of an action, and its three propositions remain to be tested.

The framework also has practical implications. For banks, it suggests

classifying each agent before deployment by the risk exposure that its capabilities, architectural access and autonomy level jointly create, and calibrating approval thresholds, human checkpoints and logging depth to that class instead of applying one governance template to all agents. For supervisors, including the Central Bank of the Islamic Republic of Iran, it suggests that expectations for agentic systems can be set in proportion to risk exposure, in line with recent supervisory guidance, while complete action logs keep agent decisions auditable.

The relationships captured in the framework imply the need for coordinated redesign of three domains: a technical architecture that provides the data infrastructure and real-time processing necessary for agentic action; a governance structure that calibrates governance intensity to the risk exposure of each application; and a regulatory framework, which is still taking shape in the jurisdictions reviewed.

Future research should first test the three propositions with deployment data (Appendix C). Further directions are case studies of the first limited agents in low-risk retail tasks in Iranian banks, studies of adoption and user trust in the Iranian context, comparison of governance models across jurisdictions, a domestic governance framework for financial agents suited to the Central Bank of the Islamic Republic of Iran, and legal analysis of liability for agent-initiated transactions.

Ethical Considerations.

This paper is a conceptual-analytical review based on documentary study and did not collect data from human subjects; accordingly, requirements for informed consent and ethics committee approval do not apply to it. The principles of scholarly publication ethics, including accurate citation and avoidance of plagiarism, have been observed.

Funding.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

This paper is based on published institutional and scholarly sources, listed in the References; no primary dataset was generated. The record list of the documented search with the screening decision for each record, the citation-tracking records and the extraction matrix are available from the corresponding author.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Akkireddy, L. H., Yadati, N. S. P. K., Balakumar, G., Vijay, V., Venganti, V., & Pochincharla, S. (2026). Securing customer data and tool-using AI agents in retail banking. In *2026 25th International Symposium on Communications and Information Technologies (ISCIT)* (pp. 1–6). IEEE. https://doi.org/10.1109/ISCIT70453.2026.11695371

Aldasoro, I., & Desai, A. (2025). *AI agents for cash management in payment systems* (BIS Working Paper No. 1310). Bank for International Settlements. https://www.bis.org/publ/work1310.htm

Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V., & Stein, M. (2024). *Intelligent financial system: How AI is transforming finance* (BIS Working Paper No. 1194). Bank for International Settlements. https://www.bis.org/publ/work1194.htm

Aldridge, I., An, J., Burke, R., Cao, M., Chien, C.-Y., Deng, K., Deng, R., Gao, Y., Guo, O., He, S., Li, Z., Lin, G., Lin, W., Lyu, P., Ng, A., Wang, Q., Xiao, H., Xu, D., Xue, Y., . . . Zheng, W. (2026). *Agentic artificial intelligence in finance: A comprehensive survey*. arXiv. https://doi.org/10.48550/arXiv.2604.21672

Anderson, D., Hardy, E., Ma, D., Murphy, T., Turk, D., & Vakili, M. (2026, September 8). *Deloitte Finance Trends 2027: Shaping the next era of stakeholder value*. Deloitte Insights. https://www.deloitte.com/global/en/insights/topics/leadership/finance-trends-leadership.html

Bandi, A., Kongari, B., Naguru, R., Pasnoor, S., & Vilipala, S. V. (2025). The rise of agentic AI: A review of definitions, frameworks, architectures, applications, evaluation metrics, and challenges. *Future Internet, 17*(9), Article 404. https://doi.org/10.3390/fi17090404

Bank for International Settlements. (2024). Artificial intelligence and the economy: Implications for central banks. In *Annual economic report 2024* (Chapter III). https://www.bis.org/publications/aer-2024/artificial-intelligence-economy-implications-central-banks

Bank of Thailand. (2025). *Guiding principles for artificial intelligence risk management* [Draft for public hearing; in Thai].https://www.bot.or.th/content/dam/bot/documents/th/laws-andrules/hearing/120668%20AI%20Guideline%20public%20hearing.pdf

Banking Industry Architecture Network. (2023). *BIAN service landscape* (Version 12.0). https://bian.org/servicelandscape-12-0-0/

de la Chica Rodriguez, J. M., Vera Díaz, J. M., & Delgado Romero, P. (2026). *Compliant with local controls, collectively discriminatory: A governance architecture for multi-agent AI in regulated finance*. arXiv. https://doi.org/10.48550/arXiv.2609.27994

