Supplementary Materials S1 and S3 (revised during the review) Materials S1 and S3

AI-Augmented Experiential Pedagogies for Transversal Competencies in Engineering and STEM Higher Education: A Systematic Review for Sustainable Development Education

Table S1. Search log (PRISMA-S)

Database: Web of Science Core Collection (institutional subscription, Lucian Blaga University of Sibiu). Editions: SCI-EXPANDED, SSCI, A&HCI, ESCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, IC. Date of search: 7 October 2026. Timespan: 2015–2026. Search field: TS (topic). No other databases, registers, websites, citation searching or contact with authors. No search filters other than the WoS refinement below. The search was not peer reviewed (PRESS). Records were exported as full records with cited references (tab-delimited, two files: 1–500 and 501–781).

Set	Concept	Search string (TS = topic)	Results
#1	AI technologies, 2015–2026	TS=(“artificial intelligence” OR AI OR “generative AI” OR GenAI OR “large language model*” OR LLM* OR ChatGPT OR chatbot* OR “conversational agent*” OR “pedagogical agent*” OR “virtual agent*” OR “intelligent tutoring system*” OR “affective computing” OR “emotion recognition” OR “AI-based” OR “AI-powered” OR “AI-enhanced” OR “AI-assisted” OR “AI-driven”) AND PY=(2015-2026)	678,471
#2	Alternative/ experiential	TS=(“role-play*” OR “role play*” OR	436,547

Set	Concept	Search string (TS = topic)	Results
	pedagogies	roleplay* OR “drama-based” OR “applied drama” OR “process drama” OR “educational drama” OR “drama pedagog*” OR theatr* OR theater* OR improvis* OR “design thinking” OR “design-based learning” OR gamif* OR “serious game*” OR “game-based learning” OR “educational game*” OR “virtual reality” OR “augmented reality” OR “mixed reality” OR “extended reality” OR metaverse OR “immersive learning” OR “immersive environment*” OR “simulation-based learning” OR “simulation-based training” OR “experiential learning” OR “problem-based learning” OR “project-based learning” OR “challenge-based learning”)	
#3	Transversal competencies	TS=(“soft skill*” OR “transversal competenc*” OR “transversal skill*” OR “generic skill*” OR “21st century skill*” OR “21st-	416,184

Set	Concept	Search string (TS = topic)	Results
		century skill*" OR "twenty-first century skill*" OR "professional skill*" OR "employability skill*" OR "interpersonal skill*" OR "communication skill*" OR "collaboration skill*" OR teamwork OR creativity OR "creative thinking" OR "emotional intelligence" OR empathy OR adaptability)	
#4	Engineering/STEM/ STEAM higher-education context	TS=("engineering education" OR "engineering student*" OR "engineering course*" OR "engineering curricul*" OR "engineering program*" OR "STEM education" OR "STEM learning" OR "STEM student*" OR "STEM course*" OR "STEM discipline*" OR "STEAM education" OR "STEAM learning" OR "STEAM approach*" OR "science technology engineering" OR "technical education" OR "technical universit*" OR polytechnic* OR "computing	87,073

Set	Concept	Search string (TS = topic)	Results
		education” OR “computer science student*”)	
#5	Combined	#1 AND (#2 OR #3) AND #4	793
	Refined	Document types: article, review article, proceedings paper, early access; language: English	781

Export check: 781 records, 781 unique WoS accession numbers, 726 with DOI, all in English, all with abstract; 354 proceedings papers, 350 articles, 45 reviews, 21 early-access articles, 11 other combinations. Deduplication: none needed (single database; 0 duplicate titles). Two retracted records (R429, R744) removed before screening.

Extended search (round 3): Scopus and IEEE Xplore

Rationale: amendment iii (Section 2.1). Concept blocks, eligibility criteria, codes and protocols unchanged from the WoS search.

Scopus (Elsevier), Advanced document search, searched 31 October 2026. Limits inside the query: publication years 2015–2026; document types article, review, conference paper; language English. Records exported (CSV, all fields): 4370.

