Trust in Human–AI Collaboration in Finance: A Bibliometric-Systematic Literature Review

Mario Mirabile[1,2], Giovanni Emanuele Corazza[2], Jose Maria Alonso-Moral[1]

[1]Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", University of Bologna, Bologna, Italy.
[2]Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, 15786 Santiago de Compostela, Spain.

Abstract

The integration of artificial intelligence (AI) in finance, encompassing algorithmic trading, robo-advisory, and credit scoring, underscores the importance of trust in human-AI collaboration. Yet, despite its importance, trust remains conceptually fragmented and inconsistently operationalized across disciplines.

To address this gap, this study presents what is, to the best of our knowledge, the first Bibliometric-Systematic Literature Review focused on trust in human-AI collaboration in finance. Following a PRISMA-guided process, we screened 430 Scopus records and identified 114 finance-specific publications (2018–2025). Using bibliographic coupling and thematic clustering, we mapped the conceptual landscape and identified six coherent research clusters: (i) AI governance in finance, (ii) eXplainable AI for finance, (iii) anthropomorphism in financial agents, (iv) user interface design for human-AI in finance, (v) robo-advisors for financial decision making, and (vi) blockchain and supply chain. Across these clusters, trust emerges as an orthogonal yet inconsistently defined construct, framed as cognitive, affective, procedural, or infrastructural, with limited integration among dimensions. Methodologically, the field is dominated by surveys, while experimental, longitudinal, and mixed-methods approaches remain scarce. The review also identifies emerging topics and key research gaps.

Building on these findings, we propose a novel, multi-level socio-technical framework that links micro-level user perceptions and behaviors, meso-level organizational and design practices, and macro-level regulatory and infrastructural conditions. We illustrate the framework's practical value through four financial use cases.

By consolidating a fragmented literature, this review provides a structured foundation for advancing human-centric and societally aligned AI systems in finance, including their development, deployment, and evaluation.

Keywords: Trust, Artificial Intelligence, Human–AI Collaboration, Finance, Bibliometric Systematic Literature Review, Trustworthy AI

1 Introduction

Artificial intelligence (AI) is becoming increasingly embedded in financial services (Giudici 2018; Maple et al. 2023). Within this evolving landscape, financial machine learning (ML) has emerged as a specialized field focused on the particular challenges of financial data (Dixon et al. 2020; Scholes 2025). Also, Generative AI (Gen-AI) and Agentic AI in financial contexts extend the capability set supporting end-to-end financial workflows, with Gen-AI strengthening generative and analytical functions and Agentic AI supporting more autonomous, goal-driven coordination for across customers and teams (Li et al. 2023b; Abou Ali et al. 2025).

These advances create substantial opportunities, but also introduce significant risks, especially in human–AI collaboration, where humans and AI jointly pursue shared goals through complementary strengths, task co-management, and adaptive feedback (Dellermann et al. 2019; Wang et al. 2019, 2020; Sowa et al. 2021; Sharma et al. 2023). Effective collaboration, however, requires more than technical coordination. It involves processes of mutual adaptation and trust-building between humans and artificial systems (Makarius et al. 2020), recognizing how AI is a tool with agency but no intelligence (Floridi 2025).

In finance, such collaboration has been studied in various domains, including robo-advisory (Jung et al. 2018; Bhatia et al. 2020; Brenner and Meyll 2020), credit scoring (Bussmann et al. 2021; Heng and Subramanian 2023), risk management (Giudici 2018; Heng and Subramanian 2023; Leitner et al. 2024), customer service (Belanche et al. 2019; Graham et al. 2025), fraud detection (Alkhalil et al. 2021), supply chains (Duong et al. 2024), and regulatory auditing (Garcia et al. 2024). While these domains differ in scope, they share a common thread: the success of human–AI collaboration depends not only on technical accuracy but on the quality of the collaboration, where trust, more than accuracy, is the sector's most vital asset, with delegating savings and investments representing the ultimate act of confidence (Loaiza and Rigobon 2025).

Yet multiple forces threaten this trust. At the technical level, biased or low-quality data, opaque algorithms, and limited auditability can erode confidence (Leitner et al. 2024). Industry-level risks such as provider concentration, and correlated model use may lead to herding, fragility, and competition challenges (Leitner et al. 2024; Dou et al. 2025). On the customer side, AI-enabled phishing, deepfakes, and data leakage raise concerns for privacy and financial safety (Giudici 2018; Alkhalil et al. 2021).

These risks might grew when interpretable models are replaced by opaque algorithms. Although these may perform better in the short term, their opacity can hide errors, biases, and inequities (Breiman 2001; Cramer et al. 2018; Gebru et al. 2021; Garcia et al. 2024).

In response, researchers have increasingly examined conceptual approaches for defining and strengthening trust. Extensions of the Technology Acceptance Model (TAM), originally proposed by Davis (1989), have incorporated trust as a key determinant of adoption (Gefen et al. 2003). Further work has analyzed emotional dynamics and trust-based reactions to AI advisors (Brenner and Meyll 2020). Also, eXplainable AI (XAI) has emerged as a key strategy to enhance interpretability and fairness (Guidotti et al. 2018; Lipton 2018; Floridi 2019; Gunning 2019; Miller 2019; Rudin 2019; Barredo Arrieta et al. 2020; Ali et al. 2023). Specifically, in finance, XAI research has classified methods by task (Černevičienė and Kabašinskas 2024), evaluated robustness in credit risk (Heng and Subramanian 2023), cataloged techniques (Weber et al. 2024), tested model-agnostic approaches (Khan et al. 2025), and demonstrated interpretable statistical methods that improve both transparency and performance (Bussmann et al. 2021).

Despite these valuable contributions, the literature remains fragmented. What is still missing is a cohesive narrative that explains how trust in human–AI collaboration in finance is conceptualized, measured, and fostered. This gap underscores the need for a systematic review that critically analyze existing works and provides a clearer foundation for research on human–AI collaboration in finance.

To address this gap, the study examines the following research questions (RQs):

RQ1: What are the most distinctive research themes related to trust in human–AI collaboration in finance?

RQ2: How have research communities conceptualized trust in the most influential studies on trust in human–AI collaboration in finance?

RQ3: What methods and tools have been proposed in the identified most influential studies to investigate specific trust dimensions in human–AI collaboration in financial contexts?

RQ4: What emerging themes and research gaps exist in research on trust in human–AI collaboration in finance across identified clusters?

To address our research questions, we conducted a Bibliometric Systematic Literature Review (B-SLR), following guidelines given by Marzi et al. (2025), but also integrating PRISMA-guided retrieval (Page et al. 2021a,b), bibliographic coupling (Zupic and Čater 2015; Donthu et al. 2021), and centrality mapping (Bonacich 2007; Fronzetti Colladon and Naldi 2020). From 430 Scopus records, we screened and retained 114 finance-specific studies. We identified document-based clusters using the weighted Leiden algorithm (Traag et al. 2019), mapped the most distinctive themes (RQ1), and performed a manual full-text review of most influential papers to examine trust conceptualization (RQ2), and methodological approaches (RQ3). Finally, we analyzed thematic structures to highlight emerging trends and research gaps (RQ4).

A key contribution of this study is the introduction of a novel, multi-level socio-technical framework of trust in financial human–AI collaboration, integrating micro-level cognitive and emotional processes, meso-level organizational design practices, and macro-level infrastructural and governance mechanisms. While our focus is finance, the framework has broader implications for AI trust research in other high-stakes domains.

Finally, the paper is organized as follows. Section 2 details our methodology. Section 3 presents the results. Section 4 discusses the implications of our findings, reflects on limitations, and proposes avenues for future research.

2 Methodology

This study implemented a B-SLR approach (Marzi et al. 2025), integrating PRISMA-guided protocol (Page et al. 2021a,b), bibliographic coupling (Zupic and Čater 2015; Donthu et al. 2021), centrality mapping (Bonacich 2007), and manual full-text review.

In this research, we use "human–AI collaboration" as an umbrella term for what other works call "partnership" or "teaming", treating these labels interchangeably in our Scopus search to capture settings where humans and AI jointly pursue shared goals (Dellermann et al. 2019; Wang et al. 2019, 2020; Sowa et al. 2021; Sharma et al. 2023). Also, to ensure accessible, high-contrast visuals, we used Tol (2021)'s colorblind-safe muted palette, maintaining consistent encoding across plots.

The methodology followed linear and sequential steps, as depicted in Figure 1.

Fig. 1 Sequential methodology integrating multiple steps.

2.1 PRISMA-guided protocol

A search was conducted in Scopus on June 9, 2025, and records were exported in CSV format. The query string was:

```
TITLE-ABS-KEY ( trust* OR distrust* OR mistrust* ) AND ( ‘‘AI’’ OR ‘‘artificial
intelligence’’ ) AND ( coworker* OR colleague* OR partner* OR collaborat* OR
team* OR group* ) AND ( ‘‘human-AI’’ OR ‘‘human-agent’’ OR ‘‘human-machine’’ OR
‘‘human-bot’’ OR ‘‘human-autonomy’’ ) AND ( hybrid* OR collect* OR augment* OR
extend* ) AND ( financ* ) AND DOCTYPE ( ar OR cp OR re ) AND ( LIMIT-TO (
LANGUAGE , "English" ) )
```

Following PRISMA 2020 (Page et al. 2021a,b), 430 records were identified in Scopus, as shown in Figure 2. After cleaning and automatic deduplication, 426 remained. Eligibility required both trust in human–AI collaboration as a primary construct, and an explicit financial application; records in which trust or finance was only incidental were excluded. We limited the corpus to English-language items indexed in Scopus as articles, conference papers, or reviews. In total, 114 records satisfied the eligibility criteria for inclusion in the bibliometric analysis.

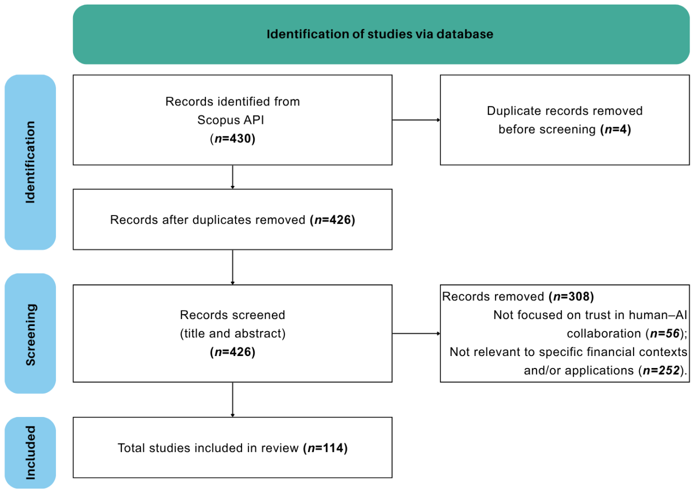

Fig. 2 PRISMA 2020 flow diagram. Adapted from Page et al. (2021a,b).

2.2 Bibliographic coupling and centrality mapping for community detection

Bibliographic coupling was performed on the 114 selected records, linking documents based on shared references (Zupic and Čater 2015; Donthu et al. 2021). Community detection was conducted using the weighted Leiden algorithm (Traag et al. 2019) in Python v3.12.9 adopting NetworkX (2025). This produced 6 clusters. Distinctiveness centrality was employed to identify the unique themes that characterize each cluster (Fronzetti Colladon and Naldi 2020). To determine the most structurally influential documents, eigenvector centrality served as the primary metric (Bonacich 2007). In addition, degree centrality, betweenness centrality, and closeness centrality were calculated to capture aspects of the network, namely, connectivity among papers, bridging roles, and overall integration within the structure, that complement the information provided by eigenvector centrality (Zhang and Luo 2017). The definitions and interpretations of these metrics are summarized in Table 1 for clarity.

Following this quantitative analysis, we conducted a manual full-text review of the 5 most central documents in each cluster, as ranked by raw and locally normalized eigenvector centrality.

2.3 B-SLR and thematic synthesis

Building on the above procedures, we structured the methodology around 4 research questions:

Table 1 Definitions of the centrality metrics used in this study.

