Vocabulary Without Infrastructure: A Bibliometric Analysis of Explainability and Equity in AI-Based Personalised Learning

Ahmed Echchoayeby[1]*, Wafae Abbaoui[1,2], Reida Ilal[1], Nassim Kharmoum[1,2], Soumia Ziti[1,2]
Intelligent Processing and Security of Systems, Faculty of Sciences Mohammed V University, Rabat, Morocco[1]
SMSD, Moroccan Society of Digital Health, Rabat, Morocco[2]

***Abstract*—Artificial intelligence has become a central paradigm in personalised learning. However, it remains unclear whether the two principles that responsible AI treats as foundational, explainability and equity, are integrated into the field's intellectual architecture or are only loosely associated with it. This question is addressed through a bibliometric analysis of 1371 publications indexed in Scopus and Web of Science between 2015 and 2025. Performance analysis and science-mapping techniques were applied using bibliometrix and VOSviewer, with cluster structure quantified through modularity and inter-cluster density metrics, and uncertainty assessed through 1000-iteration edge-resampling bootstrap. Keyword co-occurrence analysis showed that ethics-equity vocabulary connects to the applied-AI cluster at densities comparable to those of technical vocabulary (0.20 vs 0.21), but the ethics-equity and technical streams interact directly only at half that rate (0.11; bootstrap 95% CIs non-overlapping), engaging primarily through the applied-AI mediator cluster rather than with each other. Co-citation analysis showed that the foundational responsible-AI references in education do not register as cluster-forming nodes. No specialist responsible-AI venue appears in the field's Bradford core, and Sub-Saharan Africa and Latin America remain almost absent from the country-level collaboration network. To account for this pattern of vocabulary diffusion without infrastructural consolidation, we propose the Structural Integration Model (SIM). The SIM is a three-proposition framework that specifies the conditions under which explainability and equity can become structurally embedded in the field's foundation rather than only thematically present within it.**

***Keywords*—Bibliometric analysis; AI in education; personalised learning; explainable AI; educational equity; responsible AI; science mapping**

I. Introduction

Artificial Intelligence (AI)-driven personalised learning aims to individualize educational experiences to individual learners' needs, pacing, and cognitive states. It now sits at the centre of Artificial Intelligence in Education (AIED) research [1], [2]. The field's computational core has matured over three decades. Early work was anchored on Corbett and Anderson's Bayesian knowledge tracing [3] and progressed through deep knowledge tracing [4], transformer-based architectures [5], and attention-driven knowledge models [6]. The recent emergence of generative AI and large language models (LLMs) has accelerated these capabilities further, offering substantial scalability in content generation and adaptive tutoring [7], [8]. Predictive performance and instructional efficiency have risen accordingly [9], [10], and adaptive learning has been established as a central paradigm in educational technology [11], [12]. This rapid technical advancement, however, has not been matched by equivalent progress on the normative dimensions of deployment. Research in explainable AI (XAI) has established interpretability as a prerequisite for trustworthy deployment [13]. Domain-specific scholarship has emphasised the need for human-centred explanations that allow educators and learners to engage meaningfully with adaptive systems [14]. In parallel, work on algorithmic fairness has shown that machine learning models in education can reproduce and amplify socioeconomic and demographic inequalities, raising significant concerns about systemic bias [15], [16], [17].

These two dimensions are interdependent. Explainability without attention to who explanations serve risks reinforcing inequity, and fairness without transparency cannot be audited [14], [18]. Despite community-level efforts to articulate ethical governance frameworks [19], [20], [21], [22] and longstanding calls for educational AI that is inclusive and contextually grounded [23], [24], [2], explainability and equity remain only loosely embedded in the field's core technical agenda. Advances in predictive modelling and adaptive optimisation continue to develop largely in isolation from work on fairness and ethics [15], [16], [17]. Whether the field is converging toward an integrated research paradigm, or whether these dimensions persist as fragmented and unevenly connected subdomains, is therefore not self-evident from the volume of work alone. The question is not what the literature says, but how it is organised. This is a structural question, and one that science-mapping techniques are uniquely equipped to answer at scale [25], [26], [27], [28].

Recent bibliometric and tertiary reviews have provided valuable insights into AIED's overall technological development [29], [30], [31], [32], but they have not used network-based bibliometric metrics to examine the structural position of responsible-AI dimensions within the field's intellectual architecture.

We address this gap by analysing 1371 publications indexed in Scopus and Web of Science between 2015 and 2025. The analysis combines the bibliometrix R package [33] and VOSviewer [34], complemented by modularity-based clustering metrics [28], [35], and applies performance analysis and science-mapping techniques. These include co-citation analysis

*Corresponding author.

[36], co-word analysis [37], and thematic-evolution analysis [26], used together to examine productivity, collaboration networks, and intellectual structures. Particular attention is given to whether explainability and equity are structurally integrated across the field's intellectual architecture (vocabulary, citation canon, venues, collaborations, and geographic base), or whether their integration is uneven across these layers.

This study makes three contributions. First, it provides the first quantitatively grounded bibliometric map of AI-based personalised learning that explicitly examines the structural position of responsible-AI dimensions within the field's intellectual architecture, capturing the field's rapid expansion following the integration of generative AI. Second, it identifies a pattern of vocabulary diffusion without infrastructural consolidation, in which responsible-AI terms have been adopted into the field's keyword space but have not yet consolidated within its citation, venue, collaboration, or geographic infrastructure. Third, to account for this pattern, we advance the Structural Integration Model (SIM), a theoretical framework that reconceptualises explainability and equity as co-constitutive design principles for responsible AI in education.

The study is guided by five research questions:

- RQ1. What are the temporal trends and growth patterns of research on AI-based personalised learning?
- RQ2. Which authors, institutions, countries, and journals constitute the most productive and influential actors, when measured by both publication count and citation impact?
- RQ3. What intellectual structures emerge from co-citation, keyword co-occurrence, and collaboration network analysis, as quantified through modularity and inter-cluster density metrics?
- RQ4. What thematic clusters emerge from keyword co-occurrence analysis, and how have they evolved across the 2015–2021, 2022–2024, and 2025 subperiods?
- RQ5. How are explainability and equity positioned within the broader intellectual landscape: as integrated concerns, as thematically present but infrastructurally unconsolidated, or as disconnected from the mainstream?

The remainder of the study is organised as follows. Section II reviews related work and positions this study against prior bibliometric analyses, providing the conceptual bridge to the methodological strategy. Section III describes the data collection and screening process. Section IV presents the methodology. Section V reports the results. Section VI discusses the findings, addresses each research question explicitly, and advances the theoretical contribution. Section VII concludes.

II. Related Work

A. Prior Bibliometric Analyses of AI in Education

Bibliometric analysis is a quantitative methodology for evaluating, monitoring, and mapping scientific literature [25], [27], [38]. The term itself was introduced by Pritchard [38] to denote the application of mathematical and statistical methods to the study of scientific communication. By combining performance analysis (which identifies productive actors) with science mapping (which visualises how knowledge is interconnected), bibliometric studies offer a systematic view of the intellectual architecture underlying a research domain [25], [26], [27], [28].

Within AIED, prior bibliometric and systematic reviews have primarily focused on application areas, temporal trends, and technological developments. The widely cited systematic review by Zawacki-Richter et al. [1] analysed a systematic sample of empirical studies and identified personalisation, profiling, and assessment as the field's dominant application areas. The same review also drew attention to the marginalisation of pedagogical and ethical considerations, an early indication of structural imbalance in research priorities. Hinojo-Lucena et al. [30] conducted one of the first bibliometric analyses of AI in higher education and characterised the literature between 2007 and 2017 as nascent but rapidly emerging. More recently, Bahroun et al. [29] combined bibliometric and content analysis to map the rise of generative AI in educational settings, documenting its rapid expansion, though without applying network-based metrics to examine the structural position of responsible-AI dimensions. Crompton and Burke [31] surveyed AIED in higher education through 2022, while Bond et al. [32] conducted a meta-systematic review that explicitly called for greater attention to ethics, collaboration, and methodological rigour. These foundational studies provide valuable baselines, but they largely offer descriptive categorisations of the literature. They do not, however, examine the structural integration of normative dimensions such as explainability and equity into the field's intellectual architecture, a gap that motivates the present study, as developed in Section II-D.

B. The Technical Trajectory of Personalised Learning

The technical foundations of AI-based personalised learning have evolved along a well-documented trajectory. Early work on cognitive models, including Bayesian knowledge tracing [3], together with empirical evidence demonstrating the effectiveness of intelligent tutoring systems [9], validated the feasibility of adaptive learning. Subsequent conceptual contributions emphasised the transformative potential of AI for improving educational outcomes at a global scale [23], [24].

In recent years, advances in deep learning [4], [10], [5], [6] have substantially enhanced the field's predictive capabilities and enabled scalable personalisation. Parallel developments in recommender systems and adaptive learning frameworks [11], [12] further expanded the technical scope of the field.

C. The Emergence of Explainability and Equity

In contrast to this primarily technical focus, a parallel body of literature has begun to address the ethical and societal implications of AIED. In the broader AI domain, explainability has been established as a non-negotiable requirement for responsible system deployment [13]. Domain-specific frameworks for explainable AI in education [14] build on this foundation and underscore that educational environments require interpretable, human-centred explanations capable of supporting trust between educators and learners.

Empirical studies have likewise demonstrated the risks of algorithmic bias in educational systems [15], [16], showing how data-driven models can inadvertently reinforce existing socioeconomic and demographic inequalities. Conceptual analyses further warn that AIED systems may amplify inequities through structural design choices [16], [18]. In response, community-level studies have called for shared ethical frameworks [20], [22], proposing principles that explicitly integrate fairness, accountability, and transparency [21] while navigating domain-specific ethical risks [19].

D. The Structural Gap and Bridge to the Present Study

Despite the growing prominence of these ethical frameworks, the literature remains internally uneven. Research on technical performance, explainability, and educational equity continues to develop in parallel, with limited qualitative evidence of systematic integration across domains [15], [16], [18]. Existing reviews [1], [29], [30], [31], [32] have documented that these topics are present in the field, but they have not examined whether these critical dimensions are forming a cohesive, integrated research agenda, or whether their integration is uneven across the layers of the field's intellectual architecture (vocabulary, citation canon, venues, collaborations, and geographic base), a question that requires network-based bibliometric measurement rather than descriptive synthesis.

To the best of our knowledge, this study is among the first to investigate the structural position of explainability and equity in AI-based personalised learning through large-scale bibliometric mapping with quantitative network metrics, and the first to operationalise this structural integration question through modularity, intra-cluster, and inter-cluster density measures. By treating the connectivity (or lack thereof) between these dimensions as an empirical question to be answered through modularity analysis and intra- and inter-cluster density measures [28], [35], this study moves beyond descriptive literature reviews [25], [26], [27]. It provides a structural perspective on the field's intellectual organisation and its alignment with responsible-AI principles.

III. Data Collection and Screening

The study draws on a bibliometric corpus assembled from two complementary bibliographic databases: Scopus (Elsevier) and the Web of Science Core Collection (Clarivate). The two databases differ in coverage. Web of Science is valued for selective, high-quality journal indexing well suited to citation-based analyses [39], while Scopus offers broader coverage, particularly in engineering and applied sciences [40]. Each database introduces distinct biases across fields and languages. Their combined use is therefore recommended for bibliometric studies that span interdisciplinary domains [25], [40], and we adopted it on those grounds. This two-database design does not capture grey literature, preprints, or regionally-indexed and non-English scholarship, including repositories such as SciELO and AfricArXiv that carry a substantial share of Latin American and African output. Because our analysis includes a geographic-equity dimension, this coverage profile may understate contributions from underrepresented regions and is therefore likely to bias the equity-paradox finding (Section VI-D) toward greater concentration than exists in the broader literature. We return to this coverage limitation and its mitigations in Section VI-G (First). Data collection, screening, and deduplication followed the PRISMA 2020 reporting guidelines [41]. All database searches were conducted on a single date (17 March 2026) to ensure temporal consistency across the two sources. The full process is summarised in Fig. 1.

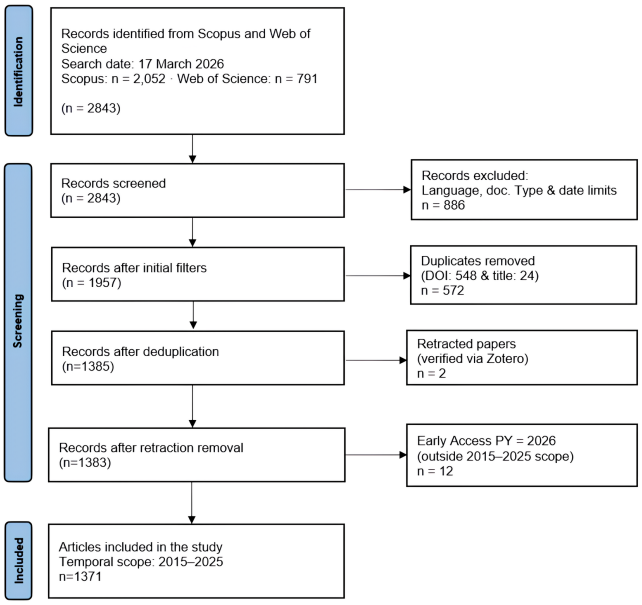

Fig. 1. PRISMA 2020 flow chart of the literature identification, screening, and inclusion process.

A. Search Strategy

The search strategy was designed to capture the intersection of three research dimensions that define the study's scope: AI methods, educational applications, and responsible-AI concepts. A three-cluster Boolean query was constructed accordingly:

- AI computational methods: artificial intelligence, machine learning, deep learning, reinforcement learning, recommender systems, and related terms.
- Educational applications: personalised learning, adaptive learning, learning pathways, intelligent tutoring systems, and related terms.
- Responsible-AI concepts: explainability, transparency, equity, fairness, trust, algorithmic bias, inclusive learning, and related terms.

The third cluster has generally not been foregrounded in the search strategies of prior bibliometric reviews of personalised learning, which have tended to privilege technical or application-oriented terms [1], [29], [31]. It is included here precisely so that the structural positioning of explainability and equity can be examined as an empirical question rather than assumed. A consequence of the three-cluster conjunctive (AND) design is that every retrieved document contains at least one responsible-AI term by construction. The corpus therefore samples the subset of AI-based personalised-learning research that engages responsible-AI concepts explicitly at the

level of title, abstract, or keywords, rather than the entire AI-based personalised-learning field. This has two implications for interpretation. First, the analysis characterises how responsible-AI vocabulary is structurally positioned *within* the responsible-AI-engaged literature; it does not estimate the overall prevalence of responsible-AI concerns across the full field, nor does it capture studies that address explainability, fairness, or equity implicitly or through implementation without using the selected terminology. Second, the structural findings that anchor this study inter-cluster density, citation-canon membership, venue concentration, and collaboration breadth are properties of connectivity within this corpus and are therefore robust to the conjunctive design, whereas raw frequency comparisons should be read as within-corpus rather than whole-field estimates. A complementary design, in which a broad two-cluster (AI + educational-application) corpus is constructed first and the responsible-AI subset is analysed against that larger baseline, would allow the prevalence and integration of responsible-AI research to be evaluated against the full field; we identify this as a priority for future replication (Section VI-G, VI-H). The full Boolean strings applied in each database are reproduced below.

Scopus (TITLE-ABS-KEY):
```
TITLE-ABS-KEY ( ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Recommender System*" OR "Computational Intelligence" OR "Reinforcement Learning" OR "Learner Model*" OR "Cognitive Model*" OR "Optimization" OR "Decision-making model*") AND ("Personalized learning" OR "Adaptive learning" OR "Learning path*" OR "Learning trajectory" OR "Adaptive path*" OR "Teaching path*" OR "Intelligent Tutoring System*") AND ("Explainab*" OR "XAI" OR "Transparen*" OR "Interpretab*" OR "Equit*" OR "Fairness" OR "Algorithmic bias" OR "Inclusive learning" OR "Trust"))
```

Web of Science (TS=):
```
TS = ( ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Recommender System*" OR "Computational Intelligence" OR "Reinforcement Learning" OR "Learner Model*" OR "Cognitive Model*" OR "Optimization" OR "Decision-making model*") AND ("Personalized learning" OR "Adaptive learning" OR "Learning path*" OR "Learning trajectory" OR "Adaptive path*" OR "Teaching path*" OR "Intelligent Tutoring System*") AND ("Explainab*" OR "XAI" OR "Transparen*" OR "Interpretab*" OR "Equit*" OR "Fairness" OR "Algorithmic bias" OR "Inclusive learning" OR "Trust"))
```

The initial search returned 2052 records from Scopus and 791 records from Web of Science, yielding a combined identification set of 2843 records.

B. Inclusion and Exclusion Criteria

Inclusion and exclusion criteria were applied prior to database export. The full set is reported in Table I.

