Mapping the Research Landscape of Transparent AI in University Assessment: A Bibliometric Investigation

Flavio Manganello[1], Alberto Nico[2] and Giannangelo Boccuzzi[1,3]

[1]Institute for Educational Technology, National Research Council, Genoa, Italy

[2]Department of Law, University “Aldo Moro”, Bari, Italy

[3]Department of Education Studies, University of Bologna “Alma Mater Studiorum”, Italy

flavio.manganello@cnr.it

alberto.nico@uniba.it

giannangeloboccuzzi@cnr.it

Abstract: This study presents a systematic bibliometric investigation of AI transparency research in university assessment contexts. Following GLOBAL recommendations (Ng et al., 2024), we examined 72 peer-reviewed publications from Scopus (2019-2025) using performance metrics and science mapping techniques. Findings reveal exponential growth from single publications in 2019 to 26 documents in 2024 ($R^2 = 0.7666$, $p = 0.0098$). The domain generated 655 citations achieving h-index of 13. China leads in output (8 documents) while Sweden demonstrates highest citation efficiency (25.50 citations per document). Science mapping identifies four primary clusters: technical transparency methodologies, educational analytics frameworks, machine learning applications, and performance prediction systems. Co-citation analysis establishes Adadi and Berrada’s XAI survey (2018) as the foundational framework (11 citations, 24 total link strength). Temporal evolution shows progression from basic concepts toward practical implementations, reflecting regulatory compliance following GDPR and EU AI Act. International collaboration reveals South-South partnerships and high-impact contributions from countries with strong data protection frameworks. These patterns provide evidence for an emerging interdisciplinary domain addressing AI accountability in higher education, offering insights for researchers, practitioners, and policymakers.

Keywords: Artificial intelligence Transparency, Higher education assessment, Bibliometric analysis, Explainable AI, Educational technology, Regulatory compliance

1. Introduction

The integration of artificial intelligence systems within university assessment practices has experienced significant changes following the implementation of comprehensive data protection regulations requiring explicit algorithmic transparency in automated decision-making processes (Voigt & Von dem Bussche, 2017). This regulatory evolution began with the General Data Protection Regulation (GDPR), which established substantial requirements for meaningful information about automated processes through Article 22, considerably altering how educational institutions must approach AI-driven assessment systems (Edwards & Veale, 2017).

The regulatory landscape has developed further with the European Union’s Artificial Intelligence Act (2024), which specifically classifies educational AI systems as “high-risk” applications requiring comprehensive transparency measures, human oversight, and detailed risk management protocols (European Commission, 2024). While these regulatory frameworks establish transparency requirements for automated decision-making in education, legal analysis suggests persistent gaps in guidance for hybrid assessment systems that integrate human oversight with algorithmic processes (Boccuzzi et al., 2025). These regulatory developments have created substantial demands for transparent, interpretable, and accountable AI systems within higher education contexts, requiring fundamental reconsideration of traditional machine learning approaches to educational assessment.

Traditional machine learning models employed in educational contexts typically operate as opaque computational systems where decision-making logic, feature importance, and prediction rationale remain hidden from educational stakeholders including students, instructors, and administrators (Rudin, 2019). This algorithmic opacity creates substantial challenges for universities where transparency, fairness, accountability, and pedagogical effectiveness constitute important institutional requirements and regulatory obligations.

Explainable Artificial Intelligence (XAI) has developed as a significant technological response to these “black box” limitations, enabling human-understandable insights into algorithmic decision-making processes while maintaining predictive performance standards (Gunning et al., 2019). XAI encompasses diverse methodological approaches including post-hoc explanation techniques, inherently interpretable models, and human-centered design principles that address transparency requirements in high-stakes decision-making contexts.

1.1 Literature Review and Knowledge Gaps

Despite growing research activity at the intersection of AI transparency and educational assessment, comprehensive systematic reviews have consistently identified significant gaps in current knowledge frameworks and methodological approaches. Altukhi and Pradhan (2024) conducted extensive analysis revealing the absence of standardised XAI definitions across educational contexts, leading to conceptual fragmentation and implementation inconsistencies that may hinder both theoretical development and practical deployment.

Bond, Zawacki-Richter and Nichols (2024) performed meta-systematic review of artificial intelligence applications in higher education, identifying apparent methodological inconsistencies in AI education research alongside insufficient attention to ethical implications and transparency requirements. Their analysis revealed patterns where technical performance metrics often superseded considerations of algorithmic accountability and stakeholder understanding.