European Parliament & Council of the European Union. (2015). Directive (EU) 2015/2366 on payment services in the internal market. *Official Journal of the European Union, L 337*, 35–127. https://eurlex.europa.eu/eli/dir/2015/2366/oj

European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union, L 119*, 1–88. https://eurlex.europa.eu/eli/reg/2016/679/oj

European Parliament & Council of the European Union. (2022). Regulation (EU) 2022/2554 on digital operational resilience for the financial sector. *Official Journal of the European Union, L 333*, 1–79. https://eurlex.europa.eu/eli/reg/2022/2554/oj

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union, L 2024/1689*. https://eurlex.europa.eu/eli/reg/2024/1689/oj

Feng, K. J. K., McDonald, D. W., & Zhang, A. X. (2025, July 28). *Levels of autonomy for AI agents* (Paper No. 25-15). Knight First Amendment Institute at Columbia University. https://knightcolumbia.org/content/levels-of-autonomy-for-ai-agents-1

Financial Stability Board. (2024). *The financial stability implications of artificial intelligence*. https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/

Financial Stability Board. (2025). *Monitoring adoption of artificial intelligence and related vulnerabilities in the financial sector*. https://www.fsb.org/2025/10/monitoring-adoption-of-artificial-intelligence-and-

related-vulnerabilities-in-the-financial-sector/

Garoussi, M., Moafi Madani, S. M., & Jamshidloo, R. (2026). Designing a model for modern banking service delivery based on the application of artificial intelligence. *Journal of Intelligent Marketing Management, 7*(2), 238–263. [In Persian]. https://www.jnabm.ir/article_735662.html

Gartner. (2026, May 26). *Gartner says applying uniform governance across AI agents will lead to enterprise AI agent failure* [Press release]. https://www.gartner.com/en/newsroom/press-releases/2026-05-26-gartner-says-applying-uniform-governance-across-ai-agents-will-lead-to-enterprise-ai-agent-failure

Gong, H. (2026). AI agents in financial markets: Architecture, applications, and systemic implications. *FinTech, 5*(2), Article 34. https://doi.org/10.3390/fintech5020034

Goss, C., Ilyasov, A., Singh, A., Srinivas, V., & Gregorie, J. (2026, March 5). *Managing the new wave of risks from AI agents in banking*. Deloitte Insights. https://www.deloitte.com/us/en/insights/industry/financial-services/agentic-ai-risks-banking.html

Grégoire, V. (2026). Agentic AI in finance: Emerging risks and the regulatory gap. *The Journal of Portfolio Management, 52*(11), 278–291. https://doi.org/10.3905/jpm.2026.054

Han, H. (2026). *Governing agentic AI in FinTech*. arXiv. https://doi.org/10.48550/arXiv.2608.11344

Huang, K., Wu, D., Ponnapalli, J., & Huang, G. (2025). AI agents in banking. In K. Huang (Ed.), *Agentic AI: Theories and practices* (pp. 237–277). Springer. https://doi.org/10.1007/978-3-031-90026-6_8

Infocomm Media Development Authority. (2026, May 20). *Model AI governance framework for agentic AI* (Version 1.5). https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf

Khosropour, H., Elyasi, M., Ghazinoori, S. S., & Taghva, M. (2024). Governance framework of digital transformation for open banking in commercial banks of Iran. *Journal of Entrepreneurship Development, 17*(2), 23–55. [In Persian]. https://doi.org/10.22059/jed.2024.374532.654350

Kshetri, N. (2025). The rise of agentic AI in finance: Opportunities, risks, and human-centric integration. *IT Professional, 27*(4), 19–24. https://doi.org/10.1109/MITP.2025.3585227

Kurshan, E., Balch, T., & Byrd, D. (2026). *The agentic regulator: Risks for AI in finance and a proposed agent-based framework for governance* [Working paper]. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6970398

Lee, Y.-C., & Yang, C. (2026). Agentic AI deployment readiness and responsible value realization in sustainable banking. *Sustainability, 18*(15), Article 7744. https://doi.org/10.3390/su18157744

McKinsey & Company. (2025, October 23). *Global banking annual review 2025: Why precision, not heft, defines the future of banking*. https://www.mckinsey.com/industries/financial-services/our-insights/global-banking-annual-review-2025

