Supplementary File S2

Search Strategies, Screening Codebook, and Reproducibility Materials

Educational AI Auditing Across the Learning Lifespan: Concepts, Methods, Evidence, and Stage-Sensitive Governance

This file documents the search strategies, deduplication procedures, screening criteria, and software environment used in the systematic mapping review reported in Section 3. Supplementary Table S1 describes the sources cited in the revised Entry, and Supplementary Table S3 lists the 981 unique included records.

S2.1. Overview of sources

Source	Role	Search executed	Records identified
Scopus	Principal database search; indexes conference proceedings alongside journals	30 August 2026	3,245
ERIC	Education-specific database; corrects for education literature not indexed in Scopus	6 September 2026	1,331 (1,068 within window)
Crossref REST API	Journal-level hand search of eleven education journals	30 August 2026	2,285 (975 title-matched)

Publication window: 1 January 2024 to 30 June 2026.

Note on date granularity. Scopus and Crossref record day-level publication dates and were filtered accordingly. ERIC records carry publication year and issue only; the ERIC search was therefore filtered by publication year, and a small number of records published in the second half of 2026 enter the review from that source alone. This asymmetry is reported rather than resolved, since resolving it would require discarding otherwise eligible records on the basis of metadata granularity rather than eligibility.

S2.2. Scopus search

S2.2.1. Query as executed

```
TITLE-ABS-KEY(
  ("artificial intelligence" OR "generative AI" OR "GenAI"
   OR "large language model*" OR "ChatGPT" OR "machine learning"
   OR "learning analytics" OR "algorithmic decision*" OR chatbot
   OR "intelligent tutor*" OR "predictive analytics")
  AND
  (educat* OR school* OR student* OR teacher* OR learner*
   OR classroom* OR curricul* OR "higher education"
   OR "vocational" OR TVET OR "early childhood")
  AND
  (audit* OR "algorithmic accountability" OR "impact assessment"
   OR "algorithmic fairness" OR "human oversight" OR contestab*
   OR "due process" OR "prompt injection" OR "data protection"
   OR accessibility OR disability OR "AI governance"
   OR "responsible AI" OR sociotechnical)
)
AND PUBYEAR > 2023 AND PUBYEAR < 2027
AND (LIMIT-TO(DOCTYPE,"ar") OR LIMIT-TO(DOCTYPE,"re") OR LIMIT-TO(DOCTYPE,"cp"))
AND (LIMIT-TO(LANGUAGE,"English"))
```

Records returned: 3,245. Export was made in three parts by publication year, because the Scopus interface caps a single export at 2,000 records: 2024 (369), 2025 (1,297), 2026 (1,579). The three parts sum to the interface total, confirming that no records were lost at export.

S2.2.2. Document types

Conference papers were retained deliberately. Of the 3,245 records, 899 (27.7 percent) are conference papers. Core work in algorithmic accountability and human-computer interaction appears at venues including FAccT, AIES, CHI, CSCW, IDC, and LAK rather than in journals, and a document-type restriction to journal articles would have excluded a substantial part of the literature the review question addresses.

S2.2.3. Query scope

A narrower query that omitted sociotechnical, algorithmic-bias, regulation, policy, predictive-analytics, and TVET terms returned 2,706 records. The broader query was retained because the omitted terms are central to the review question and both result sets required stratified screening.

S2.3. ERIC search

S2.3.1. Query as executed

```
(title:("artificial intelligence" OR "generative AI" OR "large language model"
 OR "ChatGPT" OR "machine learning" OR "learning analytics"
 OR "intelligent tutoring" OR "predictive analytics")
 OR abstract:("artificial intelligence" OR "generative AI" OR "large language model"
 OR "ChatGPT" OR "machine learning" OR "learning analytics"
 OR "intelligent tutoring" OR "predictive analytics"))
AND
abstract:("algorithmic audit" OR "AI audit" OR "algorithmic accountability"
 OR "impact assessment" OR "algorithmic fairness" OR "human oversight"
 OR contestability OR "AI governance" OR "responsible AI" OR "trustworthy AI"
 OR sociotechnical OR "algorithmic bias" OR "AI regulation" OR "AI policy"
 OR accessibility OR "data protection" OR privacy OR equity OR bias)
```

Interface filters: Peer reviewed only selected; Full text available on ERIC not selected, since that filter constrains retrieval on availability rather than eligibility.

Records returned: 1,331, of which 1,068 fall within the publication window (2024: 248; 2025: 505; 2026: 315).