TABLE I. INCLUSION AND EXCLUSION CRITERIA APPLIED PRIOR TO DATABASE EXPORT

Criterion	Inclusion	Exclusion
Publication year	2015–2025	Before 2015; 2026 or later
Language	English only	All other languages
Document type	Articles, reviews, conference proceedings papers	Book chapters, editorials, letters, notes, errata

The temporal range was restricted to 2015 to 2025, reflecting the emergence and rapid expansion of deep learning in educational applications, beginning with the introduction of deep knowledge tracing by Piech et al. [4]. This window captures the period during which output in AI-based personalised learning grew by orders of magnitude, warranting systematic examination. Conference proceedings were retained alongside journal articles and reviews, since they remain a primary dissemination channel in AI research. Influential contributions such as Piech et al. [4] and Ghosh et al. [6] originally appeared in conference venues, and excluding them would remove foundational work from the corpus. Book chapters, editorials, letters, notes, and errata were excluded because they either lack primary empirical content or fall outside the conventional scope of bibliometric coverage. Non-English records were excluded to preserve consistency in text-based analyses such as keyword co-occurrence. After these criteria were applied, 1291 records were retained from Scopus and 666 from Web of Science, for a combined screening set of 1957 records.

C. Data Cleaning and Final Corpus

The 1957 screened records were imported into R (version 4.5.2) using the bibliometrix package [33], with additional processing in dplyr. Duplicates were removed in two stages, following bibliometric methodology best practices [25]. First, exact matching on DOI identified 548 duplicate pairs across the two databases. Second, normalised title matching (lowercased, whitespace-trimmed, and punctuation-stripped) detected an additional 24 duplicates. This two-stage procedure removed 572 duplicate records in total (548 by DOI and 24 by title), leaving 1385 records, in exact agreement with Fig. 1. Two retracted publications were subsequently removed after cross-checking retraction notices in Zotero, leaving 1383 records. Finally, 12 records flagged as 2025 Early Access but formally indexed with a 2026 publication year were excluded from the quantitative corpus to prevent skewing of chronological trend analyses by an incomplete calendar year. The resulting analytical corpus comprises 1371 documents, exported to CSV for subsequent processing in VOSviewer [34]. In line with open-science and FAIR data principles [42], all R scripts used in the cleaning and analysis pipeline, including the modularity and density computation routines, will be deposited in a public Zenodo repository upon publication.

IV. METHODOLOGY

The study adopts a complementary two-tool bibliometric design [25], [27] that combines the Biblioshiny interface of the bibliometrix R package [33] with VOSviewer [34]. Bibliometrix is used for performance analysis and thematic

mapping. VOSviewer is used for network construction and visualisation, drawing on its established capability for rendering large-scale co-occurrence and co-citation networks based on the unified mapping and clustering principle of Waltman, van Eck, and Noyons [28]. Cluster-detection quality is quantified using the resolution-parameter variant of the modularity function implemented in VOSviewer's clustering algorithm [28], and cross-validated with the Louvain algorithm [43] as implemented in igraph [44] to assess the robustness of the cluster structure. Together, these components support both quantitative evaluation of productivity and detailed examination of intellectual structure.

The analysis proceeds through three sequential phases. Each phase moves the analysis one step further from description toward structural interpretation.

A. Phase 1: Corpus Overview

Phase 1 produces a descriptive profile of the dataset, comprising document types, annual scientific production, and average citations per year. These indicators establish the temporal dynamics and growth trajectory of the field. To assess the precision of the Boolean search strategy, a manual relevance audit was conducted on a random sample of 30 records (set.seed = 42), with each record classified as relevant, borderline, or irrelevant against the AI + education + responsible-AI scope. Audit results are reported in Section VI-G.

B. Phase 2: Performance Analysis

Phase 2 examines the structural and productivity characteristics of the field through established bibliometric laws and indicators. Source distribution is analysed using Bradford's Law [45] to identify the core journals contributing to the field. Phase 2 additionally identifies the most productive and influential authors, institutions, countries, and sources, reporting both publication counts and citation-impact metrics (h-index [46], total citations, M-index) following bibliometric methodology guidelines [25] which distinguish productivity from influence. The most globally and locally cited documents are also reported, and keyword frequency is used to capture dominant topics and emerging areas of interest [25].

C. Phase 3: Network and Thematic Analysis

Phase 3 uncovers the intellectual structure and thematic evolution of the field. A centrality-density thematic map is constructed using the framework of Cobo et al. [26], which classifies research themes by their centrality (external linkage to other themes) and density (internal coherence of the theme). A three-period thematic-evolution analysis (2015–2021, 2022–2024, 2025) is also conducted to address RQ4. The single-year final period reflects the volume surge in 2025 (854 documents), which warrants separate examination.

Five network visualisations are generated in VOSviewer [34]: 1) author-level co-authorship, 2) country-level co-authorship, 3) cited-reference co-citation [36], 4) cited-source co-citation [36], and 5) keyword co-occurrence [37]. Together, these networks map collaboration patterns, intellectual linkages, and thematic clustering within the field.

A minimum threshold of five occurrences or citations was applied to all VOSviewer analyses, following bibliometric methodology guidelines [25], which recommend calibrating thresholds to balance network completeness against interpretability. At n = 1371 documents, a threshold of five corresponds to approximately 0.36% of the corpus, filtering noise from sparsely connected nodes while preserving meaningful structural linkages. A sensitivity analysis was conducted at thresholds of 3, 5, 10, and 15, with full results reported in Table V (Section V-C6). Cluster count varied between five and six, while modularity (Q) remained within $Q \in [0.116, 0.192]$. Cluster counts from the Louvain implementation in igraph differ slightly from VOSviewer's smart local moving algorithm [47], reflecting differences in the underlying optimisation procedures. VOSviewer cluster assignments are used as the primary analytical basis throughout, with igraph providing an independent verification of modularity Q values. These results indicate that the keyword space is consistently more interconnected than fragmented across reasonable threshold choices, and therefore that the modular structure reported in Section V-C is stable in form even though its absolute Q values are modest.

For each VOSviewer network we report nodes, edges, modularity Q [35], overall graph density, and the top nodes by betweenness centrality where the metric is informative for the network's connectivity. Intra-cluster and inter-cluster densities are additionally reported for the keyword co-occurrence network, where they directly support the structural integration analysis (Section V-C6). Substantive cluster labels (ethics-equity, applied-AI, technical-methods, indexing-variant fragmentation, etc.) were assigned through manual qualitative inspection of the VOSviewer items export, in which the full keyword list of each cluster was reviewed against the cluster-defining vocabulary and recurring orthographic variants were flagged as fragmentation patterns. To quantify uncertainty around the inter-cluster densities that anchor Proposition 2 of the Structural Integration Model, we conducted a non-parametric bootstrap with 1000 edge-resampling iterations (set.seed = 42), reporting 95% confidence intervals alongside the point estimates. These metrics allow the structural position of the ethics-equity and technical clusters (RQ5) to be examined empirically rather than asserted.

V. Results

A. Phase 1: Corpus Overview

1) Main bibliometric information: The corpus comprises 1371 documents from 859 sources and 4167 unique authors. Collaboration is moderate (3.76 co-authors per document), but international co-authorship appears in only 13.93% of publications, a figure relevant to the country-level analysis in Section V-C3. Full descriptive statistics are reported in Table II.

2) Annual scientific production: Annual publication output is reported in Fig. 2. Output is distributed across three phases. The formative phase (2015 to 2019) comprises fewer than 15 documents per year, beginning with 5 publications in 2015 and reaching 13 in 2019. The emergence phase (2020 to 2022) records 24, 25, and 43 publications respectively. The acceleration phase (2023 to 2025) records 94 publications in 2023, 293 in 2024, and 854 in 2025. The 2025 figure is likely a lower bound owing to indexing lag (Section VI-G). The annual growth rate computed by the bibliometrix CAGR formula

TABLE II. MAIN BIBLIOMETRIC CHARACTERISTICS OF THE CORPUS (2015 TO 2025, N = 1371)

Description	Results
Main Information About Data	
Timespan	2015 : 2025
Sources (journals, books, etc.)	859
Documents	1371
Annual growth rate	67.2%
Document average age	1.81 yrs
Average citations per document	8.363
References	84847
Document Contents	
Keywords Plus (ID)	4071
Author's keywords (DE)	3840
Authors	
Authors	4167
Authors of single-authored documents	170
Author Collaboration	
Single-authored documents	179
Co-authors per document	3.76
International co-authorships	13.93%
Document Types	
Article	597
Article; early access	27
Article; proceedings paper	1
Conference paper	480
Proceedings paper	137
Review	123
Review; early access	6

Notes: A manual relevance audit ($n = 30$; `set.seed = 42`) estimated 13.3% false positives and 16.7% borderline records, implying an upward bias of up to ∼13% in raw productivity and citation-impact figures. Structural network findings (Section V-C4 and V-C6) are not materially affected, as they rely on co-occurrence and co-citation patterns rather than individual-record content. See Section VI-G (Sixth).

across the full window is 67.2%. This headline figure should be read as provisional: it is sensitive both to the small 2015 base and to the 2025 count being a lower bound affected by indexing lag (Section VI-G, Third), which inflates the apparent end-of-window growth. A more conservative compound growth rate based on a three-year rolling baseline (2015 to 2017 vs 2023 to 2025) yields approximately 61% per year. Both metrics confirm rapid field expansion.

B. Phase 2: Performance Analysis

1) Most relevant sources: The ten most productive sources by publication count are reported in Fig. 3. CEUR Workshop Proceedings leads with 23 documents, followed by IEEE Access and Lecture Notes in Networks and Systems (21 each); these raw counts carry an estimated upward bias of up to ∼13% from audit-identified false positives (Section VI-G, Sixth). Three of the top-ten entries are conference proceedings series, reflecting the AI-research norm of publishing through peer-reviewed conference venues.

The ten most locally cited sources are reported in Fig. 4. arXiv leads with 968 local citations, far ahead of IEEE Access (540) and Education and Information Technologies (402). The dominance of arXiv reflects the field's reliance on preprint citations for cutting-edge AI methodology, while the prominent positions of education-specific venues (Education and Information Technologies, Computers and Education: Artificial Intelligence, International Journal of Educational Technology in Higher Education, International Journal of Artificial Intelligence in Education) signal that the citation network draws on dedicated AIED journals despite none of them appearing in the productivity-ranked top ten.

Source distribution according to Bradford's Law [45] is reported in Fig. 5. The distribution partitions into three zones of roughly equal article counts: a core of 79 sources, a second zone of 324 sources, and a third zone of 456 sources and it conforms well to the Bradford model (R^2 = 0.91, Bradford multiplier k = 2.75). The cumulative curve opens at 23 articles for the highest-ranked source and rises steeply across the core zone before the increasingly large number of sources in the outer zones contribute only one or two documents each. Although the core includes general AI-in-education venues such as the International Journal of Artificial Intelligence in Education and Computers and Education: Artificial Intelligence, **no venue dedicated specifically to equitable, fair, or explainable AI in education appears in the core-zone sources**, a structural fact directly relevant to RQ5.

2) Most relevant institutions: Institutional productivity is reported in Fig. 6. Chitkara University (India) leads with 18 documents, followed by the Chinese Academy of Sciences (16) and the State University System of Florida. Beijing Normal University contributes 12 documents. Six institutions are tied at 11 documents each: Central China Normal University, Purdue University, the University of Patras, the University of Pennsylvania, the University of Science and Technology of China, and Zhejiang University. Five of the ten institutions are based in China. The University of Patras is the only European institution in the top ten. The State University System of Florida, Purdue University, and the University of Pennsylvania are the three North American institutions. No institution from Africa, Latin America, or Southeast Asia appears in the top ten, a finding revisited in Section VI-D in the context of the equity paradox.

3) Most relevant authors: Author-level findings are the least reliable layer of this analysis and are reported here for completeness rather than as a basis for structural inference. The surnames Wang, Zhang, Li, and Chen are among the most common in China and may correspond to multiple distinct researchers indexed under similar abbreviated forms. Where ORCID identifiers were present (n = 424 documents, 30.9% of the corpus), we verified author identity; ORCID resolution suggests that several entries in the top ten, particularly Wang J, Zhang Y, Wang Z, Zhang X, and Li X, likely represent more than one distinct researcher across different institutions. Full disambiguation would require manual reconciliation beyond the scope of bibliometric automation. We therefore treat author-level results as approximate and direct the reader to the institutional and country-level findings (Sections V-B2 and V-C3) as the more reliable structural indicators of research concentration.

With that caveat in place, the ten most productive authors range from 19 to 10 publications; detailed author-level data, including h-index, g-index, M-index, total citations, and primary

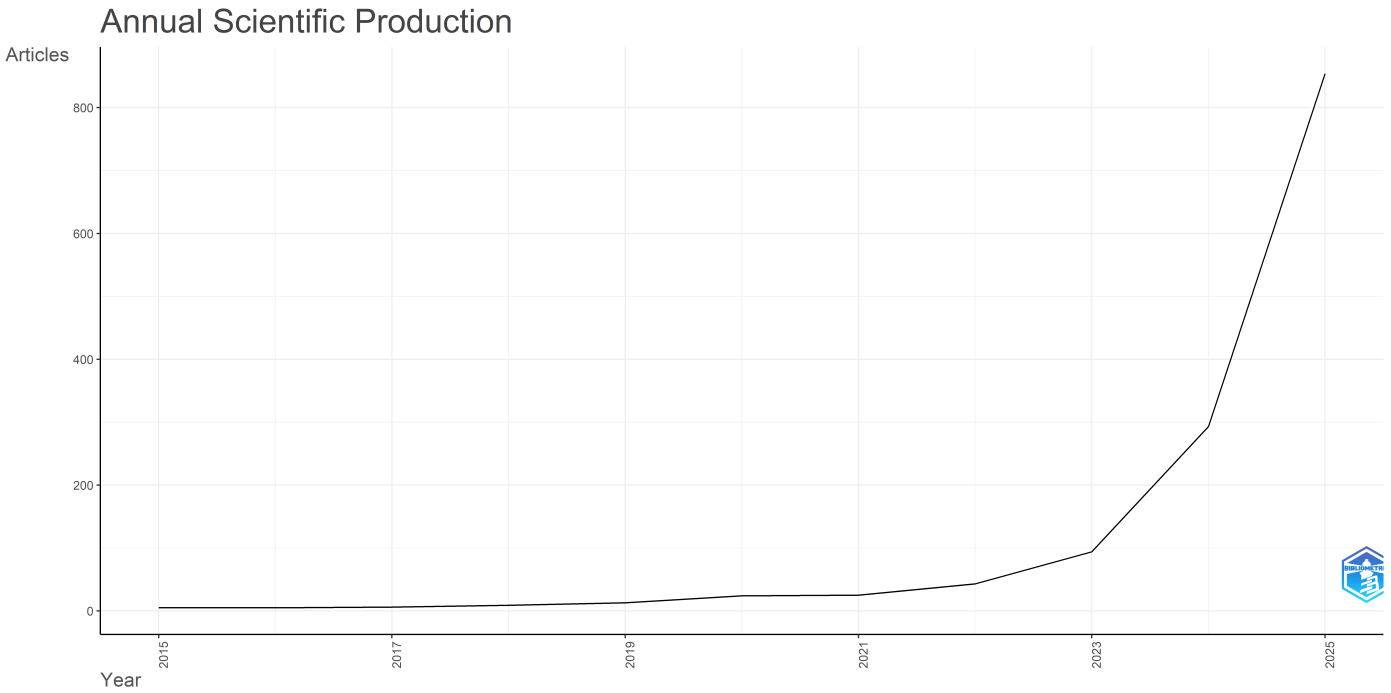

Fig. 2. Annual scientific production (2015 to 2025).