Mitchell, M., Ghosh, A., Luccioni, A. S., & Pistilli, G. (2025). *Fully autonomous AI agents should not be developed*. arXiv. https://doi.org/10.48550/arXiv.2502.02649

Monetary Authority of Singapore. (2025, November 13). *Consultation paper on proposed guidelines on artificial intelligence risk management for financial institutions* (Consultation Paper P017-2025). https://www.mas.gov.sg/-/media/mas-media-library/publications/consultations/bd/2025/final_consultation_paper_on_guidelines_on_ai_risk_management_forrelease.pdf

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

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

Okpala, I., Golgoon, A., & Kannan, A. R. (2025). *Agentic AI systems applied to tasks in financial services: Modeling and model risk management crews*. arXiv. https://doi.org/10.48550/arXiv.2502.05439

Oloumi Dodaran, M., Zareian, D., Tabibi, S. J., Abtahi, A., & Soleimani, N. (2025). Development of bank business models within smart governance. *Technology in Entrepreneurship and Strategic Management, 4*(4), 1–14. [In Persian]. https://journaltesm.com/index.php/journaltesm/article/view/342

Pathak, A., & Bansal, V. (2024). AI as decision aid or delegated agent: The effects of trust dimensions on the adoption of AI digital agents. *Computers in Human Behavior: Artificial Humans, 2*(2), Article 100094. https://doi.org/10.1016/j.chbah.2024.100094

Restrepo Amariles, D., Charlotin, D., & He-Guelton, L. (2026). AI agents in payments: Applications, risks and regulations. *European Journal of Risk Regulation, 17*(3), 884–907. https://doi.org/10.1017/err.2026.10103

Rizinski, M., & Trajanov, D. (2026). AI agents in finance and fintech: A scientific review of agent-based systems, applications, and future horizons. *Computers, Materials & Continua, 86*(1), 1–34. https://doi.org/10.32604/cmc.2025.069678

Rutten, T. (2026, April 20). *The four generations of banking technology*. Backbase. https://www.backbase.com/blog/the-four-generations-of-banking-technology

Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2026). AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges. *Information Fusion, 126*, Article 103599. https://doi.org/10.1016/j.inffus.2025.103599

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

The Financial Brand. (2025, September 15). *How the agentic AI revolution is transforming operations at 70% of banks*. https://thefinancialbrand.com/news/artificial-intelligence-banking/bankings-agentic-ai-revolution-how-70-of-institutions-are-already-transforming-operations-192250

Ul Mustafa, A., & Aysan, A. F. (2026). Measuring agentic AI adoption and control frameworks in finance. *Modern Finance, 4*(1). https://doi.org/10.61351/mf.v4i1.557

Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamin, M. (2025). *Adversarial machine learning: A taxonomy and terminology of attacks and mitigations* (NIST AI 100-2e2025). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-2e2025

Yang, H., Lin, L., She, Y., Liao, X., Wang, J., Zhang, R., Mo, Y., & Wang, C. D. (2025). *FinRobot: Generative business process AI agents for enterprise resource planning in finance*. arXiv. https://doi.org/10.48550/arXiv.2506.01423

Appendix A. Documented search and screening

Table A1. *Search strings, records and screening decisions (OpenAlex; title and abstract fields; publication dates from 1 January 2024 to 30 September 2026)*