Türkmen (2025) noted through comprehensive literature review that while XAI algorithms are increasingly recognized as potentially valuable in educational research contexts, there remains insufficient systematic analysis of how these transparency systems are being developed, validated, and implemented specifically within higher education assessment environments. This gap is particularly concerning given the high-stakes nature of educational decisions and their long-term impact on student outcomes.

Zawacki-Richter and colleagues (2019) provided foundational systematic review of AI applications in higher education, noting limited attention to ethical considerations and transparency requirements across the broader educational AI research landscape. Their work established baseline understanding while highlighting apparent needs for systematic investigation of emerging transparency-focused research domains.

1.2 Research Objectives and Framework

This investigation employs the Population-Concept-Context (PCC) framework (Peters et al., 2015) to attempt comprehensive coverage. Population encompasses researchers, practitioners, and developers working on AI transparency in higher education. Concept includes explainable AI methodologies applied to educational evaluation. Context addresses university assessment environments requiring regulatory compliance and institutional accountability.

This study addresses five research questions: (RQ1) What are temporal patterns of publication growth and associated factors? (RQ2) Who are the most productive authors, institutions, and countries, and what collaboration patterns exist? (RQ3) What are primary publication venues and highly cited contributions? (RQ4) What are main research themes and their evolution? (RQ5) What foundational works appear to underpin current research?

2. Methods

2.1 Methodological Design and Theoretical Framework

This investigation employs bibliometric analysis following GLOBAL recommendations (Ng et al., 2024) and BIBLIO requirements (Montazeri et al., 2023). The methodology integrates performance analysis examining productivity metrics and citation impact with science mapping techniques visualizing knowledge structures and collaborative networks. Scopus database was selected for its substantial coverage of peer-reviewed literature and robust indexing systems (Mongeon & Paul-Hus, 2016). The study acknowledges constraints inherent in analyzing emerging interdisciplinary domains, including small corpus size and temporal concentration that may reflect field development stage.

2.2 Search Strategy Development and Implementation

The search strategy captured literature at the intersection of explainable AI, higher education, and assessment practices using Boolean logic with three components connected by AND operators. The XAI component included “explainable AI”, “interpretable AI”, “AI transparency”, “algorithmic accountability”, “responsible AI”, and specific methods like “Shapley Additive Expla*”. The education component encompassed “higher education”, “university”, “college”, “tertiary education”. The assessment component covered “assessment”, “evaluation”, “academic performance”, “learning analytics”, “automated scoring”, “student dropout”. The search excluded medical domains and limited results to peer-reviewed English publications. Conducted on July 5, 2025, it yielded 141 documents comprising articles (54.6%), conference papers (39.0%), book chapters (4.3%), reviews (0.7%), and books (1.4%).

2.3 Data Collection and Screening Process

Systematic screening using Rayyan (Ouzzani et al., 2016) applied inclusion criteria requiring peer-reviewed publications addressing XAI in higher education assessment. Exclusion criteria eliminated non-peer-reviewed materials, K-12 studies, and research lacking explicit transparency focus. After screening, 72 studies met inclusion criteria from the initial 141 documents. Data processing employed Publish or Perish v8.0 and Excel for bibliometric indicators, VOSviewer v1.6.20 for network analysis (van Eck & Waltman, 2010), and statistical validation including linear regression and correlation analysis.

2.4 Methodological Constraints

This investigation acknowledges methodological constraints that require consideration. The small corpus size (72 documents) and substantial temporal concentration (81.9% in 2023-2025) reflect the nascent stage of AI transparency research in educational contexts, which may limit pattern generalizability and could indicate coincidental rather than systematic relationships. Additionally, the specialized search strategy may systematically exclude relevant research employing alternative terminological frameworks, while exclusive Scopus reliance may underrepresent studies in discipline-specific databases or publications with limited international indexing. These characteristics necessitate cautious interpretation of identified trends and require validation through expanded datasets and longer observation periods.

3. Results

3.1 Research Productivity and Growth Patterns

Temporal analysis of publication distribution reveals notable growth patterns that may suggest systematic research attention following regulatory and technological developments within the AI transparency domain. The field emerged with limited contributions including single publications in 2019 and 2020, representing foundational exploratory work addressing algorithmic transparency requirements within higher education contexts and establishing preliminary conceptual frameworks for GDPR Article 22 compliance in educational AI applications.