Metric	Description
Distinctiveness	Captures the uniqueness of a node by favoring connections to peripheral, low-degree neighbors over hubs, thus highlighting keywords that uniquely characterize clusters.
Eigenvector	Measures a node's influence by considering not only its connections but also the importance of the nodes it connects to. Nodes linked to other highly influential nodes receive higher scores.
Degree	Counts the number of direct links (edges) a node has, reflecting its immediate connectivity and potential influence.
Betweenness	Quantifies how often a node lies on the shortest paths between other nodes, identifying nodes that act as bridges or control points for information flow.
Closeness	Indicates how near a node is to all others in the network, based on the shortest path distances. High values mean a node can reach others quickly and efficiently.

RQ1 — **Distinctive cluster themes** (i.e., "What are the most distinctive research themes related to trust in human–AI collaboration in finance?"): keywords were extracted and standardized using a curated alias dictionary to harmonize synonyms (e.g., *"artificial intelligence", "machine learning"* → **ai**; *"eXplainable AI", "interpretable ML"* → **xai**). Then, we ranked terms using distinctiveness centrality, which favors ties to peripheral, low-degree neighbors over hubs and thus highlights keywords that uniquely characterize clusters (Fronzetti Colladon and Naldi 2020).

RQ2 — **Trust conceptualization** (i.e., "How have research communities conceptualized trust in the most influential studies on trust in human–AI collaboration in finance?"): the 5 most central documents in each cluster, sorted by node ID and eigenvector centrality, were examined to identify trust conceptualizations. Data extraction followed an interpretive synthesis procedure, as suggested by Marzi et al. (2025).

RQ3 — **Methods, tools and trust dimensions mapping** (i.e., "What methods and tools have been proposed in the identified most influential studies to investigate specific trust dimensions in human–AI collaboration in financial contexts?"): methodological details, including experimental designs, modeling approaches, and tools, were extracted from the reviewed documents. These were mapped to corresponding trust dimensions identified in the literature.

RQ4 — **Emerging topics and gap identification** (i.e., "What emerging themes and research gaps exist in research on trust in human–AI collaboration in finance across identified clusters?"): to identify emerging themes and research gaps across clusters, a full-text analysis was conducted on the top 5 most influential papers within each cluster. The interconnections among the clusters, themes and gaps were depicted using a Sankey diagram produced with SankeyMATIC (2025).

3 Results

3.1 Overview of the dataset and bibliometric structures

This subsection reports the characteristics of the retrieved dataset and the bibliometric structures identified. It presents the PRISMA-guided results, annual publication trend and journal productivity distributions. Also, it visualizes the full bibliographic coupling network, the density contour plot, and the country-level distribution of the connected papers. Specifically, the initial Scopus retrieval produced 426 records. After manual screening of titles and abstracts, 114 finance-specific documents were retained for analysis. Figure 3 illustrates the annual publication trend (left) and the distribution of publications across journals (right). The left subplot shows a consistent rise in research output between 2018 and 2025. The right subplot presents the distribution of publications across the top-10 journals.

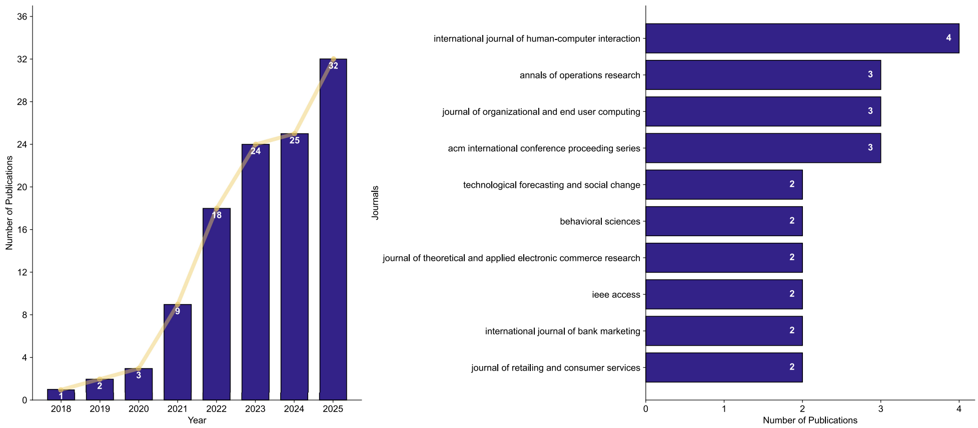

Fig. 3 Annual publication trend from 2018–2025 (left panel) and distribution of publications across journals (right panel).

Bibliographic coupling was applied to this set using the weighted Leiden community detection algorithm, which groups documents based on shared references (Traag et al. 2019). A minimum threshold of one shared reference was applied, yielding a connected network of 105 nodes and 610 coupling links, with 9 documents excluded as isolated. The resulting structure (Figure 4) reveals 6 distinct clusters ranging from 13 to 21 nodes.

As illustrated in Figure 5, the network forms 6 clusters with varying cohesion and connectivity, and an overall network modularity of 0.38 (Newman 2006). Clusters 5 and 1 are the most cohesive and form part of the network core, while the remaining clusters are smaller and less tightly connected, with weaker links to the rest of the network.

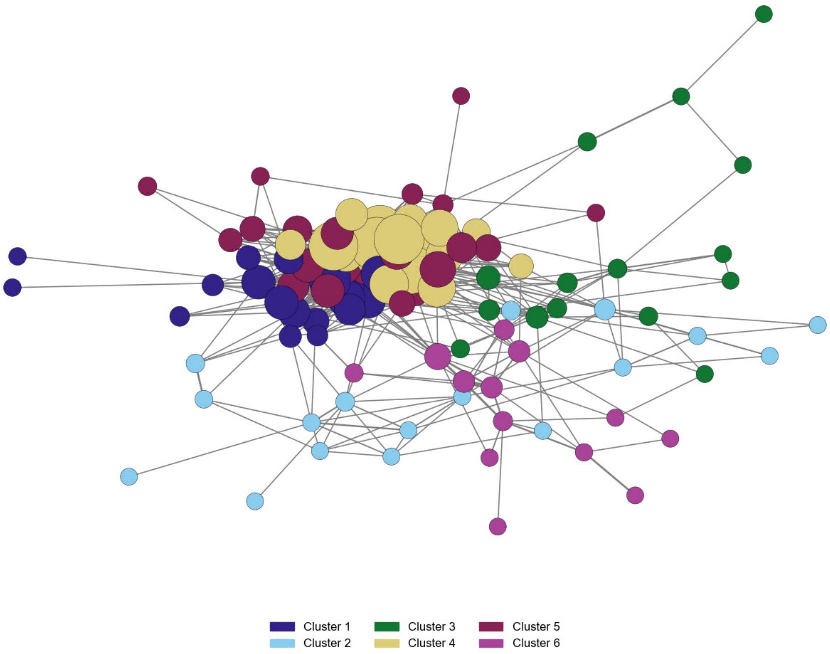

Fig. 4 Bibliographic coupling network illustrating the relationships between documents based on shared references. Node size indicates document citation count, and edge thickness reflects coupling strength.

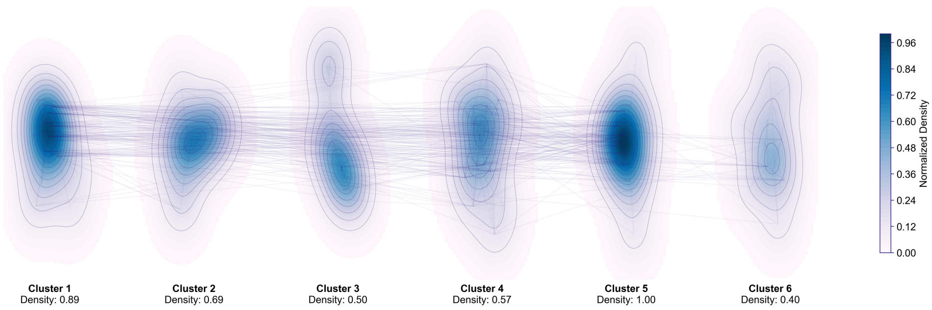

Fig. 5 Density contour plot illustrating cohesion across network clusters. Darker regions indicate stronger internal connectivity.

Furthermore, country-level affiliations for the 105 papers were analyzed using full and fractional counting (Perianes-Rodriguez et al. 2016). As shown in Figure 6, contributions are globally distributed, with China and India providing the largest shares, followed by several countries in Europe, North America, and Oceania. A broader set of countries contributes smaller proportions, reflecting a diverse but uneven international research landscape.

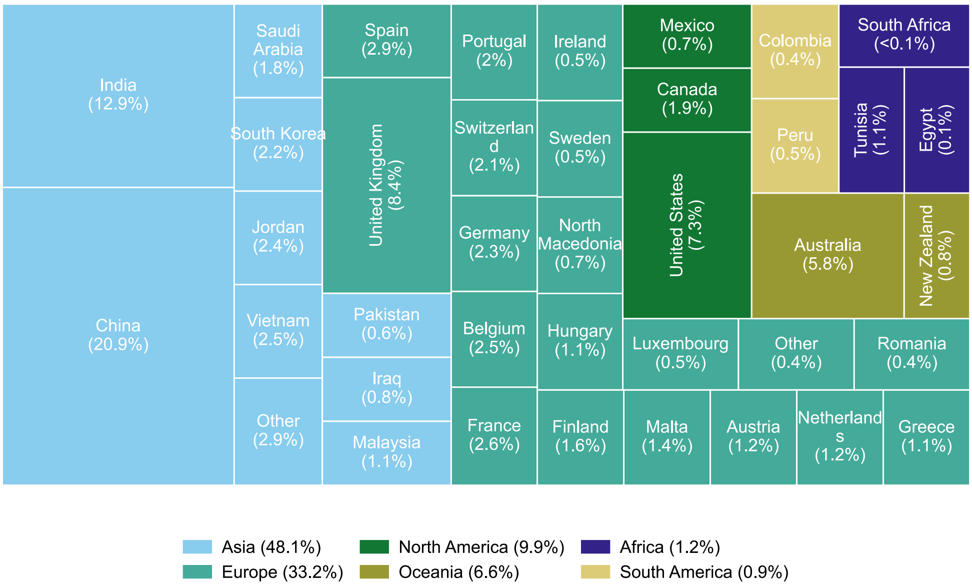

Fig. 6 Geographic distribution of publications based on country affiliation metadata from multi-country papers, using the fractional counting method. Each box is approximately sized proportionally to its global percentage share of fractionalized publication counts and annotated with its percentage. Colors denote continents.

3.2 Distinctive themes within identified clusters (RQ1)

To address RQ1 (*"What are the most distinctive research themes related to trust in human–AI collaboration in finance?"*), we first applied bibliographic coupling to organize the literature into 6 clusters. We then compute and report Global Weight Distinctiveness (GWD) values for each cluster and keywords in Table 2. Notice that, GWD is computed as defined by Fronzetti Colladon and Naldi (2020). Through this process, we identified the thematic core of each cluster and assigned 6 overarching thematic labels that capture their primary focus while highlighting the distinctive yet interconnected nature of the research areas represented in the dataset:

Cluster 1 — AI Governance in Finance: the thematic focus centers on *AI*, *AI governance*, *trust*, and *XAI*, with governance levers such as *responsible AI*, *controllability*, *intelligibility*, *AI transparency*, *AI adoption* and *finance*.

Cluster 2 — XAI for Finance: key terms include *XAI*, *AI*, and *finance*, alongside assurance terms, *ethics*, *regulatory compliance*, and *fairness*; and

Table 2 Top 10 distinctive keywords per cluster ranked by Global Weight Distinctiveness (GWD). The table highlights the most distinctive terms that characterize each research cluster, with thematic labels, top-10 higher GWD values indicating stronger distinctiveness and thematic specificity of the keyword within its cluster.

