VOLUME 01 / 2026

Designership Press

From Paper Agents to Accountable Student Scholarship: A Claim-Evidence-Contribution Ledger for K-12 Research

Dr. Freedom Cheteni

Original critical integrative analysis and prospective evaluation protocol

Research edition 1.0 | Prepared 2026-10-16

D05 / Accountable scholarship

From Paper Agents to Accountable Student Scholarship: A Claim-Evidence-Contribution Ledger for K-12 Research

Dr. Freedom Cheteni

The Research University | Volume 1, 2026 | Edition 1.0 | D05

Original critical integrative analysis and prospective evaluation protocol

Original analyses and prospective protocols. New independent human peer review is not recorded for this edition.

DOI: not assigned to this research edition.

Read online | Author ORCID: 0009-0006-3507-0094 | Editorial record

Abstract

Paper companions can make student research easier to inspect, but fluent explanations and persistent identifiers do not establish evidential reliability or authorship. This original critical integrative analysis proposes a claim-evidence-contribution ledger linking consequential manuscript claims to version-specific sources, human intellectual contributions, software operations, and editorial decisions. Drawing on original source-evaluation and retrieval research together with scholarly infrastructure standards, it distinguishes retrieval from verification, authentication from validation, and preservation from intellectual ownership. The architecture preserves original source attribution and rights, treats companions as software rather than authors, and protects the learner's option to demonstrate equivalent scholarship without public exposure. A prospective two-stage evaluation combines a 600-item, 100-manuscript benchmark of citation fidelity, coverage, answerability, abstention, and adversarial robustness with an eight-week, 24-class feasibility study of independent source-reasoning transfer and teacher workload. All sample sizes and procedures are proposed; no benchmark results, classroom observations, or completed independent human peer review are claimed. The contribution is a testable institutional design for publication as evidence of learning rather than as a proxy for prestige.

Keywords: Designership, K-12 research, paper companions, provenance, student authorship, source evaluation, assessment, ORCID, human editorial authority

1. Publication as evidence of learning

A school research publication makes two promises: that readers can inspect an intellectual contribution, and that the named learners can answer for the work attributed to them. An elegant article, a persistent identifier, or a fluent conversational interface does not establish either promise. A student may assemble accurate sentences without understanding the inferential steps between them. Conversely, an emerging researcher may offer a valuable observation in imperfect prose. The institutional challenge is to preserve the relationship between contribution, evidence, interpretation, and revision as assistance becomes more capable.

This article proposes a claim-evidence-contribution ledger for K-12 scholarship and the paper companions that make scholarship easier to examine. A ledger is a structured, revisable account of consequential decisions, not a transcript of every keystroke. It links an article's claims to version-specific evidence and identifies the human actions and software operations through which those claims were developed. Its educational purpose is to make responsible judgment visible without equating surveillance with rigor.

The proposal draws on formative assessment's attention to evidence used in subsequent learning decisions [R2] and research showing that evaluation of digital sources can be explicitly taught [R1]. It extends those concerns to a publication environment where a companion may retrieve passages, explain terminology, and help readers interrogate an article. The original contribution is an institutional design and a falsifiable evaluation plan. No classroom trial, software benchmark, or independent human peer review is reported as completed here.

2. Method and scope of the argument

This is a critical integrative analysis with a prospective evaluation protocol. Sources were selected purposively from original technical studies, original education research, and the organizations responsible for relevant scholarly standards. Their functions differ. Technical studies motivate test conditions; education research motivates learning outcomes; standards define metadata and attribution boundaries; child-centered guidance informs governance. These materials are not interchangeable evidence of an intervention effect.

The literature search and authoritative-page verification were conducted on 16 October 2026. The selection was not a systematic review: no exhaustive search, preregistered screening, pooled estimate, or claim of comprehensive coverage is made. The synthesis asks what a school publishing system would need to preserve if its outputs were to remain inspectable after assistance, republication, revision, and changes of software provider.

The unit of analysis is the publication episode: a learner or team frames a question, encounters sources, constructs an interpretation or study, responds to criticism, and decides what may become public. Public dissemination is optional. The same intellectual standard must be achievable through a confidential school assessment, accessible document, or supervised oral defense. This distinction matters because the right to demonstrate learning should not depend on accepting permanent public exposure.

