Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Global collaboration outweighs algorithmic complexity in AI-finance research: a negative binomial analysis of citation impact (2015-2025)

Luis-Felipe Arizmendi

lfarizmendi@comillas.edu

Universidad Pontificia Comillas

Research Article

Keywords: Citation analysis, Citation determinants, International collaboration, Negative binomial regression, Science mapping, Research evaluation, Artificial intelligence, FinTech

Posted Date: August 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10391182/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Identifying the determinants of citation impact is a core concern of *scientometrics*, yet these determinants are seldom studied in fast-moving, interdisciplinary technological fields. This paper examines one such field-the application of artificial intelligence (AI) in finance-by combining bibliometric science mapping with a Negative Binomial regression of citation counts for an exhaustive sample of 2,694 Scopus-indexed research articles published between 2015 and 2025. International collaboration is the strongest predictor of impact, raising expected citations by 91% (IRR = 1.91), followed by explicit funding (+ 40%) and open access (+ 17%). In a “Complexity Paradox”, the two most methodologically specialized topics are penalized: articles centered on Deep Learning—a keyword-based thematic proxy for algorithmic complexity—receive 16.8% fewer citations (IRR = 0.83), and the inherently interpretable Risk Management cluster receives an even larger penalty (-28.0%, IRR = 0.72). Because the penalty falls hardest on interpretable, regulated work, we reject an “explainability premium” interpretation and argue instead that citation impact in this field is governed by audience breadth, topic novelty, and cross-border generalizability. Interaction and mature-cohort robustness checks support this reading. The study extends citation-determinant modeling to an emerging interdisciplinary field, and cautions against using raw citation counts to compare heterogeneous research specialisms.

1. Introduction

Identifying the factors that drive the citation impact of research is a long-standing question in scientometrics, and a substantial literature has modeled citation counts as a function of collaboration, funding, open access, and topical attributes (Aksnes, 2003; Glanzel & Thijs, 2004; Tahamtan, Safipour Afshar & Ahamdzadeh, 2016). These determinants, however, have rarely been examined in fast-moving, interdisciplinary technological fields, where citation dynamics may diverge from those of established disciplines. This study addresses that gap for one of the most prominent such fields: the application of artificial intelligence in finance.

The intersection of Artificial Intelligence (AI) and Finance is no longer a futuristic promise but a central engine of transformation in the global financial sector. From high-frequency algorithmic trading to risk management using Big Data, the adoption of Machine Learning (ML) and Deep Learning (DL) has redefined both industrial practice and the academic agenda (Dwivedi et al., 2021; Fischer & Krauss, 2018). Within leading specialized outlets such as Digital Finance, scholarship has shown how ML models are reshaping financial forecasting and planning (Wasserbacher & Spindler, 2022) and how AI-driven sentiment analysis uncovers the network structure of cryptocurrency markets (Aste, 2019).

Over the last decade (2015–2025), scientific production on this phenomenon has grown exponentially, reflecting an intellectual arms race to build ever more accurate predictive models. This explosive growth has, however, produced a fragmented body of literature. While numerous Systematic Literature Reviews describe what is being published (Goodell et al., 2021), there is a notable scarcity of studies that quantitatively analyze the determinants of scientific impact in this domain. Recent landmark studies in

scientometrics have begun to map the historical diffusion of AI within broad economic subfields (Bickley, Chan & Torgler, 2022); yet the specific drivers of citation success for “black-box” algorithms versus transparent models in pure finance remain underexplored. What makes financial-AI research influential? Is it the technical complexity of the algorithms, or does the community value applicability and transparency?

This article addresses this gap with a hybrid approach. Unlike traditional reviews, it applies bibliometric econometrics to a sample of 2,694 Scopus-indexed articles. Our results challenge conventional intuition and reveal what we term the “Complexity Paradox”: a tension between algorithmic sophistication and perceived utility. Importantly, and in direct response to peer review, we show that this paradox cannot be reduced to a simple penalty on “opaque” methods, and we develop an interpretation centered on audience breadth and cross-border generalizability that is consistent with the full pattern of estimated effects. The remainder of the article is organized as follows. Section 2 details the PRISMA-based extraction protocol and the econometric methodology. Section 3 reports descriptive statistics, the intellectual structure, and the regression results. Section 4 discusses the implications, situates them within the cross-disciplinary scientometric literature, and qualifies the findings. Section 5 concludes, stating limitations and avenues for future research.

2. Methodology

2.1 Data Source and Search Strategy

To ensure representativeness, multidisciplinarity, and high peer-reviewed quality, data were extracted exclusively from Scopus, widely recognized as an optimal repository for business, management, and financial literature, offering broader coverage of specialized FinTech journals than Web of Science (WoS) while maintaining stricter quality control than Google Scholar. Extraction followed the PRISMA guidelines (Moher et al., 2009). A structured Boolean query was designed to capture the intersection of advanced computational algorithms and applied financial markets. The exact string executed was:

TITLE-ABS-KEY (("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Deep Learning" OR "Neural Networks" OR "Natural Language Processing" OR "Reinforcement Learning") AND ("Finance" OR "Banking" OR "Stock Market" OR "Algorithmic Trading" OR "FinTech" OR "Cryptocurrency" OR "Credit Risk" OR "Portfolio Management")) AND PUBYEAR > 2014 AND PUBYEAR < 2026

The initial query returned over 30,000 documents, indicating substantial noise (predominantly pure computer science and engineering work with little financial applicability). To ensure relevance to financial economics, strict inclusion/exclusion criteria were applied sequentially. First, the subject area was restricted to “Economics, Econometrics and Finance” and “Business, Management and Accounting,” purposefully excluding pure Computer Science. Second, document types were limited to peer-reviewed research articles, excluding conference proceedings, book chapters, and editorials. Finally, the language of the original document was restricted to English. The subject-area, document-type, and metadata

filters returned 2,739 records, of which 45 were in languages other than English and were excluded, yielding a final analytic sample of N = 2,694 documents. The complete identification, screening, and inclusion process is summarised in the PRISMA-style flow diagram (Fig. 1). All citation counts reflect a single Scopus census date of December 2025; because the field is young and citations accumulate over time, we explicitly report this census date to ensure the analysis is exactly reproducible. The full extraction (raw Scopus export and the analysis-ready dataset) underlies every figure reported below.