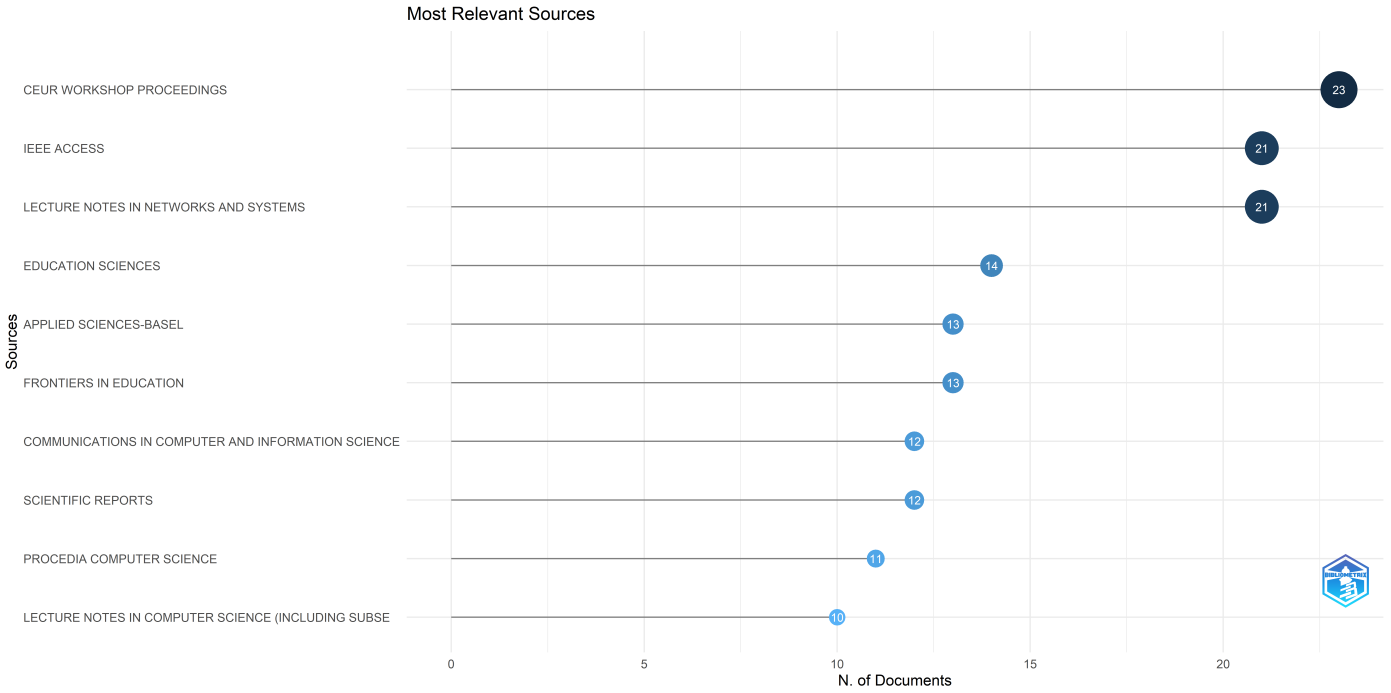

Fig. 3. Top ten most productive sources by publication count (2015 to 2025).

affiliation, are reported in Table III. The 19-document maximum is itself a low ceiling for a corpus of 1371 documents and 4167 unique authors, consistent with a highly distributed authorship pattern in which approximately 87.6% of authors contributed only a single document.

4) Most relevant keywords: The most frequent author keywords are reported in Fig. 7. Artificial intelligence leads with 335 occurrences, followed by personalized learning (213), machine learning (119), and adaptive learning and deep learning (110 each). Responsible-AI terms including explainable ai, ethics, algorithmic bias, data privacy, appear in the corpus but at substantially lower frequencies, alongside generative ai,

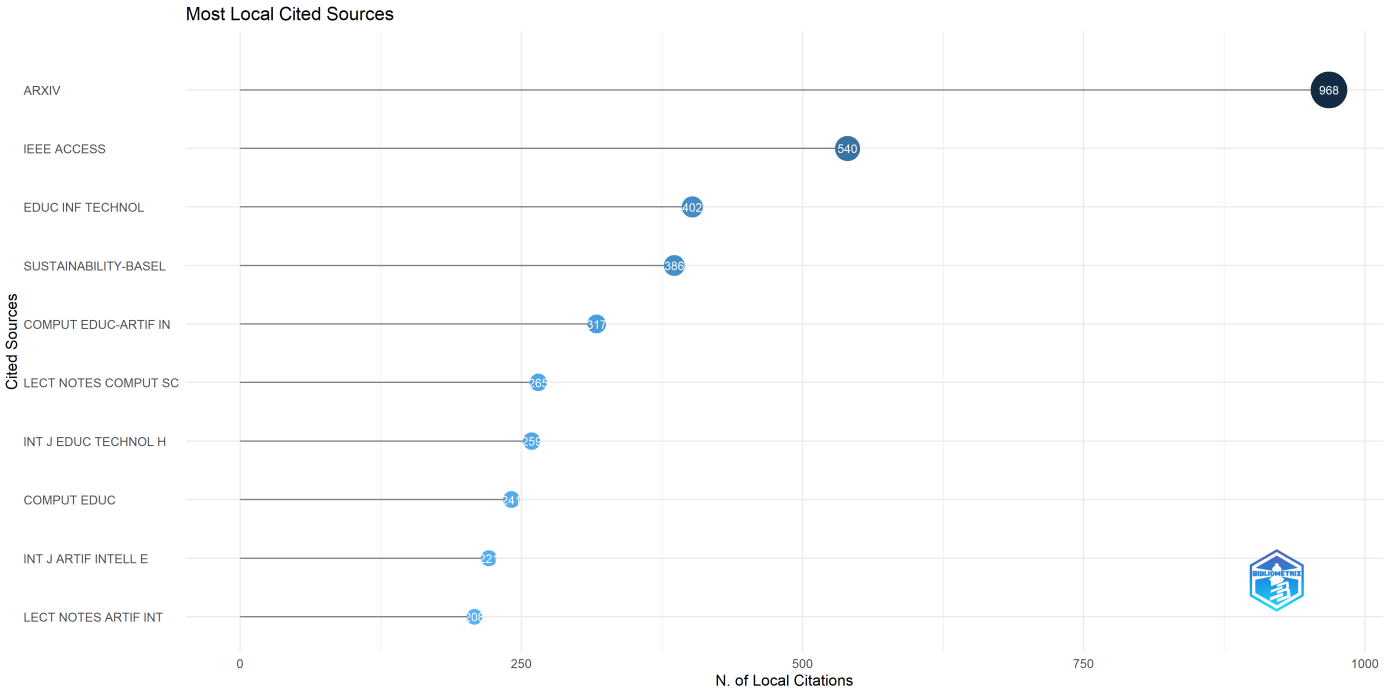

Fig. 4. Top ten most locally cited sources within the corpus, ranked by within-corpus citation count.

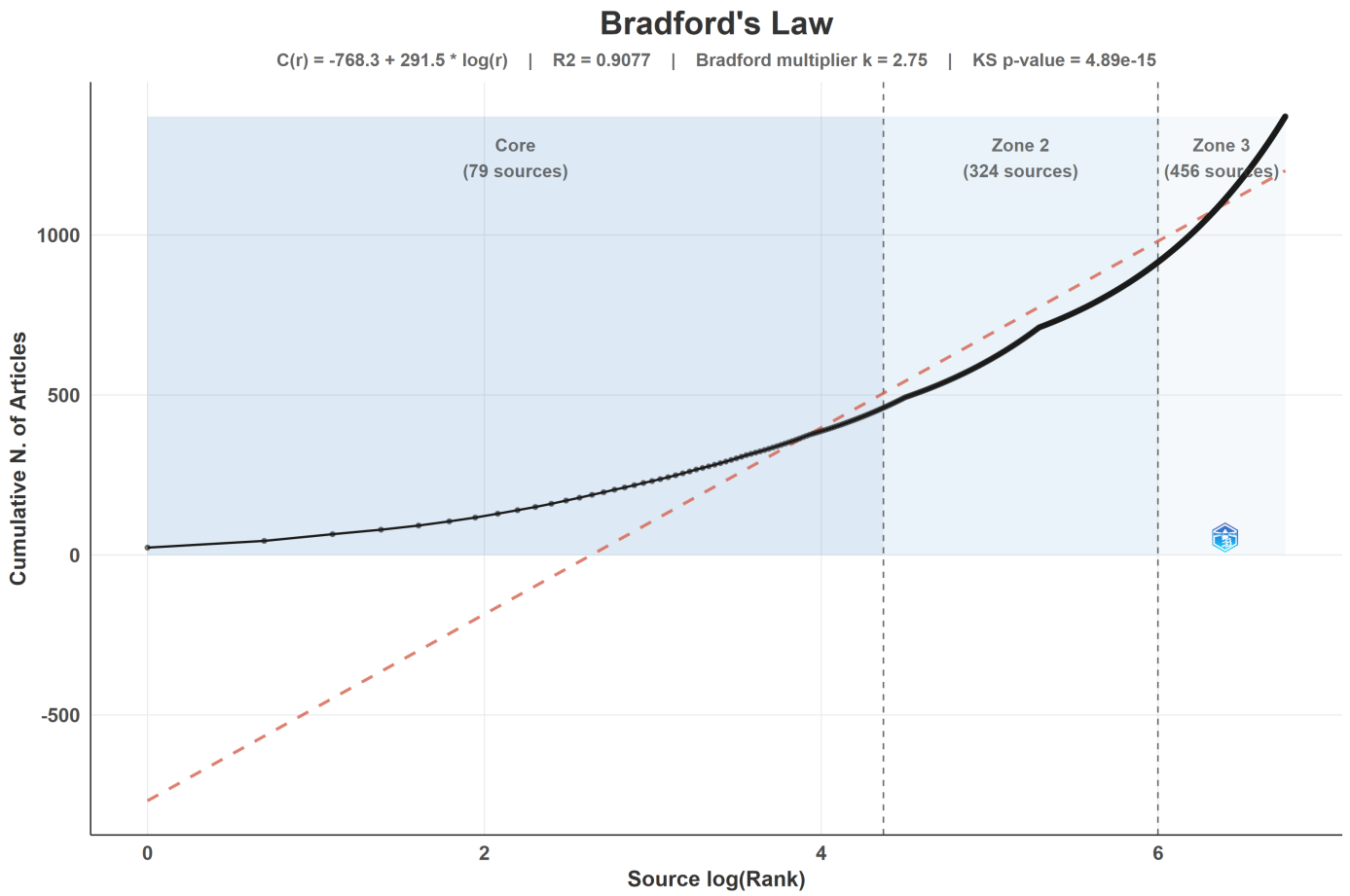

Fig. 5. Source distribution following Bradford's Law of scattering.

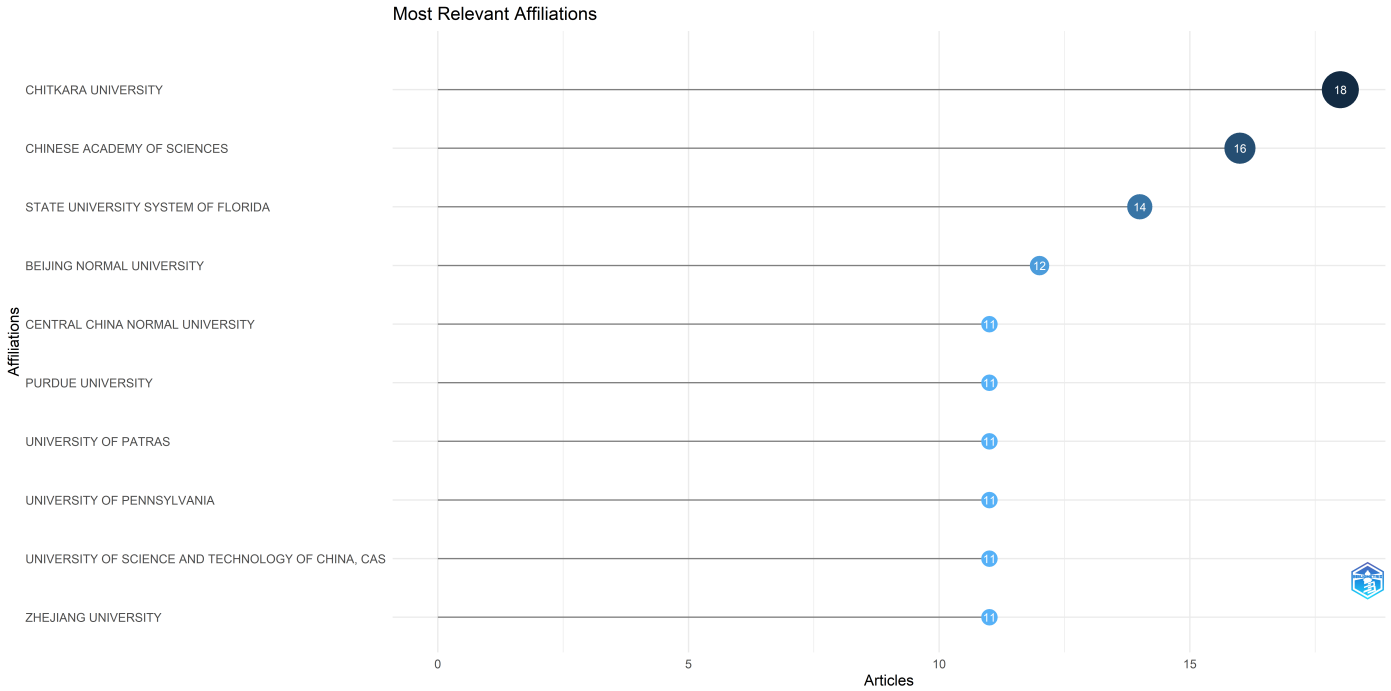

Fig. 6. Most relevant affiliations by publication count (top 10).

.

TABLE III. AUTHOR-LEVEL DATA FOR THE TEN MOST PRODUCTIVE AUTHORS IN THE CORPUS (2015 TO 2025). AUTHOR-LEVEL RESULTS ARE THE LEAST RELIABLE LAYER OF THE ANALYSIS (SEE SECTION V-B3); INSTITUTION- AND COUNTRY-LEVEL FINDINGS (SECTIONS V-B2 AND V-C3) ARE MORE ROBUST AND SHOULD BE GIVEN INTERPRETIVE PRIORITY

Rank	Author	Docs	Total citations	h-index	g-index	M-index	First pub. year	Primary affiliation (most frequent)
1	Wang J†	19	146	6	11	0.500	2015	Not consistently reported
2	Zhang Y†	18	303	6	17	0.857	2020	Wenzhou University; not consistently reported
3	Li J	15	471	5	15	0.714	2020	Aerospace Information Research Institute
4	Chen Y	14	354	7	14	1.167	2021	Binghamton University
5	Wang Z†	13	119	6	10	0.857	2020	Anhui University
6	Zhang Z	13	48	5	6	0.625	2019	Beijing University of Posts and Telecommunications
7	Zhang X†	12	18	2	4	0.333	2021	Anhui University
8	Li X†	10	120	4	10	0.500	2019	Chengdu University of Traditional Chinese Medicine
9	Wang Y	10	47	2	6	0.400	2022	Guizhou Minzu University
10	Zhang H	10	56	4	7	1.000	2023	Chang'an University

Notes: a. h-index, g-index, total citations, and M-index were computed using the bibliometrix `Hindex()` function within the 2015–2025 corpus. b. M-index = h-index ÷ years since first publication, providing a career-length-normalised impact measure. c. Authors marked with † are likely surname-collision aggregations representing more than one distinct researcher. Citation-based figures in this table are subject to the same ∼13% upwards bias noted for Table II (see Section VI-G, Sixth).

chatgpt, and large language models, whose structural position is examined in Section V-C6.

As illustrated in Fig. 7 there is a pronounced frequency gap between the top two terms (335 and 213 occurrences) and the rest of the field, where occurrence counts cluster between 70 and 120. Responsible-AI vocabulary (explainable ai, ethics, algorithmic bias) appears in the corpus, but none of these terms enter the top-ten ranking. They sit in the second tier of frequency, alongside chatgpt and generative ai. This frequency hierarchy is itself a quantitative signal of the asymmetry between technical and ethical-normative vocabulary at the level of raw term frequency *within this responsible-AI-engaged corpus* (Section III-A), although the structural integration analysis in Section V-C6 shows that the deeper picture is more nuanced.