Step	Search string or criterion	Records
Search string 1	("agentic AI" OR "AI agent" OR "AI agents" OR "autonomous agent" OR "autonomous agents") AND (bank OR banks OR banking)	523
Search string 2	"agentic banking"	177
Search string 3	("agentic AI" OR "AI agent" OR "AI agents") AND autonomy AND (finance OR financial OR banking)	289
Search string 4	("agentic AI" OR "AI agent" OR "AI agents") AND (governance OR regulation OR oversight) AND (finance OR financial OR banking)	999
Records retrieved	Sum of the four strings	1,988
Duplicates removed	Same OpenAlex identifier (394); same title (76)	470
Records screened at stage 1	Title	1,518
Excluded at stage 1	Title refers to agents but not to banking or finance (483); title refers to banking or finance but not to AI agents (197); neither (369); agent banking and financial inclusion unrelated to AI agents (92); record type other than a scholarly text (54); no title (1)	1,196
Records screened at stage 2	Title, and abstract where the title was not decisive	322
Excluded at stage 2	Another sector or function (83); technical system for a single task (83); not in English or not a scholarly text (2)	168
Candidate pool	Records that passed both stages	154
Already in the corpus	Aldasoro and Desai (2025); Aldridge et al. (2026); Gong (2026); Huang et al. (2025); Yang et al. (2025)	5
Corpus sources not retrieved	Kurshan et al. (2026): indexed in OpenAlex without an abstract, and the title lacks the agent terms of the strings; Okpala et al. (2025): indexed under another title, and its title and abstract lack the agent terms; Rizinski and Trajanov (2026): the indexed abstract is truncated and lacks the banking, autonomy and governance terms	3
Added to the corpus	Selection rule 2 (Section 3.3)	8
Not selected	Selection rules 2 and 3, judged on title and abstract: repository preprints and deposits (74); journal articles, proceedings papers, book chapters and reports (59). Lack of traceable documentation (no source listed): 8	141
Citation tracking	Reference lists and citing works of Aldridge et al. (2026), Gong (2026), Huang et al. (2025), Rizinski and Trajanov (2026) and Yang et al. (2025): 206 records; 32 met the title criterion; 10 already retrieved; of the other 22, eight outside the time window, one the conference version of a corpus source, twelve excluded at stage 2, one review preprint adding no category	0 added

Appendix B. Review corpus

Table B1. *Composition of the corpus by source type and analytical axis (a source may address more than one axis; three Persian-language articles describe the domestic literature and are not assigned to an axis)*

Source type	Sources	Axis 1: definition and distinction	Axis 2: capabilities and architecture	Axis 3: autonomy and risk	Axis 4: governance requirements
Regulatory and standard-setting documents	14	1	2	3	13
Peer-reviewed journal articles, book chapters and conference papers	15	4	4	5	6
Institutional working and research papers	3	1	2	1	0
Preprints	7	2	2	3	4
Industry reports	6	1	1	2	3
Total	45	9	11	14	26

Table B2. *Summary of the extraction matrix: sources by type, analytical axis and content used in the review (R = regulatory or standard-setting document; P = peer-reviewed publication; W = institutional working or research paper; A = preprint; I = industry report; D = domestic literature)*

Source	Type	Axis	Content used in the review
Akkireddy et al. (2026)	P	2, 4	Tool-using agents in retail banking: prompt injection, coarse authorization scopes, regulatory gaps for agent-issued actions
Aldasoro and Desai (2025)	W	2	Application of an AI agent to cash and liquidity management in payment systems
Aldasoro et al. (2024)	W	1, 2	Stages of AI in finance; agents as planners and executors; data governance and skills as preconditions
Aldridge et al. (2026)	A	1	Three defining features of financial agents; dispersion of the literature
Anderson et al. (2026)	I	3, 4	Deloitte survey of 1,434 finance leaders: comfort with autonomous agents and need for controls
Bandi et al. (2025)	P	1	Review of 143 studies; defining combination of features of agentic AI; publication trend
Bank for International Settlements (2024)	R	2, 4	Central role of data governance; logging and compliance built in from the design stage
Bank of Thailand (2025)	R	1, 4	Agentic AI inside the scope of AI risk management, rule-based automation outside it; risk-proportionate human oversight
Banking Industry Architecture Network (2023)	R	2	Service domains as boundaries for specialized agents
de la Chica Rodriguez et al. (2026)	A	3, 4	Controls on each agent do not ensure acceptable collective outcomes; governance of agent populations