Top 10 distinctive keywords across thematic clusters

1: AI Governance in Finance		**2: XAI for Finance**		**3: Anthropomorphism in Financial AI Agents**	
keyword	*GWD*	*keyword*	*GWD*	*keyword*	*GWD*
AI	74.07	XAI	96.26	finance	76.21
AI governance	48.10	AI	58.72	anthropomorphism	71.64
trust	39.97	finance	54.88	AI agents	41.44
XAI	31.60	ethics	35.38	trust	34.14
finance	18.27	regulatory compliance	27.20	trust repair	17.88
responsible AI	15.71	trust	17.49	human-AI interaction	17.56
controllability	13.47	human-centric design	15.65	attitude towards usage	12.78
AI adoption	11.19	transparency	15.07	behavioral finance	10.42
intelligibility	10.77	human experts	12.32	CASA paradigm	9.20
AI transparency	10.36	fairness	10.84	ethics	7.65

4: User Interface Design for Human-AI in Finance		**5: Robo-Advisors for Financial Decision Making**		**6: Blockchain & Supply Chain**	
keyword	*GWD*	*keyword*	*GWD*	*keyword*	*GWD*
chatbot design	97.15	finance	68.38	blockchain	120.19
fintech	74.32	AI	68.13	supply chain	93.92
chatbot	63.22	robo-advisor	66.46	agrifood traceability	88.18
trust	56.45	acceptance	52.46	transparency	25.48
acceptance	51.86	trust	41.14	trust	20.74
perception	40.41	investment decision	29.95	sustainability	17.51
privacy	27.69	overreliance	29.21	adoption	12.78
AI	26.20	decision making	27.65	privacy and security	12.40
social presence	11.42	user experience design	22.03	IoT	10.76
human-AI interaction	8.28	AI anxiety	20.93	AI	4.34

design/process cues such as *human-centric design*, *transparency*, and *human experts*. *Trust* is positioned as an outcome of transparent and auditable systems.

Cluster 3 — **Anthropomorphism in Financial AI Agents:** the emphasis lies on *finance*, *anthropomorphism*, *AI agents*, and *trust*, with calibration mechanisms like *trust repair* and *human-AI interaction* and attitude constructs such as *attitude towards usage*. Behavioral frames (*behavioral finance*, *CASA paradigm*) and *ethics* appear as supporting themes. Notice that CASA stands for Computer and Social Actors.

Cluster 4 — **User Interface Design for Human-AI in Finance:** this cluster emphasizes *chatbot design*, *fintech*, *chatbot*, and *trust*, with adoption-relevant markers, such as *acceptance* and *perception*. Constraints and enablers include *privacy*, *social presence*, and *AI* anchoring *human-AI interaction* context.

Cluster 5 — **Robo-Advisors for Financial Decision Making:** central themes include *finance*, *AI*, and *robo-advisor*, linking *acceptance* and *trust*

Table 3 Top 5 nodes per cluster ranked by raw eigenvector centrality computed on each cluster's induced subgraph H_c. Raw eigenvector values (EV_r, three decimals) and cluster-normalized values (EV_{nrm}, two decimals). Each subbox is sorted by node number; the top EV_r per cluster is highlighted.

Top 5 Nodes by Eigenvector Centrality across Clusters								
1: AI Governance in Finance			2: XAI for Finance			3: Anthropomorphism in Financial AI Agents		
Node	**EV_r**	**EV_{nrm}**	**Node**	**EV_r**	**EV_{nrm}**	**Node**	**EV_r**	**EV_{nrm}**
30	0.174	0.27	9	0.327	0.61	43	0.353	0.70
36	0.256	0.39	29	0.489	0.93	79	0.326	0.65
38	0.640	1.00	41	0.295	0.55	100	0.376	0.75
69	0.614	0.96	63	0.524	1.00	107	0.501	1.00
99	0.169	0.26	78	0.402	0.76	111	0.494	0.99
4: User Interface Design for Human-AI in Finance			5: Robo-Advisors for Financial Decision Making			6: Blockchain & Supply Chain		
Node	**EV_r**	**EV_{nrm}**	**Node**	**EV_r**	**EV_{nrm}**	**Node**	**EV_r**	**EV_{nrm}**
12	0.393	1.00	56	0.375	0.91	55	0.149	0.23
23	0.391	0.99	68	0.368	0.89	90	0.366	0.60
27	0.333	0.83	73	0.411	1.00	92	0.307	0.50
31	0.383	0.97	89	0.328	0.80	108	0.602	1.00
75	0.332	0.82	97	0.373	0.91	109	0.603	1.00

to decision constructs (*investment decision*, *decision making*). Risk and user experience factors (*overreliance*, *user experience design*, *AI anxiety*) highlight the need for calibrated reliance on algorithmic advice.

Cluster 6 — Blockchain & Supply Chain: the cluster is anchored around *blockchain*, *supply chain*, and *agrifood traceability*, framing *transparency* and *trust* as properties of decentralized provenance. Application concerns include *sustainability*, *adoption*, *privacy and security*, and *IoT*, with *AI* playing a minor role.

Building on these thematic premises, the analysis proceeds to RQ2, focusing on how trust is conceptualized by most influential nodes within the 6 identified clusters.

3.3 Conceptualizations of trust in human–AI collaboration in finance (RQ2)

To address RQ2 (*"How have research communities conceptualized trust in the most influential studies on trust in human–AI collaboration in finance?"*), this subsection reports how the 5 most influential studies within each of the 6 clusters conceptualize trust in human–AI collaboration in finance. The findings are organized by cluster, reflecting the dominant approaches observed in the most structurally influential documents (ranked by node ID and eigenvector centrality). Table 3 summarizes the top 5 most influential nodes in each cluster based on raw and cluster-normalized eigenvector centrality. Degree, betweenness, and closeness centrality scores (both raw and normalized) are reported in Table A1 in the appendix with the aim of highlighting variations in local connectivity, brokerage potential, and network accessibility, respectively.

Complementing these tabular results, Figure 7 shows the bibliographic coupling network, illustrating the spatial organization of the 6 clusters and their internal and cross-cluster link structures.

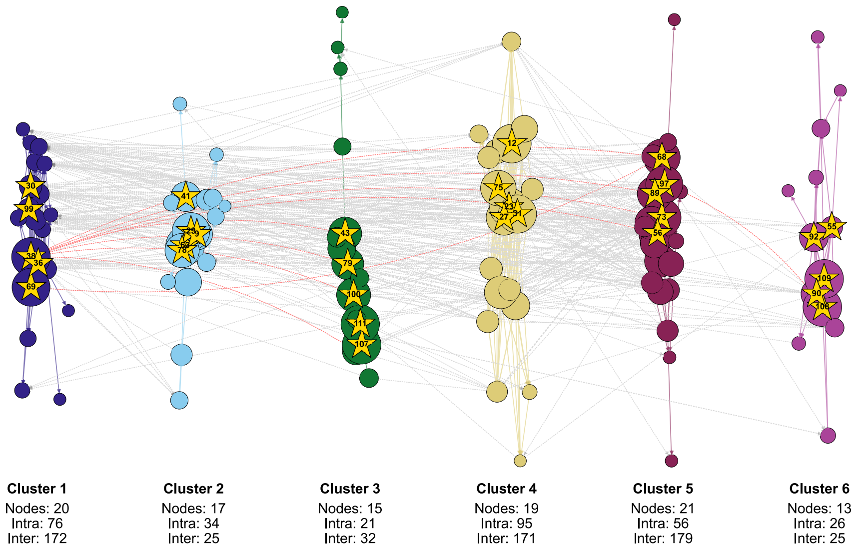

Fig. 7 Cluster-level network visualization derived from bibliographic coupling analysis, with clusters identified using the weighted Leiden algorithm. Each cluster is annotated with its top 5 nodes (marked with a ⋆) ranked by eigenvector centrality, highlighting the most structurally influential documents within their respective communities (indicated by a red dotted line, ···). Intra- and inter-cluster link densities are reported below each cluster to facilitate comparative assessment of internal cohesion and cross-cluster connectivity.

Building on these network-level insights, the subsequent paragraphs unpack how trust is conceptualized across the 6 clusters (see Table 4 for further details):

Cluster 1 — AI Governance in Finance: trust is cast as a governance outcome. It depends on competent, transparent systems that enable responsible human–AI collaboration (Khushk et al. 2024), relies on transparency and accountability as formal safeguards (Behl et al. 2023), treats algorithmic openness as the basis for managerial reliance (Chowdhury et al. 2023), equates confidence with explainable and auditable decisions aligned to oversight (Sabharwal et al. 2024), and embeds assurance within ethical control structures (Lehner et al. 2022).

Cluster 2 — XAI for Finance: trust grows from explainability that satisfies regulators and stakeholders. Transparency and fairness are emphasized (Machucho and Ortiz 2025). XAI tools reduce opacity and bolster user confidence (Saarela and Podgorelec 2024). Interpretability supports financial and strategic choices (Behera et al. 2023). Compliance and auditability anchor assurance in high-stakes settings (Khan et al. 2025), and transparent, fair, auditable insurance AI sustains both stakeholder and regulatory trust (Owens et al. 2022).

Cluster 3 — Anthropomorphism in Financial AI Agents: trust is presented as malleable and cue-driven. Apologies and human-like design help restore it after breaches (Kim and Song 2021), perceived competence and human-likeness calibrate it (Schreibelmayr et al. 2023). Some work privileges decentralized reliability over human-like features (Srinivas Aditya et al. 2021). Emotional and relational cues shape competence and integrity perceptions (Bhatia et al. 2021), and reliable algorithms tuned to investor emotions reinforce confidence in financial choices (Bhatia et al. 2020).

Cluster 4 — User Interface Design for Human-AI in Finance: trust emerges from social design choices, such as warmth–competence impressions formed through interface cues (Yan et al. 2025). Chatbot attributes, including ability, empathy, benevolence, and integrity, are especially important during recovery moments (Agnihotri and Bhattacharya 2024). Usability, coupled with anthropomorphic and emotionally intelligent signals (Wang et al. 2024), as well as aligned human-like cues that enhance social presence and perceived reliability (Chen et al. 2024), further strengthen user trust. Finally, chatbot expertise and responsiveness translate into overall brand trust (Li et al. 2023a).

Cluster 5 — Robo-Advisors for Financial Decision-Making: trust functions as the hinge between design and outcomes. Congruence, clarity, and commitment, amplified by self-efficacy and subtle human-like cues, support user acceptance (Chen et al. 2025). Human-like design features and user dispositions drive engagement, well-being, and loyalty (He and Liu 2024). Affective and ideological responses mediate how AI advice shapes attitudes and behaviors (Riedel et al. 2022), while early confidence reflects cultural factors and feature perceptions associated with adoption (Nourallah 2023). Computational–relational mechanisms enhance confidence in AI recommendations (Bawack et al. 2022).

Cluster 6 — Blockchain & Supply Chain: trust is reframed through transparency and decentralization. Blockchain visibility strengthens consumer assurance (Duong et al. 2024), while "trustless trust" orients distributed validation (Yavaprabhas et al. 2023). Fraud prevention and accountability mechanisms build confidence in sensitive supply chains (Majeed Parry et al. 2024). Transparency underpins resilient and trustworthy networks (Jellason et al. 2024), and distributed traceability enhances accountability across actors (Sri Vigna Hema and Manickavasagan 2024).

Building on these trust conceptualizations, we now address RQ3. Specifically, we focus on the methodological approaches adopted by scholars across clusters to investigate specific trust dimensions.

3.4 Methods and tools to investigate trust dimensions in human–AI collaboration in finance (RQ3)

To address RQ3 (*"What methods and tools have been proposed in the identified most influential studies to investigate specific trust dimensions in human–AI collaboration in financial contexts?"*), this subsection reports the methodological approaches employed across the most central studies within the 6 identified clusters.

Table 5 and the following paragraph report the main findings from the selected studies, specifying their clusters, nodes, authors, methods, tools, and the addressed trust dimensions. Moreover, acronyms are spelled out in the caption of the table.