Scopus query string: TITLE-ABS-KEY ("artificial intelligence" OR "AI" OR "generative AI" OR "GenAI" OR "large language model*" OR "LLM*" OR "ChatGPT" OR "chatbot*" OR "conversational agent*" OR "pedagogical agent*" OR "virtual agent*" OR "intelligent tutoring system*" OR "affective computing" OR "emotion recognition" OR "AI-based" OR "AI-powered" OR "AI-enhanced" OR "AI-assisted" OR "AI-driven") AND (TITLE-ABS-KEY ("role-play*" OR "role play*" OR "roleplay*" OR "drama-based" OR "applied drama" OR "process drama" OR "educational drama" OR "drama pedagog*" OR "theatr*" OR "theater*" OR "improvis*" OR "design thinking" OR "design-based learning" OR "gamif*" OR "serious game*" OR "game-based learning" OR "educational game*" OR "virtual reality" OR "augmented reality" OR "mixed reality" OR "extended reality" OR "metaverse" OR "immersive learning" OR "immersive environment*" OR "simulation-based learning" OR "simulation-based training" OR "experiential learning" OR "problem-based learning" OR "project-based learning" OR "challenge-based learning") OR TITLE-ABS-KEY ("soft skill*" OR "transversal competenc*" OR "transversal skill*" OR "generic skill*" OR "21st century skill*" OR "21st-century skill*" OR "twenty-first century skill*" OR "professional skill*" OR "employability skill*" OR "interpersonal skill*" OR "communication skill*" OR "collaboration skill*" OR "teamwork" OR "creativity" OR "creative thinking" OR "emotional intelligence" OR "empathy" OR "adaptability")) AND TITLE-ABS-KEY ("engineering education" OR "engineering student*" OR "engineering course*" OR "engineering curricul*" OR "engineering program*" OR "STEM education" OR "STEM learning" OR "STEM student*" OR "STEM course*" OR "STEM

discipline*" OR "STEAM education" OR "STEAM learning" OR "STEAM approach*" OR "science technology engineering" OR "technical education" OR "technical universit*" OR "polytechnic*" OR "computing education" OR "computer science student*") AND PUBYEAR > 2014 AND PUBYEAR < 2027 AND (DOCTYPE (ar) OR DOCTYPE (re) OR DOCTYPE (cp)) AND LANGUAGE (english)

IEEE Xplore (IEEE), Advanced Search > Command Search, searched 31 October 2026, publication years 2015–2026. Adaptations to IEEE Xplore limits (max. 25 terms per clause, max. 10 wildcards per search): wildcards replaced by explicit singular/plural forms, hyphens replaced by spaces. Records exported (CSV, five year-based export sets because of the 1000-record export limit): 2871.

IEEE Xplore query string: ("artificial intelligence" OR AI OR "generative AI" OR GenAI OR "large language model" OR "large language models" OR LLM OR LLMs OR ChatGPT OR chatbot OR chatbots OR "conversational agent" OR "conversational agents" OR "pedagogical agent" OR "pedagogical agents" OR "virtual agent" OR "virtual agents" OR "intelligent tutoring system" OR "intelligent tutoring systems" OR "affective computing" OR "emotion recognition") AND (("role play" OR "role playing" OR "role plays" OR roleplay OR roleplaying OR "drama based" OR "applied drama" OR "process drama" OR "educational drama" OR "drama pedagogy" OR theatr* OR theater OR improvis* OR "design thinking" OR "design based learning" OR gamif* OR "serious game" OR "serious games" OR "game based learning" OR "educational game" OR "educational games" OR "virtual reality" OR "augmented reality" OR "mixed reality" OR "extended reality" OR metaverse OR "immersive learning" OR "immersive environment" OR "immersive environments" OR "simulation based learning" OR "simulation based training" OR "experiential learning" OR "problem based learning" OR "project based learning" OR "challenge based learning") OR ("soft skill*" OR "transversal competenc*" OR "transversal skills" OR "generic skills" OR "21st century skills" OR "twenty first century skills" OR "professional skills" OR "employability skills" OR "interpersonal skills" OR "communication skills" OR "collaboration skills" OR teamwork OR creativity OR "creative thinking" OR "emotional intelligence" OR empathy OR adaptability)) AND ("engineering education" OR "engineering student" OR "engineering students" OR "engineering course" OR "engineering courses" OR "engineering curriculum" OR "engineering curricula" OR "engineering program" OR "engineering programs" OR "engineering programme" OR "engineering programmes" OR "STEM education" OR "STEM learning" OR "STEM student" OR "STEM students" OR "STEM course" OR "STEM courses" OR "STEM discipline" OR "STEM disciplines" OR "STEAM education" OR "STEAM learning" OR "STEAM approach" OR "science technology engineering" OR "technical education" OR "technical university" OR "technical universities" OR polytechnic OR polytechnics OR "computing education" OR "computer science student" OR "computer science students")

Field harmonisation (automation tool): see “Fields searched and field-harmonisation filter” below.

Deduplication by DOI and normalised title (lower-case alphanumeric, first 80 characters): Scopus 626 records already in WoS and 1 internal duplicate (627 removed; 712 to screen); IEEE Xplore 197 in WoS and Scopus, 196 in Scopus only, 1 in WoS only (394 removed; 15 to screen). Total screened in round 3: 727 (IDs S0001–S0712, I001–I015).

Screening (round 3): records divided between V.G. and A.-M.M.; AI model (Claude) screened all 727 records independently with the same instructions as below; AI decisions hidden until human decisions were complete. Agreement 705/727 (97.0%), κ = 0.94; 22 disagreements arbitrated by G.M.M. Full texts sought: 361; retrieved: 138 (retrieval closed 31 October 2026).