2.2 Variable Definition and Operationalization

Fields exported from Scopus (citation counts, affiliations, funding text, keywords) were codified into statistical variables. **Dependent variable (Y)**: Citation Count (Ci), total citations received up to the extraction date, a proxy for scientific impact. **Control variables**: Age of Article (2025 minus publication year), controlling for the natural accumulation of citations; and Open Access (dummy), controlling for the documented visibility advantage of freely available work. **Structural and institutional factors**: International Collaboration (dummy = 1 if author affiliations span more than one country) and Has Funding (dummy = 1 if the article explicitly acknowledges grant support). **Thematic factors**: four dummies (Deep Learning, Risk Management, Crypto/Blockchain, Sentiment Analysis) representing the dominant research clusters.

Construction of the thematic dummies. Each thematic indicator was built by deterministic Boolean tagging of the Author Keywords field: an article is coded 1 for a given topic if its keyword field contains any term in the corresponding lexicon (for example, the Deep Learning lexicon includes “deep learning,” “neural network,” “LSTM,” “CNN,” “convolutional”). The dummies are therefore not mutually exclusive, and each coefficient in Eq. (1) is interpreted relative to articles that do not carry the corresponding keyword. We emphasize that these indicators are keyword-based proxies for topical focus; they are not direct measures of model interpretability or methodological complexity-a distinction that becomes important in the Discussion. In particular, the Deep Learning indicator functions as a thematic proxy for algorithmic complexity: it records whether an article is topically centered on deep neural architectures, rather than measuring model complexity or opacity at the article level. Accordingly, the “Complexity Paradox” reported below is a topic-level association, and references to “algorithmic complexity” throughout the paper should be understood in this proxied, keyword-based sense.

2.3 Econometric Modeling

Beyond descriptive science mapping, we estimate a cross-sectional regression of the determinants of scientific impact. Because the dependent variable (citation count) is a non-negative integer, Ordinary Least Squares is inappropriate. Moreover, citation distributions exhibit a heavy right tail (a few papers receive most citations while many receive none). Because the empirical variance far exceeds the mean (variance > > mean), the *equidispersion* assumption of the Poisson model is severely violated. We therefore employ a Negative Binomial regression, which introduces an unobserved heterogeneity parameter (alpha) to account for overdispersion. The negative binomial, as the canonical remedy for

overdispersed counts relative to Poisson, follows standard treatments of count-data regression (Cameron & Trivedi, 1998, 2013), and the model is estimated by maximum likelihood. The full model is:

ln(μi) = β0 + β1·Age + β2·Intl_Collab + β3·Funding + β4·OpenAccess

+ β5·DeepLearning + β6·RiskMgmt + β7·Crypto + β8·Sentiment + ε (1)

where μi is the expected number of citations for article i. The exponentiated coefficients (e^β) give the Incidence Rate Ratio (IRR), allowing each factor to be read as a percentage change in expected citations, holding all else constant.

Specification choice and interaction terms. We adopt a parsimonious additive (main effects) specification in Eq. (1) so that each IRR retains a direct interpretation. Because funding and open access could plausibly operate differently across thematic clusters, we do not simply assume additivity: we estimate augmented specifications with Funding-by-Topic and OpenAccess-by-Topic interactions and report them as a robustness check in Section 3.5. We retain the additive model as the baseline because, as shown there, the main effects are unchanged by the inclusion of these terms.

3. Results

3.1 Distributional Analysis and Temporal Trends

Before estimation, the distributional properties of the data were examined. Figure 2 compares the observed citation counts with a theoretical Poisson distribution. The extreme deviation of the observed data from the Poisson curve, particularly the excess of zeros and the heavy tail, visually confirms overdispersion (alpha > 0) and justifies the Negative Binomial model.

Regarding temporal evolution (Fig. 3), Deep Learning exhibits a “hype-cycle” pattern: it dominated the early literature (around 40% of tagged articles in 2015–2016) and has since eased to roughly a quarter of annual output, even as international collaboration has risen steadily over the decade.

3.2 Descriptive Statistics

Table 1 summarises central tendency and dispersion for the sample (N = 2,694). The dependent variable, Citation Count, is highly skewed, as is characteristic of scientometric data. The average article receives 22.13 citations, but the median article receives only 5; the standard deviation is very high (81.39), and the maximum reaches 2,987 citations. This immense variance confirms a “winner-takes-all,” Pareto-like distribution in which a small fraction of “blockbuster” papers accumulate the bulk of recognition, while a large tail remains barely cited, reinforcing the choice of a Negative Binomial model. The gap between the mean (22.13) and the median (5) is itself a compact measure of this skew.

The structural and institutional variables describe a young, globalized, and well-supported domain. The average Age is low (2.12 years), underscoring that AI-and-Finance is a recently exploded frontier rather

than a mature discipline. Some 36% of articles acknowledge explicit funding, 46% are open access, and 31% involve international collaboration, suggesting heavy reliance on cross-border knowledge spillovers. Among thematic dummies, Deep Learning is the dominant specific topic (27%), slightly ahead of Risk Management (26%), while Crypto/Blockchain (10%) and Sentiment Analysis (8%) are consolidated but more niche sub-fields.

Table 1
Descriptive statistics of the variables (N = 2,694)

Variable	Obs	Mean	Median	Std. Dev.	Min	Max
Citations (Count)	2,694	22.13	5	81.39	0	2,987
Age (years since pub.)	2,694	2.12	1	2.22	0	10
Intl. Collaboration (dummy)	2,694	0.31	0	0.46	0	1
Has Funding (dummy)	2,694	0.36	0	0.48	0	1
Open Access (dummy)	2,694	0.46	0	0.50	0	1
Topic: Deep Learning	2,694	0.27	0	0.44	0	1
Topic: Risk Management	2,694	0.26	0	0.44	0	1
Topic: Crypto/Blockchain	2,694	0.10	0	0.30	0	1
Topic: Sentiment Analysis	2,694	0.08	0	0.27	0	1

Note. *The median citation count (5) is far below the mean (22.13), indicating a strong right-skew in the data. For the binary indicators, the median is 0, as all sample proportions are below 0.5.*

3.3 The Intellectual Structure (Network Analysis)

Figure 4 presents the keyword co-occurrence network, mapping the intellectual structure of the AI-finance domain. The map was produced from a co-occurrence matrix of the most frequent author keywords using a modularity-based community-detection routine (van Eck & Waltman, 2010); node size is proportional to keyword occurrence, and edge width to co-occurrence frequency. We did not apply a partitional method such as k-means. The map organizes the literature into four thematic communities (Table 2), color-coded in Fig. 4, that correspond directly to the four thematic indicators used in the regression: an algorithmic/predictive core, a FinTech ecosystem, a risk-management community, and a market-sentiment community. The network partition (which groups co-occurring keywords) and the regression indicators (which are deterministic Boolean tags of individual articles, Section 2.2) are produced by different procedures but tell a consistent story.