S2.3.2. Query refinement

An initial formulation permitted ERIC subject descriptors to satisfy the second concept block:

```
... AND (abstract:( ... audit and governance terms ... )
 OR descriptor:("Artificial Intelligence" OR "Ethics" OR "Educational Policy"
 OR "Equal Education" OR "Privacy" OR "Accountability" OR "Access to Education"))
```

The initial query returned 8,786 records and 6,544 after peer-review and date filters. The descriptor "Artificial Intelligence" appeared in 6,362 records and made the second concept block non-constraining. The final query therefore restricted the second block to abstract fields, producing the 1,331 records reported in S2.3.1. Both formulations are documented to explain the difference in retrieval counts.

S2.3.3. Export procedure

ERIC exports in MEDLINE/PubMed-style .nbib format, capped at 200 records per file. Seven sequential exports were made using the Start from result # parameter with values 1, 201, 401, 601, 801, 1001, and 1201, retrieving 1,224 records under a Since 2022 filter. Records were parsed from the .nbib format into tabular form; the parser is reported in S2.7. All 1,224 records carry abstracts (AB field), so ERIC records were screened on title and abstract.

S2.3.4. Scope of ERIC retrieval

ERIC indexes a substantial body of non-peer-reviewed material, including departmental guidance, institutional white papers, and grantee submissions carrying ED accession numbers. These were excluded by the peer-review filter. Some of this material is relevant to the governance questions the Entry addresses and is cited, where used, as a policy or standards source rather than as part of the mapped corpus. The distinction is recorded in the source_category field of Supplementary Table S1.

S2.4. Journal-level hand search

S2.4.1. Procedure

Journal metadata were retrieved through the Crossref REST API using the rcrossref package for R, filtering on ISSN, on type: journal-article, and on a published date between 1 January 2024 and 30 June 2026.

Note on the Crossref date field. The from_pub_date and until_pub_date filters operate on the Crossref published date, which for most publishers corresponds to the online-first date and may precede the print issue date. An article carried in a 2026 issue but posted online in 2025 therefore falls in the 2025 window. This is a property of the field and not an error, but it must be stated for the counts to be reproducible.

S2.4.2. ISSN register

Several journals resolve under only one of their two registered ISSNs in Crossref. Retrieval was attempted with the online ISSN first and the print ISSN as fallback. The ISSN that returned records for each journal is recorded below; using the alternative ISSN returns zero records for the journals so marked.

Journal	ISSN used	Note
International Journal of Educational Technology in Higher Education	2365-9440	
Computers and Education: Artificial Intelligence	2666-920X	
Smart Learning Environments	2196-7091	
Computers & Education	0360-1315	Print; online ISSN 1873-782X returns no records

Journal	ISSN used	Note
British Journal of Educational Technology	1467-8535	
Computers and Education Open	2666-5573	
The Internet and Higher Education	1096-7516	Print; online ISSN 1873-5525 returns no records
Educational Research Review	1747-938X	Print; online ISSN 1878-0385 returns no records
Interactive Technology and Smart Education	1758-8510	
Review of Educational Research	1935-1046	Online; print ISSN 0034-6543 returns no records
Journal of Educational Computing Research	1541-4140	

S2.4.3. Title terms

Nineteen terms, matched case-insensitively against normalised titles:

artificial intelligence, generative ai, genai, chatgpt, large language model, intelligent tutor, machine learning, deep learning, learning analytics, automated, adaptive learning, algorithm, chatbot, robot, predictive, ai literacy, facial recognition, natural language processing, \bai\b

The nineteenth term, the abbreviation bounded as a whole word, is necessary because title-level searching without it systematically misses articles that use the abbreviation and no other search term. Under the eighteen-term set the query returns 747 records; under the nineteen-term set it returns 975.

S2.4.4. Title normalisation

Publisher metadata in Crossref carries markup that defeats naive string matching. Titles from Wiley journals in particular contain SGML small-caps tags and embedded line breaks, so that a phrase such as large language model may appear in the metadata as large <scp>language model</scp> or be split across lines by whitespace padding. Titles were therefore normalised before matching by removing markup tags, replacing line breaks and tabs with spaces, unifying the hyphen, non-breaking hyphen, figure dash, en dash, and em dash to a single character, collapsing repeated whitespace, and lower-casing.

The effect is material and unevenly distributed. Without normalisation the British Journal of Educational Technology returns 54 title matches; with normalisation it returns 75.

S2.4.5. Retrieval reliability

The Crossref API returns HTTP 429 under sustained querying. Initial retrieval returned zero records for three journals for this reason, which is indistinguishable at the call site from a genuine absence of records. Retrieval was therefore repeated with a three-second interval between journals, a five-second interval before ISSN fallback, and up to five attempts per ISSN with backoff intervals of 15, 30, 45, and 60 seconds. Counts reported in the Entry are from a single complete run under this configuration.