Cluster 1 — AI Governance in Finance: this stream mixes evidence syntheses and empirical/technical studies: PRISMA-based SLRs (Khushk et al. 2024), cross-sectional modeling with Warp PLS-SEM (Behl et al. 2023), a supervised-ML implementation with LIME, an H2O deep classifier and precision–recall evaluation (Chowdhury et al. 2023), ADR with SHAP across classic ML techniques (Sabharwal et al. 2024), and a hermeneutic narrative review coded in ATLAS.ti using Rest's ethical framework (Lehner et al. 2022). Trust is parsed into fairness, accountability, sustainability and transparency (Behl et al. 2023); transparency/explainability with reliability and fairness (Chowdhury et al. 2023; Sabharwal et al. 2024); cognitive/affective/integrity facets with capability, compassion and honesty (Khushk et al. 2024); and ethical–regulatory facets such as objectivity, privacy, and accountability (Lehner et al. 2022).

Cluster 2 — XAI for Finance: most contributions are PRISMA SLRs that foreground explainability as a compliance and assurance lever (Owens et al. 2022; Saarela and Podgorelec 2024; Khan et al. 2025; Machucho and Ortiz 2025), complemented by a survey analyzed via CB-SEM in AMOS with Harman's test for bias (Behera et al. 2023). Tools concentrate on PRISMA and SEM suites, while the focal dimensions converge on transparency, explainability and fairness, joined by reliability/trustworthiness, integrity, compliance, responsibility, and reliance in decisions (Owens et al. 2022; Behera et al. 2023; Saarela and Podgorelec 2024; Khan et al. 2025; Machucho and Ortiz 2025).

Cluster 3 — Anthropomorphism in Financial AI Agents: methods span a 2×2 scenario experiment on apology attribution and anthropomorphism analyzed with ANOVA/ANCOVA and related post hoc/correlation tests (Kim and Song 2021); mixed scenario–survey designs with nonparametric inference and qualitative text analysis (Schreibelmayr et al. 2023); a SLR on decentralized reliability (Srinivas Aditya et al.

Table 4 Comparative mapping of trust conceptualizations across studies by cluster, node, and author(s) year).

Cluster	Node	Author(s) year	Trust conceptualization
1	30	Khushk et al. (2024)	Rooted in competence and transparency, framed as essential to responsible human–AI collaboration.
1	36	Behl et al. (2023)	Anchored in transparency and accountability as a governance mechanism ensuring responsible AI.
1	38	Chowdhury et al. (2023)	Rooted in algorithmic transparency supporting managerial reliance on AI outputs.
1	69	Sabharwal et al. (2024)	Confidence in AI decisions that remain explainable, auditable, and aligned with responsible governance.
1	99	Lehner et al. (2022)	Driven by transparency, accountability, and oversight in ethical AI governance.
2	9	Machucho and Ortiz (2025)	Regulator-aligned transparency and fairness.
2	29	Saarela and Podgorelec (2024)	Built via XAI tools reducing opacity and boosting confidence.
2	41	Behera et al. (2023)	Explainability-driven confidence in AI-supported financial and strategic decisions.
2	63	Khan et al. (2025)	Rooted in compliance and auditing within high-stakes financial AI.
2	78	Owens et al. (2022)	Transparent, fair, and auditable insurance AI ensuring stakeholder and regulatory confidence.
3	43	Kim and Song (2021)	Repairable via apologies and anthropomorphic tactics post-trust breach.
3	79	Schreibelmayr et al. (2023)	Calibrated by perceived competence and human-likeness shaped through anthropomorphic cues.
3	100	Srinivas Aditya et al. (2021)	Rooted in decentralized reliability over anthropomorphism.
3	107	Bhatia et al. (2021)	Perceived competence and integrity shaped by human-like emotional and relational cues.
3	111	Bhatia et al. (2020)	Algorithmic reliability with human-like sensitivity to investor emotions in financial decisions.
4	12	Yan et al. (2025)	Arises from warmth and competence perceptions shaped by the design of social cues.
4	23	Agnihotri and Bhattacharya (2024)	Linked to chatbot traits: ability, empathy, benevolence, and integrity in recovery.
4	27	Wang et al. (2024)	Links ease-of-use with relational cues (anthropomorphism and emotional intelligence).
4	31	Chen et al. (2024)	Aligned anthropomorphic cues enhancing perceived social presence and reliability.
4	75	Li et al. (2023a)	chatbot expertise, anthropomorphism, and responsiveness reinforcing brand trust.
5	56	Chen et al. (2025)	Stemmed from congruence, clarity, commitment; reinforced by self-efficacy, anthropomorphic cues.
5	68	He and Liu (2024)	Rooted in human-like design and user traits, trust underpins engagement, well-being, and loyalty.
5	73	Riedel et al. (2022)	Affective and ideological mediator linking AI financial advice to user attitudes and behaviors.
5	89	Nourallah (2023)	Initial and culturally shaped confidence linking robo-advisor features to user adoption.
5	97	Bawack et al. (2022)	Computational–relational mechanism enhancing confidence in AI-driven recommendations.
6	55	Duong et al. (2024)	Confidence in blockchain transparency reinforcing consumer assurance.
6	90	Yavaprabhas et al. (2023)	Defined as "trustless trust" through blockchain.
6	92	Majeed Parry et al. (2024)	Built through fraud prevention and accountability in sensitive supply chains.
6	108	Jellason et al. (2024)	Confidence in blockchain transparency fostering trustworthy and resilient supply chains
6	109	Sri Vigna Hema and Manickavasagan (2024)	Distributed trust from blockchain transparency enhancing accountability.

Table 5 Summary of top influential papers by methods, tools, and trust dimensions. Acronyms: ADR = Action Design Research; AMOS = Analysis of Moment Structures; ANCOVA = Analysis of Covariance; ANOVA = Analysis of Variance; CB-SEM = Covariance-Based Structural Equation Modeling; FVT = Face Validity Test; H2O DL = H2O Deep Learning; LIME = Local Interpretable Model-agnostic Explanations; ML = Machine Learning; NCA = Necessary Condition Analysis; PEO = Population–Exposure–Outcome; PLS-SEM = Partial Least Squares Structural Equation Modeling; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RPYS = Reference Publication Year Spectroscopy; SEM = Structural Equation Modeling; SHAP = SHapley Additive exPlanations; SLR = Systematic Literature Review; SOIPSVM = Selective Organizational Information Privacy and Security Violations Model; SPC = Search Path Count; Warp PLS = Warp Partial Least Squares; WMW = Wilcoxon–Mann–Whitney test; WSE = Within-Subject Experiment; RM-ANOVA = Repeated-Measures ANOVA.

Cluster	Node	Author(s) Year	Methods	Tools	Dimensions
1	30	Khushk et al. (2024)	SLR	PRISMA	cognitive, affective, integrity, transparency
1	36	Behl et al. (2023)	Survey; FVT, SOIPSVM	Warp PLS; PLS-SEM	fairness, accountability, sustainability, transparency
1	38	Chowdhury et al. (2023)	ML framework	LIME; H2O DL	transparency, explainability, fairness, accountability, reliability
1	69	Sabharwal et al. (2024)	ADR; ML experiments	SHAP	transparency, explainability, accountability, reliance, fairness
1	99	Lehner et al. (2022)	Semi-systematic review	ATLAS.ti; Rest's ethics framework	cognitive, privacy, transparency, integrity
2	9	Machucho and Ortiz (2025)	SLR	PRISMA	transparency, fairness, integrity, reliance, compliance
2	29	Saarela and Podgorelec (2024)	SLR	PRISMA	transparency, explainability, reliability, integrity
2	41	Behera et al. (2023)	Survey	CB-SEM; Harman test	transparency, explainability, reliance, compliance
2	63	Khan et al. (2025)	SLR	PRISMA	transparency, explainability, fairness, accountability, compliance
2	78	Owens et al. (2022)	SLR	PRISMA	cognitive, relational, compliance, transparency
3	43	Kim and Song (2021)	Scenario experiment	ANOVA/ANCOVA; Duncan; Pearson	cognitive, behavioral, reparation
3	79	Schreibelmayr et al. (2023)	Survey	WMW; Spearman	cognitive, affective, relational, transparency
3	100	Srinivas Aditya et al. (2021)	SLR	None	integrity, transparency, decentralized reliability
3	107	Bhatia et al. (2021)	Expert interviews	Deductive/latent coding	cognitive, relational, transparency, situational
3	111	Bhatia et al. (2020)	Focus groups	NVivo	cognitive, affective, transparency, reliance
4	12	Yan et al. (2025)	Scenario experiments	Mediation analysis	cognitive, affective, relational
4	23	Agnihotri and Bhattacharya (2024)	Scenario experiments	SEM; MPLUS; bootstrap	ability, benevolence, integrity
4	27	Wang et al. (2024)	Survey	SEM; mediation	usability, affective, anthropomorphism
4	31	Chen et al. (2024)	WSE	RM-ANOVA (SPSS)	cognitive, social presence, satisfaction
4	75	Li et al. (2023a)	Survey	SEM (AMOS); regression; reliability	competence, interactivity, usability, anthropomorphism, brand trust, human support; risk; security
5	56	Chen et al. (2025)	Survey	PLS-SEM; NCA	competence; communication/commitment/congruence
5	68	He and Liu (2024)	Science mapping	SciMAT; SPC	competence, dependability, explainability
5	73	Riedel et al. (2022)	Scenario experiments	PROCESS	affective, ideology, reliability
5	89	Nourallah (2023)	Survey	PLS-SEM; bootstrap; k-means	performance expectancy, hedonic, facilitating conditions, social influence, trust propensity, culture
5	97	Bawack et al. (2022)	Bibliometric review	Louvain; RPYS	cognition, reliance
6	55	Duong et al. (2024)	Survey	PROCESS	transparency, relational trust, integrity
6	90	Yavaprabhas et al. (2023)	SLR	Inductive/deductive coding	cognitive, institutional, affective, temporal, trustor–trustee
6	92	Majeed Parry et al. (2024)	Narrative review	None	transparency, traceability, authenticity, integrity, accountability, consumer trust
6	108	Jellason et al. (2024)	SLR	PEO	transparency, traceability, accountability, integrity, governance
6	109	Sri Vigna Hema and Manickavasagan (2024)	Comprehensive review	None	transparency, traceability, integrity, authenticity, accountability

2021); and two phenomenological programs-expert interviews with deductive content analysis (Bhatia et al. 2021) and focus groups coded in NVivo (Bhatia et al. 2020). Dimensions traverse cognitive competence and understandability, behavioral reliance, affective responses (benevolence, uncanniness), relational human-likeness/agency, transparency/explainability, data integrity and security, plus situational and human-in-the-loop reliance factors (Bhatia et al. 2020, 2021; Kim and Song 2021; Srinivas Aditya et al. 2021; Schreibelmayr et al. 2023).

Cluster 4 — **User Interface Design for Human-AI in Finance:** evidence comes from scenario experiments with mediation tests (Yan et al. 2025), dual-study scenarios with SEM in Mplus and bootstrapping (Agnihotri and Bhattacharya 2024), survey-based SEM with mediation (Wang et al. 2024), a within-subject eye-tracking study analyzed via repeated-measures ANOVA and gaze metrics (Chen et al. 2024), and survey models in AMOS with additional regressions and reliability/validity checks (Li et al. 2023a). Reported dimensions emphasize warmth–competence and relational trust (Yan et al. 2025), ability–benevolence–integrity in service recovery (Agnihotri and Bhattacharya 2024), usability and anthropomorphic/emotional cues (Wang et al. 2024), social presence and satisfaction alongside cognitive trust (Chen et al. 2024), responsiveness, ease, anthropomorphism, brand trust, reliance on human support, risk, and privacy/security moderation (Li et al. 2023a).

Cluster 5 — **Robo-Advisors for Financial Decision-Making:** approaches include scenario surveys with PLS-SEM and NCA (Chen et al. 2025), bibliometric main-path/science mapping using SciMAT (He and Liu 2024), serial-mediation experiments with PROCESS (Riedel et al. 2022), a cross-national PLS-SEM study with bootstrapping and k-means segments (Nourallah 2023), and a bibliometric review leveraging Bibliometrix, Louvain clustering, and RPYS (Bawack et al. 2022). Trust is broken down into cognitive competence with relational communication/commitment/congruence (Chen et al. 2025), dependability and explainability (He and Liu 2024), emotion-based and ideological mediators with perceived reliability (Riedel et al. 2022), adoption-related cognitions (performance/effort), hedonic and situational supports, social/dispositional tendencies, and cultural orientations (Nourallah 2023), plus core cognition–reliance linkages (Bawack et al. 2022).