Fields searched and field-harmonisation filter

Database	Field searched by the query	Controlled or derived terms included	Records exported	Pass filter (title, abstract, author keywords)	Marked ineligible by automation tool
WoS Core Collection	TS (topic): title, abstract, author keywords, Keywords Plus	Keywords Plus (generated from the titles of cited references)	781	694 of 779 screened (validation only; filter not applied to WoS)	0
Scopus	TITLE-ABS-KEY: title, abstract, author keywords, index keywords	Index keywords (controlled vocabularies, e.g., Compendex, plus indexer terms)	4370	1339 (4281 when index keywords are included)	3031
IEEE Xplore	Command Search without field tags ("All Metadata"): title, abstract, author keywords, index terms and other metadata	IEEE Terms, INSPEC controlled and non-controlled terms	2871	409	2462

Why a filter: the same query run on index keywords retrieves many records in which the concepts are not substantive (e.g., "engineering education" or "artificial intelligence" assigned as broad controlled terms). The filter makes the searched field equivalent across Scopus and IEEE Xplore (title, abstract, author keywords). It is stricter than the WoS topic field, which keeps Keywords Plus; the filter was therefore not applied to WoS but validated on it.

Executable procedure: File S8, field_filter.py (regular-expression re-implementation of blocks #1 AND (#2 OR #3) AND #4 with the same truncations; hyphen and space interchangeable; acronyms AI, GenAI, LLM*, STEM*, STEAM* case-sensitive; word-boundary matching; unit tests with --selftest). The output file lists the decision and the matched blocks for every exported record (record identifier, DOI, title). Re-running the script on the exports of 31 October 2026 reproduces the counts above exactly.

Validation on WoS (File S8, filter_validation_wos.csv): 694 of 779 screened records passed; 140 of the 145 studies included in rounds 1–2 passed (96.6%); the five missed studies (R048, R356, R474, R529, R664) matched the query only through Keywords Plus.

Validation on the removed records (Table S2/S4, sheets S4b and S4b2; File S8, draw_filter_sample.py and data/filter_sample_300_screened.csv): random sample of 300 of the 5493 removed records (seed 20261104; 166 Scopus, 134 IEEE Xplore, proportional allocation). V.G. and A.-M.M. screened all 300 independently with the title/abstract criteria below; G.M.M. resolved disagreements. Result: V.G. forwarded 12 and A.-M.M. 10 records; agreement 292/300 (97.3%), κ = 0.62; exclusion code agreed in 279/285 joint exclusions; 8 disagreements resolved by G.M.M. (2 forwarded, 6 excluded). 9 of 300 records (3.0%; Wilson 95% CI 1.6–5.6%) would

have been forwarded to full text (Scopus 6/166; IEEE Xplore 3/134), corresponding to about 165 records (95% CI 87–308) across the 5493 removed records, compared with 361/727 (49.7%) among records that passed the filter. Interpretation: the filter removed mostly irrelevant records but also an estimated 19–46% of the Scopus/IEEE Xplore records that would have reached full text; this is reported as a limitation of coverage (Section 4.7 of the article).

Later retrieval: repeated individual searches

Period: 1–5 November 2026. Scope: all 317 reports sought but not retrieved after rounds 1–3 (WoS 94; Scopus/IEEE Xplore 223). Methods: repeated search of each report through the institutional subscriptions of Lucian Blaga University of Sibiu, the Zotero open-access look-up (Unpaywall-based), publisher websites, institutional and subject repositories, ResearchGate pages posted by the authors, and conference proceedings archives; only legally accessible versions were used. Study investigators were not contacted.

Result: 35 further reports obtained (5 from the WoS search: R093, R154, R332, R516, R543; 30 from the Scopus/IEEE Xplore search). Full-text assessment with the unchanged criteria FT1–FT4: 25 included, 10 excluded (FT1 = 1, FT2 = 7, FT3 = 1, FT4 = 1; Table S2/S4, sheet S2b4). After round 4: 365 of 647 reports retrieved (56.4%), 282 not retrieved (WoS 89, Scopus/IEEE Xplore 193; sheet S4a).

Later retrieval: library requests

Period: 5 November 2026. Scope: the 282 reports still not retrieved after round 4. Methods: requests for the remaining reports through the library of Lucian Blaga University of Sibiu; the reports obtained were delivered as a Zotero export with attached PDF files. Of the 93 PDF files received, 71 corresponded to reports in the not-retrieved list (matched by DOI or normalised title), 5 were duplicate files and 17 were reports already assessed in round 4; only the 71 new reports were assessed. Study investigators were not contacted.