5) Most globally cited documents: The ten most globally cited documents are reported in Fig. 8 (These citation figures carry the same ∼13% upwards bias noted in Section VI-G, Sixth). Sallam [48] leads with 1588 citations for a systematic review of ChatGPT in healthcare education, followed by Ghosh

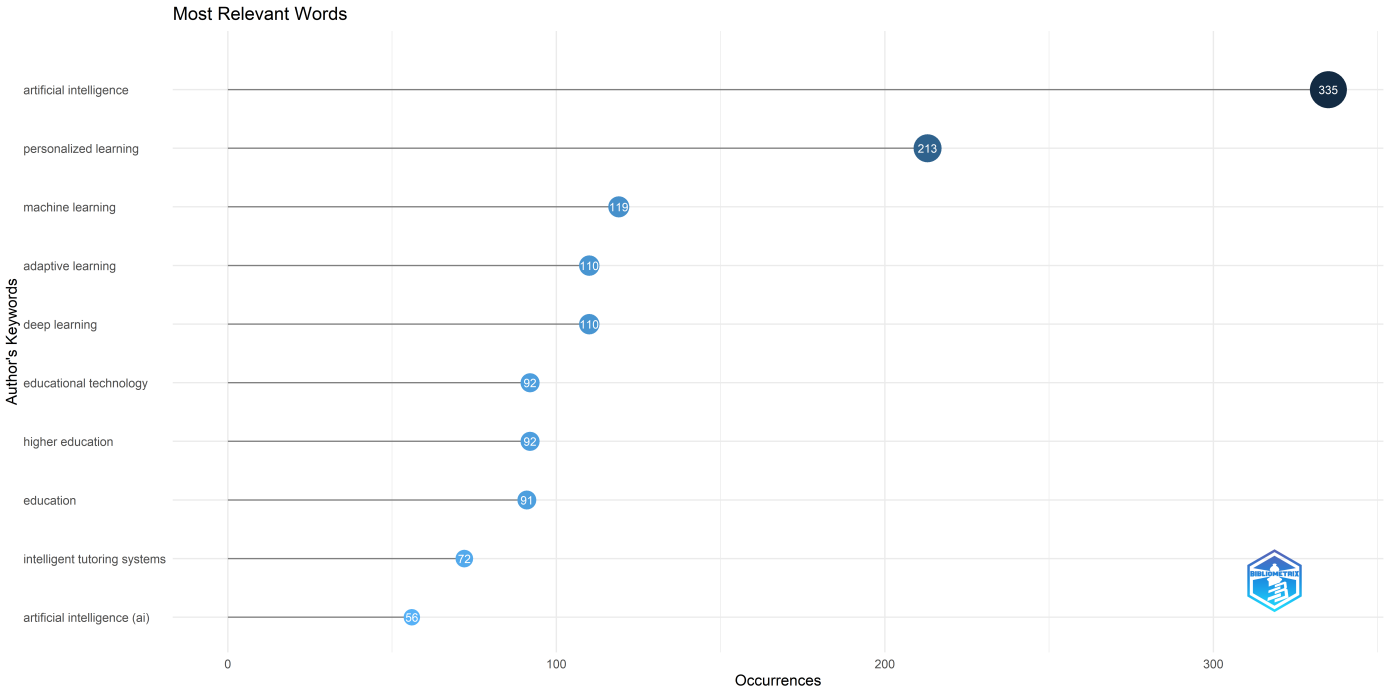

Fig. 7. Most relevant author keywords by occurrence (top 10).

et al. [6] (568, knowledge tracing), Bahroun et al. [29] (325, generative AI), and Abulibdeh et al. [49] (256, sustainability and AI). The top-cited entries span venues across geoscience, remote sensing, the Internet of Things, and general computing; however, **none of the ten most globally cited focuses primarily on algorithmic fairness, equity, or explainability in personalised learning**, a finding directly relevant to RQ5.

C. Phase 3: Science Mapping, Network and Thematic Analysis

A consolidated summary of network metrics across all five VOSviewer networks is provided in Table IV to enable comparison at a glance. The detailed structural findings for each network are then developed in turn.

1) Thematic map: The centrality-density thematic map produced using the framework of Cobo et al. [26] is reported in Fig. 9. The motor-themes quadrant (upper-right) contains no themes. The niche-themes quadrant (upper-left) contains reinforcement learning, federated learning, and blockchain. The basic-themes quadrant (lower-right) contains two groupings: 1) artificial intelligence, higher education, and education; and 2) personalised learning, adaptive learning, and educational technology. The emerging-or-declining-themes quadrant (lower-left) contains intelligent tutoring systems, personalization, and smart education. A further cluster comprising machine learning, deep learning, and explainable ai is positioned in the upper-central region of the map, near the centrality midline and at above-average density.

Fig. 10 traces thematic continuity and transformation across the two sub-periods using the inclusion-index methodology of Cobo et al. [26], with ribbon thickness representing thematic continuity between periods. Several themes that were prominent in the early period, including trust, intelligent tutoring systems, neural networks, and e-learning, fragment or fade in the later period, while deep learning, reinforcement learning, and artificial intelligence consolidate as dominant themes. New themes appearing in 2025 include generative ai, data privacy, and systematic review. Explainable AI, present as a distinct theme in 2015 to 2021, does not persist as a standalone cluster into the 2022 to 2025 period. This thematic-evolution pattern is consistent with the diffusion of the term into adjacent vocabularies rather than its consolidation as a dedicated research stream.

2) Co-authorship network at the author level: The author-level co-authorship network, constructed at a minimum of three publications per author, is reported in Fig. 11. Two connected dyads are visible: Tairi and Jeghal, and Ogata and Flanagan. Each dyad is linked by a single co-authorship edge. The remaining visible nodes, including Uglev Viktor, Munjal Kalpana, Saddhono Kundharu, Elsayary Areej, Embarak Ossama, and Bala B. Kiran, appear as isolated individuals with no connecting edges. No edges link the two dyads to one another. This is a notable null result. In a corpus of 1371 documents authored by 4167 researchers, only ten authors meet the three-publication threshold, and only four of these ten participate in any co-authorship edge.

Network statistics for this graph are: nodes = 10, edges = 2, modularity Q = 0.50 (reflecting the trivial multi-component structure), average degree = 0.40, density = 0.044. At a relaxed threshold of two publications per author, the network expands but remains extremely sparse, with overall density well below 0.05. This is a low value for a domain of this size and growth rate. As illustrated in Fig. 11, this is a graph of essentially disconnected components. The visual sparsity is consistent with the broader productivity pattern, in which approximately 87.6% of the 4167 authors contributed only one document and

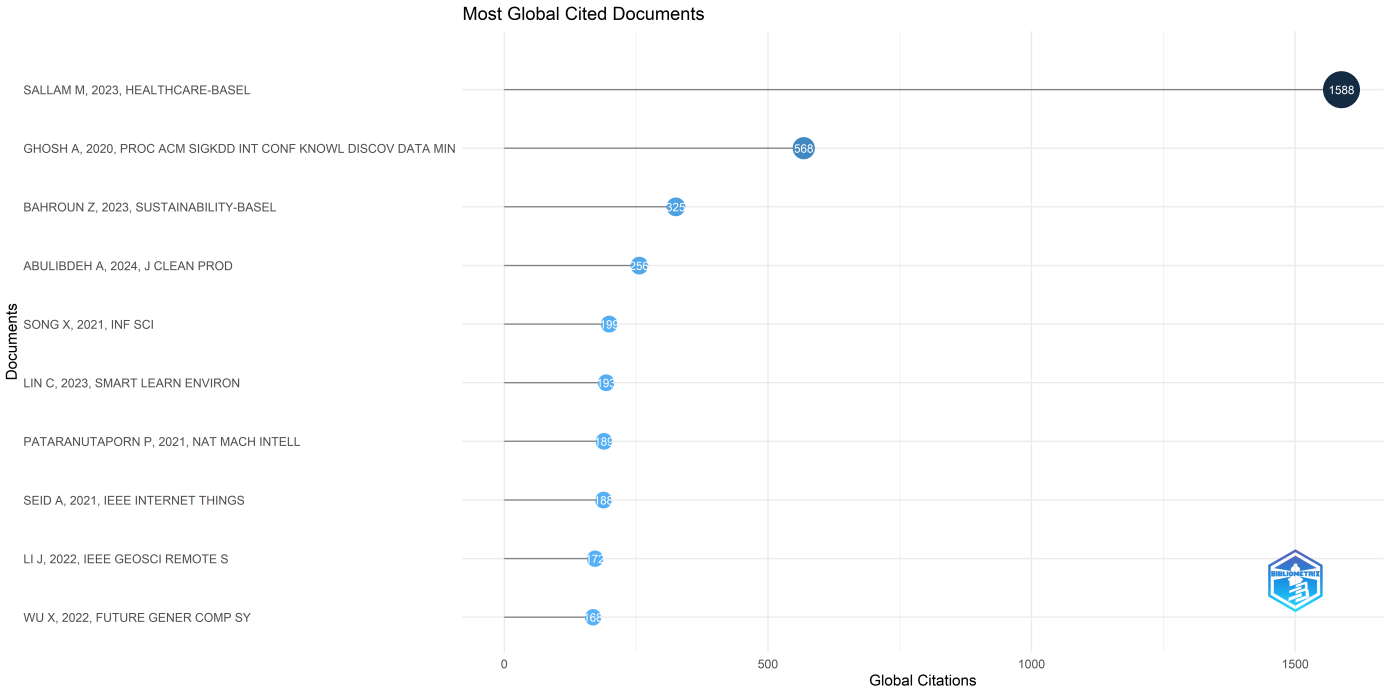

Fig. 8. Top 10 most globally cited documents in the corpus.

TABLE IV. NETWORK METRICS ACROSS THE FIVE VOSVIEWER NETWORKS AT MINIMUM-FIVE THRESHOLDS, REPORTING NODES, EDGES, MODULARITY Q, AVERAGE DEGREE, DENSITY, AND CLUSTER COUNT

Network	Nodes	Edges	Modularity Q	Avg degree	Density	Clusters	Notes
Author co-authorship (min 3)	10	2	0.500	0.40	0.044	8	Two dyads + six isolated nodes (counted as singleton clusters); sparse
Country co-authorship (min 3)	39	134	0.275	6.87	0.181	5	Top betweenness: Indonesia (0.180), USA (0.176), India (0.165)
Co-citation of references (min 5)	37	139	0.346	7.51	0.209	3	Ethics-equity references not cluster-forming
Co-citation of journals (min 5)	559	21350	0.340	76.39	0.137	7	*AI Ethics* peripheral
Keyword co-occurrence (min 5)	88	902	0.164	20.50	0.236	4	Ethics-equity-to-technical inter-cluster density 0.114

Notes: a. Modularity Q and cluster counts in Table IV are computed via the Louvain implementation in igraph applied to the Pajek exports. VOSviewer's smart local moving algorithm produces slightly different cluster counts on the same networks (e.g., 4 clusters for co-citation references, 7 for keyword co-occurrence at threshold 5), reflecting differences in optimisation procedures. VOSviewer cluster assignments are used as the basis for the substantive cluster-content analysis in Sections V-C4–V-C6, while the Louvain Q values reported here provide an independent quantitative validation of modular structure. b. The author co-authorship network's "8 clusters" comprises two connected dyads and six isolated nodes treated as singleton clusters. c. Density refers to overall graph density.

the maximum observed output is 19 documents — a heavily right-skewed authorship distribution typical of young, rapidly growing interdisciplinary fields [50]. This is a field in which most contributors publish once or twice and then leave, and no dense recurring author-collaboration backbone has yet formed. The implications of this finding are addressed in Section VI-C.

3) Co-authorship network at the country level: The country-level co-authorship network, constructed at a minimum of three publications per country, is reported in Fig. 12. India displays the largest node size, with China occupying the central position in the network and the United States anchoring a third major hub. India anchors a cluster extending across South Asia, the Gulf, and parts of North and Sub-Saharan Africa. A European cluster comprises Germany, Italy, the Netherlands, and other Western European nodes; the United Kingdom bridges this cluster to the Asian sub-clusters. Australia and Japan maintain connections to both communities. South Africa is the only Sub-Saharan African node visible, and Morocco is the only North African node visible; no Latin American nodes appear at the three-publication threshold.

Country-network statistics are: nodes = 39, edges = 134, modularity Q = 0.275, average degree = 6.87, density = 0.181. The top five countries by betweenness centrality are Indonesia (0.180), USA (0.176), India (0.165), Portugal (0.089), and Kazakhstan (0.080). The relatively low modularity score indicates that the country network is more interconnected than fragmented at the cluster level. The geographic distribution of

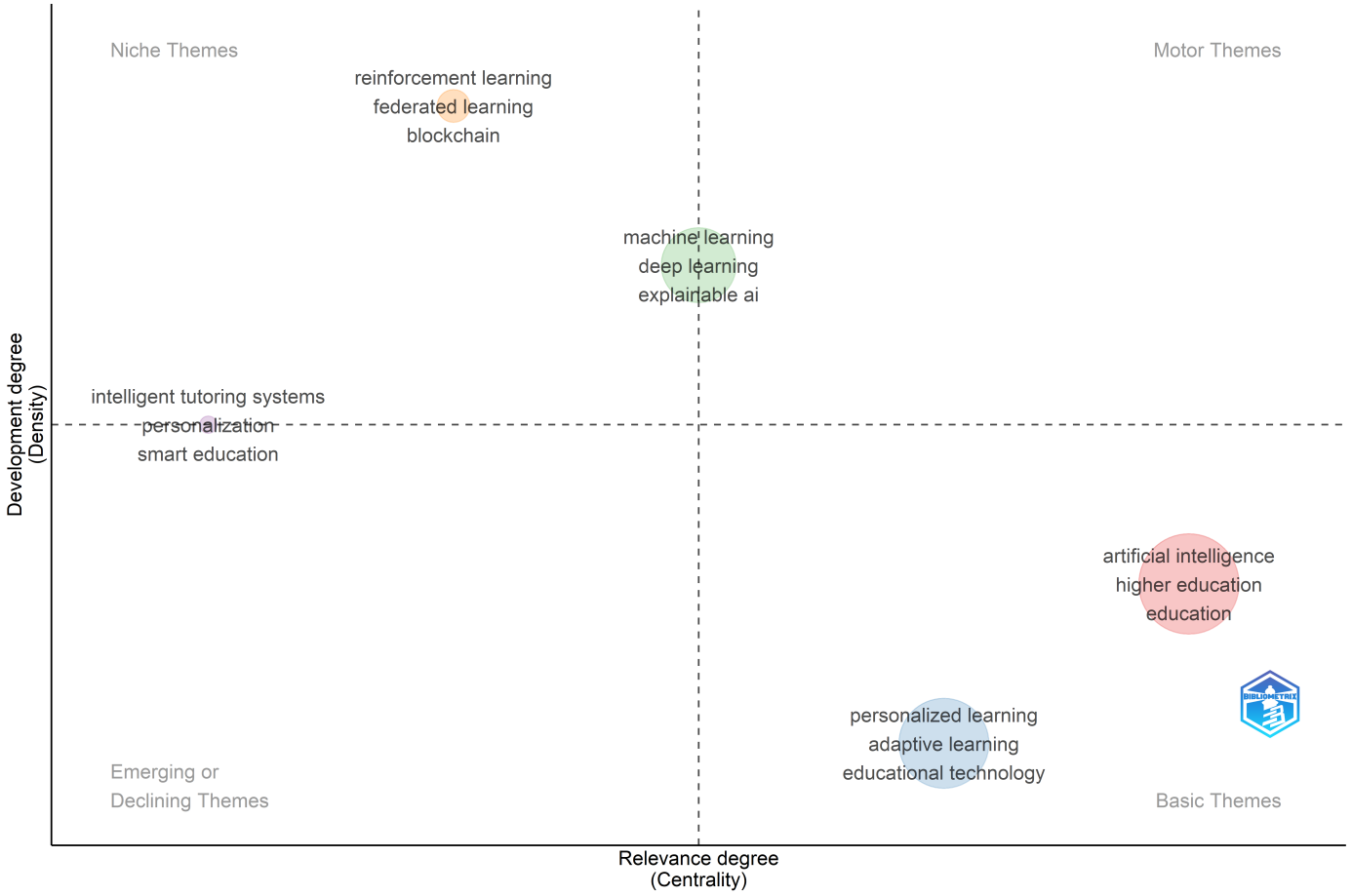

Fig. 9. Thematic map of the 1371-document corpus showing clusters positioned by centrality (x-axis) and density (y-axis) across the four classical Cobo quadrants.

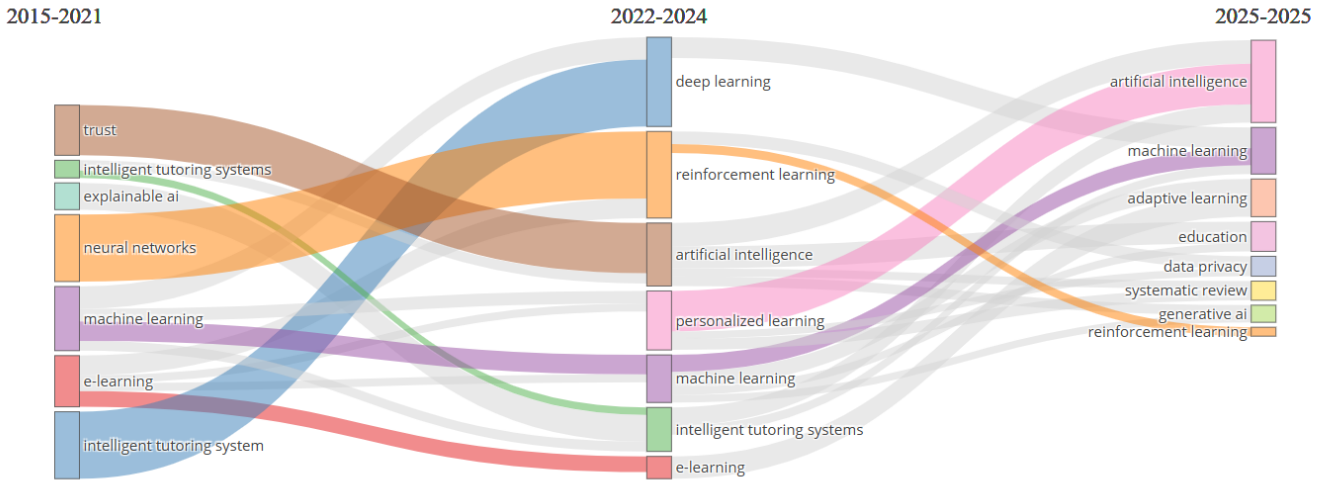

Fig. 10. Thematic evolution across three periods (2015–2021, 2022–2024, 2025).

nodes, however, and the near-absence of Sub-Saharan African and Latin American countries, produces a structurally uneven map that is interpreted in detail in Section VI-D.