Table 2
Thematic clusters and representative keywords (network statistics)

Cluster	Label	Dominant focus	Top representative keywords
1	Algorithmic Core	Predictive modelling	Deep Learning, Neural Networks, Forecasting, LSTM, CNN, Optimization
2	FinTech Ecosystem	Decentralized finance	Blockchain, Cryptocurrency, Bitcoin, Smart Contracts, Ethereum, Digital Assets
3	Risk Management	Banking applications	Credit Risk, Credit Scoring, Bankruptcy Prediction, Decision Trees, Default
4	Market Sentiment	Behavioural finance	Sentiment Analysis, Social Media, Twitter, Text Mining, Investor Sentiment

Note. *Four thematic communities were detected from the keyword co-occurrence network; they align one-to-one with the four thematic indicators used in the regression.*

3.4 Determinants of Scientific Impact (Regression Analysis)

Table 3 reports the Negative Binomial estimates. The model fits well: the dispersion parameter alpha (1.865) differs significantly from zero ($p < 0.001$), rejecting Poisson equidispersion; the Wald chi-square statistic (1236.1, $p < 0.001$) confirms joint significance; and the information criteria (AIC = 18,874.1; HQC = 18,895.4) indicate an appropriate balance between parsimony and fit. Following the reviewer's suggestion, we also report 95% confidence intervals for the Incidence Rate Ratios.

Table 3
Negative Binomial regression results (dependent variable: citations)

Predictor	Coef. (β)	Std. Err.	Sig.	IRR (e^β)	95% CI (IRR)
Control variables					
Intercept (Constant)	1.157	0.064	***	3.181	[2.81, 3.60]
Age of Article	0.521	0.016	***	1.683	[1.63, 1.74]
Structural factors					
Intl. Collaboration	0.647	0.059	***	1.909	[1.70, 2.14]
Has Funding	0.339	0.057	***	1.403	[1.26, 1.57]
Open Access	0.158	0.055	**	1.171	[1.05, 1.30]
Thematic factors					
Topic: Deep Learning	-0.184	0.062	**	0.832	[0.74, 0.94]
Topic: Risk Management	-0.328	0.063	***	0.720	[0.64, 0.82]
Topic: Crypto/Blockchain	0.077	0.093	n.s.	1.080	[0.90, 1.30]
Topic: Sentiment Analysis	0.093	0.101	n.s.	1.098	[0.90, 1.34]

Note. *N = 2,694; Log-Likelihood = -9427.0; Alpha (α) = 1.865 (p < 0.001); Wald chi-square(8) = 1236.1 (p < 0.001); AIC = 18,874.1; Hannan-Quinn = 18,895.4. Significance: ** p < 0.01, *** p < 0.001, n.s. = not significant. IRR = Incidence Rate Ratio; IRR > 1 indicates a positive effect. 95% CIs are computed as exp(β ± 1.96·SE). For Crypto and Sentiment the interval contains 1, consistent with non-significance; for Deep Learning and Risk Management it lies entirely below 1.*

The estimates confirm the “Complexity Paradox”: articles focused on Deep Learning reduce the expected citation rate by 16.8% (IRR = 0.832), holding all else constant. The *largest* thematic penalty, however, falls on Risk Management (-28.0%, IRR = 0.720, p < 0.001)-a result we interpret carefully in Section 4.

3.5 Robustness Checks

We probed the stability of the estimates in two ways.

Interaction effects. We estimated the model, adding Funding-by-Topic and OpenAccess-by-Topic interactions for the two penalized clusters. The main effects in Table 3 are essentially unchanged, and most interactions are insignificant; the one exception is a positive Funding-by-Risk-Management interaction (β = 0.390, p = 0.002), indicating that the Risk Management penalty is substantially attenuated for funded work. A likelihood-ratio test of the four interaction terms is jointly significant (LR = 11.41, 4 df, p = 0.022) but yields only a marginal improvement in fit (AIC 18,870.7 vs 18,874.1). We therefore retain the additive model as the baseline and treat the funding-by-risk moderation as a substantive finding that we revisit in Section 4.1.

Sensitivity to the recent cohort. Because the field is young and roughly half of the sample was published in 2024–2025, we re-estimated the model on the sub-sample of articles with at least two full years of citation exposure (Age > = 2; N = 1,338). As Table 4 shows, the international collaboration premium and the Risk Management penalty remain strong and significant, whereas the Deep Learning penalty attenuates and becomes only marginally significant (IRR = 0.88, p = 0.089). This pattern suggests that part of the Deep Learning penalty reflects the recency of the deep-learning surge-very recent papers have had little time to accumulate citations-while the Risk Management penalty is structural rather than a cohort artifact.

Table 4
Robustness: full sample versus mature-cohort sub-sample (IRR)

Predictor	Full sample (N = 2,694)	Age > = 2 (N = 1,338)
Intl. Collaboration	1.909 ***	1.704 ***
Has Funding	1.403 ***	1.379 ***
Topic: Deep Learning	0.832 ***	0.878 (p = 0.089)
Topic: Risk Management	0.720 ***	0.767 ***

Note. *Incidence Rate Ratios from the Negative Binomial model. *** p < 0.001. The collaboration premium and the Risk Management penalty are robust to excluding the youngest cohort; the Deep Learning penalty weakens to marginal significance, consistent with a hype-cycle interpretation.*

4. Discussion

4.1 The Complexity Paradox and the Limits of an "Explainability" Interpretation

A central and statistically robust result is that the two most methodologically specialized topics in the field both attract significantly fewer citations than the rest of the literature. Articles centered on Deep Learning show a 16.8% reduction in expected citations (IRR = 0.832; 95% CI [0.74, 0.94]; p < 0.01) and, contrary to what a simple opacity-based account would predict, articles in the Risk Management cluster show an even larger and more significant penalty (IRR = 0.720; 95% CI [0.64, 0.82]; p < 0.001). We label the headline observation-that greater algorithmic sophistication does not translate into greater scholarly impact-the "Complexity Paradox."