3. Four boundaries that a paper agent must respect

First, retrieval is not verification. Retrieval-augmented generation connects language generation to an external information store [R6], but retrieval alone does not establish that a cited passage supports a particular assertion. Gao and colleagues separate correctness, fluency, and citation quality in their ALCE benchmark [R3]. The proposed ledger therefore records the evidence relation itself rather than treating the existence of a citation as sufficient.

Second, inspectability is not authorship. A learner may inspect a generated explanation without having produced the contribution described by it. Authorship must follow actual intellectual participation and accountability. CRediT supplies a standardized vocabulary for describing contributions [R8]; the ledger uses that vocabulary where appropriate while retaining a plain-language account that students and families can understand.

Third, authentication is not validation. An authenticated ORCID connection establishes the connection between a record holder and an identifier through ORCID's authorization workflow [R10]. It does not establish that the person conducted an experiment, wrote an article, or reviewed another person's work. Fourth, access is not permission. A publicly readable document may lack the rights required for republication. The ledger must keep rights evidence separate from availability and from the educational merits of a source [R12].

These boundaries prevent a predictable category error: allowing the sophistication of infrastructure to stand in for the quality of the scholarship it carries. A companion should make unsupported claims easier to discover. It should not become an additional surface through which unsupported authority is manufactured.

4. The claim-evidence-contribution ledger

The proposed ledger contains five linked records. A work record identifies the article, its authors, venue, version, dates, identifiers, and disclosure statements. A source record identifies each exact input version, its creator, source location, retrieval date, license evidence, and file checksum. A claim record gives a stable identifier to a consequential proposition and classifies it as observation, source-supported assertion, interpretation, hypothesis, or recommendation. A contribution record describes who performed the relevant intellectual work. A decision record preserves the criticism received, the response, and the authority responsible for release.

The architecture adapts the distinction among entities, activities, and responsible agents in W3C PROV-DM [R7]. Its additional educational distinction is between a source relation and an ownership relation. A student can reason from an existing study without becoming its author. A software maintainer can create the retrieval interface without becoming the author of the studies it retrieves. A publisher can preserve the resulting object without acquiring the learner's intellectual contribution.

For a consequential claim, the minimum evidence relation includes a source-version identifier, a passage locator, the relevant excerpt or allowable reference to it, the asserted relation, and a human review state. Relations include supports, contradicts, limits, motivates, and does not resolve. The last category is important: an unresolved relation can be educationally productive. Where support requires several passages, the record identifies the combination and the inference connecting them. A single plausible sentence is not allowed to conceal a missing premise.

The ledger does not claim to prove truth. A checksum establishes whether bytes match a recorded object; it does not establish that the object's claims are correct. A dated teacher decision establishes who made a judgment; it does not make that judgment infallible. Records must therefore support correction, counterevidence, and appeal. Earlier decisions remain legible while superseding decisions are clearly marked. Public readers see a concise evidence trail; authorized reviewers can inspect the richer record needed to assess it.

5. A worked hypothetical example

Consider a hypothetical secondary-school team investigating shade and playground surface temperature. Its initial claim is that planting trees would reduce heat exposure at its school. A source describing a relationship in another city may motivate the project, but it does not establish the local intervention effect. The ledger classifies that claim as a hypothesis, records the source's setting and limitations, and identifies the students' contribution as question formation and study design rather than discovery of an already established local result.

The team then collects temperature measurements under a documented sampling plan. Observation records distinguish instrument readings from calculated summaries. A student who notices that shaded measurements were taken earlier in the day contributes a consequential validity challenge. The resulting revision-matching observation times or narrowing the conclusion-receives a contribution entry. This is more informative than counting paragraphs written, prompts entered, or pages in a final portfolio.

A companion asked whether the project proves that tree planting causes lower school-wide exposure should locate the design and limitations passages and explain their scope. If the manuscript reports only observational measurements, the companion must not manufacture randomized evidence. If asked for a student's home address or an unpublished participant name, it must not search private records. If asked to explain a technical term absent from the paper, it may offer a clearly labeled general explanation only in an authorized mode, separated from assertions about the paper itself.

The example is illustrative and contains no actual school observations. Its purpose is to specify the ledger's discriminations: what came from prior scholarship, what learners observed, what they inferred, what software helped express, and what a reviewer accepted as appropriately bounded.