The geographic distribution of absolute output is reported in Fig. 13. Shading identifies high-output countries (China, USA, India, UK, Australia, Saudi Arabia, parts of continental

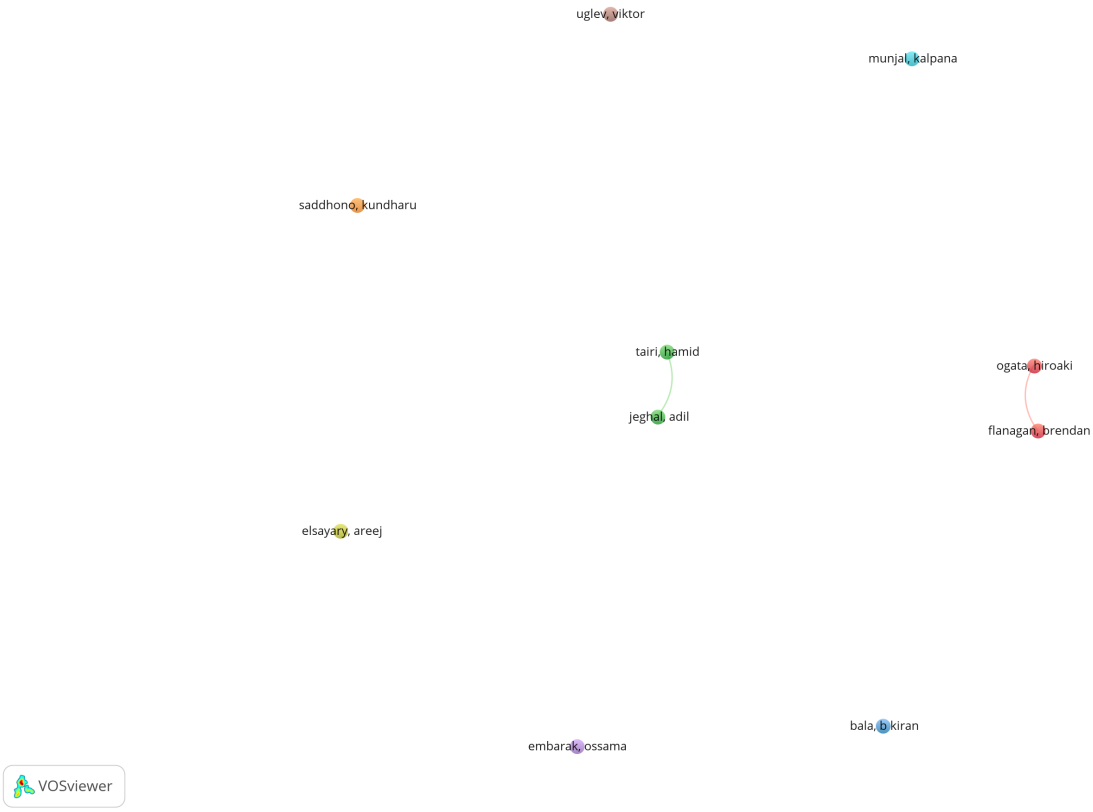

Fig. 11. Co-authorship network at the author level (minimum 3 publications per author).

Europe), while Sub-Saharan Africa, Central Asia, and much of Latin America remain unshaded or lightly shaded. The top-50 collaboration arcs reinforce this asymmetry: dense link bundles connect North America, Western Europe, China, India, and Australia, while incoming arcs to Sub-Saharan Africa and Latin America are sparse.

4) Co-citation network of cited references: The co-citation network of cited references, constructed at a minimum of five co-citations, is reported in Fig. 14. The largest and most central node, indexed in VOSviewer as 'Lin Z., 2020, IEEE Access' due to last-author label conventions, corresponds to Chen, Chen and Lin [51], the most frequently co-cited reference within the corpus documents. Four substantively interpretable clusters emerge from the VOSviewer partition of the network. The Louvain implementation in igraph applied to the same network yields modularity Q = 0.346 and three clusters, with two of the four VOSviewer clusters merging in the Louvain solution. Substantive analysis below uses the VOSviewer four-cluster partition. The first and largest cluster (n = 13) anchors the foundational AIED and explainability canon, organised around Chen, Chen and Lin [51] (the most central node, indexed in VOSviewer as 'Lin Z., 2020' due to last-author label conventions), the *Intelligence Unleashed* lineage (Luckin et al. [24]), Gašević et al. on explainable AI in education, VanLehn [9] on tutoring effectiveness, Sohl-Dickstein on deep knowledge tracing [4], and Berrada's XAI survey, alongside critiques such as Selwyn on AI–teacher substitution and Shawe-Taylor on educational inequality. The second cluster (n = 9) comprises broad-coverage AIED literature reviews (Du, Bilyatdinova, Cheng, Franzsen) together with the deep-learning methodological backbone (Hinton's *Deep Learning*, Schmidhuber's LSTM, Barto's reinforcement-learning textbook) and Clarke's thematic-analysis methodology paper. The third cluster (n = 6) captures adaptive-learning and generative-AI higher-education reviews, including Gligorea et al. [12] (indexed in VOSviewer as 'Tudorache P.' due to last-author labelling), Zawacki-Richter et al. [1] (indexed as 'Gouverneur F.'), Kasneci et al. [8] on ChatGPT, and Bahroun et al. [29] on generative AI in education. The fourth cluster (n = 9) groups applied higher-education and personalised-learning systems, including Akgun and Greenhow [19] on K-12 ethics (the most central node in this cluster), Jiao on online higher-education AI, Lee on classroom chatbots, Kamalov et al. [7] on the sustainable AI-in-education revolution, Human-Hendricks on machine-learning-based personalised adaptive learning, and Nikolopoulou on generative AI in higher education.

Crucially, **Baker and Hawn [15], Khosravi et al. [14], and Holmes et al. [20] do not appear as cluster-forming nodes in the network**. This indicates that the foundational responsible-AI references in education are not part of the field's structurally central co-citation core.

Co-citation network statistics are: nodes = 37, edges = 139, modularity Q = 0.346 (Louvain) across the VOSviewer-partitioned four clusters described above, average degree =

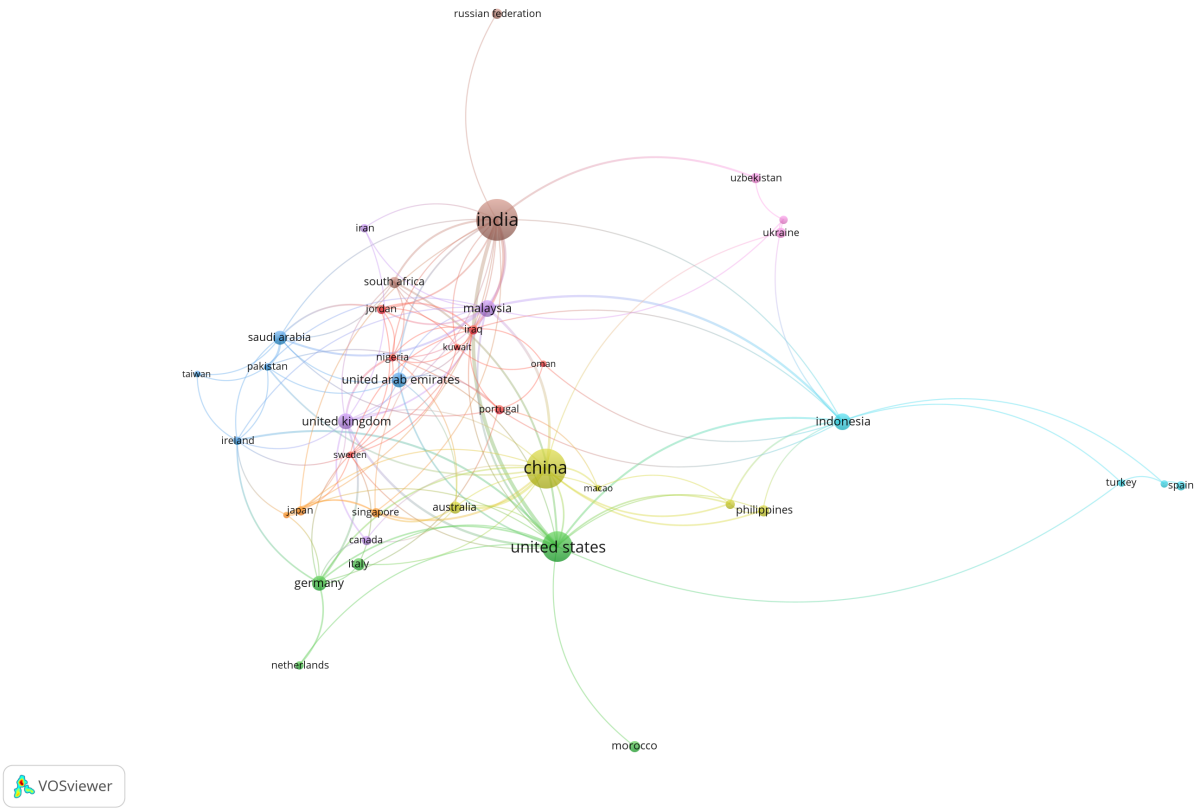

Fig. 12. Country co-authorship network (minimum 3 publications per country).

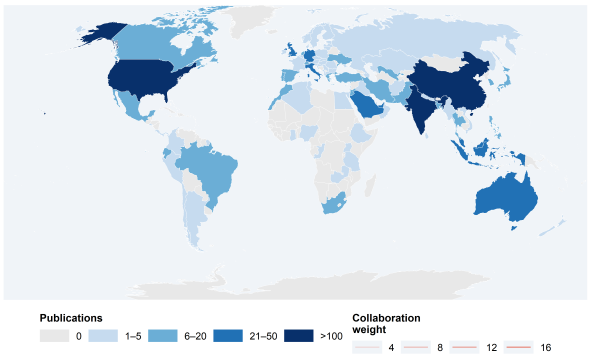

Fig. 13. Geographic distribution of research output on AI-based personalised learning, 2015 to 2025.

7.51, density = 0.209. The top five cited references by betweenness centrality are Chen et al. [51] (0.266; indexed as 'Lin Z., 2020'), Gligorea et al. [12] (0.110; indexed as 'Tudorache P.'), Du et al. (0.078), Hinton et al. (0.058), and Gašević et al. (0.057). As visualised in Fig. 14, these nodes form four connected hubs across the network: Chen et al. anchors the foundational AIED canon (Cluster 1); Hinton et al. and Gašević et al. bridge the deep-learning methodological backbone (Cluster 2) into the foundational canon; Gligorea et al. and Du et al. occupy bridging positions between the adaptive-learning review cluster (Cluster 3) and the applied higher-education cluster (Cluster 4). The absence of foundational responsible-AI references from the structural centre of this network anchors the citation-infrastructure argument developed in Section VI-B.

5) Co-citation network of cited journals: The journal-level co-citation network is reported in Fig. 15. IEEE Access and Sustainability display the largest node sizes and occupy the most prominent positions. Four substantively interpretable cluster groupings emerge: an education-and-applied-sciences cluster (Education Sciences, Sustainability, Frontiers in Education); a general science and computing cluster (IEEE Access, Scientific Reports, PLOS One, Nature); an arXiv-anchored preprint and open-access education cluster (including Smart Learning Environments); and a biomedical cluster (NPJ Digital Medicine, JAMA, Medical Image Analysis), reflecting the strong representation of healthcare-education applications in the corpus. A small AI Ethics node sits in a peripheral cluster alongside computer-security venues. No specialist journal dedicated to algorithmic fairness or explainability in education appears as a cluster-forming node.

Journal co-citation network statistics are: nodes = 559, edges = 21350, modularity Q = 0.340 across seven clusters, average degree = 76.39, density = 0.137. AI Ethics registers as a peripheral node with low total link strength relative to the cluster-forming journals. The peripheral position of AI Ethics and the absence of any specialist responsible-AI-in-education

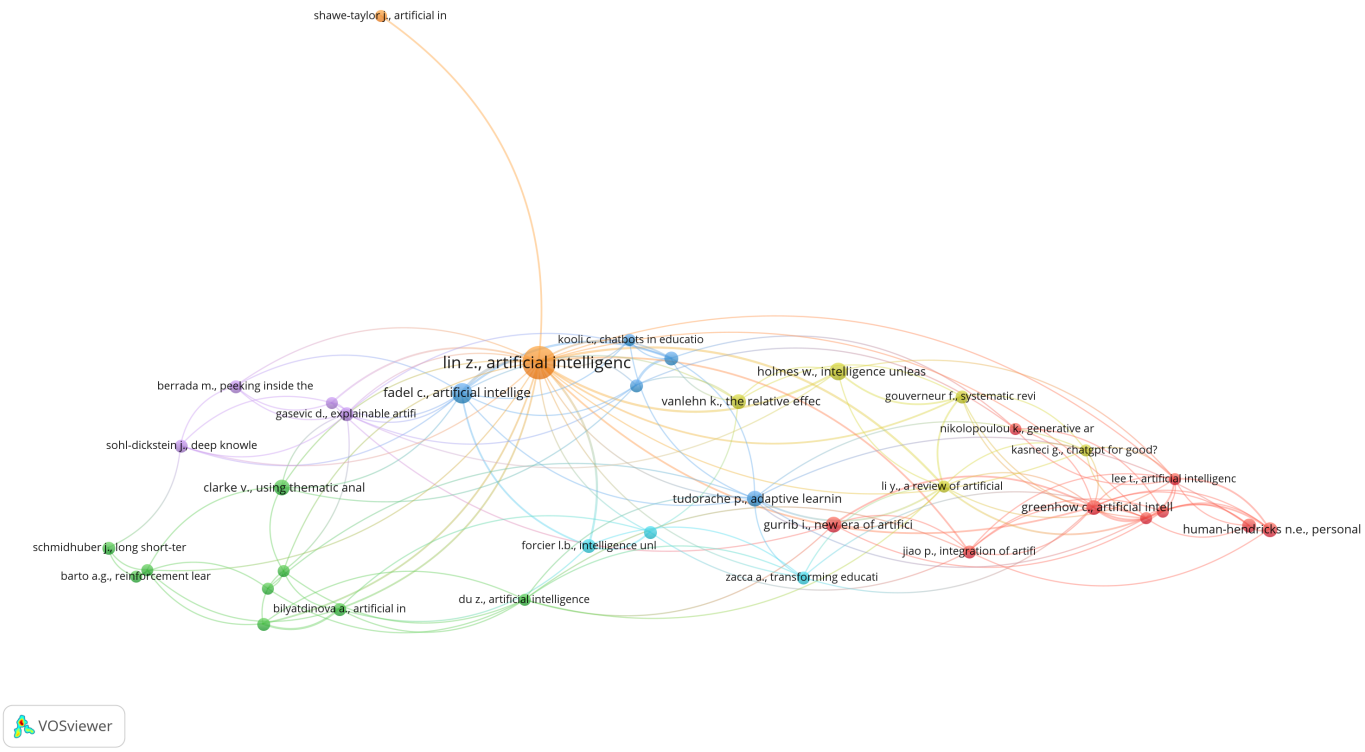

Fig. 14. Co-citation network of cited references (minimum five co-citations).

venue from the cluster cores reinforce the venue-infrastructure argument developed in Sections VI-C and VI-E.