The period spanning 2021-2022 witnessed moderate expansion with 6 and 5 documents respectively, coinciding with increased institutional awareness of data protection regulations and growing recognition of transparency requirements for educational AI systems. This developmental phase appears to have marked transition from theoretical exploration toward methodological development as researchers began addressing practical implementation challenges for explainable assessment systems within university environments.

Considerable growth occurred during 2023-2024, with publications increasing from 16 to 26 documents, representing over 300% growth and indicating accelerated research attention to this specialized intersection. This expansion may reflect convergence of multiple factors including maturation of XAI methodological approaches, increased availability of educational datasets suitable for transparent analysis, growing institutional demand for accountable automated assessment systems, and heightened awareness of regulatory compliance requirements.

Linear regression analysis provides statistical validation for observed growth trends, achieving $F(1,5) = 16.42$ with $p = 0.0098$ and $R^2 = 0.7666$, indicating that approximately 77% of variance in publication numbers can be explained by temporal progression. This statistical significance suggests the possibility of systematic rather than random growth patterns, though the limited temporal baseline requires cautious interpretation of trend extrapolation given the early stage nature of this research domain.

The dataset spanning six citation years (2019-2025) demonstrates building scholarly influence with 655 total citations distributed across 72 papers, yielding average citation rate of 109.17 citations per year. This citation accumulation may reflect growing academic engagement, while the finding that 61.1% of papers have received at least one citation suggests broad research community recognition. Seven papers achieving 10 or more citations indicate the possible emergence of a core set of influential contributions establishing theoretical and methodological foundations for the field.

3.2 Geographic Distribution and International Collaboration

International engagement analysis reveals substantial global reach with research contributions spanning 43 countries across six continents, indicating widespread international interest in AI transparency applications for educational contexts. This geographic diversity suggests that concerns about educational AI transparency transcend specific national regulatory contexts and appear to reflect common challenges in implementing accountable automated assessment systems.

Research concentration patterns show moderate distribution with the top 5 contributing countries (China: 8 documents, Italy: 7 documents, Australia: 7 documents, Germany: 7 documents, Bangladesh: 6 documents) accounting for 33.0% of total publications, while the remaining 67.0% is distributed among 38 additional nations. This balanced distribution suggests relatively distributed global participation rather than extreme concentration in dominant research centers.

Continental analysis reveals different regional engagement characteristics with Europe emerging as the leading contributor, providing 43.8% of total publications (49 documents from 19 countries). This European prominence may reflect strong regulatory frameworks and institutional responses following GDPR implementation, possibly creating research incentives for developing compliant educational AI systems requiring transparency and accountability measures.

Asia demonstrates substantial momentum as the second-largest regional contributor with 29.5% of publications (33 documents from 9 countries), suggesting growing recognition of AI transparency requirements within rapidly expanding higher education sectors across Asian nations. This engagement pattern may reflect both regulatory influences from international frameworks and domestic policy development addressing educational AI governance.

Citation analysis reveals interesting patterns diverging from simple productivity rankings. Sweden emerges as a strong performer in citation metrics, achieving 25.50 citations per document despite contributing only 4 documents, demonstrating quality-focused research strategies that achieve substantial international influence. Italy and Australia demonstrate consistent performance combining substantial output (7 documents each) with strong citation performance (11.14 and 11.00 citations per document respectively), suggesting both nations function as key regional hubs that appear to successfully integrate productive research output with scholarly impact.

International collaboration network analysis reveals varied partnership structures, particularly among developing economy institutions. Malaysia and Saudi Arabia both demonstrate high collaborative intensity (total link strength of 4.00), though Saudi Arabia shows higher citation performance (10.80 citations per document) compared to Malaysia (5.83 citations per document), indicating different collaborative strategies and research focus areas.

3.3 Key Contributors and Scholarly Networks

Author collaboration analysis using fractional counting methodology reveals different productivity and impact patterns within the research community. Individual researcher analysis identifies Dragan Gasevic as a highly productive contributor with 5 documents and 61 citations, achieving 12.20 citations per document alongside substantial collaborative integration (total link strength of 5.00). This positioning suggests sustained research leadership spanning multiple collaborative projects and theoretical framework development within the domain.