6. Designing the course-to-publication workflow

At project opening, learners receive the assessment criteria, source-use expectations, allowed forms of assistance, publication choices, and routes for correction. Teachers select a small set of consequential checkpoints rather than requiring continuous activity capture. A practical sequence includes question justification, source appraisal, design or argument review, one substantive revision, and a final explanation of contribution. The ledger should be generated from these ordinary instructional events wherever possible.

Before public submission, the learner chooses several claims and explains why their evidence is adequate, where it is weak, and how a plausible counterexample would change the conclusion. The teacher can ask for a comparable explanation using a new passage. This creates a distinction between a polished product and transferable judgment. Digital-source evaluation instruction provides a relevant educational precedent [R1], but whether the proposed publication workflow improves transfer remains an empirical question.

The final assessment includes the artifact, the evidence trail, and a short defense. It must not reward unnecessarily extensive logging. A multilingual learner can demonstrate judgment through a supported oral explanation, translated planning notes, or accessible visual annotations. Assistance with expression must remain distinguishable from assistance that supplies the reasoning being assessed. The

relevant boundary is the intended learning outcome, established before the task, rather than a retrospective guess based on writing style.

Publication follows a separate editorial decision. A strong course performance need not imply that an article is ready for a public journal; an editorial revision need not invalidate the learning demonstrated. Journals should accept negative findings, constrained interpretations, and failed designs when the reporting is responsible. An institution that celebrates only successful-looking projects will encourage learners to hide the very revisions through which expertise becomes visible.

7. Source-grounded companions as accountable interfaces

A companion should initially operate over a bounded, versioned manuscript package. Every paper-specific answer should identify the passages on which it relies and expose a route to the original text. Interpretive summaries, direct quotations, and suggested questions require different labels. The system should disclose whether an answer draws solely from the manuscript, from a verified supplementary source, or from a general explanation that has not been established by the paper.

Unanswerability must be a normal state. The SQuAD successor introduced by Rajpurkar and colleagues deliberately includes plausible questions unsupported by the supplied paragraph [R4]. This motivates testing questions such as an unreported sample size, an unmeasured long-term outcome, or a nonexistent approval date. A good companion says what is missing and points to the relevant methods or disclosure section, rather than treating every question as an invitation to complete the record imaginatively.

The interface must also keep document content separate from system authority. A sentence inside a paper telling software to ignore its instructions is source text, not permission to change access rules. Benchmark conditions should include irrelevant passages and misleading contextual cues: Shi and colleagues demonstrated distractibility using an arithmetic benchmark with irrelevant context [R5]. That result is not a direct estimate of error in school paper companions; it supplies a reason to test this failure mode locally.

8. Authorship, identifiers, and credit without appropriation

The system should maintain separate lists for article authors, acknowledged contributors, editorial decision makers, software creators, and platform stewards. A student who originated and defended a study must not disappear because an adult edited the prose. An adult's supervisory position does not automatically justify authorship. Contribution statements should be negotiated early, revisited after substantive changes, and approved by all named human authors.

COPE's position treats AI tools as ineligible for authorship because they cannot accept the responsibilities required of authors; it also places responsibility for assisted content on human authors [R9]. Accordingly, the companion is described as software with a version, provider, documented function, and limitations. Its role can be disclosed in methods or acknowledgments and cited as software when appropriate. It is not given a human identity or an invented credential.

ORCID identifiers should be obtained through authentication, stored with the provenance of that authentication, and associated only with the actual person concerned [R10]. A publisher's founder may be credited for platform conception or software stewardship where those contributions are real. That does not justify adding every imported source article to the founder's publication list or replacing its original authors. If a separately released companion constitutes a legitimate software output, its own creators and contribution record require review before any authorized ORCID update.

Persistent identifiers likewise identify distinct objects rather than confer quality. Existing source identifiers remain attached to their original works. Relationships among manuscripts, versions, data, and software can be represented using the typed relationships supported by Crossref [R11]. Registration decisions require the relevant publisher authority; the ledger is designed to make those decisions auditable, not to substitute for them.

9. Rights, version integrity, and intellectual custody

Each source package should preserve the original title, complete author attribution, publisher or repository, date, exact source version, original identifier, and authoritative license evidence. Where a permissible copy is preserved, its checksum and file format are recorded. The extraction used by a companion receives its own checksum, extraction-method version, and passage mapping. An updated source becomes a new linked version; an extraction change is not silently represented as an unchanged original.