6) Keyword co-occurrence network: The keyword co-occurrence network, constructed at a minimum of five co-occurrences, is reported in Fig. 16. The VOSviewer clustering algorithm partitions the network into seven clusters of varying size. The three largest substantive clusters (Clusters 1, 2, and 3; n = 20, 18, 18) carry the bulk of the network's analytical weight and are the basis for the intra- and inter-cluster density analysis below. Cluster 4 (n = 14) is anchored on indexing-variant forms of central terms and is treated as a vocabulary-fragmentation cluster rather than a substantive thematic grouping. The remaining three clusters (5, 6, 7) extend the fragmentation pattern visible in Cluster 4 and are described at the end of the list.

a) Cluster 1, Ethics, equity, and inclusivity vocabulary (20 nodes): ai ethics, equity, fairness, inclusivity, ethical ai, ethics, transparency, trust, inclusive education, ai literacy, academic integrity, generative ai, sustainability, personalised learning (UK spelling), and related terms.

b) Cluster 2, Applied AI in education with bias and privacy concerns (18 nodes): artificial intelligence, personalized learning (US spelling), algorithmic bias, data privacy, digital divide, educational equity, ethical concerns, k-12 education, personalized education, predictive analytics, and related terms.

c) Cluster 3, Technical methods and analytics (18 nodes): machine learning, deep learning, neural networks, adaptive learning, learning analytics, federated learning, blockchain, lstm, explainable artificial intelligence (xai), and related terms.

d) Cluster 4, Vocabulary fragmentation (14 nodes): ai, artificial intelligence (ai), explainable artificial intelligence, xai, chatgpt, large language models, big data, data mining, education, higher education, learning, teaching, technology, systematic review. This cluster is anchored on indexing-variant forms of foundational terms (ai/artificial intelligence and xai/explainable artificial intelligence) alongside generic concept-level terms.

e) Cluster 5, Applied and explainable AI methods (9 nodes): accessibility, artificial intelligence in education, chatbots, digital transformation, engineering education, explainable ai, lime, natural language processing, and reinforcement learning.

f) Cluster 6, Tutoring systems and emerging AI (8 nodes): affective computing, generative artificial intelligence, intelligent tutoring system, intelligent tutoring systems, knowledge graph, knowledge tracing, personalization, and stem education.

g) Cluster 7, Sustainable education (1 node): sustainable education.

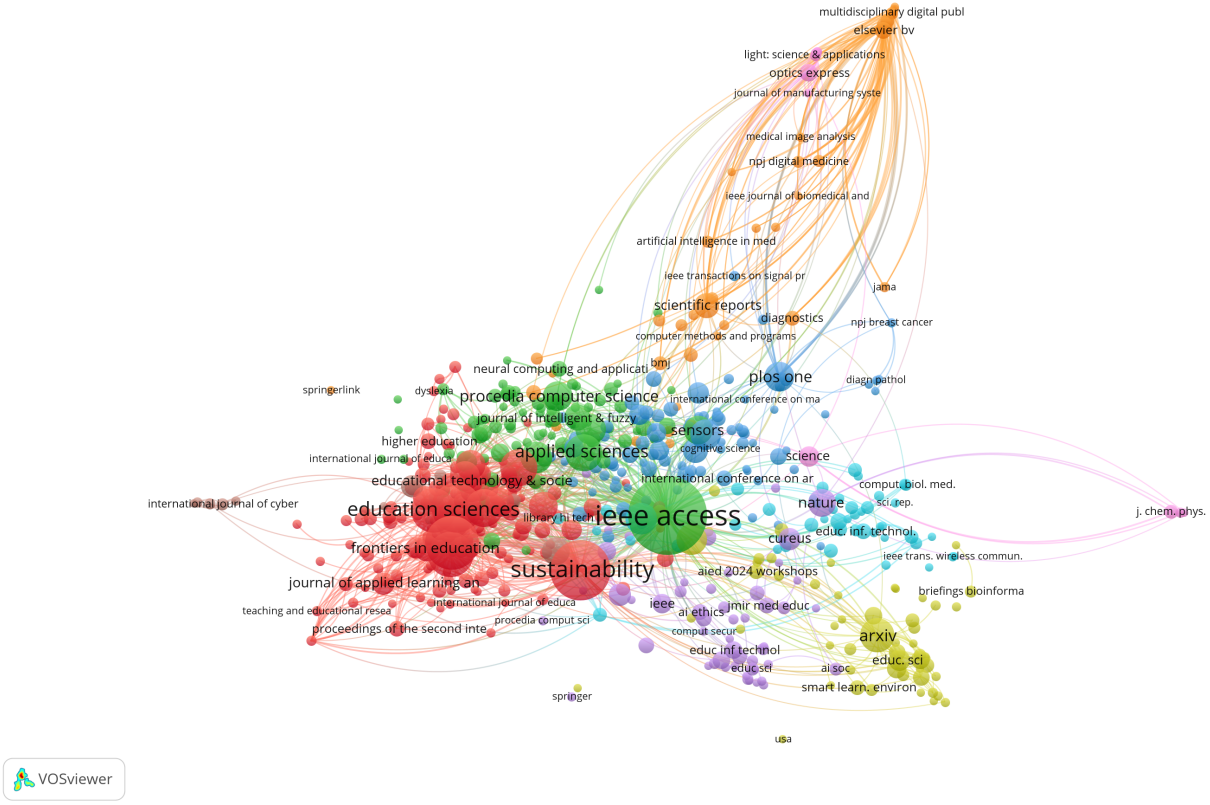

Fig. 15. Co-citation network of cited journals (minimum five co-citations).

Network statistics are: nodes = 88, edges = 902, modularity Q = 0.164 (Louvain) / 0.176 (VOSviewer smart local moving algorithm, seven clusters), average degree = 20.50, density = 0.236. Substantive cluster analysis below uses the VOSviewer seven-cluster partition.

Intra-cluster density and inter-cluster density were computed for the three largest clusters. The intra-cluster density of the ethics-equity cluster (C1) is 0.37, of the applied-AI cluster (C2) is 0.55, and of the technical-methods cluster (C3) is 0.44. The ethics-equity cluster is therefore internally less cohesive than either the applied-AI or technical-methods clusters, consistent with the dispersion of ethics-and-equity vocabulary across multiple cluster boundaries. Inter-cluster densities show that the ethics-equity and technical clusters connect to the applied-AI cluster at comparable rates (C1↔C2 = 0.20; C2↔C3 = 0.21), but the ethics-equity and technical clusters connect to each other directly at approximately half that density (C1↔C3 = 0.11). The ratio of C1↔C2 to C2↔C3 density is 0.95, indicating essentially equal integration of both streams with the applied-AI cluster, while the direct ethics-equity-to-technical channel remains substantially weaker.

Betweenness centrality identifies the keywords that mediate connections across clusters. The top-five bridge nodes in the network are adaptive learning (0.060, Cluster 3), higher education (0.058, Cluster 4), deep learning (0.050, Cluster 3), generative ai (0.048, Cluster 1), and machine learning (0.045, Cluster 3). Four of the five top bridge nodes belong to the technical-methods or indexing-variant clusters; only generative ai belongs to the ethics-equity cluster. No core ethics-equity term (fairness, equity, ai ethics, trust, inclusivity) appears in the top-five betweenness ranking, reinforcing the finding that the ethics-equity cluster connects to the rest of the network primarily through mediation rather than serving as a bridge in its own right.

This pattern indicates that ethics-equity vocabulary is not peripheral to the field's everyday discourse. Its connection to the applied-AI cluster (0.20) is comparable to the connection between the technical and applied-AI clusters (0.21). The structurally distinctive feature is instead the weak direct cross-talk between the ethics-equity and technical streams, combined with the lower internal cohesion of the ethics-equity cluster itself. The two streams engage with each other primarily through the mediation of the applied-AI cluster rather than directly. This finding provides the empirical foundation for the mediated disconnection argument developed in Section VI-E and for the framing of Proposition 2 of the Structural Integration Model.

To assess the robustness of the mediated disconnection pattern, we performed a non-parametric bootstrap (1000 edge-resampling iterations, set.seed = 42) on the keyword network. The bootstrap distribution preserves the rank ordering and approximately halves the C1–C3 density relative to C2↔C3: the 95% confidence interval for C1↔C3 (ethics-equity ↔technical) is [0.053, 0.086], while the 95% CI

for C2↔C3 (applied-AI ↔technical) is [0.111, 0.157]. The two intervals do not overlap, confirming that the ethics-equity-to-technical channel is statistically distinguishable from the applied-AI-to-technical channel under edge-level perturbation. The corresponding 95% CI for C1↔C2 (ethics-equity ↔applied-AI) is [0.108, 0.150], statistically indistinguishable from C2↔C3 and consistent with the comparable-integration claim above. Bootstrap point estimates are systematically lower than the observed densities (a known property of edge-resampling), but ratios and ordering are preserved.

Three patterns of vocabulary fragmentation are visible across the cluster structure. First, personalized learning (US spelling) anchors Cluster 2, while personalised learning (UK spelling) anchors Cluster 1, splitting the field's defining concept across two cluster boundaries by orthography alone. Second, the field's most central explainability term appears in four different surface forms across three clusters: explainable artificial intelligence (xai) and explainable ai (xai) in Cluster 3, explainable artificial intelligence and xai in Cluster 4, and explainable ai in Cluster 5. Third, artificial intelligence (Cluster 2), artificial intelligence (ai) (Cluster 4), ai (Cluster 4), artificial intelligence in education (Cluster 5), and generative artificial intelligence (Cluster 6) likewise distribute the field's foundational term across cluster boundaries. These patterns indicate that vocabulary consolidation in the field has not yet stabilised, and that the structural integration analysis above must be read in light of this fragmentation.

The sensitivity analysis shows that cluster count remains within five to six and modularity remains within $Q \in [0.116, 0.192]$ across all four thresholds. These values fall below the $Q \approx 0.3$ threshold that Newman [35] suggests for "significant community structure," but two methodological considerations qualify this reading. First, dense co-occurrence networks are known to exhibit lower modularity than sparser social or citation networks because the resolution-limit problem [52] suppresses Q in highly connected graphs. This constraint motivated Waltman, van Eck, and Noyons [28] to introduce the resolution-parameter variant of modularity used by VOSviewer's smart local moving algorithm [47]. Second, low Q in an interdisciplinary research field is not evidence of analytical failure; it is itself a substantive finding indicating that the field's keyword space is more interconnected than fragmented and does not separate cleanly into highly modular communities at any reasonable threshold. This structural property, in which vocabulary diffuses across cluster boundaries rather than consolidating within them, constitutes the empirical pattern interpreted by the Structural Integration Model in Section VI-E.

VI. Discussion

The findings reported in Section V reveal a more nuanced picture than simple peripherality: technical advancement has outpaced the infrastructural consolidation of responsible AI, even as responsible-AI vocabulary has diffused into the field's keyword space at densities comparable to those of technical vocabulary. The discussion below addresses each finding in turn, engages with rival interpretations, and advances the Structural Integration Model.

A. Growth Outpaces Infrastructural Integration

The growth trajectory documented in Section V-A signals a structural transformation of the research area rather than incremental accumulation. The distinction drawn by Roll and Wylie [2] between evolutionary and revolutionary change in AI in education offers a useful reading of this pattern. The observed expansion is predominantly evolutionary in their sense. It extends the existing technical capability set anchored on knowledge tracing [3], [4], [10] and transformer-based modelling [5], and it broadens the application surface of adaptive systems [11], [12]. It is not revolutionary in the sense of embedding AI within diverse socio-cultural learning contexts along the lines proposed by Luckin et al. [24].

Recent reviews [48], [29], [53], [49], [31], [32] map this expanding application surface across healthcare, generative AI, intelligent tutoring, sustainability, and higher education. The post-2022 acceleration is plausibly attributable to generative AI and large language models, consistent with Kasneci et al. [8] on ChatGPT's research-reshaping influence and the sustainable-revolution framing of Kamalov et al. [7].

Normative dimensions are present in this expansion at the level of vocabulary, as the keyword co-occurrence analysis confirms. They have not yet consolidated, however, at the level of citation, venue, or collaboration infrastructure. This pattern extends Hinojo-Lucena et al. [30]. Technical maturation and infrastructural integration of responsible-AI scholarship now appear to be advancing at visibly different rates, with the integration deficit converging on the diagnoses already independently advanced by Bond et al. [32] and the AI4People framework of Floridi et al. [22].

B. The Citation Canon Reproduces a Technical Orientation

The co-citation analysis (Section V-C4) shows a foundational AIED canon organised around Chen, Chen and Lin [51], the *Intelligence Unleashed* lineage (Luckin et al. [24]), VanLehn [9] on tutoring effectiveness, and Sohl-Dickstein's deep knowledge tracing [4], together with a deep-learning methodological backbone comprising Hinton's *Deep Learning*, Schmidhuber's LSTM, and Barto's reinforcement-learning textbook. The broader foundational lineage of Corbett and Anderson [3], Zhang et al. [10], Vaswani et al. [5], and Ghosh et al. [6] is intellectually central to the field but does not appear as cluster-forming nodes at the five-citation threshold, indicating that these methodological foundations are diffused across many citing documents rather than concentrated as repeatedly co-cited pairs. Works addressing explainability, fairness, and responsible AI (Baker and Hawn [15], Khosravi et al. [14], Holstein and Doroudi [16], Holmes et al. [20], Mehrabi et al. [17], and Selbst et al. [18]) are likewise present in the corpus but do not register as cluster-forming nodes. The broader explainable-AI reference of Barredo Arrieta et al. [13], which establishes transparency as a prerequisite for trustworthy deployment, is similarly absent from the co-citation core. This pattern is consistent with transparency not yet being treated as a general system requirement in the field.

Citation structures shape the entry points through which new researchers encounter a domain and the reference set they are likely to adopt [25], [27]. A canon dominated by technical works therefore tends to reinforce a technical orientation and

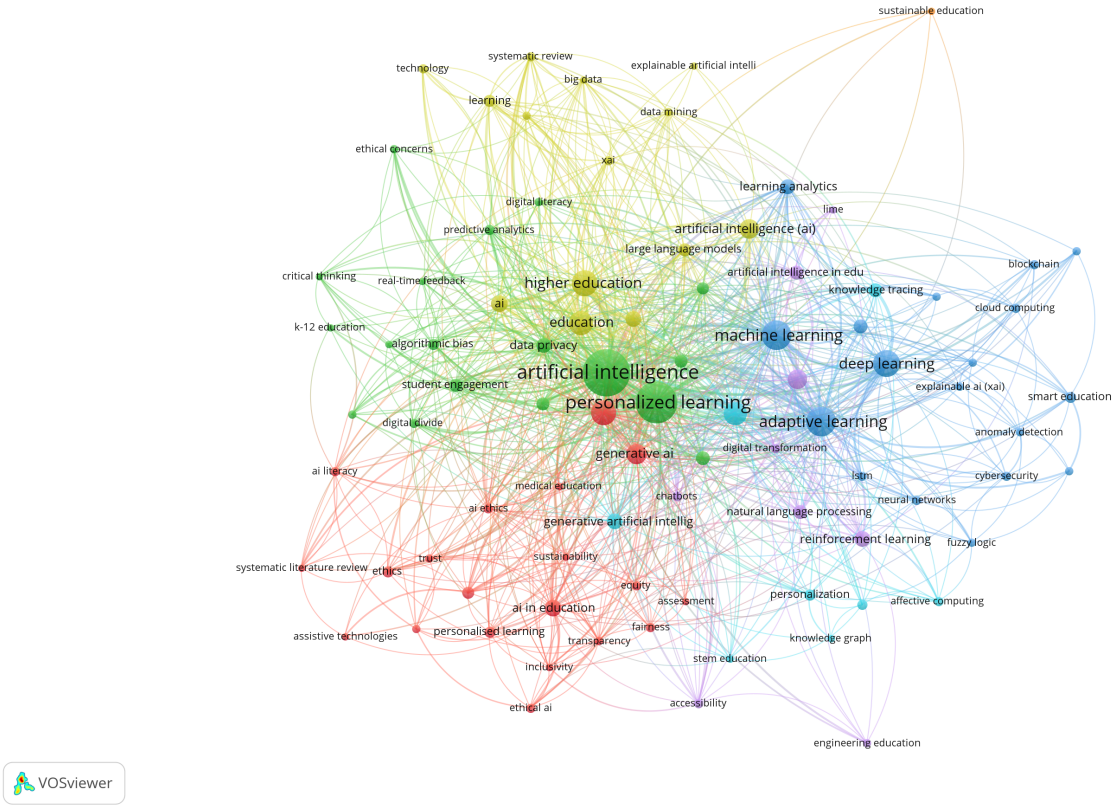

Fig. 16. Keyword co-occurrence network (minimum five co-occurrences).

TABLE V. SENSITIVITY ANALYSIS OF CLUSTER STRUCTURE ACROSS ALTERNATIVE VOSVIEWER THRESHOLDS IN THE KEYWORD CO-OCCURRENCE NETWORK. MODULARITY Q IS COMPUTED VIA THE LOUVAIN IMPLEMENTATION IN IGRAPH

Threshold	Nodes	Edges	Modularity Q	Clusters	Smallest cluster size	Note
3	578	17120	0.192	5	76	Larger graph; consistent low Q
5	302	9709	0.161	5	29	Used in main analysis
10	141	4422	0.133	6	9	Slight over-fragmentation
15	82	2091	0.116	5	5	Stable structure

Note: Modularity Q and cluster counts are computed via the Louvain implementation in igraph (see Table IV, Note a).

constrain the structural visibility of ethical and equity-focused scholarship. This is true even when, as the keyword analysis shows, that scholarship's vocabulary has been adopted into the surrounding discourse. The present finding extends Zawacki-Richter et al. [1] in one important respect. The structural under-anchoring of responsible-AI scholarship persists through a roughly 170-fold expansion in annual output. This indicates that growth alone does not consolidate the citation, venue, and collaboration scaffolding that integrative research requires, a pattern consistent with the asymmetries Popenici and Kerr [23] and Zawacki-Richter et al. [1] observed in the earlier, smaller literature.