Cluster 6 — **Blockchain & Supply Chain:** methods range from surveys modeled with PROCESS (Duong et al. 2024) to SLRs with inductive/deductive coding (Yavaprabhas et al. 2023), narrative/case syntheses (Majeed Parry et al. 2024), and PEO-guided thematic SLRs (Jellason et al. 2024), alongside a comprehensive review (Sri Vigna Hema and Manickavasagan 2024). The mapped dimensions consistently spotlight transparency and traceability, extended by authenticity, integrity/data immutability, accountability/compliance, provenance governance, temporal (swift vs. long-term) and multi-perspective trust facets spanning cognitive, institutional and affective layers (Yavaprabhas et al. 2023; Duong et al. 2024; Majeed Parry et al. 2024; Jellason et al. 2024; Sri Vigna Hema and Manickavasagan 2024).

Having established the methods, tools, and trust dimensions, we now proceed to RQ4, which explores the key emerging trends and gaps that may guide future research.

3.5 Emerging themes and research gaps in trust in human–AI collaboration in finance (RQ4)

To address RQ4 (*"What emerging themes and research gaps exist in research on trust in human–AI collaboration in finance across identified clusters?"*), this subsection reports the results of the full-text review of the 5 most influential papers within each of the 6 identified clusters. The aim is to map the emerging themes and the corresponding research gaps identified by the authors related to trust in human–AI collaboration in finance. The resulting visualization (Figure 8) synthesizes the overall structure of the analysis. Each cluster is positioned on the left, followed by the derived themes in the center and the associated research gaps on the right. The numbers adjacent to each node indicate the relative frequency of occurrence of that element within the corpus, while the width of each colored flow (by cluster) reflects the intensity of connection between clusters, themes, and gaps.

The identified emerging themes and related research gaps are summarized as follows:

Cluster 1 — AI Governance in Finance: themes span organizational human–AI interaction, transparent financial systems, XAI-based governance, and ethical decision-making in accounting. Though, gaps center on stronger empirical tests, moderators of trust and team dynamics, transition frameworks to transparent AI, AI literacy, and models for human–machine moral collaboration (Lehner et al. 2022; Behl et al. 2023; Chowdhury et al. 2023; Khushk et al. 2024; Sabharwal et al. 2024).

Cluster 2 — XAI for Finance: work highlights XAI for business model innovation, cross-domain XAI, information system value and strategy, and insurance compliance/trust. However, missing are standardized metrics/benchmarks, longitudinal and cross-domain validations, causal evidence for growth advantage, as well as an agreed balance between interpretability, scalability and accuracy (Owens et al. 2022; Behera et al. 2023; Saarela and Podgorelec 2024; Khan et al. 2025; Machucho and Ortiz 2025).

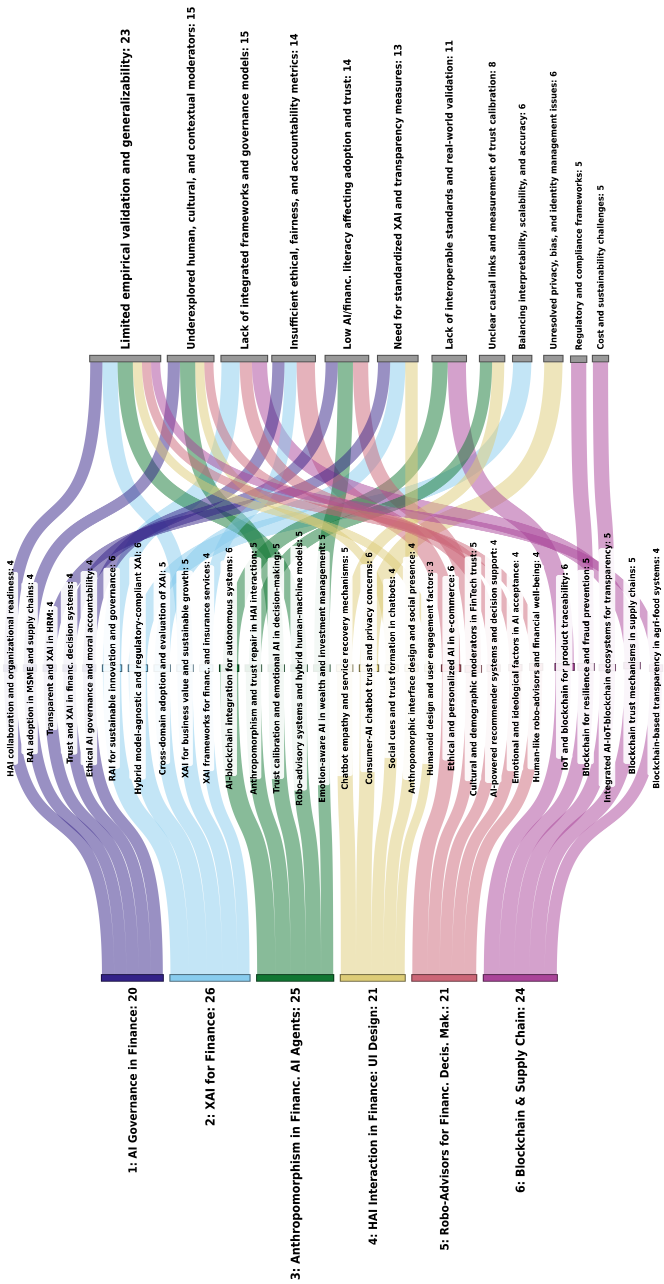

Fig. 8 Sankey diagram showing the distribution and interconnection of research themes and gaps across the 6 identified clusters. Clusters are shown on the left, emerging themes in the center, and corresponding research gaps on the right. Numbers beside each node indicate their relative frequency, while flow thickness (colored by cluster) represents the strength of thematic linkage.

Cluster 3 — Anthropomorphism in Financial AI Agents: themes address trust repair processes, calibration via transparency/emotion, decentralized reliability (AI–blockchain), and hybrid robo-advisory models with emotion-aware decisions. Nevertheless, gaps involve precise anthropomorphism dimensions and boundary conditions, validated multi-item trust measures, real-world testbeds/standards for blockchain-based robotics, quantitative validation, richer data integration, and larger/diverse samples (Bhatia et al. 2020, 2021; Kim and Song 2021; Srinivas Aditya et al. 2021; Schreibelmayr et al. 2023).

Cluster 4 — User Interface Design for Human-AI in Finance: emerging themes focus on social-cue design, warmth–competence pathways, humanoid traits (ease, emotional intelligence, personalization), visual–conversational interplay, and multi-factor chatbot trust (privacy, risk, brand, human support). But, needed are longitudinal and cross-cultural tests, broader cue taxonomies, guidance for empathetic design under failure, field methods beyond labs, and mapping trust across exposure–adoption with clarity on human support vs. autonomy (Li et al. 2023a; Agnihotri and Bhattacharya 2024; Chen et al. 2024; Wang et al. 2024; Yan et al. 2025).

Cluster 5 — Robo-Advisors for Financial Decision-Making: themes include human-like features and self-efficacy as trust antecedents, optimization/personalization and sentiment, affect–attitude mechanisms, cross-cultural initial trust and Unified Theory of Acceptance and Use of Technology (UTAUT), and recommenders. However, gaps point to longitudinal trust–loyalty dynamics, cultural/demographic moderators, real-world validation of humanoid cues and recommenders, ethical and sustainable XAI adoption, discrete emotions and ideology effects, multi-country panels, and voice-assistant support (Bawack et al. 2022; Riedel et al. 2022; Nourallah 2023; He and Liu 2024; Chen et al. 2025).

Cluster 6 — Blockchain & Supply Chain: themes emphasize blockchain traceability/transparency in food and wine chains, trustor–trustee lenses, IoT/smart contracts, resilience for small agri-food firms, and cross-chain, sustainability, and ethical sourcing uses. Nonetheless, gaps include cross-cultural/longitudinal validation, reconciling trustless vs trust-enabling views, standards/interoperability and scalability, governance/compliance and cost barriers, energy impacts, stakeholder roles, and long-term safety/efficiency assessments (Yavaprabhas et al. 2023; Duong et al. 2024; Jellason et al. 2024; Majeed Parry et al. 2024; Sri Vigna Hema and Manickavasagan 2024).

4 Discussion and future directions

4.1 Discussion of results

Guided by the results reported in Section 3, we next discuss RQ1–RQ4 and propose a novel, multi-level socio-technical framework, accompanied by four financial use cases.

First, bibliographic coupling and distinctiveness analysis revealed 6 interrelated thematic clusters structuring the research landscape of trust in human–AI collaboration in finance: i) AI governance in finance; ii) XAI for finance; iii) anthropomorphism in financial AI agents; iv) user interface design for Human-AI in Finance; v) robo-advisors for financial decision making; and vi) blockchain and supply chain. These clusters highlight how trust acts as an integrative bridge between behavioral outcomes (e.g., acceptance, reliance), technical affordances (e.g., transparency, explainability), and institutional and infrastructural assurances (e.g., governance, blockchain). Also, this structure indicates a coherent yet fragmented field with present but limited cross-fertilization among clusters.

Second, across clusters, trust is conceptualized through complementary yet fragmented lenses. In the governance and XAI clusters, trust is predominantly cognitive and procedural, anchored in transparency, accountability, and explainability as governance levers for assurance (Khushk et al. 2024; Sabharwal et al. 2024; Khan et al. 2025; Machucho and Ortiz 2025). In anthropomorphism and interaction clusters, trust takes a socio-affective form, shaped by warmth–competence perceptions, empathy, and anthropomorphic cues that enhance social presence and perceived reliability (Bhatia et al. 2021; Agnihotri and Bhattacharya 2024; Yan et al. 2025). In robo-advisor research, trust functions as a behavioral mediator linking design attributes and user adoption, balancing confidence and overreliance (Riedel et al. 2022; Chen et al. 2025). Finally, blockchain studies reframe trust as infrastructural, embedded in transparency, traceability, and decentralized verification (Duong et al. 2024; Sri Vigna Hema and Manickavasagan 2024).

Third, methodologically, survey-based SEM dominates the literature, especially within the governance, XAI, interaction, and robo-advisor clusters (Riedel et al. 2022; Behl et al. 2023; Wang et al. 2024). Scenario experiments, qualitative interviews, and eye-tracking studies provide complementary but less frequent evidence, often in user-interface and anthropomorphism contexts (Bhatia et al. 2021; Yan et al. 2025; Kim and Song 2021). Technical research integrates XAI tools such as SHAP and LIME for model interpretability (Chowdhury et al. 2023; Sabharwal et al. 2024), while blockchain research employs thematic reviews and process frameworks focused on transparency and accountability (Yavaprabhas et al. 2023; Jellason et al. 2024). The methodological landscape remains cross-sectional, with limited triangulation across behavioral and technical evidence.

Fourth, across clusters, 5 broad research frontiers emerge:

1. Governance and ethics research calls for frameworks translating Responsible AI principles into auditable practice (Behl et al. 2023; Khushk et al. 2024).
2. XAI requires standardized metrics, longitudinal validation, and scalable model-agnostic tools (Khan et al. 2025; Machucho and Ortiz 2025).

3. Human–AI interaction and anthropomorphism need empirically validated cue taxonomies, longitudinal and cross-cultural testing, and real-world deployments (Kim and Song 2021; Yan et al. 2025).
4. Robo-advisor studies demand examination of cultural and emotional moderators of trust and long-term user loyalty (Riedel et al. 2022; Nourallah 2023; Chen et al. 2025).
5. Blockchain research should address the challenges of interoperability, sustainability, and governance, particularly the trade-offs between trustless and trust-enabling architectures (Duong et al. 2024; Jellason et al. 2024).