Why the simple "explainability premium" fails. A tempting interpretation is that the community penalizes "black-box" models because financial institutions operate under stringent transparency requirements (GDPR, the US Equal Credit Opportunity Act, the Basel guidelines on risk-data aggregation), thereby creating an "explainability premium" for interpretable work. We examined this conjecture closely and now regard it, in its strong form, as inconsistent with our own evidence. If an explainability premium were the dominant mechanism, the Risk Management cluster, dominated by inherently interpretable techniques

(credit scoring, decision trees, bankruptcy prediction) and the sub-field most tightly bound to regulatory frameworks, should enjoy a citation *advantage*, not the largest penalty in the model. Because the data point the opposite way, we explicitly do not claim that interpretability alone drives the topic penalties. This directly addresses the central concern raised in peer review.

A more consistent account: audience breadth, maturity, and generalizability. We propose a single mechanism that fits the full pattern of coefficients. The penalized topics-Deep Learning and credit-risk-are both mature, methodologically saturated, and frequently single-market or narrowly applied: deep-learning price-prediction and credit-scoring papers typically speak to specialized, bounded audiences and are often validated on one national exchange or one credit registry. By contrast, the reference literature and the (positively signed but non-significant) emerging topics, crypto/blockchain and sentiment/NLP, are newer, more cross-disciplinary, and connect to broader readerships spanning economics, computer science, behavioral finance, and media studies. On this reading, what is rewarded is not interpretability per se but **reach and generalizability**: the breadth of the audience a paper can address and the transferability of its findings across settings. Critically, this same mechanism explains the dominant positive predictor in the model, international collaboration (+ 91%)-which mechanically broadens both the audience and the empirical generality of a study. The Complexity Paradox is thus better understood as a *generalizability premium* than as an explainability premium.

Two robustness results (Section 3.5) sharpen this reading. First, the two penalties behave differently once the youngest cohort is removed: the Risk Management penalty is stable (it is structural), whereas the Deep Learning penalty attenuates to marginal significance, indicating that part of it reflects the recency of the deep-learning surge rather than a durable discount on complexity. Second, the Risk Management penalty is significantly *attenuated for funded work* (a positive Funding-by-Risk interaction, $p = 0.002$). Both results fit the generalizability account: funded credit-risk projects tend to be larger, better resourced and more broadly validated, and therefore travel further, while the bulk of unfunded, single-market applied risk work remains narrowly cited. Neither result is consistent with an interpretability-based story, under which regulated, interpretable risk research should be rewarded regardless of funding.

Qualifiers and disclaimers. Three caveats discipline these interpretations. First, our thematic variables are keyword-based proxies for topical focus, not direct measures of algorithmic complexity, model interpretability, audience size, or geographic scope; the Deep Learning indicator proxies "algorithmic complexity" at the topic level only, and cannot, on its own, separate an "opacity" effect from confounds such as topic age and audience breadth. Second, because the dummies are not mutually exclusive, each penalty is measured relative to articles lacking that keyword, and the reference category is heterogeneous. Third, the design is cross-sectional, and the estimates are associational; no causal claim about "why" a topic is cited less can be sustained without article-level controls. Accordingly, we present the generalizability account as the most parsimonious reading consistent with the evidence, and we treat the explainability conjecture as a bounded hypothesis about the opacity dimension of Deep Learning that future, article-level work should test directly (see Section 5.4).

4.2 Cross-Disciplinary Context

The reviewer rightly asked how these patterns compare with regression-based studies of citation impact in other domains. Two comparisons are instructive.

Collaboration. Our international-collaboration premium is qualitatively consistent with a large cross-disciplinary scientometric literature documenting that internationally co-authored papers are cited more than purely domestic ones across physics, medicine, engineering, and the social sciences (Glanzel & Thijs, 2004; Aksnes, 2003; Tahamtan, Safipour Afshar & Ahamdzadeh, 2016). That literature attributes the effect largely to an "audience effect"-more authors embedded in more national networks broaden dissemination, while noting that the magnitude varies substantially by field. Our + 91% sits at the upper end of reported effects, which is plausible given the youth and intrinsically global character of FinTech research; it also reinforces the audience-breadth mechanism advanced in Section 4.1.

Interpretability across regulated domains. The tension between predictive accuracy and interpretability is not unique to finance. In medicine, another high-stakes, heavily regulated field, Rudin (2019) argues forcefully against deploying black-box models for high-stakes decisions when inherently interpretable models exist, and Caruana et al. (2015) document interpretable models being preferred for clinical deployment precisely because clinicians and regulators must audit them. By contrast, core computer science venues frequently reward marginal gains in accuracy on benchmark tasks, even at the cost of opacity. This cross-domain contrast originally motivated our explainability conjecture in finance. The Risk Management result, however, is a useful corrective: it shows that *topic-level* citation patterns in finance are shaped by more than interpretability, and that a genuine cross-domain test of an explainability premium would require coding model interpretability at the article level (e.g., interpretable vs. black-box methods within each topic) rather than relying on topic proxies. We regard this as one of the most promising extensions of the present work.

4.3 The Globalization of Financial Intelligence

Beyond thematic preferences, the model identifies structural network characteristics as the primary drivers of impact. International collaboration is the strongest predictor, nearly doubling expected citations (+ 91%, IRR = 1.909, $p < 0.001$). This premium can be understood through the dual lens of knowledge spillover and epistemic diversity: multi-country teams typically combine diverse proprietary datasets, thereby overcoming the overfitting of algorithms to a single national market and prompting the design of more generalizable, robust models that perform across macroeconomic environments. From a sociological perspective, internationally co-authored papers also benefit from a structurally amplified dissemination network, with each co-author acting as a bridge to a distinct geographic and institutional citation cluster. This network effect aligns with the positive effect of explicit funding (+ 40%, IRR = 1.403), since international consortia are often the recipients of large transnational grants that provide both computational resources and the institutional prestige that attracts early citations.

4.4 Addressing Potential Endogeneity and Network Biases

Interpreting structural variables in bibliometric regressions requires caution because of potential endogeneity, particularly a "self-fulfilling prophecy" driven by self-citations and team size. A common

critique is that the collaboration premium (IRR = 1.909) might be an artifact of larger teams repeatedly citing their own prior work. To address this, we computed the Spearman rank-order correlation between team size (number of authors) and the international-collaboration dummy, obtaining a moderate coefficient (rho = 0.35). This confirms that, while international papers tend to have somewhat larger teams, internationalization is not collinear with group size; the 91% premium cannot be reduced to a mechanical self-citation effect. Drawing on social-network theory, publishing across borders lets researchers tap non-overlapping scientific communities, creating “structural holes” in the global citation network that speed the diffusion of novel information. The inclusion of controls for article age and open access (+ 17.1%, IRR = 1.171) further ensures that the effects are not merely driven by accessibility or exposure time. We therefore read the impact of global networks as reflecting genuine methodological reach and broader validation rather than self-promotion, consistent with the generalizability account of Section 4.1.