A group of four authors comprising Muhammad Afzaal, Uno Fors, Jalal Nouri, and Aayesha Zia demonstrates strong coordination with identical metrics of 3 documents and 101 citations each, achieving high citation performance of 33.67 citations per document. Their consistent collaborative integration (total link strength of 3.00 each) suggests these researchers function as a closely integrated research unit, possibly co-authoring the same influential publications focused on explainable AI applications for educational feedback and student self-regulation systems.

Gabriella Casalino represents productive European engagement contributing 3 documents with 51 citations, achieving 17.00 citations per document and strong collaborative integration (total link strength of 3.00). This performance suggests sustained contribution to methodological frameworks bridging computer science and educational applications within the AI transparency domain.

Institutional analysis reveals distributed engagement patterns with 121 institutions contributing to the research landscape. Productivity concentration analysis shows balanced distribution, with top 5 contributing institutions (Monash University: 5 documents, King Abdulaziz University: 3 documents, Stockholm University: 3 documents, Universiti Teknologi Malaysia: 3 documents, University of Bari Aldo Moro: 3 documents) accounting for only 11.6% of total publications. This low institutional concentration suggests widespread scholarly engagement rather than dominance by specialized research centers, indicating that AI transparency research addresses challenges that appear to resonate across diverse institutional contexts globally.

3.4 Publication Venues and Scholarly Impact

Publication venue analysis demonstrates the interdisciplinary character of AI transparency research spanning computer science, educational technology, policy studies, and applied ethics domains. This diversity may reflect the complex nature of implementing transparent AI systems within educational contexts, requiring integration of technical expertise, pedagogical understanding, and regulatory compliance knowledge.

Conference proceedings dominate the publication landscape with "ACM International Conference Proceeding Series" leading with 6 publications, followed by "CEUR Workshop Proceedings" with 4 publications. This conference prevalence may reflect the rapidly evolving nature of the research domain, where conference presentations enable faster knowledge dissemination compared to traditional journal publication cycles, potentially facilitating timely sharing of methodological innovations and practical implementation experiences.

Journal venue diversity includes "IEEE Access" with 3 publications and "International Journal of Artificial Intelligence in Education" with 2 publications, representing established channels for methodologically rigorous research addressing technical innovations and educational applications respectively. The presence of diverse publication venues across multiple disciplines suggests integration of AI transparency research within established scholarly communication networks.

Citation impact analysis identifies influential foundational contributions establishing theoretical frameworks and practical methodologies for the emerging field. Susnjak, Ramaswami and Mathrani (2022) achieved substantial impact with 103 citations (34.33 cites per year) for their work on learning analytics dashboards providing actionable insights to learners, developing integrated theoretical and practical frameworks for transparent learning analytics interfaces focused on actionable feedback mechanisms supporting student self-regulation.

Afzaal and colleagues (2021) generated considerable influence with 77 citations (19.25 cites per year) for their research on explainable AI applications for data-driven feedback and intelligent action recommendations supporting student self-regulation. Their work established approaches to XAI-driven educational feedback systems with explicit focus on student agency and transparent algorithmic recommendations.

Technical methodology contributions include Baranyi, Nagy and Molontay (2020) achieving 61 citations (12.20 cites per year) for developing methodological frameworks for transparent student retention prediction using interpretable deep learning approaches. Their work addresses important institutional needs for understanding and explaining automated early warning systems while maintaining predictive accuracy requirements.

3.5 Thematic Structure and Intellectual Evolution

Comprehensive keyword co-occurrence analysis reveals the conceptual architecture underlying AI transparency research in higher education through systematic examination of author-assigned terminology and thematic relationships (Figure 1). Following extensive thesaurus development that standardized 220 author keywords into 147 core concepts to address terminology fragmentation, VOSviewer's clustering algorithm identified 22 thematic clusters based on co-occurrence patterns and link strength relationships.

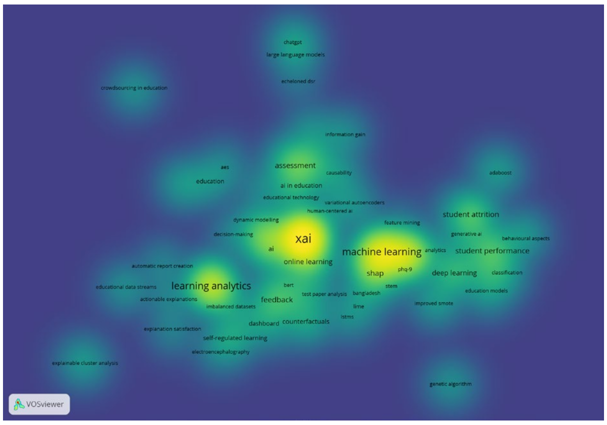

Figure 1: Co-occurrence analysis of author keywords. Density visualization

Statistical validation demonstrates thematic organization achieving structural properties with silhouette coefficient of 0.68, indicating cluster formation and thematic boundaries. To enhance interpretability and identify broader research themes, the 22 clusters were systematically aggregated into four primary macro-clusters based on conceptual similarity, temporal evolution patterns, and theoretical coherence.