C. Vocabulary Diffusion Without Infrastructural Consolidation

The keyword co-occurrence analysis (Section V-C6) reveals two patterns that together complicate a simple peripherality reading. First, ethics-equity vocabulary is integrated with the applied-AI cluster at densities comparable to those of technical vocabulary (0.20 vs 0.21), but the ethics-equity and technical streams connect to each other directly at roughly half that density (0.11); they engage through the applied-AI mediator rather than with one another. Second, the ethics-equity cluster is internally less cohesive (intra-density 0.37) than either the applied-AI (0.55) or technical (0.44) clusters, suggesting that ethics-equity vocabulary remains thematically diffuse even where it has diffused into the field.

Read alongside the co-citation, source, and country-level findings, this picture sharpens. The keyword space shows ethics-equity vocabulary entering the field's everyday discourse, but the citation canon does not include the foundational responsible-AI references in education as cluster-forming nodes (Section V-C4). The Bradford core contains no

specialist venue dedicated to fair, equitable, or explainable AI in education (Section V-B1). The author-level co-authorship network is extraordinarily sparse, with only ten authors meeting the three-publication threshold and only four of those participating in any co-authorship edge (Section V-C2). We characterise this combined pattern as vocabulary diffusion without infrastructural consolidation: the field has adopted the language of responsible AI at the level of keywords, but the citation, venue, and collaboration scaffolding required to translate that language into integrated practice has not yet formed.

The author-network null result (Section V-C2) deserves particular attention. Two readings are compatible with the data. First, the field may be populated by a wide base of single-contribution or two-contribution authors who do not return to the topic, a "drive-by" pattern characteristic of fields that attract opportunistic crossover from neighbouring computer-science research areas (consistent with the single-document profile noted in Section V-C2). Second, the absence of dense recurring author collaborations may reflect the absence of structurally anchoring research groups, which in mature fields function as the institutional carriers of integrative research agendas. In either reading, the absence of a dense author-collaboration backbone weakens the conditions under which integrative explainability and equity research could consolidate, complementing the citation- and venue-infrastructure findings directly.

The diagnosis here is therefore not that the field lacks relevant knowledge, nor that responsible-AI vocabulary is missing from its discourse. Rather, the field lacks the structural infrastructure that would allow knowledge already in circulation to consolidate into integrative practice. This extends Holmes et al. [20] and Bond et al. [32], who diagnosed responsible-AI commitments in education as parallel normative programmes, by adding structural evidence: the parallelism appears not only at the level of principles community members espouse but in the citation, venue, and collaboration infrastructures within which those principles would have to operate.

D. The Equity Paradox in Knowledge Production

The country-level collaboration network (Section V-C3) concentrates output in high-income regions, with limited participation from Sub-Saharan Africa and much of Latin America. This concentration corroborates concerns raised by Zawacki-Richter et al. [1] regarding the lack of diversity in AI-in-education research, and by Holstein and Doroudi [16] regarding whose contexts shape the design of educational AI. It has direct design consequences. As Holstein and Doroudi [16] argue, cultural assumptions from dominant contexts tend to become embedded in the AI systems those contexts produce. The geographic concentration reported here suggests that the conditions for such embedding are currently at their strongest. This concentration is visible at both the country level (Section V-C3) and the institutional level (Section V-B2), where no institution from Africa, Latin America, or Southeast Asia appears in the top ten by output. The pattern also intensifies the algorithmic-bias risks documented by Baker and Hawn [15] and Mehrabi et al. [17], because the training data, feature choices, and evaluation benchmarks that bias analyses target are themselves generated within a narrow set of institutional settings.

The present findings extend these arguments in one important respect. Inequities in AI in education are typically framed as outcomes of deployed systems [15], [19], [20], [16]. The collaboration-network evidence here suggests that inequities are also embedded in the production of the knowledge that shapes those systems, so that research directed at equitable learning is itself produced within structurally unequal global research networks. We refer to this pattern as the equity paradox of knowledge production: technologies intended to promote equitable learning are disproportionately developed within contexts that are least representative of global educational diversity.

In light of the corrected keyword-network findings, the equity paradox now stands as the dominant structural peripherality finding in this study. Ethics-equity vocabulary is integrated into the field's applied-AI discourse, but the contexts of production are not. It is the geographic concentration of knowledge production, not the keyword-level marginalisation of equity terms, that constitutes the most clearly peripheral structural feature of the field.

Three rival readings of the broader structural argument deserve consideration.

The first is that the patterns observed here are the normal early-stage property of any rapidly growing interdisciplinary field, a developmental lag rather than a structural pathology. We acknowledge this possibility but note that the structural under-anchoring persists across a 170-fold growth in annual output, with no indication of converging cluster connectivity in the thematic-evolution analysis (Section V-C1). The time-window evidence is therefore inconsistent with a transient lag.

The second rival reading is that the keyword analysis is distorted by vocabulary fragmentation. Synonym proliferation in a young field (for example, "explainable AI" vs "XAI" vs "interpretable AI") may inflate apparent cluster boundaries. The data partially support this. Explainable AI, XAI, and explainable artificial intelligence appear in three different clusters simultaneously (Section V-C6), and a distinct vocabulary-fragmentation cluster (Cluster 4) is anchored on indexing-variant forms of central terms. However, the ethics-equity cluster also contains substantively distinct terms (equity, fairness, trust, inclusivity, digital divide) not reducible to vocabulary fragmentation, and the citation-canon finding is independent of vocabulary effects entirely.

The third rival reading is that responsible-AI work is published in separate venues not captured by the search. This is partly true (FAccT and FAT* proceedings, for example), but it is mitigated by the inclusion of explainab*, fairness, equity, and algorithmic bias in the Boolean query, which would retrieve any indexed work using these terms regardless of venue.

Taken together, the rival readings are partial at best. The vocabulary diffusion without infrastructural consolidation interpretation remains the most parsimonious account of the combined evidence across all four analytical layers.

The structural picture across the four network analyses provides the consolidated answer to RQ3. Modularity values

across the field's networks remain modest (Q = 0.275 for country co-authorship, Q = 0.346 for cited-reference co-citation, Q = 0.340 for cited-journal co-citation, Q = 0.164 for keyword co-occurrence), indicating that the field's intellectual architecture is consistently more interconnected than fragmented at the cluster level. Within this interconnected architecture, three structurally distinctive features stand out. First, the foundational responsible-AI references in education (Baker and Hawn [15], Khosravi et al. [14], Holmes et al. [20]) do not appear as cluster-forming nodes in the co-citation network, indicating their structural under-anchoring in the field's citation canon. Second, the ethics-equity and technical-methods clusters in the keyword co-occurrence network connect to each other directly at approximately half the density of either's connection to the applied-AI mediator cluster (0.11 vs 0.20–0.21; bootstrap 95% CIs non-overlapping), indicating mediated rather than direct cross-stream engagement. Third, the country-level collaboration network is anchored on high-income contexts, with Sub-Saharan Africa and Latin America almost absent from the structural map. Together, these three features constitute the structural pattern that the next subsection interprets through the Structural Integration Model.

E. Theoretical Contribution: The Structural Integration Model

Beyond identifying the pattern of vocabulary diffusion without infrastructural consolidation, this study advances a theoretical account of why the integration deficit persists and what resolving it requires. We propose the Structural Integration Model (SIM), which reconceptualises explainability and equity as co-constitutive design principles whose integration depends on the structural organisation of the field's knowledge infrastructure (Fig. 17). The model rests on three propositions, each linked to a testable hypothesis. The SIM is presented as a hypothesis-generating framework: its propositions describe associations observed in the present corpus together with theoretically motivated directional relationships that require independent longitudinal or interventional validation beyond the current bibliometric data.

1) Proposition 1 (Mutual dependency): Explainability and equity are not independent design goals but mutually constitutive constraints. A system interpretable only to technically sophisticated users is likely to fail the equity objectives articulated by Baker and Hawn [15], Holstein and Doroudi [16], and Mehrabi et al. [17]. A system that pursues fairness without the transparency that Barredo Arrieta et al. [13] and Khosravi et al. [14] treat as foundational cannot be audited or trusted [18]. The two dimensions must therefore be jointly operationalised. *Testable hypothesis (H1):* in a future bibliometric replication, integrated explainability and equity papers (defined as documents whose author-keyword set spans both clusters) will exhibit higher mean citation impact than papers confined to a single cluster, after controlling for publication year and venue.

2) Proposition 2 (Mediated disconnection): The integration deficit between the ethics-equity research stream and the technical-methods stream is structural. Even where ethics-equity vocabulary is present in the applied-AI cluster at densities comparable to technical vocabulary, the two streams engage with each other only through the mediation of the applied-AI cluster rather than directly. In the present corpus, ethics-equity-to-technical inter-cluster density (0.11; bootstrap 95% CI [0.053, 0.086]) is approximately half the density of either stream's connection to the applied-AI cluster (0.20 to 0.21; bootstrap 95% CIs [0.108, 0.150] and [0.111, 0.157] respectively, non-overlapping with the ethics-equity-to-technical interval). This mediated disconnection, combined with the absence of foundational responsible-AI references from the co-citation core (Section V-C4) and the absence of specialist responsible-AI venues from the Bradford core (Section V-B1), is associated with the weak direct cross-stream interaction that integrative research practice would require. *Testable hypothesis (H2):* the ethics-equity-to-technical inter-cluster density will not exceed half of the ethics-equity-to-applied-AI inter-cluster density without targeted policy intervention (dedicated funding calls requiring integration, integrative-research journals, or structured cross-cluster collaboration programmes).

3) Proposition 3 (The Equity paradox as structural amplifier): Geographic concentration of research production is associated with the mediated disconnection. When knowledge is produced within relatively homogeneous institutional and cultural contexts, the blind spots Holstein and Doroudi [16] document remain unchallenged, and the absence of the contexts that would most clearly motivate direct ethics-equity-to-technical engagement is reinforced rather than corrected. *Testable hypothesis (H3):* as Sub-Saharan African and Latin American participation in the country-level collaboration network increases (operationalised as their combined share of total publications rising above 5%), the ethics-equity-to-technical inter-cluster density will increase commensurately.

Taken together, the three propositions position explainability and equity as structurally interdependent dimensions whose integration requires conceptual, infrastructural, and institutional transformation. The SIM differs from existing responsible-AI frameworks in three specific ways. First, where the AI4People framework of Floridi et al. [22] articulates responsible-AI principles as ethical commitments that ought to guide deployment, the SIM treats explainability and equity as empirically measurable structural properties of the field's knowledge infrastructure, with bibliometric quantities (inter-cluster densities, citation-canon membership, venue concentration, collaboration breadth) serving as observable proxies for integration. Second, where the community-wide framework of Holmes et al. [20] identifies what responsible-AI commitments the AIED community should adopt, the SIM identifies why those commitments have not yet consolidated, locating the explanation in the mediated disconnection between the ethics-equity and technical streams (Proposition 2) and the geographic concentration of knowledge production (Proposition 3). Third, where the ethical-principles inventory of Nguyen et al. [21] enumerates principles to be applied, the SIM links each proposition to a falsifiable hypothesis (H1–H3) with concrete bibliometric thresholds, allowing structural integration to be tracked over time rather than asserted. The SIM also complements the revolution-versus-evolution diagnosis of Roll and Wylie [2] by specifying a mechanism through which evolutionary expansion alone is unlikely to deliver revolutionary change in responsible AI in education: vocabulary diffusion can extend the surface of the field without consolidating its citation, venue, or collaboration scaffolding. Fig. 17 presents the SIM as a proposed chain of associations: P3 (geographic concentration) co-occurs with P2 (mediated disconnection), which is in turn associated with weaker P1 (mutual depen-

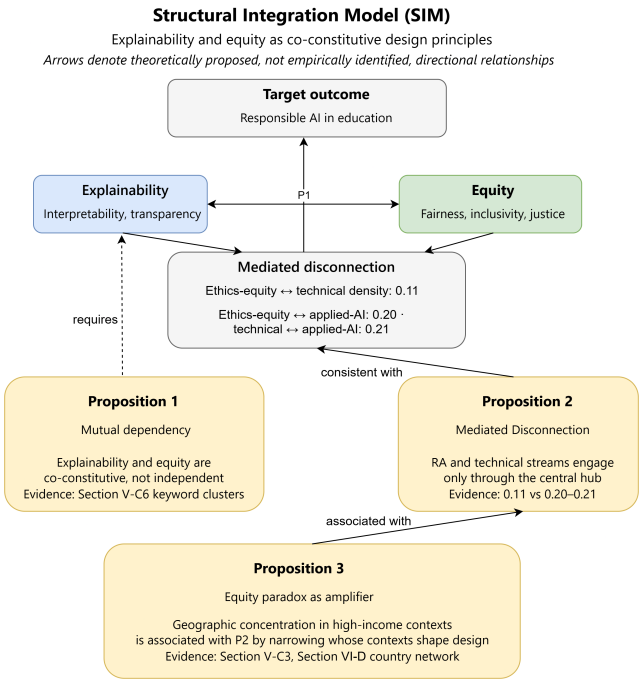

Fig. 17. The Structural Integration Model (SIM), positioning explainability and equity as co-constitutive design principles whose integration is associated with structural barriers (P2) and geographic concentration (P3). Arrows denote theoretically proposed, not empirically identified, directional relationships.

dency), a configuration consistent with an unrealised target outcome of Responsible AI in Education. The directionality shown is theoretically motivated and is not established by the present cross-sectional data. Each proposition is annotated with its empirical anchor in the corpus, allowing the model to be read as a directly testable framework rather than a purely conceptual claim.

The SIM is derived inductively from a single bibliometric corpus and is therefore proposed as a hypothesis-generating framework rather than a validated causal model. Its status as an explanatory account is established not by fit to the present data, which it is built from, but by the falsifiability of H1–H3 against independent and future data, for which concrete bibliometric thresholds are specified. Confirmatory testing on an independent corpus, and comparison against competing explanatory models on that corpus, remain necessary steps that the present study defines but does not itself perform.

F. Implications for Practice and Policy

The convergence of findings across the four analytical layers identifies a consistent infrastructural gap. Foundational components are available: adaptive systems [3], [4], [10], explainability frameworks [14], [13], and equity-oriented analyses [15], [19], [20], [21], [16], [17], [18]. The citation, venue, collaboration, and geographic infrastructure that would translate vocabulary-level diffusion into integrated practice is not yet in place. Three concrete implications follow.

1) For funders and research councils: Targeted funding calls should require that proposals integrate explainability, fairness, and educational pedagogy as core design requirements rather than as auxiliary work packages. Calls of this kind, modelled on the integrated-research priorities of the AI4People framework [22], would directly target the mediated disconnection diagnosed by Proposition 2 by creating direct funding-level incentives for ethics-equity and technical collaboration.

2) For journal editors: The Bradford-core analysis (Section V-B1) shows that no specialist venue dedicated to fair, equitable, or explainable AI in education currently sits in the field's core. Editors of established AIED journals should consider dedicated special issues that explicitly require integration across the three streams. The establishment of a Responsible AI in Education journal would also lower a current structural barrier.

3) For the AIED research community: Participatory research approaches that engage underrepresented learner populations [24] are likely to provide critical insights into how explanations function across diverse contexts, and which fairness trade-offs are acceptable to the learners and educators concerned. Building dense recurring author collaborations across institutional and geographic boundaries would also begin to address the author-network sparsity documented in Section V-C2.

G. Limitations

Eight limitations qualify the findings.

1) First (database coverage): The corpus is restricted to Scopus and Web of Science. Although these databases together provide the most comprehensive coverage of peer-reviewed scientific literature [54], [40], they underrepresent grey literature, preprints, and non-English-language scholarship. This coverage profile may disproportionately remove work from underrepresented regions, and may therefore bias the equity-paradox finding toward greater concentration than exists in the broader literature. Future replications could mitigate this bias by drawing on additional indices of the kind evaluated by Martín-Martín et al. [54], and on regional repositories such as SciELO and AfricArXiv. They could also align data-sharing practices with the FAIR principles of Wilkinson et al. [42].