Result: 71 reports obtained (14 from the WoS search, 56 Scopus, 1 IEEE Xplore). Full-text assessment and extraction with the unchanged criteria FT1–FT4 and the protocol v4 below: 55 proposed for inclusion, 16 for exclusion (FT1 = 4, FT2 = 9, FT3 = 1, FT4 = 2; Table S2/S4, sheet S2b5). One excluded report (S0256) had been retracted by IEEE (authorship offered for sale) and was coded FT4. After round 5: 436 of 647 reports retrieved (67.4%), 211 not retrieved (WoS 75, Scopus/IEEE Xplore 136; sheet S4a).

Verification: V.G. and A.-M.M. assessed the 16 proposed exclusions independently (agreement 14/16 decisions, 12/14 codes; G.M.M. excluded the two disputed reports); one author verified all 904 extracted and coded values of the 55 included studies (37 corrected) and a second author double-checked 12 studies (35/41 values agreed). All corrections are listed in Table S2, sheet S2l.

Amendments to the methods

(i) Additional full-text retrieval for reports not obtained in the first search: a second open-access search of WoS reports (22 October 2026), repeated individual searches of all non-retrieved reports (1–5 November 2026) and requests through the university library (5 November 2026).
(ii) 22 October 2026: systematic reviews excluded from the outcome synthesis to avoid double counting.

(iii) 31 October 2026: extension of the search to Scopus and IEEE Xplore with the field-harmonisation filter; reliability benchmark of the AI-assisted screening; description of non-retrieved reports.
(iv) 1–5 November 2026: independent assessment of 300 records removed by the filter; coding of all primary studies with explicit rules for scope class, outcome records (measure type, contrast, time point, AI at assessment, inferential test, direction), AI-use structure and MMAT ratings (protocol v4 in Document S3); GRADE-based certainty rating of the main findings. Eligibility criteria and screening codes were not changed.

Title/abstract screening criteria and codes (also used as the AI screening instructions)

Review topic: AI-augmented alternative/experiential pedagogies for developing transversal competencies in engineering / STEM / STEAM higher education (systematic review, PRISMA 2020).

Judge ONLY from title, abstract, keywords, source, doc type. When the abstract leaves a criterion unclear but plausible, choose UNCERTAIN (it goes to full-text). Do not exclude on a guess.

INCLUDE when ALL four hold

- **C1 Context**: tertiary / higher education (university, college, polytechnic, tertiary vocational) or professional/continuing training of engineers, in engineering, computing/IT, or other STEM/STEAM disciplines (architecture and design count as STEAM). Mixed samples that include tertiary STEM students are fine. “STEM education” with no stated level and no sign of K-12 = UNCERTAIN on C1.
- **C2 AI component**: an AI technology is used as part of teaching/learning (generative AI, LLMs, ChatGPT, chatbots, conversational/pedagogical/virtual agents, intelligent tutoring systems, AI-driven feedback or learning analytics used to support learners/teachers, affective computing/emotion recognition, AI-driven simulations/VR/game characters), or the paper is a review/framework about such use.
- **C3 Pedagogy or competency link**: the AI is embedded in, or studied together with, an alternative/experiential pedagogy (role-play, drama/theatre/improvisation, design thinking, gamification/serious games/game-based learning, VR/AR/MR/XR/metaverse/immersive learning, simulation-based learning, experiential learning, problem-/project-/challenge-based learning, hackathons/studio learning), OR the study targets, measures, or substantively discusses transversal competencies (communication, collaboration/teamwork, creativity, emotional intelligence, empathy, adaptability, soft/professional/21st-century/employability skills, critical thinking, generic problem-solving, self-regulation, leadership, ethics/reflection).
- **C4 Publication type**: empirical study (any design), design/implementation report, systematic/scoping/bibliometric review, or substantive conceptual/framework paper.

EXCLUDE codes (give the FIRST that applies)

- **E1** Not about education/learning (AI in industry, business, clinical practice, robotics engineering itself, etc.).

- **E2** Wrong population/context: only K-12/school pupils, or only non-STEM disciplines (medicine, nursing, health professions, languages, business, humanities, law), or general public.
- **E3** AI absent or only peripheral (mentioned in passing/future work), OR AI is only the subject being taught (AI/ML course or AI-literacy curriculum) with no AI tool used in the pedagogy.
- **E4** No link to alternative/experiential pedagogy or transversal competencies (e.g., AI only for automated grading of technical answers, dropout/performance prediction, code-correctness assessment, plagiarism detection, purely technical knowledge/skill acquisition, generic usage/acceptance surveys without competency or pedagogy focus).
- **E5** Ineligible publication type (editorial, proceedings front matter, abstract-only, erratum, book review, retracted).