Creative Commons recommends attribution through title, author, source, and license, together with identifying modifications where appropriate [R12]. For the proposed journal workflow, version-specific rights review should precede preservation and public reuse. Repository metadata licenses must not be presumed to license the article itself. Excluded third-party figures, restrictive notices, or conflicting evidence require a hold or a narrower use, rather than an automatic ingestion decision.

The ledger should preserve the provenance of a license assertion as carefully as the assertion itself. It also needs a correction route when source metadata conflict with the document. Data-documentation practice supplies a useful analogy: datasheets describe how a dataset was assembled and where it should or should not be used [R16]. The proposed manuscript package applies that discipline to the inputs,

transformations, and limits of a paper companion without claiming that documentation alone resolves rights or accuracy.

10. Stage A: a prospective technical benchmark

Before classroom deployment, the proposed study would construct a held-out benchmark from 100 rights-cleared research manuscripts across disciplines and reporting formats. Six questions per manuscript would produce 600 items: three answerable questions, two plausible but unsupported questions, and one adversarial or misleading-context question. This is a proposed target, not an assembled dataset. Exact sampling rules, exclusion reasons, passage annotations, and question categories would be frozen before system comparison. A separate development set would support prompt and retrieval tuning.

Answerable questions would include an explicit fact, a methods limitation, and a bounded inference requiring more than one passage. Unsupported questions would ask for absent information or contain a plausible false premise. Adversarial items would include an irrelevant near-match source or an instruction embedded in source text. Two trained human assessors-one with information-evaluation expertise and one with relevant subject expertise-would independently label answerability, acceptable claims, and supporting passages. Disagreements would be adjudicated before outputs were scored.

The comparison would evaluate the same locked model and settings under three conditions: unrestricted generation without the manuscript package, retrieval with citations, and retrieval with the ledger's claim boundaries and abstention rules. The primary contrast is ledger-constrained retrieval versus citation-only retrieval on unsupported-answer rate. The unrestricted condition is a benchmark comparator, not a proposed student-facing service. A deterministic passage-search interface would provide an additional practical baseline. Generation parameters, retrieval index, source versions, prompts, and software dependencies would be archived. Multiple seeded runs would expose variability where the provider supports reproducibility controls.

Citation fidelity is the proportion of claim-citation links whose cited evidence actually supports the linked claim. Citation coverage is the proportion of source-verifiable claims supplied with adequate supporting evidence. Unsupported-answer rate is scored separately on unanswerable items; refusal on answerable items measures lost utility. A prespecified guardrail would require the one-sided 95% lower confidence bound for ledger-minus-citation-only answerable-item accuracy to exceed minus five percentage points; this is a proposed institutional tolerance, not an established educational threshold. Other contrasts are secondary, with multiplicity adjustment specified before testing. Assessors would be blinded to system labels and receive outputs in randomized order. Automated entailment scores may assist auditing but

cannot constitute the sole reference standard. Document-level paired bootstrap intervals would retain dependence among questions about the same paper.

11. Stage B: a prospective classroom feasibility study

Only a system meeting the technical and safeguarding gates would enter a school study. The proposed feasibility design recruits 24 teacher-class units across 12 secondary schools, with two comparable units per school and approximately 25 students per unit. Participation, sample size, access agreements, ethics review, and family consent procedures would all require confirmation before recruitment. These numbers are planning targets. No power claim or enrolled sample is implied.

Within each school, units would be randomly allocated to an eight-week ledger-supported publication sequence or an equally resourced comparison sequence using ordinary research portfolios. Different teachers would deliver the two conditions to reduce direct within-teacher crossover. Both groups would receive the same curriculum content, source-evaluation instruction, project time, assessment criteria, and access to the same permitted assistance. The intervention contrast is the structured ledger and its instructional use, not unequal access to an attractive new tool.

The primary educational outcome would be an individually completed transfer task using an unfamiliar research packet, without the companion. Twelve claims would require judgments about support, scope, and uncertainty, plus short explanations anchored to source passages. A preregistered five-level rubric, scored 0-4 per item for a maximum total of 48, would assess evidence selection, inferential restraint, and response to counterevidence. Two blinded raters would score a calibration set and a randomly sampled overlapping subset; reliability, disagreements, and adjudication would be reported.