“AI Ethics and Educational Technology” aggregates clusters focusing on responsible implementation and human-centered approaches, combining research on responsible AI frameworks, emerging generative technologies including large language models and ChatGPT applications, and core XAI theoretical foundations with human-centred computing principles. This macro-cluster encompasses research addressing ethical implications of transparent AI systems, stakeholder engagement in algorithmic decision-making, and value-aligned design principles for educational contexts requiring accountability and fairness considerations.

“Deep Learning and Predictive Analytics” consolidates technically-oriented clusters focusing on advanced methodological approaches, integrating predictive analysis methodologies with SHAP interpretability techniques, machine learning foundations addressing student performance prediction, student attrition modeling with optimization approaches, and algorithmic implementations including fuzzy clustering and prototype-based methods. This macro-cluster represents the technical infrastructure supporting transparent educational AI applications, addressing challenges of maintaining predictive performance while providing explanations for high-stakes educational decisions.

“Learning Analytics and Feedback Systems” synthesizes application-focused clusters addressing practical implementation within educational environments, combining learning analytics infrastructure with feedback mechanisms and counterfactual explanations, higher education contextual applications including degree completion studies, online learning environments with adaptive student modeling, and real-time educational data stream processing. This macro-cluster emphasizes actionable insights and student-centered design principles, bridging technical capabilities with educational practice requirements through transparent interfaces and decision support systems.

“Assessment and Educational Evaluation” encompasses evaluation-oriented clusters addressing formal measurement and monitoring systems, integrating assessment methodologies with behavioral analysis, student wellbeing monitoring through interpretable deep learning approaches, international large-scale assessment frameworks, quality assurance mechanisms, actionable explanation generation, user trust evaluation, and evaluation methodologies for interpretable systems. This macro-cluster focuses on transparency requirements for high-stakes assessment decisions, fairness considerations in automated evaluation, and institutional accountability frameworks requiring algorithmic explanation capabilities.

Temporal evolution analysis reveals four developmental phases characterizing the field’s intellectual progression (Figure 2). Foundation establishment (2021-2022) introduced core XAI concepts and basic learning analytics frameworks, establishing theoretical groundwork for educational applications. Methodological expansion (2023) integrated advanced interpretability techniques and developed specialized applications addressing specific educational contexts and stakeholder needs. Application maturation (2024) focused on predictive analytics combined with feedback systems, demonstrating integration of technical advancement with practical educational requirements.

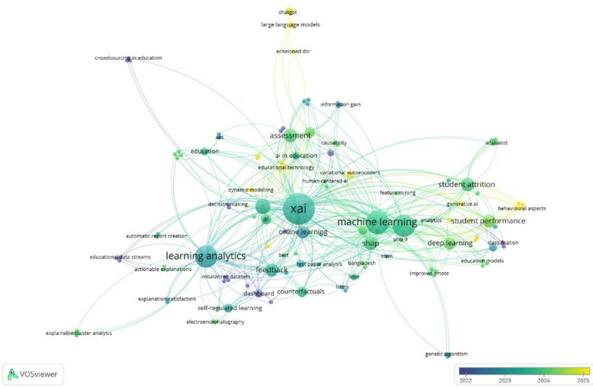

Figure 2: Co-occurrence analysis of author keywords. Overlay visualization

3.6 Intellectual Foundations and Theoretical Frameworks

Co-citation analysis provides systematic identification of intellectual foundations supporting current research development within the AI transparency domain. Statistical validation reveals that 30% of reference pairs exhibit strong co-citation relationships ($r > 0.5$), which may exceed typical bibliometric baselines and suggests theoretical integration across previously distinct knowledge domains including foundational XAI theory, educational AI applications, and technical implementation methodologies.