2) Second (structural inference): Bibliometric analysis measures structural properties of a field [25], [37], [36] and cannot adjudicate the substantive quality or pedagogical impact of individual studies. This limitation has already been flagged by Donthu et al. [25] and illustrated in the AI-in-education context by Hinojo-Lucena et al. [30] and Bond et al. [32]. The under-anchoring of responsible-AI references in the citation network is evidence of structural under-consolidation, not of scientific weakness. Substantive quality assessment remains the province of targeted systematic reviews along the lines of Zawacki-Richter et al. [1] and Nabizadeh et al. [11].

3) Third (indexing lag): The 2025 publication count of 854 is a lower bound. Indexing in Scopus and WoS continues for 6 to 12 months after the publication year, and the search was conducted on 17 March 2026. The true 2025 count will only stabilise by late 2026. The 67.2% CAGR headline therefore reflects a partially incomplete final year and should be read alongside the more conservative 61% three-year-rolling-base estimate. The 12 early-access records with a 2026 publication

year that were excluded from the quantitative corpus also contribute to this provisional character. The centrality-density positions reported in Section V-C1 should be revisited as indexing matures, ideally within the thematic-evolution framework of Cobo et al. [26].

4) Fourth (author-name disambiguation): Author-level findings are subject to surname collision, particularly for common Chinese surnames (Wang, Zhang, Li, Chen). ORCID-based partial disambiguation (n = 424 ORCID-linked records, 30.9%) suggests that several entries in the top ten, including Wang J, Zhang Y, Wang Z, Zhang X, and Li X, likely represent more than one distinct researcher. Full disambiguation would require manual reconciliation. Institutional and country-level findings are not subject to this limitation and should be treated as more reliable.

5) Fifth (threshold sensitivity): All VOSviewer networks were constructed at a minimum five-occurrence threshold. The sensitivity analysis at thresholds of 3, 5, 10, and 15 (Table V) showed cluster count remained stable at five (or five to six), with modularity within $Q \in [0.116, 0.192]$. This confirms structural stability in the form of the cluster structure. The consistently low modularity values across thresholds, however, indicate that the keyword space is more interconnected than it is fragmented. This is itself a substantive finding, interpreted in Section V-C6 and Section VI-C: the field's vocabulary integrates well, even where its citation, venue, and collaboration infrastructures do not.

6) Sixth (search precision): A manual relevance audit of a random sample of 30 records (set.seed = 42) classified 21 records (70%) as clearly relevant to the AI + education + responsible-AI scope, 5 records (16.7%) as borderline (e.g., AI bias mitigation work without an explicit educational application, or pedagogy-focused work without an explicit AI component), and 4 records (13.3%) as false positives, primarily healthcare and finance applications of AI in which the abstracts contained education-adjacent terminology (e.g., "predictive modelling," "personalised medicine") that triggered cluster co-occurrence with the educational query terms. The structural findings reported in Sections V-C4 and V-C6 are unlikely to be sensitive to false positives at this rate, since the analyses rely on co-occurrence and co-citation patterns rather than on individual record content. However, raw productivity and citation-impact figures (Sections V-B1 and V-B5) carry a small upward bias of approximately 13%. Future replications could further refine search precision through full-text relevance screening, although this would substantially increase manual workload at corpus sizes of this magnitude.

7) Seventh (single-corpus derivation of the SIM): The Structural Integration Model is induced from, and illustrated on, the same corpus used to identify the integration gaps it addresses. It has not been validated on an independent dataset, nor tested against a competing explanatory model of the same data. This is an inherent limitation of theory-building from a single bibliometric case. We mitigate it by specifying three falsifiable hypotheses (H1–H3, Section VI-E) with concrete thresholds testable on independent and future corpora, converting the model from a post-hoc description into a set of empirical predictions; confirmatory and comparative validation remain future work (Section VI-H).

8) Eighth (conjunctive corpus design): The three-cluster conjunctive query conditions the corpus on the presence of responsible-AI terminology (Section III-A). The structural findings are properties of connectivity within this corpus, but prevalence and frequency comparisons are necessarily within-corpus rather than whole-field. A complementary two-stage design, constructing a broad AI-plus-education corpus and then analysing the responsible-AI subset against it, would permit whole-field prevalence estimation and is a priority for future replication.

H. Future Research Directions

Three research directions follow from the structural gaps identified here.

1) Integrative system design: Future research should develop and empirically evaluate AI-based personalised-learning systems that embed explainability mechanisms and equity considerations as core design requirements rather than as post hoc additions. This direction is consistent with Khosravi et al. [14], Baker and Hawn [15], Holstein and Doroudi [16], and the sociotechnical framing of Selbst et al. [18]. Hypothesis H1 (Section VI-E) provides a directly testable empirical entry point.

2) Longitudinal structural analysis: Bibliometric analyses of the kind reported here should be conducted at regular intervals so that infrastructural consolidation can be monitored as it persists, narrows, or evolves. This is particularly relevant given the continuing expansion of LLM-driven research [55], [7], [8]. The thematic-evolution framework of Cobo et al. [26] provides a natural methodological basis, and the bibliometric guidelines of Donthu et al. [25] and Zupic and Čater [27] offer corresponding protocol standards. A five-year replication window would allow direct comparison against the present baseline. This would also provide the first independent test of the SIM's propositions, including comparison against alternative explanatory accounts of the same structural pattern. Hypothesis H2 (Section VI-E) operationalises a specific quantitative target.

3) Global knowledge diversification: Further research is needed to understand the structural barriers limiting participation from underrepresented regions, and to evaluate concrete strategies for broadening participation. These strategies include targeted funding mechanisms (for example, joint North-South research grants), regional consortia (an African AI-Education Network, a Latin American AIED Observatory), open-access mandates, and inclusion of language-diverse databases (SciELO, AfricArXiv) in bibliometric infrastructure. This direction is motivated by Zawacki-Richter et al. [1], Baker and Hawn [15], Holstein and Doroudi [16], Akgun and Greenhow [19], and the sociotechnical critique of Selbst et al. [18]. Hypothesis H3 (Section VI-E) provides the testable target.

VII. Conclusion

This study set out to examine whether explainability and equity, the two foundational principles of responsible AI, are structurally embedded in the intellectual architecture of AI-based personalised learning, or whether they remain only loosely associated with it. A bibliometric analysis of 1371 publications indexed in Scopus and Web of Science between 2015 and 2025 has shown that the field has expanded at a pace few

educational technology areas have matched. This expansion has been absorbed predominantly by a technically cohesive core spanning adaptive learning, knowledge tracing, and deep-learning research. The structural position of explainability and equity within this expanding architecture, however, is more nuanced than a simple peripherality reading would suggest.

At the level of vocabulary, ethics-equity terms are integrated into the field's applied-AI discourse at densities comparable to those of technical vocabulary (0.20 vs 0.21). At the level of citation, venue, collaboration, and geographic infrastructure, however, this integration has not consolidated. Foundational responsible-AI references do not register as cluster-forming nodes in the co-citation network, no specialist responsible-AI venue appears in the Bradford core, the ethics-equity and technical streams interact directly only at half the density of either's connection to the applied-AI cluster (0.11 vs 0.20–0.21; bootstrap 95% CIs non-overlapping), and the country-level collaboration network is anchored on high-income contexts with Sub-Saharan Africa and Latin America almost absent. We characterise this combined pattern as vocabulary diffusion without infrastructural consolidation.

The study's principal contribution is the Structural Integration Model. By formalising explainability and equity as structural properties of the field's knowledge infrastructure rather than as ethical commitments alone, the SIM specifies what would have to change at the level of citation canon, venues, collaborations, and geography, for responsible-AI principles to consolidate into integrated research practice. Three falsifiable hypotheses (H1–H3) operationalise the model with concrete bibliometric thresholds, providing direct empirical targets for future replications and a measurable benchmark against which the field's structural integration can be tracked over time.

The implication for the field is direct. Responsible AI in education will not be achieved through volume of vocabulary alone. Vocabulary integration, on the evidence presented here, is already underway; what remains absent is the surrounding infrastructure. Progress requires deliberate construction of that infrastructure: funding instruments that mandate cross-stream integration, editorial commitments that establish dedicated responsible-AI-in-education venues in the field's Bradford core, participatory research approaches that broaden the geographic base of knowledge production beyond high-income contexts, and longitudinal bibliometric monitoring that allows the field's structural integration to be tracked rather than assumed. Whether the next decade of AI in education delivers on its responsible-AI commitments in practice, and not only in vocabulary, will depend on whether the research community treats explainability and equity as the structural design conditions the SIM identifies them to be.

REFERENCES

[1] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "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. 39, 2019.

[2] I. Roll and R. Wylie, "Evolution and revolution in artificial intelligence in education," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 582–599, 2016.

[3] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User Modeling and User-Adapted Interaction*, vol. 4, no. 4, pp. 253–278, 1994.

[4] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing," in *Advances in Neural Information Processing Systems*, 2015, pp. 505–513, arXiv:1506.05908.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems 30*. Curran Associates, 2017, pp. 5998–6008, arXiv:1706.03762.

[6] A. Ghosh, N. Heffernan, and A. S. Lan, "Context-aware attentive knowledge tracing," in *Proc. 26th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining*. ACM, 2020, pp. 2330–2339.

[7] F. Kamalov, D. Santandreu Calonge, and I. Gurrib, "New era of artificial intelligence in education: Towards a sustainable multifaceted revolution," *Sustainability*, vol. 15, no. 16, p. 12451, 2023.

[8] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, and G. Kasneci, "ChatGPT for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.

[9] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," *Educational Psychologist*, vol. 46, no. 4, pp. 197–221, 2011.

[10] J. Zhang, X. Shi, I. King, and D.-Y. Yeung, "Dynamic key-value memory networks for knowledge tracing," in *Proc. 26th Int. Conf. World Wide Web*. ACM, 2017, pp. 765–774.

[11] A. H. Nabizadeh, J. P. Leal, H. N. Rafsanjani, and R. R. Shah, "Learning path personalization and recommendation methods: A survey of the state-of-the-art," *Expert Systems with Applications*, vol. 159, p. 113596, 2020.

[12] I. Gligorea, M. Cioca, R. Oancea, A.-T. Gorski, H. Gorski, and P. Tudorache, "Adaptive learning using artificial intelligence in e-learning: A literature review," *Education Sciences*, vol. 13, no. 12, p. 1216, 2023.

[13] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, 2020.

[14] H. Khosravi, S. Buckingham Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, S. Knight, R. Martinez-Maldonado, S. Sadiq, and D. Gašević, "Explainable artificial intelligence in education," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100074, 2022.

[15] R. S. Baker and A. Hawn, "Algorithmic bias in education," *International Journal of Artificial Intelligence in Education*, vol. 32, no. 4, pp. 1052–1092, 2022.

[16] K. Holstein and S. Doroudi, "Equity and artificial intelligence in education: Will 'AIEd' amplify or alleviate inequities in education?" in *The Ethics of Artificial Intelligence in Education*, W. Holmes and K. Porayska-Pomsta, Eds. Routledge, 2022, pp. 151–173.

[17] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–35, 2021.

[18] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi, "Fairness and abstraction in sociotechnical systems," in *Proc. 2019 ACM Conf. Fairness, Accountability, and Transparency (FAT* '19)*. ACM, 2019, pp. 59–68.

[19] S. Akgun and C. Greenhow, "Artificial intelligence in education: Addressing ethical challenges in K–12 settings," *AI and Ethics*, vol. 2, no. 3, pp. 431–440, 2022.

[20] W. Holmes, K. Porayska-Pomsta, K. Holstein, E. Sutherland, T. Baker, S. Buckingham Shum, O. C. Santos, M. T. Rodrigo, M. Cukurova, I. I. Bittencourt, and K. R. Koedinger, "Ethics of AI in education: Towards a community-wide framework," *International Journal of Artificial Intelligence in Education*, vol. 32, no. 3, pp. 504–526, 2022.

[21] A. Nguyen, H. N. Ngo, Y. Hong, B. Dang, and B.-P. T. Nguyen, "Ethical principles for artificial intelligence in education," *Education and Information Technologies*, vol. 28, no. 4, pp. 4221–4241, 2023.

[22] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, and E. Vayena, "AI4People – an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds and Machines*, vol. 28, no. 4, pp. 689–707, 2018.

[23] S. A. D. Popenici and S. Kerr, "Exploring the impact of artificial intelligence on teaching and learning in higher education," *Research and Practice in Technology Enhanced Learning*, vol. 12, no. 22, 2017.

[24] R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, *Intelligence Unleashed: An Argument for AI in Education*. Pearson Education, 2016.

[25] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *Journal of Business Research*, vol. 133, pp. 285–296, 2021.

[26] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field," *Journal of Informetrics*, vol. 5, no. 1, pp. 146–166, 2011.

[27] I. Zupic and T. Čater, "Bibliometric methods in management and organization," *Organizational Research Methods*, vol. 18, no. 3, pp. 429–472, 2015.

[28] L. Waltman, N. J. van Eck, and E. C. M. Noyons, "A unified approach to mapping and clustering of bibliometric networks," *Journal of Informetrics*, vol. 4, no. 4, pp. 629–635, 2010.

[29] Z. Bahroun, C. Anane, V. Ahmed, and A. Zacca, "Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis," *Sustainability*, vol. 15, no. 17, p. 12983, 2023.

[30] F.-J. Hinojo-Lucena, I. Aznar-Díaz, M.-P. Cáceres-Reche, and J.-M. Romero-Rodríguez, "Artificial intelligence in higher education: A bibliometric study on its impact in the scientific literature," *Education Sciences*, vol. 9, no. 1, p. 51, 2019.

[31] H. Crompton and D. Burke, "Artificial intelligence in higher education: The state of the field," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, p. 22, 2023.

[32] M. Bond, H. Khosravi, M. De Laat, N. Bergdahl, V. Negrea, E. Oxley, P. Pham, S. W. Chong, and G. Siemens, "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. 4, 2024.

[33] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *Journal of Informetrics*, vol. 11, no. 4, pp. 959–975, 2017.

[34] N. J. Van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010.

[35] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[36] H. Small, "Co-citation in the scientific literature: A new measure of the relationship between two documents," *Journal of the American Society for Information Science*, vol. 24, no. 4, pp. 265–269, 1973.

[37] M. Callon, J. P. Courtial, W. A. Turner, and S. Bauin, "From translations to problematic networks: An introduction to co-word analysis," *Social Science Information*, vol. 22, no. 2, pp. 191–235, 1983.

[38] A. Pritchard, "Statistical bibliography or bibliometrics?" *Journal of Documentation*, vol. 25, no. 4, pp. 348–349, 1969.

[39] A. Martín-Martín, E. Orduna-Malea, M. Thelwall, and E. Delgado López-Cózar, "Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories," *Journal of Informetrics*, vol. 12, no. 4, pp. 1160–1177, 2018.

[40] P. Mongeon and A. Paul-Hus, "The journal coverage of Web of Science and Scopus: A comparative analysis," *Scientometrics*, vol. 106, no. 1, pp. 213–228, 2016.

[41] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, and D. Moher, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, no. n71, 2021.

[42] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, and B. Mons, "The FAIR guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, no. 160018, 2016.

[43] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.

[44] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal, Complex Systems*, vol. 1695, 2006, https://igraph.org.

[45] S. C. Bradford, "Sources of information on specific subjects," *Engineering*, vol. 137, pp. 85–86, 1934.

[46] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences*, vol. 102, no. 46, pp. 16 569–16 572, 2005.

[47] L. Waltman and N. J. van Eck, "A smart local moving algorithm for large-scale modularity-based community detection," *European Physical Journal B*, vol. 86, p. 471, 2013.

[48] M. Sallam, "ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns," *Healthcare*, vol. 11, no. 6, p. 887, 2023.

[49] A. Abulibdeh, E. Zaidan, and R. Abulibdeh, "Navigating the confluence of artificial intelligence and education for sustainable development in the era of industry 4.0: Challenges, opportunities, and ethical dimensions," *Journal of Cleaner Production*, vol. 437, p. 140527, 2024.

[50] A. J. Lotka, "The frequency distribution of scientific productivity," *Journal of the Washington Academy of Sciences*, vol. 16, no. 12, pp. 317–323, 1926.

[51] L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: A review," *IEEE Access*, vol. 8, pp. 75 264–75 278, 2020.

[52] S. Fortunato and M. Barthélemy, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences*, vol. 104, no. 1, pp. 36–41, 2007.

[53] C.-C. Lin, A. Y. Q. Huang, and O. H. T. Lu, "Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review," *Smart Learning Environments*, vol. 10, no. 41, 2023.

[54] A. Martín-Martín, M. Thelwall, E. Orduna-Malea, and E. Delgado López-Cózar, "Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations," *Scientometrics*, vol. 126, no. 1, pp. 871–906, 2021.

[55] C. K. Y. Chan and W. Hu, "Students' voices on generative AI: Perceptions, benefits, and challenges in higher education," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 43, 2023.