Secondary outcomes would include article-level citation errors, revision quality, students' ability to distinguish their contribution from assistance, and teacher minutes spent on feedback and documentation. A four-week delayed task would explore retention. Surveys would assess perceived autonomy and publication pressure, but would not substitute for performance evidence. Final publication counts and page views would be descriptive institutional outcomes, never the primary indicators of learning.

12. Analysis, decision rules, and falsification

The classroom estimand is the intention-to-treat effect on the equally weighted average participating class, adjusted for class-average baseline performance and school pairing. Each teacher-class unit contributes its mean transfer score with equal weight, so larger classes do not dominate the primary estimate; a student-weighted estimate is a separate sensitivity analysis. Because assignment occurs by teacher-class unit, analysis must respect that unit. With only 12 pairs, small-sample uncertainty is material. Randomization inference over the paired assignments and a prespecified paired analysis would accompany an appropriately adjusted multilevel sensitivity model; a large number of student records cannot repair a small number of randomized units.

The protocol would specify missing-data assumptions, attrition by arm, scoring exclusions, and sensitivity analyses before outcomes are accessed. No student would be excluded for failing to complete a public article. Baseline comparisons would describe the recruited sample, not test whether randomization succeeded. A simulation using plausible cluster correlations, score variability, and missingness would determine whether a subsequent efficacy trial is feasible; this pilot would not be relabeled a definitive effectiveness test after inspection of favorable results.

Three falsification criteria distinguish the design from a promotional demonstration. First, a ledger that improves formatted attribution but fails to improve unassisted transfer has not demonstrated the proposed learning mechanism. Second, a companion that raises citation coverage while retaining unsupported claims has not demonstrated improved evidential reliability. Third, a workflow that improves scores by imposing substantial additional uncompensated teacher work has not demonstrated sustainable institutional value.

Technical release thresholds must be fixed locally before testing and reported with confidence intervals, severity classifications, and residual risks. Zero observed privacy leakage in a small synthetic-canary test does not establish zero risk. Safety failures trigger remediation and fresh testing on held-out items. Outcome families, exploratory subgroup analyses, and protocol deviations would be distinguished explicitly so that selective reporting cannot manufacture success.

13. Privacy, accessibility, and the right not to publish

UNESCO's guidance emphasizes human-centered, age-appropriate educational governance [R13], while UNICEF's 2025 guidance places children's rights, privacy, safety, inclusion, and accountability at the center of AI policy [R14]. For this proposal, those commitments become design constraints: collect the minimum evidence needed for the learning decision; separate instructional records from public artifacts; prevent public indexing of private drafts; and make disclosure understandable to the learner.

Publication consent should be separate from course enrollment and research-study participation. Students need a meaningful choice among named publication, an appropriate alternative attribution arrangement, restricted school sharing, or private assessment. An ORCID requirement cannot erase age, account-eligibility, consent, accessibility, or safety considerations. A persistent researcher identifier is a scholarly convenience, not a justification for coercive exposure.

The proposed platform would provide keyboard-accessible reading, structured headings, usable citations, accessible downloads, captioned or transcribed defenses, and a nondigital evidence route. Dictation and translation assistance should be available without requiring disclosure of disability to a public audience. Teachers need a manageable alternative when network access fails. Audit access should be role-limited and logged, with explicit retention periods and a process for correcting or removing inappropriate student material.

The NIST AI Risk Management Framework provides a voluntary structure for governing, mapping, measuring, and managing AI risk [R15]. Applying such a framework does not itself certify legal compliance or educational benefit. Institutional staff remain responsible for applicable requirements, supplier review, incident response, and changes in the population or context of use.

14. Editorial quality gates and limits

A signature student research journal should earn authority through the quality of its decisions. The first gate checks source identity, rights, attribution, and disclosure completeness. The second checks whether the question, method, evidence, and conclusion align. The third checks the learner's contribution and capacity to explain it. The fourth is an independent human editorial decision about public release, with conflicts of interest managed separately from course grading. Automated checks may identify problems; they do not become peer reviewers.

Editorial outcomes should include acceptance, revision, redirection to a different article type, and an educationally constructive decision not to publish. Reviews should identify the most consequential validity issue, request changes proportionate to the claim, and distinguish a limitation from a fatal defect. A prospective protocol should be judged as a protocol. It must not be dressed as completed primary research because the journal seeks a substantial inaugural volume.