Extraction fields (fill even for excluded records when evident; "" if not stated)

- `level`: higher_ed | K12 | mixed | professional | unclear
- `discipline`: e.g., engineering (general), mechanical eng., computing/CS, STEM (general), STEAM, architecture/design, other
- `ai_type`: e.g., generative AI/LLM, chatbot/agent, ITS, AI feedback/analytics, affective computing, AI in VR/simulation, other
- `pedagogy`: comma list from: role-play, drama/theatre/improvisation, design thinking, gamification/games, immersive VR/AR/XR, simulation, experiential, PBL/project/challenge-based, other, none
- `competencies`: comma list from: communication, collaboration/teamwork, creativity, emotional intelligence/empathy, adaptability, critical thinking, problem solving, soft skills (general), self-regulation, other, none
- `study_type`: empirical-quantitative | empirical-qualitative | mixed-methods | design/implementation | review | conceptual | other

Document S3. Protocols given to the AI model

The AI model (Claude, Anthropic) received the following written instructions. All AI outputs were verified by the authors as described in Sections 2.4–2.7 of the article.

Full-text retrieval and eligibility protocol (AI model)

Review: PRISMA 2020 systematic review of AI-augmented alternative/experiential pedagogies for transversal competencies in engineering / STEM / STEAM higher education. These records already passed title/abstract screening (verified by three human authors). Your job per record: (1) try to obtain the full text legally and freely, (2) if obtained, extract eligibility facts and propose a full-text decision. Human authors will verify everything you propose.

1. Retrieval (use ONLY the WebFetch and WebSearch tools; never curl/wget/python HTTP)

1. WebFetch `https://doi.org/<DOI>`. If it reports a redirect, WebFetch the redirect URL. Ask the fetch prompt whether the FULL TEXT (methods + results, not only the abstract) is on the page, and to extract the facts in section 2.
2. If the publisher page has only the abstract (typical for IEEE Xplore, ACM, Springer chapters, Wiley, T&F behind paywall) or the fetch fails (403, timeout), run ONE WebSearch with the exact title in quotes (add "pdf" if useful) to look for a legitimately free version: arXiv, institutional/university repository, author homepage, conference open proceedings, PubMed Central, Semantic Scholar PDF, publisher free-access page. Do NOT use Sci-Hub, LibGen or similar pirate sites. ResearchGate/Academia.edu: only if WebFetch actually returns the full text; otherwise skip them.
3. If a candidate URL looks promising, WebFetch it and confirm it really contains the full paper (methods and results) and that it is the same paper (title + first author).
4. Spend at most ~4 tool calls per record. If no full text is confirmed, mark `access = "NOT FOUND"` and move on.

`access` values: "PUBLISHER" (full text on publisher page), "PREPRINT/REPOSITORY" (author/repository/arXiv version), "NOT FOUND". `fulltext_url`: the URL where you confirmed the full text ("" if NOT FOUND).

2. Facts to extract when full text is available ("" when not reported)

- `country`, `level` (undergraduate / postgraduate / mixed HE / K-12 / professional / unclear), `discipline`, `n_participants` (number or "n/a")
- `design`: experiment/RCT | quasi-experiment | pre-post single group | survey | qualitative | mixed methods | case study/experience report | design/implementation (no evaluation) | systematic/scoping review | conceptual
- `ai_tool` (named tool/model) and `ai_role`: one of "learner-facing tutor/assistant", "role-play / simulated partner", "scenario/content generator", "feedback/assessment of competencies", "learning analytics / affective sensing", "co-creative partner (ideation/design)", "other"
- `pedagogy`: role-play, drama/theatre/improvisation, design thinking, gamification/games, immersive VR/AR/XR, simulation, experiential, PBL/project/challenge-based, other
- `competencies`: communication, collaboration/teamwork, creativity, emotional intelligence/empathy, adaptability, critical thinking, problem solving, self-regulation, ethics, other
- `competency_measure`: how the competency was measured (validated scale / self-report survey / performance rubric / qualitative / not measured)
- `key_finding`: max 35 words, factual, from the results section

3. Proposed full-text decision

`ft_decision`: "INCLUDE" | "EXCLUDE" | "NO FULL TEXT" Exclude only if the full text shows one of: - FT1 population/context is not tertiary STEM/engineering/computing/STEAM (e.g., K-

12 only, medicine only, non-STEM only) - FT2 AI is not used in the teaching/learning process (only the subject taught, only a researchers' analysis tool, only mentioned or proposed for future work) - FT3 no alternative/experiential pedagogy AND no transversal competency is targeted, measured or substantively discussed - FT4 not a usable report: abstract/extended abstract/poster without methods or results, workshop description, editorial - FT5 duplicate report of a study already in the set (name the other ID if you can tell) Reviews and conceptual frameworks are eligible if they meet FT1–FT3. `ft_reason`: "" or FT1..FT5, plus `ft_rationale` (max 30 words).