Across the 6 clusters, our bibliometric and qualitative synthesis reveals a consistent multi-level pattern in how trust is shaped within financial human–AI collaboration. Cluster 1 (AI Governance in Finance) and Cluster 2 (XAI for Finance) predominantly engage meso–macro mechanisms, emphasizing organizational governance practices, regulatory alignment, and system-level assurance structures. Cluster 3 (Anthropomorphism in Financial AI Agents), Cluster 4 (User Interface Design for Human-AI in Finance), and Cluster 5 (Robo-Advisors for Financial Decision Making) operate chiefly at the micro–meso interface, centering on individual cognition, affect, social presence, and how organizational design choices shape user experience. Cluster 6 (Blockchain & Supply Chain), by contrast, is rooted primarily in macro-level infrastructures of trust with AI serving a secondary yet conceptually relevant role. These patterns demonstrate that trust emerges from different configurations of micro-, meso-, and macro-level mechanisms depending on the domain, providing the empirical foundation for the integrative socio-technical framework proposed in the next subsection.

4.2 Introducing a multi-level socio-technical framework of trust

Synthesizing evidence across the four RQs, this study conceptualizes trust in financial human–AI collaboration as a dynamic, layered, and emergent socio-technical phenomenon operating across three interconnected levels: micro, meso, and macro. This framing draws on complex adaptive systems theory, which explains how local interactions among heterogeneous agents can generate system-level coherence over time (Holland 1992). In this view, trust arises through recursive feedback loops linking individual experiences, organizational practices, and institutional and infrastructural solutions along with systemic governance. Furthermore, this framework extends complexity thinking to account for how, in mixed-motive and interdependent interaction settings, participation, cooperation, and trust emerge from adaptive processes of communication and commitment among human and artificial agents, embedded within evolving technical and institutional environments (Lee et al. 2019; Dafoe et al. 2020; Reuel et al. 2025).

In this sense, at the micro level, trust originates from user behavior and perception to AI performance and interaction quality. Confidence in competence, reliability, and fairness combines with emotional resonance or warmth, particularly salient in social applications such as robo-advisory and conversational agents. These patterns align with socio-cognitive models and the CASA theory (Nass et al. 1994; Castelfranchi and Falcone 2010; Pelau et al. 2021), emphasizing that both rational evaluations and affective responses co-produce user trust.

At the meso level, trust becomes embedded in organizational procedures and design practices. Here, XAI tools (e.g., SHAP, LIME) and interface design decisions operationalize transparency, understandability, and recoverability, translating organizational processes, controls, and development practices into user experience. However, as Hoffman (2024) notes, trust cannot be engineered solely through technical transparency. The notion of *trustability* expands this focus toward experiential qualities, such as usability, learnability, and perceived usefulness, that sustain trust development within organizational contexts.

At the macro level, regulatory systems and blockchain technologies might encode verification and governance mechanisms promoting public and infrastructural trust by delivering consistent, systemic assurances. This add a further interpretative layer to trust as a property of the socio-technical environment, influenced by systemic governance and intrinsic technology features.

These levels interact dynamically rather than hierarchically, as depicted in Figure 9. Micro-level perceptions inform meso-level design adaptations; meso-level practices are shaped by macro-level infrastructures; and macro-level norms, in turn, influence individual expectations and behaviors. The resulting system reflects trust as a complex adaptive state that scales across layers of the socio-technical ecosystem.

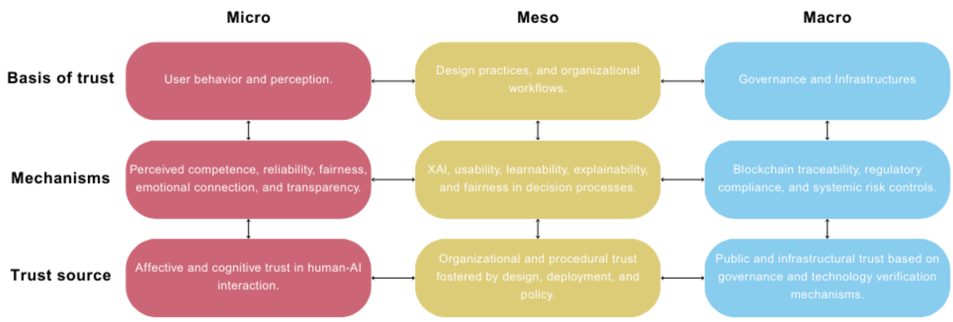

Fig. 9 Multi-level socio-technical framework of trust in human–AI collaboration in finance, derived from bibliometric and conceptual synthesis. Columns represent analytical levels (Micro ↔ Meso ↔ Macro), while rows denote thematic dimensions. Horizontal and vertical bidirectional arrows indicate reciprocal influences, positioned within the gaps between cells for clarity. The micro level shapes meso-level design through user perceptions and behaviors; the meso level translates these needs into system features and aligns them with macro-level infrastructures; the macro level establishes the enabling environment for meso-level implementation and shapes micro-level perceptions.

To critically exemplify this dynamic interplay, Table 6 outlines four finance-oriented cases: i) wire transfer verification, ii) financial education, iii) stock market participation, and iv) financial reporting. Each case demonstrates how trust emerges through interactions across the micro, meso, and macro levels as an emergent and distributed property across the socio-technical system. To further support conceptual integration, Figure 10 visually summarizes the framework, simplifying the understanding of cross-level interactions and feedback loops that shape trust formation and calibration over time.

Table 6 Four examples of dynamic interactions across micro, meso, and macro levels in financial contexts.

Context	Micro	Meso	Macro	Dynamic interactions
Verification of Payee (VoP)	Payers verify that beneficiary names match account identifiers before confirming transfers.	Payment providers perform real-time name verification between payer and payee institutions.	The European Payments Council defines common rules to enhance security and trust in credit transfers.	Coordination among users, providers, and scheme governance reinforces reliability and confidence in payments.
Financial education	Individuals seek to understand and manage financial choices.	Organizations use digital tools and adaptive learning systems to deliver guidance.	National strategies and public policies promote inclusion and literacy.	Feedback between learners, providers, and policy design improves effectiveness and trust.
Stock market participation	Investors interpret market signals and assess risk.	Trading platforms employ algorithmic analytics and advisory systems.	Regulators and exchanges oversee standards, transparency, and stability.	Investor actions influence platform design and regulatory responses over time.
Financial reporting	Users interpret and rely on disclosed financial data.	Firms employ automated reporting and analysis systems.	Standards and oversight bodies ensure consistency and accountability.	Mutual adaptation between users, firms, and regulators sustains transparency and trust.

4.3 Limitations and boundary conditions

Even if we have reported results from a methodologically solid and grounded B-SLR, there are different boundary conditions that must be acknowledged.

First, the retrieval is Scopus-based and time-bounded. As a result, relevant contributions indexed in other databases or preprint repositories may be omitted. Furthermore, backward and forward snowballing was not conducted, which may limit the inclusion literature. Additionally, potential cultural or linguistic biases may arise from the specific Scopus query keywords used (e.g., *trust* vs. *safety*), which could affect the comprehensiveness of the retrieved studies.

Second, bibliometric structures are intrinsically sensitive to database scope, citation density, and temporal dynamics. Thus, changing the threshold for bibliographic coupling would reduce the number of papers included in the coupling network, since

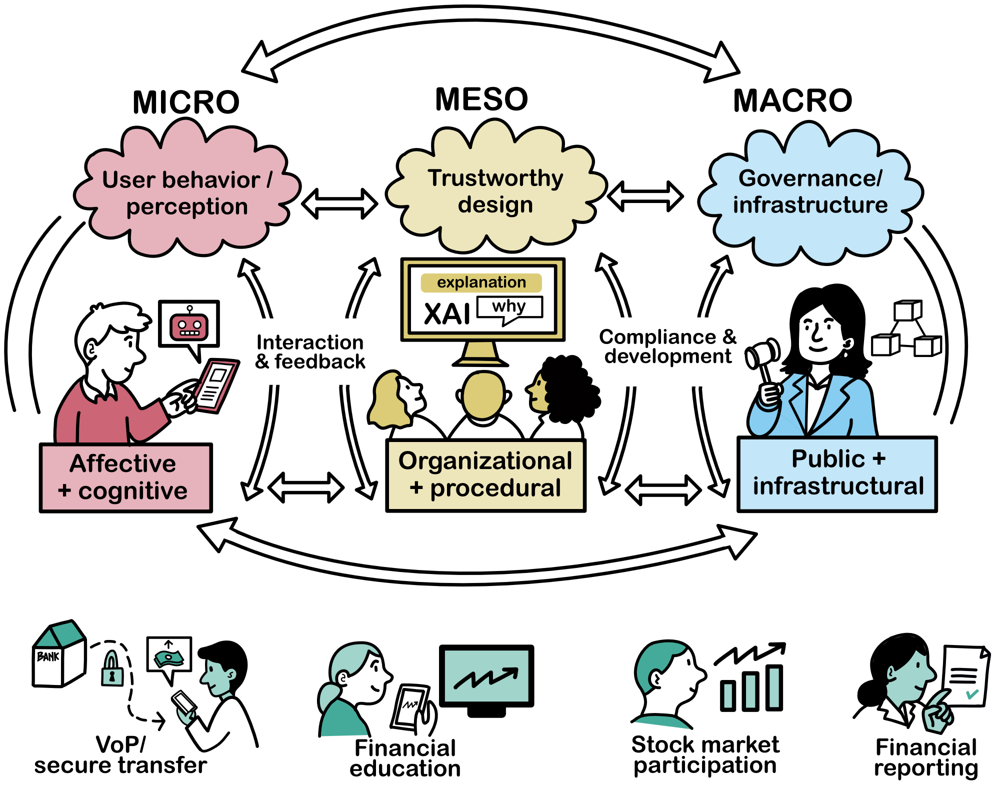

Fig. 10 Visual sketch of the multi-level socio-technical framework of trust in human-AI collaboration. The illustration simplifies the conceptual relationships between micro-, meso-, and macro-level dynamics, emphasizing feedback loops and adaptive interactions across individuals, organizations, and institutional infrastructures. It serves as an accessible visual complement to the theoretical model presented in Figure 9 and to the examples provided in Table 6.

higher thresholds exclude documents with weaker reference overlaps, thereby decreasing connectivity and potentially biasing cluster identification toward densely cited works. In our case, this threshold was set to the minimum (one shared reference) to preserve network connectivity and avoid excluding relevant but less densely connected papers. Additionally, citation-based metrics tend to favor already well-cited or older works, potentially underrepresenting novel or interdisciplinary studies (van den Besselaar and Sandström 2019). To mitigate such biases, we adopted eigenvector centrality rather than raw citation counts as the primary measure of influence, as it emphasizes connection to other influential works, rather than sheer citation volume. This approach aligns with best practices in contemporary bibliometric research (Belter 2015) and provides a more balanced representation of intellectual influence.

Third, the interpretation of clusters and thematic boundaries reflects both algorithmic and human judgment. Although the weighted Leiden algorithm and full-text review enhance robustness, the resulting themes depend on parameter choices and coding interpretations. Hence, cluster delineations should be viewed as indicative rather than absolute, providing a conceptual rather than taxonomic map of the field.

Finally, the results are temporally bound and context-dependent. Rapid advances in AI, new regulatory frameworks, and evolving financial technologies may reshape the research landscape beyond this study. Therefore, while the present synthesis offers a baseline, it should be revisited and updated to capture the dynamic evolution of human–AI trust research in finance.

4.4 Contributions and future research directions

This review advances the field in three ways. First, as far as we know, it offers the first B-SLR of trust in human–AI collaboration in finance. Second, it clarifies how trust is conceptualized, and identifies methodological contributions, emerging themes and research gaps. Third, it introduces a novel, multi-level socio-technical framework explaining how trust is emergent and distributed across levels and actors, with four exemplifying use cases.