Adadi and Berrada's (2018) survey of explainable artificial intelligence emerges as a frequently cited theoretical reference with 11 citations and 24 total link strength, providing conceptual foundations that appear to be consistently applied across educational contexts. Their systematic framework for understanding XAI methodologies, explanation types, and evaluation approaches functions as a theoretical anchor for educational applications requiring algorithmic transparency.

Additional foundational works include Alamri and Alharbi (2021) with 5 citations and 11 total link strength, providing systematic review of explainable student performance prediction models that bridge general XAI theory with educational applications. Baranyi and colleagues (2020) contribute technical foundations with 4 citations and 57 total link strength, establishing methodological approaches for interpretable deep learning in educational contexts.

Educational psychology foundations include Hattie and Timperley (2007) with 3 citations and 17 total link strength, providing theoretical frameworks for effective feedback systems that inform transparent AI applications in educational contexts. Technical methodology references include Ribeiro, Singh and Guestrin (2016) for LIME methodology (3 citations, 39 total link strength), Chen and Guestrin (2016) for XGBoost approaches (3 citations, 15 total link strength), and Molnar (2020) for interpretable machine learning frameworks (4 citations, 23 total link strength).

4. Discussion

4.1 Theoretical Contributions and Domain Development

This systematic bibliometric investigation provides evidence for the emergence of an interdisciplinary research domain that appear to address AI transparency within higher education assessment contexts. The identification of four thematically organized clusters (silhouette coefficient of 0.68) suggests theoretical coherence within this specialized intersection, though expanded datasets and longitudinal validation remain necessary for examining domain stability and conceptual boundaries.

The documented progression from foundational XAI concepts toward education-specific applications may indicate development beyond technical application toward theoretical frameworks that appear to address educational contexts. This evolution suggests possible synthesis of technical transparency requirements with pedagogical effectiveness principles, regulatory compliance demands, and institutional accountability frameworks, potentially offering models for other high-stakes application domains requiring AI explainability.

Co-citation analysis revealing 30% of reference pairs exhibiting strong relationships ($r > 0.5$) suggests intellectual integration across previously separate domains including computer science, educational psychology, regulatory compliance, and applied ethics. This interdisciplinary synthesis may indicate theoretical development emerging from domain intersection rather than application of existing frameworks to educational contexts.

4.2 Practical Implications for Implementation

For educational practitioners and institutional administrators, the concentration of highly cited works addressing transparent learning analytics interfaces, explainable feedback systems, and interpretable prediction models may provide evidence-based guidance for implementation strategies. The prevalence of conference proceedings over traditional journals suggests active knowledge translation between research and practice communities, indicating that research findings may be rapidly incorporating practical implementation experiences and stakeholder feedback.

The identification of collaborative models, particularly coordinated multi-institutional initiatives among developing economy institutions, may offer viable frameworks for international knowledge sharing and capacity building in AI transparency implementation. These partnership patterns could provide practical approaches for institutions seeking to develop XAI capabilities while addressing local regulatory requirements and pedagogical contexts.

Geographic concentration of high-impact research in countries with strong data protection frameworks suggests possible relationships between regulatory environments promoting AI transparency and research excellence. While causal relationships require systematic investigation, these patterns indicate that regulatory frameworks mandating algorithmic accountability may facilitate rather than hinder research innovation and practical implementation.

4.3 Policy and Regulatory Implications

For policymakers developing AI governance frameworks, the documented research community responsiveness to regulatory developments following GDPR implementation suggests that targeted transparency requirements may stimulate relevant academic investigation and practical innovation. The growth pattern ($R^2 = 0.7666$) following major regulatory milestones suggests that research communities may adapt rapidly to address policy-driven transparency demands.

The emergence of specialized research addressing educational AI transparency indicates that sector-specific regulations may be necessary and effective in driving research attention and implementation standards. The observed relationship between countries with robust data protection frameworks and research excellence warrants systematic investigation for policy design implications.

4.4 Theoretical Synthesis and Framework Development

The co-citation analysis reveals how XAI theory, educational assessment literature, and regulatory frameworks appear to be converging into a distinctive theoretical synthesis. Rather than simple application of existing XAI methodologies to educational contexts, the evidence suggests development of what might be characterized as "Educational Algorithmic Accountability" frameworks. These emerging approaches appear to integrate technical transparency requirements with pedagogical effectiveness principles, institutional accountability demands, and regulatory compliance standards.

This theoretical convergence may represent a significant departure from traditional educational technology adoption patterns, where technical solutions are often retrofitted to educational contexts. The identified macro-clusters suggest instead that educational requirements are fundamentally shaping XAI development, potentially contributing novel approaches to the broader AI transparency field.