European Parliament and Council (2015), PSD2	R	4	Authorization of payment transactions; refund of unauthorized transactions
European Parliament and Council (2016), GDPR	R	4	Purpose limitation and data minimization; solely automated decisions
European Parliament and Council (2022), DORA	R	4	Digital operational resilience; responsibility for third-party ICT services
European Parliament and Council (2024), AI Act	R	4	Classification by intended purpose; human oversight; duties of deployers
Feng et al. (2025)	W	3	Levels of autonomy for AI agents
Financial Stability Board (2024)	R	4	Vulnerabilities that AI may amplify, including third-party dependencies and model risk
Financial Stability Board (2025)	R	4	Monitoring of AI adoption and cooperation among authorities
Garoussi et al. (2026)	P	D	Model of AI-based banking services (domestic)
Gartner (2026)	I	4	Analyst forecast on the failure of uniform governance of agents
Gong (2026)	P	1, 3	Definition of agentic finance; bounded autonomy; dependence of systemic outcomes on how agents are coupled and governed
Goss et al. (2026)	I	4	Deloitte article: agents create risks that existing bank risk frameworks may not fully address
Grégoire (2026)	P	3, 4	Five emerging risks; autonomy-proportional governance; regulatory divergence
Han (2026)	A	4	Verifiability as the binding constraint; reproducibility of agent decisions
Huang et al. (2025)	P	2	Types of banking agents and their architecture
Infocomm Media Development Authority (2026)	R	3, 4	Governance framework for agentic AI; autonomy conditional on error tolerance, data access and reversibility
Khosropour et al. (2024)	P	D	Governance of digital transformation for open banking (domestic)
Kshetri (2025)	P	4	Conditions for trustworthy use of agentic AI in financial services
Kurshan et al. (2026)	A	3, 4	Multi-layered risks of agents; limits of traditional model-risk frameworks; four-layer governance
Lee and Yang (2026)	P	2, 4	Simulated readiness conditions for moving agentic projects in banks from pilot to production
McKinsey & Company (2025)	I	2	Count of AI use cases announced by large banks; cost scenario
Mitchell et al. (2025)	A	3	Position that fully autonomous agents should not be developed
Monetary Authority of Singapore (2025)	R	4	Proposed AI risk management guidelines covering AI agents; proportionality
NIST (2023)	R	3, 4	Definition of risk; AI risk management framework
NIST (2024)	R	4	Generative AI profile of the risk management framework
Okpala et al. (2025)	A	1, 2, 4	Difference between agents and fixed-rule automation; agent crews for modeling and model risk management
Oloumi Dodaran et al. (2025)	P	D	Bank business models under smart governance (domestic)
Pathak and Bansal (2024)	P	3	Decision aid and delegated agent; three dimensions of trust
Restrepo Amariles et al. (2026)	P	3, 4	Agentic payments; liability gaps; multi-layered governance
Rizinski and Trajanov (2026)	P	1	Review of agent-based systems in finance and fintech
Rutten (2026)	I	1	Four generations of banking technology (industry narrative, Backbase)
Sapkota et al. (2026)	P	1, 2	Taxonomy of AI agents and agentic AI; characteristic challenges
The Financial Brand (2025)	I	3	Reported shares of deployment and piloting; executives' views on readiness for autonomy
Ul Mustafa and Aysan (2026)	P	3, 4	Disclosure-based measurement of agentic AI adoption and control language in 2,500 filings
Vassilev et al. (2025)	R	3, 4	Taxonomy of adversarial attacks on AI systems and of mitigations
Yang et al. (2025)	A	2	Agent-based business-process workflows in financial institutions

Appendix C. Operationalization of the propositions

Table C1. *Suggested constructs, indicators, data sources and empirical methods for the three propositions*

Proposition	Constructs and indicators	Data source	Suitable empirical method
P1	Transactional access: whether the agent can call payment, transfer or account-change interfaces, and its value limits. Exposure: severity (loss per incident, number of customers affected); scope (number of systems and accounts the agent can reach) is recorded as a descriptor. Held constant: autonomy level and capability set	Agent inventory and access-rights register; incident and near-miss logs of deployed agents	Matched comparison of agents with and without transactional access inside one bank; multi-case field study across banks
P2	Autonomy-only rating (levels 1 to 4); rating on each of the four dimensions of Table 6; observed incidents and near misses per 1,000 agent actions, with the controls in place recorded as a covariate	Risk assessments made by the bank's risk function; expert ratings of described agent profiles; incident logs	Comparison of the two ratings with observed incidents (rank correlation; regression with autonomy, capabilities and access as predictors, with interaction terms); expert-panel rating of agent vignettes
P3	Governance calibration: whether human checkpoints, approval thresholds and logging depth differ by exposure class (tiered) or are the same for all agents (uniform). Outcomes: rate of incidents attributed to missing controls; processing delay and share of actions escalated to humans	Survey of banks that deploy agents; internal-audit or supervisory records	Structural equation modeling of survey data, with governance calibration as predictor and the two outcomes as dependent constructs; comparative multi-case field study of banks with tiered and uniform governance; before-and-after comparison when a bank changes its policy