Data extraction and MMAT appraisal protocol (AI model)

Review: PRISMA 2020 systematic review — how AI augments alternative/experiential pedagogies (role-play, drama/improvisation, design thinking, games/gamification, immersive VR/AR/XR, simulation, experiential, PBL/project/challenge-based) to develop transversal competencies in engineering/STEM/STEAM higher education. Each study below already passed full-text eligibility (proposal). Read the full text carefully and extract ONLY what the text reports. Leave "" when not reported. Never infer numbers.

Output keys (one JSON object per study)

- `id`
- `confirm_include`: "yes" | "no" — second look at eligibility (C1 tertiary STEM/engineering/computing/STEAM context; C2 AI used in teaching/learning, not only as subject or researcher tool; C3 experiential/alternative pedagogy or transversal competency targeted/measured/substantively discussed; C4 usable report with methods). If "no", give `exclude_reason` (FT1..FT5 as in fulltext_protocol.md) and say why in `notes`.
- `country`: country/countries where the study took place (for reviews: "review")
- `level`: undergraduate | postgraduate | mixed HE | unclear
- `discipline`: main discipline(s), short
- `n`: number of participants analysed (integer as string) or "n/a"
- `duration`: intervention length (e.g., "1 session", "6 weeks", "1 semester") or ""
- `design`: experiment/RCT | quasi-experiment (with comparison group) | pre-post single group | cross-sectional survey | qualitative | mixed methods | case study/experience report | design/implementation (no learner data) | systematic/scoping review | conceptual
- `comparison_group`: "yes" | "no" | "n/a"
- `ai_tools`: named tools/models (e.g., "ChatGPT (GPT-4o)", "custom RAG chatbot")
- `ai_role`: one primary from: tutor/assistant | role-play or simulated partner | scenario/content generator | feedback/assessment | co-creative partner | analytics/affective sensing | immersive agent (AI character/agent inside VR/AR/game) | other
- `ai_mechanism`: 1–2 sentences: HOW the AI augments the pedagogy in this study (what the AI does, at which step, for whom)

- `pedagogy`: semicolon list from: role-play; drama/theatre/improvisation; design thinking; games/gamification; immersive VR/AR/XR; simulation; experiential; PBL/project/challenge-based; collaborative learning; flipped/blended; other
- `competency_outcomes`: list of objects, one per transversal competency the study targets or measures: {"competency": one of communication | collaboration/teamwork | creativity | emotional intelligence/empathy | adaptability | critical thinking | problem solving | self-regulation/metacognition | ethics/responsibility | leadership | other, "measure": validated scale | self-report survey | performance rubric/test | qualitative | not measured, "result": positive | mixed | no effect | negative | not measured, "evidence": max 25 words with the key statistic or quote-paraphrase}
- `other_outcomes`: max 25 words (engagement, motivation, knowledge, usability…)
- `risks`: semicolon list of risks/challenges the authors REPORT from: over-reliance; accuracy/hallucination; academic integrity; privacy/data; equity/access; cost/infrastructure; teacher readiness/workload; bias/ethics; reduced human interaction; technical problems; none reported
- `sustainability_link`: "yes" if the paper explicitly links to sustainability/SDGs/ESD, else "no"; if yes add 10 words in `sustainability_note`
- `limitations`: max 25 words, authors' own stated limitations
- `mmat_category`: qualitative | quantitative RCT | quantitative non-randomized | quantitative descriptive | mixed methods | not applicable (review/conceptual/design without data)
- `mmat`: object with S1, S2, and the five criteria of the chosen category (keys like "1.1"..."1.5" / "2.1"..."2.5" / "3.1"..."3.5" / "4.1"..."4.5" / "5.1"..."5.5"), each "Yes" | "No" | "Can't tell". Empty object if not applicable.
- `notes`: anything a human verifier should know (max 30 words)

MMAT 2018 criteria (Hong et al., 2018)

Screening: S1 Are there clear research questions? S2 Do the collected data allow to address the research questions? 1 Qualitative: 1.1 approach appropriate to answer the RQ; 1.2 data collection methods adequate; 1.3 findings adequately derived from the data; 1.4 interpretation sufficiently substantiated by data; 1.5 coherence between data sources, collection, analysis and interpretation. 2 RCT: 2.1 randomization appropriately performed; 2.2 groups comparable at baseline; 2.3 complete outcome data (≥80%); 2.4 outcome assessors blinded; 2.5 participants adhered to the assigned intervention. 3 Non-randomized: 3.1 participants representative of the target population; 3.2 measurements appropriate regarding outcome and intervention; 3.3 complete outcome data; 3.4 confounders accounted for in design and analysis; 3.5 intervention administered as intended during the study period. 4 Quantitative descriptive: 4.1 sampling strategy relevant to address the RQ; 4.2 sample representative of the target population; 4.3 measurements appropriate; 4.4 risk of nonresponse bias low; 4.5 statistical analysis appropriate. 5 Mixed methods: 5.1 adequate rationale for using mixed methods; 5.2 components effectively integrated; 5.3 outputs of integration adequately interpreted; 5.4 divergences between quantitative and qualitative results adequately addressed; 5.5 components adhere to the quality criteria of each tradition. Pre-post single-group studies → category 3 (non-randomized). Cross-sectional surveys → category 4. Be honest and conservative: "Yes" only when the paper clearly reports it.