4.5 Future Directions

Future research should prioritize developing evaluation frameworks for XAI effectiveness in educational contexts, expanding investigation from performance prediction toward engagement analytics, and systematically investigating relationships between regulatory environments and research excellence. Longitudinal studies tracking domain evolution and empirical validation of identified collaborative models will be essential for determining whether observed patterns represent field characteristics or transitional phenomena.

4.6 Conclusions

This bibliometric investigation of 72 peer-reviewed publications spanning 2019-2025 provides systematic empirical evidence for the emergence of an interdisciplinary research domain that appears to address AI transparency and accountability within higher education assessment contexts. The analysis reveals growth trajectories with publication output expanding from limited contributions in 2019 to 26 documents in 2024, demonstrating statistically significant trends ($R^2 = 0.7666$) that may suggest systematic rather than random research attention development.

The domain has generated scholarly impact with 655 total citations achieving h-index of 13 and g-index of 24, indicating academic engagement despite the field's recent emergence. Geographic distribution analysis reveals international collaboration patterns including South-South partnerships among developing economy institutions alongside strong citation performance in countries with established data protection frameworks, suggesting possible relationships between regulatory environments and research excellence.

Science mapping through systematic keyword analysis identified four thematically organized clusters with structural properties, providing evidence for theoretical coherence within this specialized intersection of technical transparency requirements and educational applications. Co-citation analysis suggested intellectual integration across computer science, educational psychology, and regulatory compliance domains, indicating possible development of frameworks rather than application of existing approaches.

For academic researchers, this systematic mapping identifies collaboration opportunities and research gaps while highlighting needs for standardized terminological frameworks and expanded methodological approaches. For educational practitioners and administrators, the concentration of influential works may provide evidence-based guidance for implementation strategies while identified collaborative models may offer frameworks for international knowledge sharing. For policymakers, the observed research community responsiveness to regulatory developments suggests that targeted transparency requirements may stimulate relevant academic investigation and innovation.

These findings should be interpreted as systematic evidence for an emerging research domain requiring validation through expanded datasets, longer observation periods, and complementary analytical approaches as the field continues developing. Future research should prioritize developing evaluation frameworks, expanding investigation toward engagement analytics, and systematically investigating relationships between regulatory environments and research excellence to support continued domain maturation.

Ethics declaration: This study involved analysis of publicly available bibliometric data retrieved from Scopus database and did not require institutional ethics approval. All analyzed publications were already published and accessible through standard academic databases.

AI declaration: AI tools were used minimally for manuscript formatting and reference style verification. All research design, data analysis, interpretation of results, and scientific conclusions were developed entirely by the authors without AI assistance.

References

Adadi, A. and Berrada, M. (2018) “Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)”, *IEEE Access*, Vol. 6, pp. 52138-52160. doi:10.1109/ACCESS.2018.2870052

Afzaal, M., Nouri, J., Zia, A., Papapetrou, P., Fors, U., Wu, Y., Li, X. and Weegar, R. (2021) “Explainable AI for data-driven feedback and intelligent action recommendations to support students self-regulation”, *Frontiers in Artificial Intelligence*, Vol. 4, Article 723447. doi:10.3389/frai.2021.723447

Alamri, R. and Alharbi, B. (2021) “Explainable student performance prediction models: a systematic review”, *IEEE Access*, Vol. 9, pp. 33132-33143. doi:10.1109/ACCESS.2021.3061368

Altukhi, Z.M. and Pradhan, S. (2024) “Systematic literature review: Explainable AI definitions and challenges in education”, *arXiv* preprint arXiv*:2504.02910*. doi:10.48550/arXiv.2504.02910

Baranyi, M., Nagy, M. and Molontay, R. (2020) “Interpretable deep learning for university dropout prediction”, in *Proceedings of the 21st Annual Conference on Information Technology Education*, pp. 13-19. doi:10.1145/3368308.3415382

Boccuzzi, G., Nico, A. and Manganello, F. (2025) “Harmonizing human and algorithmic assessment: Legal reflections on the right to explainability in education”, *Proceedings of the 17th International Conference on Education and New Learning Technologies*, pp. 9489-9493. doi:10.21125/edulearn.2025.2446

Bond, M., Zawacki-Richter, O. and Nichols, M. (2024) “A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour”, *International Journal of Educational Technology in Higher Education*, Vol. 21, No. 1, pp. 1-25. doi:10.1186/s41239-024-00444-7

Chen, T. and Guestrin, C. (2016) “XGBoost: A scalable tree boosting system”, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794. doi:10.1145/2939672.2939785

Edwards, L. and Veale, M. (2017) “Slave to the algorithm? Why a ‘right to an explanation’ is probably not the remedy you are looking for”, *Duke Law & Technology Review*, Vol. 16, pp. 18-84.