5. Conclusions

5.1 Summary of Main Findings

This study provides early econometric evidence on the determinants of scientific impact in the AI-Finance nexus. Analyzing 2,694 articles published between 2015 and 2025, we map an intellectual structure that has evolved rapidly from isolated predictive models toward decentralized FinTech ecosystems. The clearest result is that structural factors-above all international collaboration-are the strongest drivers of citation success, raising expected citations by 91%. We also document a “Complexity Paradox”: manuscripts centered on the most sophisticated and specialized topics are cited less often, with Deep Learning carrying a 16.8% penalty and the inherently interpretable Risk Management cluster carrying an even larger penalty (-28.0%). Because both penalties point the same way, we interpret the paradox as evidence of a generalizability and audience-breadth premium rather than a simple penalty on opaque methods.

5.2 Theoretical Implications

Theoretically, the study challenges the assumption that methodological sophistication intrinsically yields stronger academic influence. Rather than positing an unqualified “explainability premium,” we argue that the economics-and-finance community rewards work that is broadly applicable and globally connected. Explainability may still matter for the opacity dimension of black-box methods, but our evidence shows it cannot, by itself, account for the observed topic-level citation patterns; the more general constructs of reach and generalizability fit the data better and unify the thematic and structural results.

5.3 Practical and Policy Implications

For practitioners and policymakers, the academic premium on generalizable, well-connected research mirrors operational realities. Banks, asset managers, and FinTech firms face compliance requirements (credit-scoring justifications, trading audits) that constrain the deployment of wholly opaque systems,

and they value models that transfer across markets and customer segments. To maximize both academic impact and industrial relevance, researchers should favor cross-border collaboration, broad empirical validation, and, where high-stakes decisions are involved, transparent, auditable, explainable AI (XAI) frameworks, while recognizing that transparency is one of several routes to influence rather than a guarantee of it.

For scientometrics and research evaluation, the results carry a specific caution: raw citation counts should not be read as direct quality signals when comparing heterogeneous specialisms. Mature, applied, or regionally-scoped topics - such as credit-risk modeling - appear to be systematically under-cited relative to emerging, cross-disciplinary ones, independent of methodological merit, so evaluators and funders should normalize for topic, age, and audience before drawing conclusions. From a science-policy standpoint, the large premia on international collaboration and competitive funding indicate that instruments promoting cross-border consortia and open grant support are also, in effect, instruments for amplifying research impact.

5.4 Limitations and Future Research

This study has several limitations that also define an agenda. First, the extraction was confined to Scopus and to English-language publications; integrating Web of Science or regional databases would broaden the geopolitical spectrum. Second, citations are only one proxy for impact and need not reflect industrial adoption or patents; altmetrics and practitioner interviews could validate the mechanisms proposed here. Third, and most importantly in light of peer review, our thematic variables are keyword-based proxies that conflate interpretability with topic maturity, audience size, and geographic scope; a decisive test of the explainability-versus-generalizability question requires coding interpretability at the article level and estimating it within topics. Fourth, while we report interaction effects and a mature-cohort robustness check (Section 3.5), the design remains cross-sectional and associational; controls for team size, journal prestige, and single-versus multi-country empirical scope would further sharpen the estimates. Finally, as Generative AI and Large Language Models permeate financial analysis, future bibliometric models should assess whether these architectures incur the same complexity penalty as those observed in traditional Deep Learning.

Declarations

Author Contribution

LFA wrote the main manuscript text and prepared all figures and tables.

Data Availability and Reproducibility

The analysis is based on 2,694 English-language Scopus articles (2015-2025), with citation counts taken at a December 2025 census date. To respect the Scopus/Elsevier database license, the openly available

Zenodo repository (CC BY 4.0; https://doi.org/10.5281/zenodo.20738261) does not redistribute the raw Scopus records; it contains the analysis-ready dataset of the nine coded variables, a record-identifier file (Scopus EID and DOI) for independent re-retrieval, the exact search query and inclusion criteria, and the estimation code. All descriptive statistics, the Negative Binomial estimates in Table 3, the 95% confidence intervals, the interaction and mature-cohort robustness checks in Section 3.5, and the Spearman correlation between team size and international collaboration (rho = 0.35) are fully reproducible from these materials. Econometric results were obtained using GRETL 2026b and Python 3.14.2, both of which are freely available to all readers.

Funding information on this research

No funding has been provided for this research.

References

1. Aksnes, D. W. (2003). Characteristics of highly cited papers. Research Evaluation, 12(3), 159-170.
2. Arner, D. W., Barberis, J., & Buckley, R. P. (2020). FinTech, RegTech, and the reconceptualization of financial regulation. Northwestern Journal of International Law & Business, 37(3), 371-413.
3. Aste, T. (2019). Cryptocurrency market structure: connecting emotions and economics. Digital Finance, 1, 5-21. https://doi.org/10.1007/s42521-019-00008-9
4. Belanche, D., Casalo Arino, L. V., & Flavian, C. (2019). Artificial Intelligence in FinTech: understanding robo-advisors adoption among customers. Industrial Management & Data Systems. https://doi.org/10.1108/IMDS-08-2018-0368
5. Bickley, S. J., Chan, H. F., & Torgler, B. (2022). Artificial intelligence in the field of economics. Scientometrics, 127(4), 2055-2084. https://doi.org/10.1007/s11192-022-04294-w
6. Cameron, A. C., & Trivedi, P. K. (1998). Regression Analysis of Count Data. Cambridge University Press (Econometric Society Monographs No. 30); 2nd ed., 2013.
7. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of KDD 2015, 1721-1730. https://doi.org/10.1145/2783258.2788613
8. Cheng, M., & Qu, Y. (2020). Does bank FinTech reduce credit risk? Evidence from China. Pacific-Basin Finance Journal, 63, 101398. https://doi.org/10.1016/j.pacfin.2020.101398
9. Dowling, M., & Lucey, B. (2023). ChatGPT for (finance) research: The Bananarama Conjecture. Finance Research Letters, 53, 103662. https://doi.org/10.1016/j.frl.2023.103662
10. Dwivedi, Y. K., Hughes, L., Ismagilova, E., et al. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