Additions used in the second retrieval round (34 reports)

Calibrated MMAT decision rules (agreed by the human authors at calibration; apply strictly)

- "Yes" only when the paper explicitly reports the item.
- Blinding of outcome assessors (2.4): "Can't tell" when not mentioned (not "No"), except for direct self-assessment.
- When a qualitative analysis method is not described at all, 1.3 and 1.5 = "No".
- Mixed methods 5.5 = "No" if any component fails its own tradition's criteria (any "No").
- 4.4 (non-response) = "Yes" only if response/retention >= 80% is reported.
- 3.5 = "No" if the paper reports the intervention was delivered at a clearly lower dose than planned.
- 3.2 = "No" when outcomes are only estimates without instruments or data.
- 4.5 = "No" when no statistical analysis at all is reported for a quantitative descriptive study.
- Reports without learner outcome data (reviews, conceptual, design/implementation, activity reports with no learner data): mmat_category = "not applicable", mmat = {}.

Working rules

- Read ONLY the text file given for each ID (it is the full text extracted from the PDF). Do not use the web.
- Output one JSON object per line (JSONL) with exactly the keys listed in "Output keys"; competency_outcomes is a list of objects; mmat is an object.
- Also add key "ft_rationale" (max 30 words) explaining the eligibility decision.
- Be conservative and factual; never infer numbers.

Re-coding protocol v4 (applied to all primary studies)

The following instructions were given to the AI model for the re-coding of every primary study from its full text (rounds 1–4) and, combined with the full-text eligibility and extraction protocol above, for the reports obtained in round 5. Every value was verified by one author against the full text, with an independent double check of a random 20% of studies by a second author (Section 2.6 of the article).

Context: PRISMA 2020 systematic review of AI-augmented experiential pedagogies for transversal competencies in engineering/STEM higher education. The study below is ALREADY INCLUDED. You re-code only the fields defined here, from the full text. Human authors verify everything you produce. Extract ONLY what the text reports. Never infer or compute numbers yourself; copy them exactly as printed (means, SDs, n, test statistics, p, effect sizes, CIs). Use "" when not reported.

Transversal competencies (use these labels): communication | collaboration/teamwork | creativity | critical thinking | problem solving | self-regulation/metacognition | ethics/responsibility | emotional intelligence/empathy | adaptability | leadership | systems thinking | other (name it in competency_detail).

A. Study-level fields

id

study_type: primary empirical | review | conceptual | design/implementation report (no learner outcome data)

scope_class:

"core" = AI is used in teaching/learning INSIDE an experiential/alternative pedagogy (role-play, drama/improvisation, design thinking, games/gamification, immersive VR/AR/XR, simulation, experiential learning, PBL/project/problem/challenge-based, hackathon/studio) AND at least one transversal competency is targeted or measured;

"competency-only" = AI used in teaching/learning and a transversal competency targeted/measured, but NO experiential pedagogy;

"pedagogy-only" = AI inside an experiential pedagogy, but no transversal competency targeted or measured;

"contextual" = reviews, conceptual papers, design reports without learner data, or surveys of general AI use not tied to a designed learning activity.

design_detail: one line: e.g., "quasi-experiment, two intact classes, post-test only", "RCT, individual randomisation, pre/post", "historical cohort comparison".

comparator: no-AI control | alternative AI condition (e.g., restricted vs unrestricted, different tools) | historical cohort | none

baseline_equivalence: reported equivalent | reported non-equivalent | not reported | n/a

ai_use_structure: structured | unconstrained | mixed/both conditions | unclear (definitions below)

structure_features: semicolon list of the features present, from: phase-restricted | solution-withholding (hints/Socratic/scaffolding) | verification-required (prompt logs, justify/defend, compare with own work) | sequenced-after-own-attempt | usage-limits (turn caps, individual accounts, time limits) | instructor-curated AI output | none

structure_evidence: max 30 words quoting/paraphrasing what the paper says about how AI use was organised.

Operational definitions (AI-use structure)

Structured: the design imposes at least one explicit constraint on HOW, WHEN or FOR WHAT learners use AI (any feature in the list above other than "none"), reported in the methods.

Unconstrained: learners have open, on-demand access during the task without such constraints, OR the AI produces the core generative/problem-solving output (ideas, solution, code, design) that students then use.

Mixed/both conditions: the study compares structured with unconstrained use, or different groups receive different structures.

Unclear: the paper does not describe the organisation of AI use sufficiently.