These insights point to several avenues for future research. There is a need to move beyond survey-based trust measures toward behavioral, task-level, and longitudinal approaches that better capture calibrated reliance and trust repair. More work is also required to develop actionable XAI for finance, including standardized explanation-quality metrics and evidence of how explanations shape decisions. Research on anthropomorphism and interface design should identify when human-like cues help or hinder collaboration, especially in recovery scenarios. At the institutional level, questions of provenance, auditability, and interoperability remain central to fostering trust. Future studies should also broaden geographic and contextual diversity to understand cultural, regulatory, and sectoral influences. Finally, the rise of Gen-AI and multi-agent workflows brings new challenges for compliance, oversight, and explainability, underscoring the need for longitudinal, multi-level designs that trace how trust evolves across actors, tasks, and settings.

5 Conclusions

This study presents, as far as we know, the first B-SLR of trust in human–AI collaboration in finance, synthesizing 114 studies (2018–2025). The evidence reveals a rapidly expanding yet fragmented field structured into 6 clusters in which trust consistently functions as a multidimensional bridge linking socio-cognitive mechanisms, technical properties and organizational practices, and institutional and infrastructural assurances. Despite notable advances in XAI and adoption research, integration across trust conceptualizations remains limited, hindering translation into practice. Addressing these gaps calls for cross-disciplinary work, multi-database retrievals, and longitudinal or field designs that move beyond intentions to use and instead capture calibration, resilience after failure, and institution-level assurance in real financial workflows. Anchoring our review, the proposed multi-level socio-technical framework consolidates dispersed findings into a coherent account of how trust is formed, maintained, repaired and calibrated across micro, meso, and macro levels. Finally, the framework offers a testable roadmap for longitudinal and field studies that causally link perceptions, practices, and assurances. In doing so, it becomes an actionable guide for designing and governing trust-aware, human- and societally aligned AI.

Appendix A Cluster-level centrality metrics

Table A1 Top-5 nodes per cluster ranked by raw eigenvector on each cluster's induced subgraph H_c. Raw centrality values (eigenvector with three decimals; others with two) and, in parentheses, the corresponding cluster-normalized scores (min–max within cluster to $[0,1]$, two decimals).

Cluster	Node ID	Eigenvector (raw; local norm.)	Degree (raw; local norm.)	Betweenness (raw; local norm.)	Closeness (raw; local norm.)
1	30	0.174 (0.27)	0.63 (0.92)	0.03 (0.14)	1.22 (0.88)
1	36	0.256 (0.39)	0.53 (0.75)	0.04 (0.18)	1.30 (0.99)
1	38	0.640 (1.00)	0.68 (1.00)	0.17 (0.70)	1.31 (1.00)
1	69	0.614 (0.96)	0.47 (0.67)	0.04 (0.18)	1.27 (0.95)
1	99	0.169 (0.26)	0.53 (0.75)	0.03 (0.15)	1.13 (0.75)
2	9	0.327 (0.61)	0.50 (1.00)	0.45 (1.00)	0.74 (0.84)
2	29	0.489 (0.93)	0.38 (0.71)	0.04 (0.09)	0.81 (1.00)
2	41	0.295 (0.55)	0.38 (0.71)	0.13 (0.30)	0.64 (0.62)
2	63	0.524 (1.00)	0.44 (0.86)	0.18 (0.40)	0.81 (0.98)
2	78	0.402 (0.76)	0.38 (0.71)	0.12 (0.28)	0.69 (0.72)
3	43	0.353 (0.70)	0.36 (0.67)	0.55 (0.86)	0.62 (0.97)
3	79	0.326 (0.65)	0.21 (0.33)	0.04 (0.06)	0.62 (1.00)
3	100	0.376 (0.75)	0.50 (1.00)	0.64 (1.00)	0.61 (0.96)
3	107	0.501 (1.00)	0.29 (0.50)	0.14 (0.22)	0.45 (0.51)
3	111	0.494 (0.99)	0.21 (0.33)	0.00 (0.00)	0.45 (0.50)
4	12	0.393 (1.00)	0.72 (0.77)	0.07 (0.40)	2.30 (0.88)
4	23	0.391 (0.99)	0.83 (0.92)	0.01 (0.08)	2.46 (1.00)
4	27	0.333 (0.83)	0.89 (1.00)	0.16 (1.00)	2.29 (0.88)
4	31	0.383 (0.97)	0.78 (0.85)	0.08 (0.51)	2.29 (0.88)
4	75	0.332 (0.82)	0.67 (0.69)	0.04 (0.23)	2.20 (0.81)
5	56	0.375 (0.91)	0.50 (1.00)	0.07 (0.30)	1.04 (1.00)
5	68	0.368 (0.89)	0.20 (0.33)	0.02 (0.06)	0.79 (0.59)
5	73	0.411 (1.00)	0.50 (1.00)	0.08 (0.32)	1.05 (1.00)
5	89	0.328 (0.80)	0.40 (0.78)	0.04 (0.15)	1.04 (0.99)
5	97	0.373 (0.91)	0.25 (0.44)	0.03 (0.12)	0.84 (0.68)
6	55	0.149 (0.23)	0.50 (0.83)	0.30 (0.71)	1.09 (0.49)
6	90	0.366 (0.60)	0.58 (1.00)	0.42 (1.00)	1.39 (0.82)
6	92	0.307 (0.50)	0.58 (1.00)	0.17 (0.39)	1.56 (1.00)
6	108	0.602 (1.00)	0.50 (0.83)	0.09 (0.20)	1.51 (0.95)
6	109	0.603 (1.00)	0.58 (1.00)	0.08 (0.18)	1.50 (0.93)

References

Abou Ali M, Dornaika F, Charafeddine J (2025) Agentic AI: a comprehensive survey of architectures, applications, and future directions. Artif Intell Rev 59(1):11. https://doi.org/10.1007/s10462-025-11422-4

Agnihotri A, Bhattacharya S (2024) Chatbots' effectiveness in service recovery. Int J Inf Manag 76:102679. https://doi.org/10.1016/j.ijinfomgt.2023.102679

Ali S, Abuhmed T, El-Sappagh S, et al (2023) Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. Inf Fusion 99:101805. https://doi.org/10.1016/j.inffus.2023.101805

Alkhalil Z, Hewage C, Nawaf L, et al (2021) Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. Front Comput Sci 3. https://doi.org/10.3389/fcomp.2021.563060

Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al (2020) Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. Inf Fusion 58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bawack RE, Wamba SF, Carillo KDA, et al (2022) Artificial intelligence in E-Commerce: a bibliometric study and literature review. Electron Mark 32(1):297–338. https://doi.org/10.1007/s12525-022-00537-z

Behera RK, Bala PK, Rana NP (2023) Creation of sustainable growth with explainable artificial intelligence: An empirical insight from consumer packaged goods retailers. J Clean Prod 399:136605. https://doi.org/10.1016/j.jclepro.2023.136605

Behl A, Sampat B, Pereira V, et al (2023) The role played by responsible artificial intelligence (RAI) in improving supply chain performance in the MSME sector: an empirical inquiry. Ann Oper Res https://doi.org/10.1007/s10479-023-05624-8

Belanche D, Casaló LV, Flavián C (2019) Artificial Intelligence in FinTech: understanding robo-advisors adoption among customers. Ind Manag & Data Syst 119(7):1411–1430. https://doi.org/10.1108/IMDS-08-2018-0368

Belter CW (2015) Bibliometric indicators: opportunities and limits. J Med Libr Assoc 103(4):219–221. https://doi.org/10.3163/1536-5050

van den Besselaar P, Sandström U (2019) Measuring researcher independence using bibliometric data: A proposal for a new performance indicator. PLOS ONE 14(3). https://doi.org/10.1371/journal.pone.0202712

Bhatia A, Chandani A, Chhateja J (2020) Robo advisory and its potential in addressing the behavioral biases of investors — A qualitative study in Indian context. J Behav Exp Finance 25:100281. https://doi.org/10.1016/j.jbef.2020.100281

Bhatia A, Chandani A, Atiq R, et al (2021) Artificial intelligence in financial services: a qualitative research to discover robo-advisory services. Qual Res Financ Mark 13(5):632–654. https://doi.org/10.1108/QRFM-10-2020-0199

Bonacich P (2007) Some unique properties of eigenvector centrality. Soc Netw 29(4):555–564. https://doi.org/10.1016/j.socnet.2007.04.002

Breiman L (2001) Statistical modeling: the two cultures. Stat Sci Rev J Inst Math Stat 16(3):199–231. https://doi.org/10.1214/ss/1009213726

Brenner L, Meyll T (2020) Robo-advisors: A substitute for human financial advice? J Behav Exp Finance 25:100275. https://doi.org/10.1016/j.jbef.2020.100275

Bussmann N, Giudici P, Marinelli D, et al (2021) Explainable Machine Learning in Credit Risk Management. Comput Econ 57(1):203–216. https://doi.org/10.1007/s10614-020-10042-0

Castelfranchi C, Falcone R (2010) Socio-Cognitive Model of Trust: Basic Ingredients. In: Trust Theory: A Socio-Cognitive and Computational Model. John Wiley & Sons, Ltd, p 35–94, https://doi.org/10.1002/9780470519851.ch2

Černevičienė J, Kabašinskas A (2024) Explainable artificial intelligence (XAI) in finance: a systematic literature review. Artif Intell Rev 57(8):216. https://doi.org/10.1007/s10462-024-10854-8

Chen J, Guo F, Ren Z, et al (2024) Effects of anthropomorphic design cues of chatbots on users' perception and visual behaviors. Int J Hum-Comput Interact 40(14):3636–3654. https://doi.org/10.1080/10447318.2023.2193514

Chen Y, Aw ECX, Tan GWH (2025) Financial empowerment through robo-advisors: understanding the keys to trust and loyalty. Ind Manag & Data Syst 125(6):2178–2205. https://doi.org/10.1108/IMDS-05-2024-0461

Chowdhury S, Joel-Edgar S, Dey PK, et al (2023) Embedding transparency in artificial intelligence machine learning models: managerial implications on predicting and explaining employee turnover. Int J Hum Resour Manag 34(14):2732–2764. https://doi.org/10.1080/09585192.2022.2066981

Cramer H, Garcia-Gathright J, Springer A, et al (2018) Assessing and addressing algorithmic bias in practice. Interact 25(6):58–63. https://doi.org/10.1145/3278156

Dafoe A, Hughes E, Bachrach Y, et al (2020) Open problems in cooperative ai. URL https://arxiv.org/abs/2012.08630

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart 13(3):319–340. URL http://www.jstor.org/stable/249008

Dellermann D, Calma A, Lipusch N, et al (2019) The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. In: Bui TX (ed) Proceedings of the 52nd Annual Hawaii International Conference on System Sciences: January 8-11, 2019, Maui, Hawaii. University of Hawaii at Manoa Hamilton Library ScholarSpace, Honolulu, HI, pp 274–283, URL http://hdl.handle.net/10125/59468

Dixon MF, Halperin I, Bilokon P (2020) Machine Learning in Finance: From Theory to Practice. Springer, Cham, https://doi.org/10.1007/978-3-030-41068-1

Donthu N, Kumar S, Mukherjee D, et al (2021) How to conduct a bibliometric analysis: An overview and guidelines. J Bus Res 133:285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Dou WW, Goldstein I, Ji Y (2025) AI-Powered Trading, Algorithmic Collusion, and Price Efficiency. Working Paper 34054, National Bureau of Economic Research, https://doi.org/10.3386/w34054

Duong CD, Nguyen TH, Ngo TVN, et al (2024) Blockchain technology and consumers' organic food consumption: a moderated mediation model of blockchain-based trust and perceived blockchain-related information transparency. J Asia Bus Stud 19(1):54–78. https://doi.org/10.1108/JABS-07-2024-0387

Floridi L (2019) Establishing the rules for building trustworthy AI. Nat Mach Intell 1(6):261–262. https://doi.org/10.1038/s42256-019-0055-y

Floridi L (2025) AI as agency without intelligence: On artificial intelligence as a new form of artificial agency and the multiple realisability of agency thesis. Philos & Technol 38(1):30. https://doi.org/10.1007/s13347-025-00858-9