11. Fischer, T. G., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669. https://doi.org/10.1016/j.ejor.2017.11.054
12. Geng, R., Bose, I., & Chen, X. (2015). Prediction of financial distress: An empirical study of listed Chinese companies using data mining. European Journal of Operational Research, 241(1), 236-247. https://doi.org/10.1016/j.ejor.2014.08.016
13. Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. Energy Economics, 81, 709-727. https://doi.org/10.1016/j.eneco.2019.05.006
14. Glanzel, W., & Thijs, B. (2004). The weight of collaboration on the citation impact of a scientist. Scientometrics, 60(3), 425-444.
15. Goodell, J. W., et al. (2021). Artificial intelligence and financial markets: A systematic review. Expert Systems with Applications, 178, 115020.
16. Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: survey, opportunities, and challenges. Journal of Big Data, 6(1), 44. https://doi.org/10.1186/s40537-019-0206-3
17. Hedayati Moghaddam, A. H., Moghaddam, M. H., & Esfandyari, M. (2016). Stock market index prediction using artificial neural network. Journal of Economics, Finance and Administrative Science, 21(41), 89-93. https://doi.org/10.1016/j.jefas.2016.07.002
18. Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. European Journal of Operational Research, 259(2), 689-702. https://doi.org/10.1016/j.ejor.2016.10.031
19. Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. Information Sciences, 467, 312-322. https://doi.org/10.1016/j.ins.2018.08.019
20. Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. Information Processing & Management, 57(5), 102212. https://doi.org/10.1016/j.ipm.2020.102212
21. Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. Knowledge-Based Systems, 164, 163-173. https://doi.org/10.1016/j.knosys.2018.10.034
22. Ma, X., Sha, J., Wang, D., et al. (2018). Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGBoost algorithms. Electronic Commerce Research and Applications, 28, 45-56. https://doi.org/10.1016/j.elerap.2018.08.002
23. Manela, A., & Moreira, A. (2017). News implied volatility and disaster concerns. Journal of Financial Economics, 123(1), 137-162. https://doi.org/10.1016/j.jfineco.2016.01.032
24. Mhlanga, D. (2020). Industry 4.0 in finance: the impact of artificial intelligence (AI) on digital financial inclusion. International Journal of Financial Studies, 8(3), 45. https://doi.org/10.3390/ijfs8030045
25. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ, 339, b2535.

26. Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. Journal of Banking & Finance, 84, 25-40. https://doi.org/10.1016/j.jbankfin.2017.07.002
27. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x
28. Sohangir, S., Wang, D., Pomeranets, A., & Khoshgoftaar, T. M. (2018). Big Data: Deep Learning for financial sentiment analysis. Journal of Big Data, 5(1), 3. https://doi.org/10.1186/s40537-017-0111-6
29. Tahamtan, I., Safipour Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: a comprehensive review of the literature. Scientometrics, 107(3), 1195-1225. https://doi.org/10.1007/s11192-016-1889-2
30. Teo, A.-C., Tan, G. W.-H., Ooi, K.-B., et al. (2015). The effects of convenience and speed in m-payment. Industrial Management & Data Systems, 115(2), 311-331. https://doi.org/10.1108/IMDS-08-2014-0231
31. van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538.
32. Wallin, J. A. (2005). Bibliometric methods: pitfalls and possibilities. Basic & Clinical Pharmacology & Toxicology, 97(5), 261-275.
33. Wasserbacher, H., & Spindler, M. (2022). Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls. Digital Finance, 4(1), 63-88. https://doi.org/10.1007/s42521-021-00046-2
34. Zhu, Y., Zhou, L., Xie, C., et al. (2019). Forecasting SMEs credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. International Journal of Production Economics, 211, 22-33. https://doi.org/10.1016/j.ijpe.2019.01.032
35. Zupic, I., & Cater, T. (2015). Bibliometric methods in management and organization. Organizational Research Methods, 18(3), 429-472.