S2.5. Deduplication

Operation	Records removed
Within Crossref (DOI)	0
Within Scopus (DOI, or normalised title where DOI absent)	1
Scopus records already retrieved by Crossref (DOI)	27
ERIC records already retrieved by Scopus or Crossref (DOI)	110
ERIC records already retrieved by Scopus or Crossref (normalised title)	25

Title-based deduplication was required for ERIC because 377 of its records within the window carry no DOI. Normalisation for this purpose removed all non-alphanumeric characters, collapsed whitespace, and lower-cased; matching was on the full normalised string, with a prefix comparison on the first sixty characters used to identify near matches for manual inspection. The prefix comparison returned no additional matches beyond the exact ones.

Within-Crossref deduplication returned zero removals because records were retrieved journal by journal under non-overlapping ISSN filters. The operation was performed nonetheless and its null result is reported, since a zero is informative only if the check was made.

S2.6. Stratification patterns

Scopus records were stratified by disciplinary field to sequence screening across the fields the review question spans. Stratification determined the order in which records were screened, not which records were screened; all records in all strata were assessed against the eligibility criteria.

Patterns were applied, case-insensitively, to the concatenation of title, abstract, author keywords, index keywords, and source title. A record was assigned to the first matching field in the order listed.

Stratum	Pattern	Records
algorithmic_accountability	fairness|accountab|audit|impact assessment|facct|aies|algorithmic bias	1,424
ai_ethics	ethic|moral|value alignment|responsib	687
law_regulation	\blaw\b|legal|regulat|ai act|gdpr|due process|compliance|policy instrument	260
security	security|prompt injection|adversarial|jailbreak|threat model|data leak	181
disability_accessibility	disabilit|accessib|assistive|neurodiver|universal design|screen reader	162
hci	human-computer|human computer|\bchi\b|cscw|interaction design|user experience|usability	77
learning_analytics	learning analytics|educational data mining|\blak\b|student model	31
public_administration	public administration|public sector|government|bureaucra|civic	16
unclassified	no pattern matched	379

Records in the unclassified stratum were re-examined with a supplementary pattern covering framings the primary patterns did not capture:

sociotechnical | socio-technical | human oversight | national ai strateg | imaginar

A wider supplementary pattern that additionally included governance, ai strateg, and stakeholder trust matched 300 of the 379 unclassified records and was not used, since governance alone is too frequent in this literature to sequence screening usefully. The narrower pattern matched 267 records. The remaining unclassified records were screened in their entirety.

S2.7. Screening codebook

S2.7.1. Inclusion criteria

A record was included when all four criteria were satisfied.

1. Date and type. Published between 1 January 2024 and 30 June 2026 as a peer-reviewed journal article, review, or conference paper.
2. Language. English.
3. Setting and object. The object of study is an AI system, its users, or its governance arrangements, in a formal, non-formal, or informal educational setting.
4. Evidence. The record reports either (a) empirical evidence concerning a specific deployed or tested AI system, or (b) analysis of a documented governance instrument, meaning a policy document, regulation, standard, or institutional guideline.

S2.7.2. Exclusion codes

Code	Ground	Applied when
E1	Not a research report	Editorial, commentary, correction, retraction notice, conference abstract, workshop proposal, laboratory or symposium description, or book review
E2	Publication mapping	Bibliometric, scientometric, or topic-mapping study of publication output, in which AI systems are not the object of study
E3	No educational setting	Clinical, corporate, public-sector, or general-population study; or education appears only as an illustration, with no educational-setting data and no institutional analysis
E4	Not retrievable	Full text could not be obtained
E5	Outside window	Language or publication date outside the specified range
E6	Outcome outside scope	Set in education, but the object of study is user attitudes, perceptions, anxiety, or adoption intention, and no evidence is reported on system behaviour, outcomes, or governance arrangements
E8	Not anchored to evidence or instrument	Conceptual framework, curriculum proposal, competency model, trend review, or system description not anchored to evaluation evidence or to a documented governance instrument

S2.7.3. Application notes for E3, E6, and E8

E3, E6, and E8 required the greatest screening judgment. The following rules were applied consistently across sources.

E3. Clinical and patient-facing studies were excluded even where they concern the education of health professionals, unless the object of study was an educational deployment rather than a clinical one. A study evaluating an AI system used to assess medical students is in scope; a study evaluating the same system used to assess patients is not.

E6. This is an exclusion of outcome scope, not of quality. Technology-acceptance, adoption-intention, and AI-anxiety studies are numerous in this literature and frequently well conducted, but they report on dispositions toward systems rather than on systems. Where such a study additionally reported on institutional policy or governance arrangements, it was included on that basis.