B. Outcome records (`outcomes`: list)

One record per competency × measure type × contrast. If the same competency is measured with two different instruments (e.g., a validated scale AND a rubric), create two records. If several items/subscales of the SAME instrument measure the same competency, aggregate them into one record and use the aggregation rule below. Fields:

competency, competency_detail (instrument subscale or wording, max 12 words)
measure_type:

"performance" = rubric, test, expert/teacher rating of a product, task or observed behaviour;

"validated self-report" = published questionnaire/scale with a citation to its source or reported validity/reliability evidence (e.g., alpha) in this study;

"non-validated self-report" = author-made survey or single items, no validity/reliability evidence;

"qualitative" = interviews, reflections, open responses, observation notes analysed qualitatively;

"not measured" = competency targeted or discussed but no data.

contrast:

"between-group" = AI condition compared with a comparator (post-test, gain or adjusted means);

"within-group" = change over time inside the AI group only (pre-post);

"correlational" = association between AI use and the competency;

"descriptive" = post-only description without any contrast (e.g., % agreeing);

"none" (not measured).

If a study reports BOTH a between-group and a within-group result for the same competency/measure, create TWO records.

timepoint: immediate (end of intervention) | delayed follow-up (state interval in competency_detail) | during intervention | n/a
ai_at_assessment: "with AI" (the assessed task/product was produced while AI was available) | "without AI" (assessed without AI support, e.g., closed test, transfer task, delayed test without AI) | unclear | n/a. This separates AI-assisted task performance from learners' own competency.
inferential_test: yes (a statistical test or CI is reported for this contrast) | no
result (apply these rules strictly):

"positive": between-group: statistically significant difference favouring the AI condition ($p < .05$ or CI excluding null); within-group: significant improvement; descriptive/qualitative/no test: authors report improvement or benefit without contrary evidence → still "positive", but inferential_test = no.

"negative": significant difference favouring the comparator, or significant decline; or (no test / qualitative) authors report deterioration or harm without contrary evidence.

"non-significant": a test was reported and was not significant (this is NOT evidence of no effect).

"no difference (equivalence)": ONLY when an equivalence/non-inferiority test or an explicit CI-based argument shows no meaningful difference.

"mixed": within one record, the aggregated subscales/items/time points/groups point in different directions (e.g., some significant positive, some non-significant or negative), or qualitative data report both benefits and harms for that competency.

"not determinable": data reported but direction cannot be determined; or "not measured".

Multiple time points: code the immediate post-test in the main record; add a separate record for any delayed follow-up.

stats: copy exactly: group means, SDs and n (per group), test statistic, df, p, effect size and CI, e.g. "AI M=8.5 SD=1.2 n=36; control M=7.4 SD=1.6 n=35; t(69)=2.30, p=.012; d=0.55". "" if none.
effect_size_reported: the effect size as printed with its metric (e.g., "d=1.35", "eta2=0.114", "r=0.34"), "" if none.
ci_reported: CI as printed, "" if none.
evidence: max 25 words, the key finding in the authors' words.

C. MMAT 2018 re-appraisal with basis (only for primary empirical studies with learner outcome data)

mmat_category: qualitative | quantitative RCT | quantitative non-randomized | quantitative descriptive | mixed methods
mmat: S1, S2 and the five criteria of the category, each "Yes" | "No" | "Can't tell".
For MIXED-METHODS studies ALSO rate the components: mmat_qual_component (1.1–1.5) and mmat_quant_component with the criteria of the matching quantitative category (2.x RCT, 3.x non-randomized, or 4.x descriptive) and mmat_quant_component_category. Criterion 5.5 = "Yes" only if both components meet their criteria.
Rating rules (MMAT guidance):

- "Yes": the report gives information showing the criterion is met.
- "No": the report gives information showing the criterion is NOT met (a demonstrated weakness: e.g., confounders reported and not controlled; non-response 45%; instrument shown to be inappropriate; analysis reported and inappropriate).
- "Can't tell": the report does not give enough information to judge (reporting gap). Missing information is ALWAYS "Can't tell", never "No". (Exception: 2.4 blinding not mentioned = "Can't tell".)

mmat_basis: for every "No", one short phrase per criterion: {"3.4": "groups differed at baseline, no adjustment", ...}.
mmat_notes: max 30 words.

D. Output

One JSON object per study (JSON Lines). Keys: id, study_type, scope_class, design_detail, comparator, baseline_equivalence, ai_use_structure, structure_features, structure_evidence, outcomes (list of records with the keys in B), mmat_category, mmat, mmat_qual_component, mmat_quant_component_category, mmat_quant_component, mmat_basis, mmat_notes, notes (max 30 words for the human verifier). Reviews/conceptual/design reports: fill study_type, scope_class, outcomes = [] and mmat fields empty ({}, "").