Figures

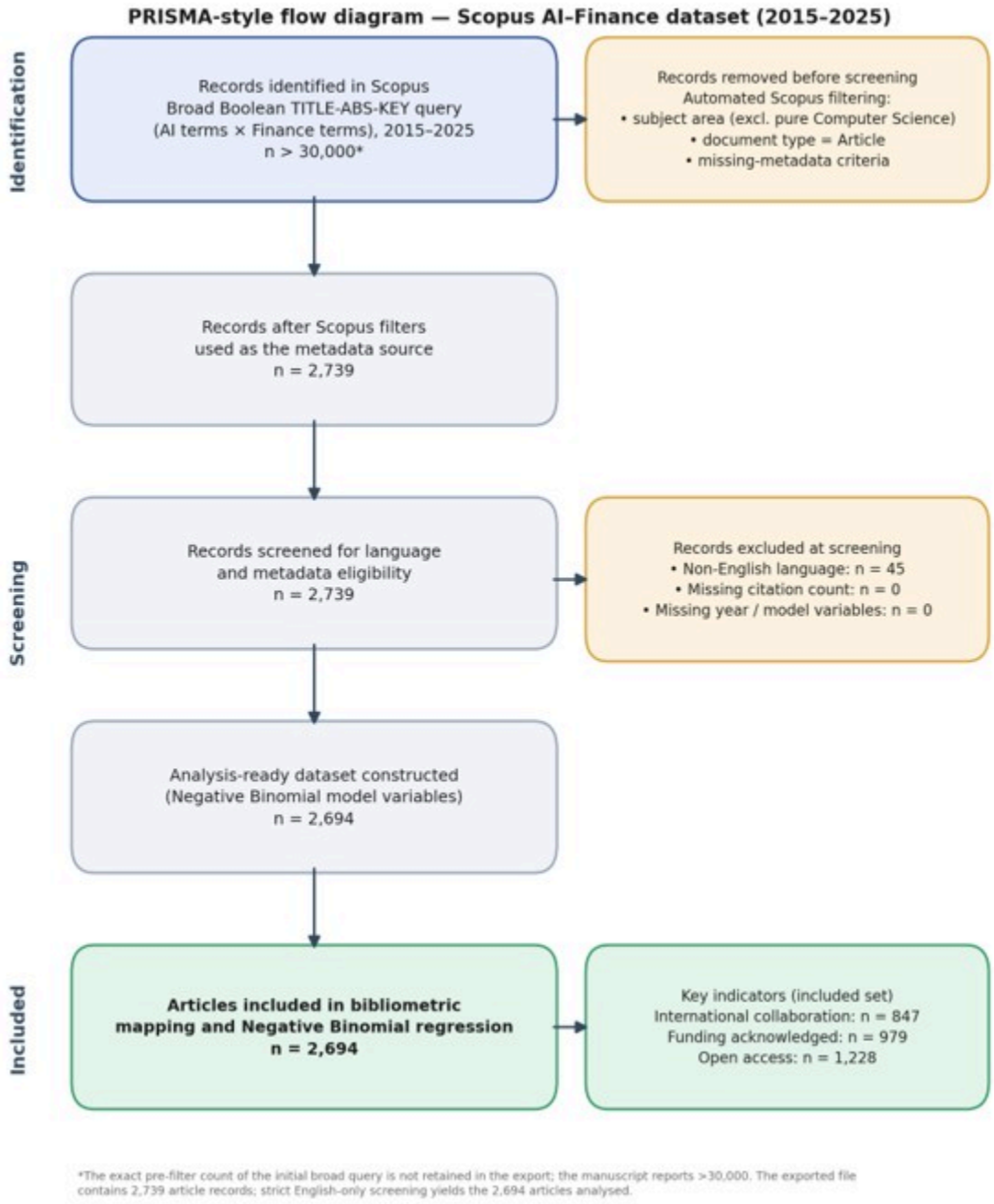


Figure 1

PRISMA-style flow diagram for the Scopus AI–Finance dataset (2015–2025). The initial Scopus search used a broad Boolean TITLE-ABS-KEY query combining AI-related and finance-related terms. After applying Scopus filters for subject area (Economics, Econometrics and Finance; Business, Management and Accounting), document type (research articles) and metadata availability, 2,739 records were exported and converted into an analysis-ready dataset. The export contained no EID or DOI duplicates and no missing values in the model variables. An audit of the Scopus "Language of Original Document" field identified 2,694 English-language records and 45 records in other original-language categories; the latter were excluded at screening, yielding a final analytic sample of N = 2,694 articles for bibliometric mapping and Negative Binomial regression.

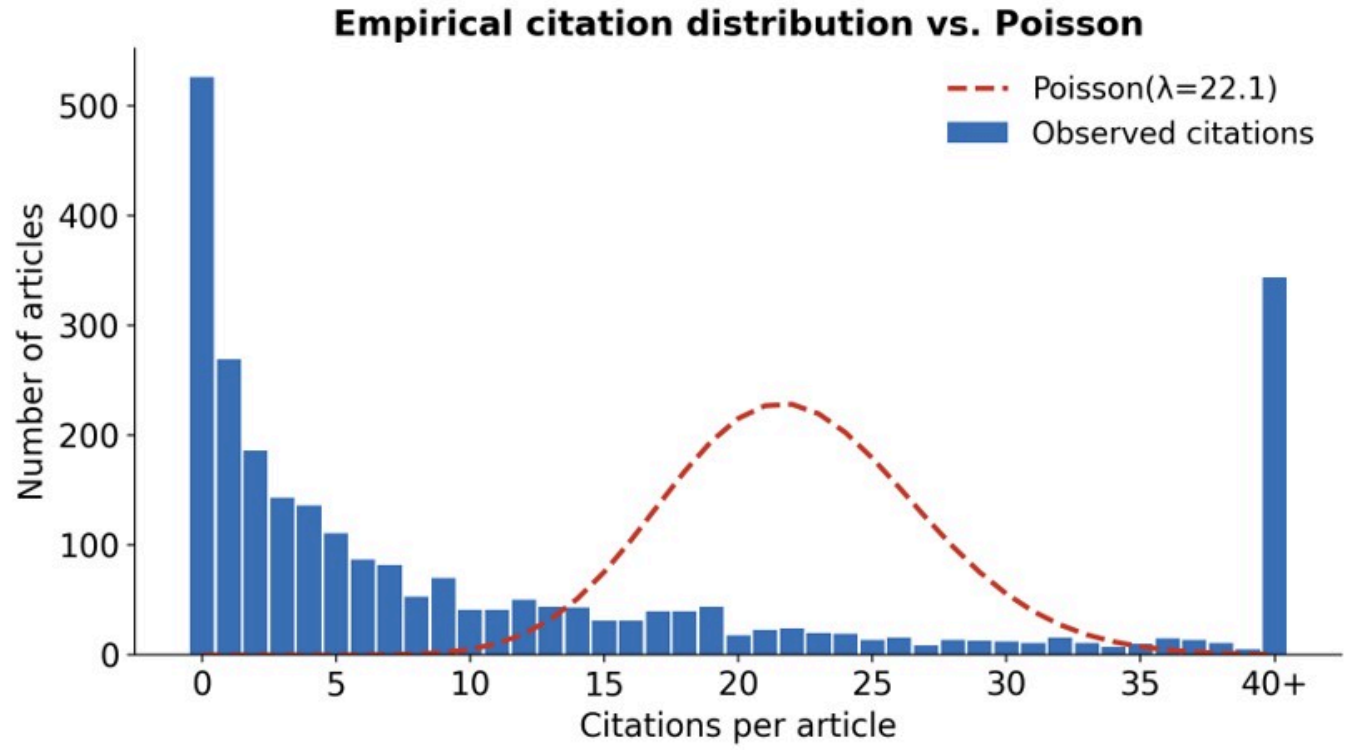


Figure 2

Empirical distribution of citations versus a theoretical Poisson distribution. The discrepancy (excess zeros and heavy right tail) confirms overdispersion.