E8. This is the operative boundary of the review. The distinction applied was whether the record reports on something that exists and can be examined. A paper proposing an ethical framework for AI in a discipline is excluded; a paper analysing the AI policies actually adopted by a set of institutions is included. A paper describing a system architecture without evaluating it is excluded; the same system with reported accuracy, agreement, or comparison against a human baseline is included.

S2.7.4. Screening code revision

An eighth code, E7, was used during an early screening pass to mark records that satisfied the eligibility criteria but belonged to a design cluster already represented in the corpus. It was withdrawn and all records so marked were returned to the pool and reassessed against the criteria in S2.7.1. Excluding eligible studies on grounds of redundancy is a judgment about corpus balance rather than about eligibility, and it is not available to a systematic review. No record in the final corpus carries this code.

S2.7.5. Screening depth by source

Set	Records screened	Depth
ERIC	933	Title and abstract for every record
Scopus: law and regulation, disability, security, HCI, learning analytics, public administration, unclassified	1,793	Title and abstract for every record
Scopus: algorithmic accountability, AI ethics	2,111	Title for every record; abstract consulted where the title was not decisive
Crossref journal hand search	975	Title only; Crossref returns no abstract

S2.7.6. Reviewer arrangements

Screening was conducted by a single reviewer. No second reviewer independently screened a subsample, and inter-rater reliability is therefore not reported. Title and abstract screening was carried out with the assistance of a large language model, which produced a first-pass classification against the criteria in S2.7.1 that the author reviewed and confirmed. All final inclusion decisions were made by the author. No quality appraisal of included studies was undertaken.

S2.8. Software environment

Retrieval, deduplication, and record management were carried out in R. Parsing of ERIC .nbib exports and tabulation of screening decisions were carried out in Python.

R packages: rcrossref (Crossref API), dplyr, tidyr, stringr, purrr, readr, tibble (data manipulation), janitor (column-name normalisation), stringdist (cross-source title matching), irr (agreement statistics), readxl, writexl (screening sheets), DiagrammeR, DiagrammeRsvg, rsvg (flow diagram).

Python packages: pandas.

A run log recording the search date, the date filter field used, per-journal record counts, per-stage counts, and the full sessionInfo() output was written for each retrieval run and is available from the author.

S2.8.1. Retrieval script

The Crossref retrieval used the following configuration. CFG$date_filter was set to pub; the alternative values issued and online select the Crossref issued and published-online fields respectively and return different counts.

```
flt <- c(issn = issn, type = "journal-article",
       from_pub_date = "2024-01-01", until_pub_date = "2026-06-30")
rcrossref::cr_works(filter = flt, limit = 1000,
                    cursor = "*", cursor_max = 30000)
```

Title normalisation, as described in S2.4.4:

```
clean_title <- function(x) {
  x |> as.character() |>
    stringr::str_remove_all("<[^>]+>") |>
    stringr::str_replace_all("[\r\n\t]", " ") |>
    stringr::str_replace_all("[\u2010\u2011\u2012\u2013\u2014]", "-") |>
    stringr::str_squish() |>
    stringr::str_to_lower()
}
```

S2.9. Record counts

Stage	Crossref	Scopus	ERIC	Total
Records identified	2,285	3,245	1,331	6,861
Within publication window	2,285	3,245	1,068	6,598
Matched on title terms (Crossref only)	975	—	—	—
Duplicates removed, within source	0	1	0	1
Duplicates removed, across sources	0	27	135	162
Records screened	975	3,217	933	5,125
Records excluded	789	2,595	752	4,136
Included at source-level screening	186	622	181	989
Final duplicates removed	—	8	—	8
Unique included records	—	—	—	981

The aggregate flow is reported in Figure 1. Exclusion codes are defined above; source-level totals are reported in Table S2.7.

S2.10. Protocol amendments

The review was not prospectively registered, and the following decisions were taken during rather than before conduct. They are listed so that a reader can judge their effect.

1. The nineteenth Crossref title term was added after inspection showed that title-level matching without it missed articles using the abbreviation alone.
2. The ERIC query was reformulated after the initial descriptor-inclusive version was found to be non-constraining, as described in S2.3.2.
3. The E7 code was used and withdrawn, as described in S2.7.4.
4. Title normalisation was introduced after inspection showed publisher markup defeating string matching in one journal family, as described in S2.4.4.

None of these decisions was taken with knowledge of which individual records would be affected. All four are corrections of retrieval or classification failures rather than adjustments to the scope of the review question.