European Commission (2024) *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence*, Official Journal of the European Union, L 1689.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S. and Yang, G. Z. (2019) “XAI—Explainable artificial intelligence”, *Science Robotics*, Vol. 4, No. 37, Article eaay7120. doi:10.1126/scirobotics.aay7120

Hattie, J. and Timperley, H. (2007) “The power of feedback”, *Review of Educational Research*, Vol. 77, No. 1, pp. 81-112. doi:10.3102/003465430298487

Molnar, C. (2020) *Interpretable Machine Learning*, Lulu.com.

Mongeon, P. and Paul-Hus, A. (2016) “The journal coverage of Web of Science and Scopus: a comparative analysis”, *Scientometrics*, Vol. 106, No. 1, pp. 213-228. doi:10.1007/s11192-015-1765-5

Montazeri, A., Mohammadi, S., Hesari, P. M., Ghaemi, M., Riazi, H. and Sheikhi-Mobarakeh, Z. (2023) “Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): a minimum requirements”, *Systematic Reviews*, Vol. 12, No. 1, Article 239. doi:10.1186/s13643-023-02410-2

Nagy, M. and Molontay, R. (2024) “Interpretable dropout prediction: towards XAI-based personalized intervention”, *International Journal of Artificial Intelligence in Education*, Vol. 34, No. 2, pp. 274-300. doi:10.1007/s40593-023-00331-8

Ng, J. Y., Liu, H., Masood, M., Syed, N., Stephen, D., Ayala, A. P., Sabé, M., Solmi, M., Waltman, L., Haustein, S. and Moher, D. (2024) "Guidance for the Reporting of Bibliometric Analyses: A Scoping Review", *medRxiv.* doi:10.1101/2024.08.26.24312538

Ouzzani, M., Hammady, H., Fedorowicz, Z. and Elmagarmid, A. (2016) "Rayyan—a web and mobile app for systematic reviews", *Systematic Reviews*, Vol. 5, No. 1, Article 210. doi:10.1186/s13643-016-0384-4

Peters, M. D. J., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D. and Soares, C. B. (2015) "Guidance for conducting systematic scoping reviews", *International Journal of Evidence-Based Healthcare*, Vol. 13, No. 3, pp. 141-146. doi:10.1097/XEB.0000000000000050

Ribeiro, M. T., Singh, S. and Guestrin, C. (2016) "'Why should I trust you?' Explaining the predictions of any classifier", in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135-1144. doi:10.1145/2939672.2939778

Rudin, C. (2019) "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead", *Nature Machine Intelligence*, Vol. 1, No. 5, pp. 206-215. doi:10.1038/s42256-019-0048-x

Susnjak, T., Ramaswami, G. S. and Mathrani, A. (2022) "Learning analytics dashboard: a tool for providing actionable insights to learners", *International Journal of Educational Technology in Higher Education*, Vol. 19, No. 1, Article 12. doi:10.1186/s41239-021-00313-7

Türkmen, G. (2025) "The review of studies on explainable artificial intelligence in educational research", *Journal of Educational Technology & Society*, Vol. 28, No. 1, pp. 45-62. doi:10.1177/07356331241310915

van Eck, N. J. and Waltman, L. (2010) "Software survey: VOSviewer, a computer program for bibliometric mapping", *Scientometrics*, Vol. 84, No. 2, pp. 523-538. doi:10.1007/s11192-009-0146-3

Voigt, P. and Von dem Bussche, A. (2017) *The EU General Data Protection Regulation (GDPR): A Practical Guide*, Springer International Publishing. doi:10.1007/978-3-319-57959-7

Zawacki-Richter, O., Marín, V. I., Bond, M. and Gouverneur, F. (2019) "Systematic review of research on artificial intelligence applications in higher education – where are the educators?", *International Journal of Educational Technology in Higher Education*, Vol. 16, No. 1, pp. 1-27. doi:10.1186/s41239-019-0171-0