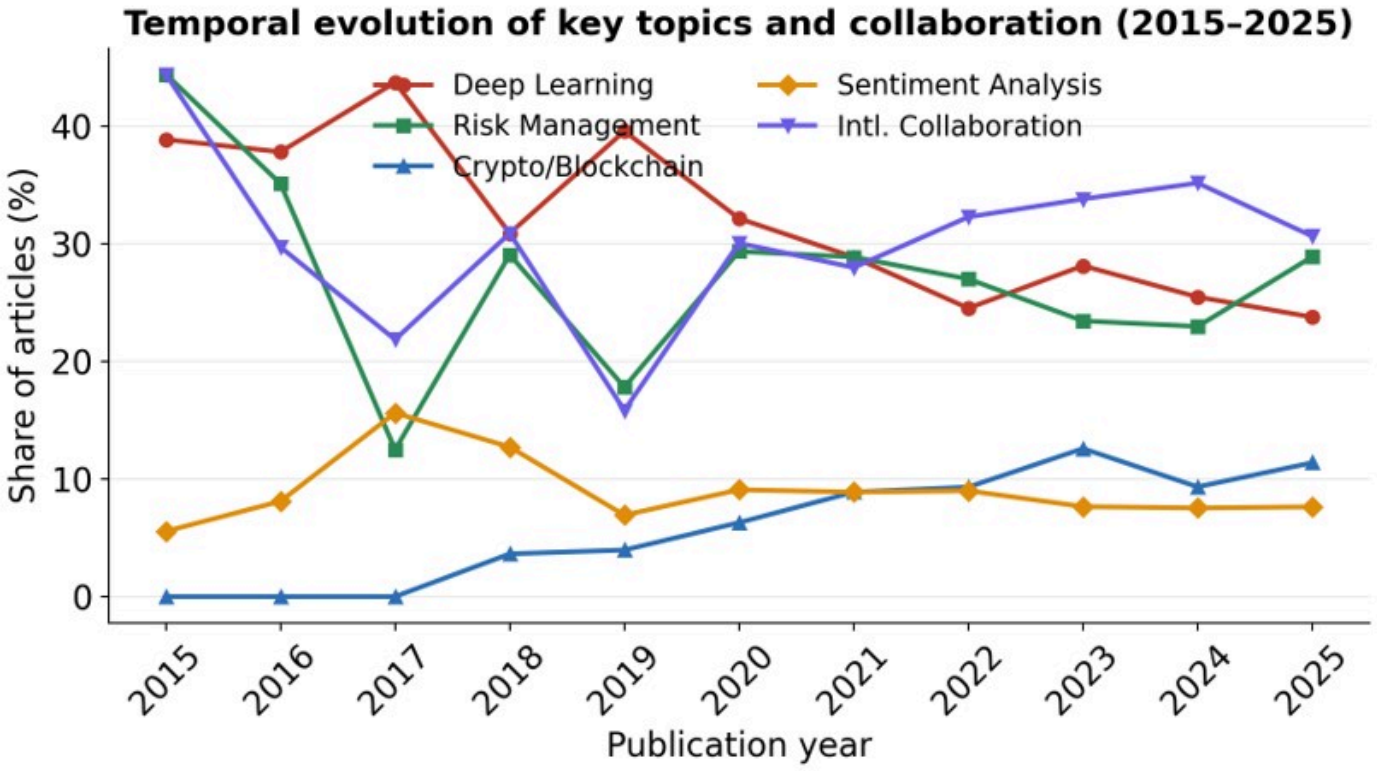


Figure 3

Temporal evolution of key topics and international collaboration (2015-2025).

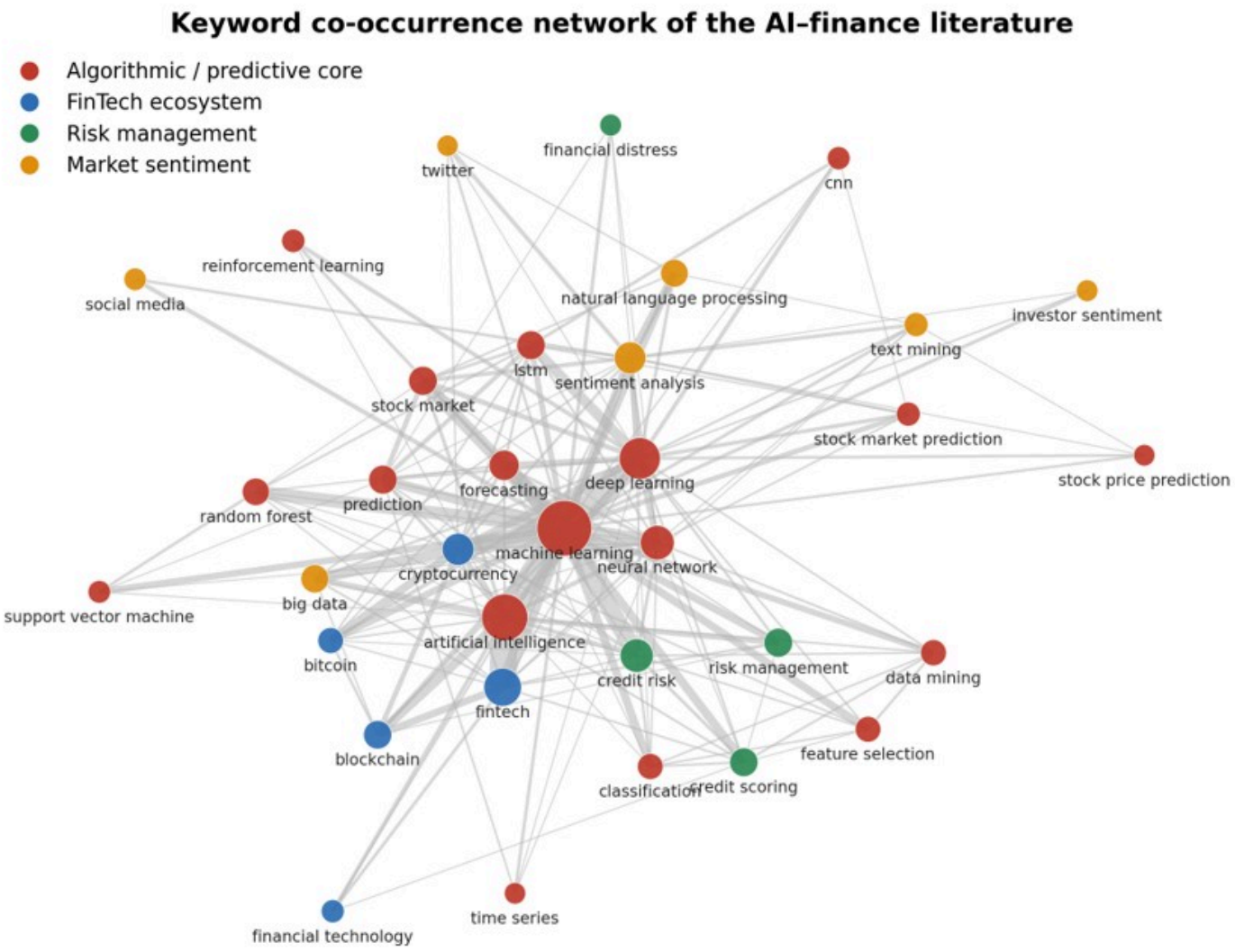


Figure 4

Keyword co-occurrence network of the AI-finance literature (most frequent author keywords). Node size is proportional to keyword occurrence and edge width to co-occurrence frequency. Colours denote the four thematic communities, matching the four regression indicators: ***red****= algorithmic/predictive core (e.g., deep learning, neural networks, forecasting);* ***blue*** *= FinTech ecosystem (blockchain, cryptocurrency, Bitcoin);* ***green*** *= risk management (credit risk, credit scoring, default);* ***orange*** *= market sentiment (sentiment analysis, text mining, social media).*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AIFinancerecordidentifiers.csv
- AIFinanceanalyticdataset.csv