Fronzetti Colladon A, Naldi M (2020) Distinctiveness centrality in social networks. PLOS ONE 15(5):1–21. https://doi.org/10.1371/journal.pone.0233276

Garcia ACB, Garcia MGP, Rigobon R (2024) Algorithmic discrimination in the credit domain: what do we know about it? AI & Soc 39(4):2059–2098. https://doi.org/10.1007/s00146-023-01676-3

Gebru T, Morgenstern J, Vecchione B, et al (2021) Datasheets for datasets. Commun ACM 64(12):86–92. https://doi.org/10.1145/3458723

Gefen D, Karahanna E, Straub DW (2003) Trust and Tam in Online Shopping: An Integrated Model. MIS Q 27. https://doi.org/10.2307/30036519

Giudici P (2018) Fintech Risk Management: A Research Challenge for Artificial Intelligence in Finance. Front Artif Intell 1. https://doi.org/10.3389/frai.2018.00001

Graham G, Nisar TM, Prabhakar G, et al (2025) Chatbots in customer service within banking and finance: Do chatbots herald the start of an AI revolution in the corporate world? Comput Hum Behav 165:108570. https://doi.org/10.1016/j.chb.2025.108570

Guidotti R, Monreale A, Ruggieri S, et al (2018) A Survey of Methods for Explaining Black Box Models. ACM Comput Surv 51(5). https://doi.org/10.1145/3236009

Gunning D (2019) DARPA's explainable artificial intelligence (XAI) program. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. Association for Computing Machinery, New York, NY, USA, IUI '19, p ii, https://doi.org/10.1145/3301275.3308446

He X, Liu Y (2024) Knowledge evolutionary process of artificial intelligence in e-commerce: Main path analysis and science mapping analysis. Expert Syst Appl 238:121801. https://doi.org/10.1016/j.eswa.2023.121801

Heng YS, Subramanian P (2023) A Systematic Review of Machine Learning and Explainable Artificial Intelligence (XAI) in Credit Risk Modelling. In: Arai K (ed) Proceedings of the Future Technologies Conference (FTC) 2022, Volume 1. Springer International Publishing, Cham, pp 596–614, https://doi.org/10.1007/978-3-031-18461-1_39

Hoffman RR (2024) Trusting as an Emergent: Implications for Design. Ergon Des 0(0):10648046241295270. https://doi.org/10.1177/10648046241295270

Holland JH (1992) Complex Adaptive Systems. Daedalus 121(1):17–30. URL http://www.jstor.org/stable/20025416

Jellason NP, Ambituuni A, Adu DA, et al (2024) The potential for blockchain to improve small-scale agri-food business' supply chain resilience: a systematic review. Br Food J 126(5):2061–2083. https://doi.org/10.1108/BFJ-07-2023-0591

Jung D, Dorner V, Weinhardt C, et al (2018) Designing a robo-advisor for risk-averse, low-budget consumers. Electron Mark 28(3):367–380. https://doi.org/10.1007/s12525-017-0279-9

Khan FS, Mazhar SS, Mazhar K, et al (2025) Model-agnostic explainable artificial intelligence methods in finance: a systematic review, recent developments, limitations, challenges and future directions. Artif Intell Rev 58(8):232. https://doi.org/10.1007/s10462-025-11215-9

Khushk A, Zhiying L, Yi X, et al (2024) Navigating human-AI dynamics: implications for organizational performance (SLR). Int J Organ Anal 33(5):1293–1311. https://doi.org/10.1108/IJOA-04-2024-4456

Kim T, Song H (2021) How should intelligent agents apologize to restore trust? Interaction effects between anthropomorphism and apology attribution on trust repair. Telemat Inform 61:101595. https://doi.org/10.1016/j.tele.2021.101595

Lee MK, Kusbit D, Kahng A, et al (2019) WeBuildAI: Participatory framework for algorithmic governance. Proc ACM Hum-Comput Interact 3(CSCW):181. https://doi.org/10.1145/3359283

Lehner OM, Ittonen K, Silvola H, et al (2022) Artificial intelligence based decision-making in accounting and auditing: ethical challenges and normative thinking. Account Audit Account J 35(9):109–135. https://doi.org/10.1108/AAAJ-09-2020-4934

Leitner G, Singh J, van der Kraaij A, et al (2024) The rise of artificial intelligence: benefits and risks for financial stability. Finance Stab Rev 1:2. URL https://www.ecb.europa.eu//press/financial-stability-publications/fsr/special/html/ecb.fsrart202405_02~58c3ce5246.en.html

Li J, Wu L, Qi J, et al (2023a) Determinants affecting consumer trust in communication with ai chatbots: The moderating effect of privacy concerns. J Organ End User Comput 35(1):1–24. https://doi.org/10.4018/JOEUC.328089

Li Y, Wang S, Ding H, et al (2023b) Large language models in finance: A survey. In: Proceedings of the Fourth ACM International Conference on AI in Finance. Association for Computing Machinery, New York, NY, USA, ICAIF '23, p 374–382, https://doi.org/10.1145/3604237.3626869

Lipton ZC (2018) The mythos of model interpretability. Commun ACM 61(10):36–43. https://doi.org/10.1145/3233231

Loaiza I, Rigobon R (2025) The Limits of AI in Financial Services. https://doi.org/10.2139/ssrn.5196350

Machucho R, Ortiz D (2025) The Impacts of Artificial Intelligence on Business Innovation: A Comprehensive Review of Applications, Organizational Challenges, and Ethical Considerations. Syst 13(4):264. https://doi.org/10.3390/systems13040264

Majeed Parry G, Revolidis I, Ellul J, et al (2024) Bottling Up Trust: A Review of Blockchain Adoption in Wine Supply Chain Traceability. IEEE Access 12:178320–178344. https://doi.org/10.1109/ACCESS.2024.3505428

Makarius EE, Mukherjee D, Fox JD, et al (2020) Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. J Bus Res 120:262–273. https://doi.org/10.1016/j.jbusres.2020.07.045

Maple C, Szpruch L, Epiphaniou G, et al (2023) The AI Revolution: Opportunities and Challenges for the Finance Sector. https://doi.org/10.48550/arXiv.2308.16538

Marzi G, Balzano M, Caputo A, et al (2025) Guidelines for Bibliometric-Systematic Literature Reviews: 10 steps to combine analysis, synthesis and theory development. Int J Manag Rev 27(1):81–103. https://doi.org/10.1111/ijmr.12381

Miller T (2019) Explanation in artificial intelligence: Insights from the social sciences. Artif Intell 267:1–38. https://doi.org/10.1016/j.artint.2018.07.007

Nass C, Steuer J, Tauber ER (1994) Computers are social actors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, CHI '94, pp 72–78, https://doi.org/10.1145/191666.191703

NetworkX (2025) eigenvector_centrality — NetworkX documentation. URL https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.eigenvector_centrality.html

Newman MEJ (2006) Modularity and community structure in networks. Proc Natl Acad Sci 103(23):8577–8582. https://doi.org/10.1073/pnas.0601602103

Nourallah M (2023) One size does not fit all: Young retail investors' initial trust in financial robo-advisors. J Bus Res 156:113470. https://doi.org/10.1016/j.jbusres.2022.113470

Owens E, Sheehan B, Mullins M, et al (2022) Explainable artificial intelligence (xai) in insurance. Risks 10(12). https://doi.org/10.3390/risks10120230

Page MJ, McKenzie JE, Bossuyt PM, et al (2021a) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372. https://doi.org/10.1136/bmj.n71

Page MJ, Moher D, Bossuyt PM, et al (2021b) Prisma 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 372. https://doi.org/10.1136/bmj.n160

Pelau C, Dabija DC, Ene I (2021) What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. Comput Hum Behav 122:106855. https://doi.org/https://doi.org/10.1016/j.chb.2021.106855

Perianes-Rodriguez A, Waltman L, Eck NJv (2016) Constructing bibliometric networks: A comparison between full and fractional counting. J Informetr 10(4):1178–1195. https://doi.org/10.1016/j.joi.2016.10.006

Reuel A, Bucknall B, Casper S, et al (2025) Open problems in technical AI governance. arXiv preprint arXiv:2407.14981. URL https://arxiv.org/abs/2407.14981

Riedel A, Mulcahy R, Northey G (2022) Feeling the love? How consumer's political ideology shapes responses to AI financial service delivery. Int J Bank Mark 40(6):1102–1132. https://doi.org/10.1108/IJBM-09-2021-0438

Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1(5):206–215. https://doi.org/10.1038/s42256-019-0048-x

Saarela M, Podgorelec V (2024) Recent Applications of Explainable AI (XAI): A Systematic Literature Review. Appl Sci 14(19). https://doi.org/10.3390/app14198884

Sabharwal R, Miah SJ, Wamba SF, et al (2024) Extending application of explainable artificial intelligence for managers in financial organizations. Ann Oper Res https://doi.org/10.1007/s10479-024-05825-9

SankeyMATIC (2025) Home. https://sankeymatic.com/, (Accessed Oct. 23, 2025)

Scholes MS (2025) Artificial intelligence and uncertainty. Risk Sci 1:100004. https://doi.org/10.1016/j.risk.2024.100004

Schreibelmayr S, Moradbakhti L, Mara M (2023) First impressions of a financial ai assistant: differences between high trust and low trust users. Front Artif Intell Volume 6 - 2023. https://doi.org/10.3389/frai.2023.1241290

Sharma A, Lin IW, Miner AS, et al (2023) Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nat Mach Intell 5(1):46–57. https://doi.org/10.1038/s42256-022-00593-2

Sowa K, Przegalinska A, Ciechanowski L (2021) Cobots in knowledge work: Human – AI collaboration in managerial professions. J Bus Res 125:135–142. https://doi.org/10.1016/j.jbusres.2020.11.038

Sri Vigna Hema V, Manickavasagan A (2024) Blockchain implementation for food safety in supply chain: A review. Compr Rev Food Sci Food Saf 23(5):e70002. https://doi.org/10.1111/1541-4337.70002

Srinivas Aditya U, Singh R, Singh PK, et al (2021) A survey on blockchain in robotics: Issues, opportunities, challenges and future directions. J Netw Comput Appl 196:103245. https://doi.org/10.1016/j.jnca.2021.103245

Tol P (2021) Colour Schemes. Tech. Rep. SRON/EPS/TN/09-002, SRON Netherlands Institute for Space Research, URL https://sronpersonalpages.nl/~{}pault/

Traag VA, Waltman L, van Eck NJ (2019) From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 9(1):5233. https://doi.org/10.1038/s41598-019-41695-z

Wang D, Weisz JD, Muller M, et al (2019) Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI. Proc ACM Hum-Comput Interact 3(CSCW). https://doi.org/10.1145/3359313

Wang D, Churchill E, Maes P, et al (2020) From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, CHI EA '20, pp 1–6, https://doi.org/10.1145/3334480.3381069

Wang Y, Mohsin M, Jamil K (2024) Customer trust and willingness to use shopping assistant humanoid chatbot. Serv Ind J 0(0):1–25. https://doi.org/10.1080/02642069.2024.2432935

Weber P, Carl KV, Hinz O (2024) Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature. Manag Rev Q 74(2):867–907. https://doi.org/10.1007/s11301-023-00320-0

Yan H, Wei Y, Xiong H (2025) How Do Initial and Interactive Social Cues Increase Customers' Continuance Usage Intention of Chatbots? Int J Hum-Comput Interact 41(8):4700–4717. https://doi.org/10.1080/10447318.2024.2352928

Yavaprabhas K, Pournader M, Seuring S (2023) Blockchain as the "trust-building machine" for supply chain management. Ann Oper Res 327(1):49–88. https://doi.org/10.1007/s10479-022-04868-0

Zhang J, Luo Y (2017) Degree centrality, betweenness centrality, and closeness centrality in social network. In: Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017), pp 300–303, https://doi.org/10.2991/msam-17.2017.68

Zupic I, Čater T (2015) Bibliometric methods in management and organization. Organ Res Methods 18(3):429–472. https://doi.org/10.1177/1094428114562629