The proposal has limits. Ledger completion can become bureaucratic, human ratings can encode disagreement, and learners can rehearse defenses without deeper understanding. Companion benchmarks may underestimate real-world prompting and multilingual variation. A well-documented source can still be wrong. These limitations motivate the proposed transfer tasks, adversarial tests, workload measures, and correction process; they are not eliminated by them.

15. Designership as accountable intellectual participation

The proposed contribution is a way to connect scholarly infrastructure to the learner's development of judgment. Publication becomes more educationally meaningful when a reader can move from a claim to its evidence, from evidence to its limits, and from an artifact to the human contribution that produced it. A paper companion is valuable when it assists those movements while preserving uncertainty and credit.

Designership, in this formulation, is the practice of designing conditions under which people can make consequential intellectual contributions and answer for them. Its success would be demonstrated by better independent reasoning, more faithful use of evidence, equitable participation, and sustainable teacher work. The ledger and the evaluation protocol offer a testable starting architecture. They do not report that those outcomes have already been achieved.

References and verification record

References retain their original authorship and publication identity. Verification dates below record source and metadata checking; they do not confer review or approval on this article.

[R1] Sarah McGrew (2020). Learning to evaluate: An intervention in civic online reasoning. Computers & Education, 145, 103711.

DOI: https://doi.org/10.1016/j.compedu.2019.103711

Source: https://www.sciencedirect.com/science/article/pii/S0360131519302647

Verified 2026-10-16. Publisher search record and original research-group page verified; original article, not a finding about the proposed ledger. Research-group evidence: https://cor.inquirygroup.org/research/learning-to-evaluate-intervention/

[R2] Paul Black; Dylan Wiliam (1998). Assessment and Classroom Learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7-74.

DOI: https://doi.org/10.1080/0969595980050102

Source: https://www.tandfonline.com/doi/abs/10.1080/0969595980050102

Verified 2026-10-16. Original review article; publisher issue-year1998 verified; online platform posting2006 is not used as the original publication year.

[R3] Tianyu Gao; Howard Yen; Jiatong Yu; Danqi Chen (2023). Enabling Large Language Models to Generate Text with Citations. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 6465-6488.

DOI: https://doi.org/10.18653/v1/2023.emnlp-main.398

Source: https://aclanthology.org/2023.emnlp-main.398/

Verified 2026-10-16. ACL authoritative metadata and abstract verified; the article does not evaluate this school platform.

[R4] Pranav Rajpurkar; Robin Jia; Percy Liang (2018). Know What You Don’t Know: Unanswerable Questions for SQuAD. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 784-789.

DOI: https://doi.org/10.18653/v1/P18-2124

Source: https://aclanthology.org/P18-2124/

Verified 2026-10-16. ACL authoritative metadata and abstract verified; abstention benchmark motivates proposed unanswerability tests.

[R5] Freda Shi; Xinyun Chen; Kanishka Misra; Nathan Scales; David Dohan; Ed H. Chi; Nathanael Schärli; Denny Zhou (2023). Large Language Models Can Be Easily Distracted by Irrelevant Context. Proceedings of the 40th International Conference on Machine Learning, PMLR 202, 31210-31227.

Source: https://proceedings.mlr.press/v202/shi23a.html

Verified 2026-10-16. PMLR authoritative authors, title, pages, year and abstract verified; original experiment uses arithmetic problems, not school paper agents.

[R6] Patrick Lewis; Ethan Perez; Aleksandra Piktus; Fabio Petroni; Vladimir Karpukhin; Naman Goyal; Heinrich Küttler; Mike Lewis; Wen-tau Yih; Tim Rocktäschel; Sebastian Riedel; Douwe Kiela (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Advances in Neural Information Processing Systems, 33.

Source: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

Verified 2026-10-16. NeurIPS proceedings author list, year and abstract verified. The proceedings citation is used; no fabricated DOI assigned.

[R7] Luc Moreau (editor); Paolo Missier (editor) (2013). PROV-DM: The PROV Data Model. W3C Recommendation, 30 April 2013.

Source: https://www.w3.org/TR/2013/REC-prov-dm-20130430/

Verified 2026-10-16. Normative recommendation version and editors verified; other contributors are listed by W3C. The educational ledger is this article’s adaptation, not a W3C-endorsed product.

[R8] National Information Standards Organization (2022). ANSI/NISO Z39.104-2022, CRediT, Contributor Roles Taxonomy. ANSI/NISO standard.

DOI: https://doi.org/10.3789/ansi.niso.z39.104-2022

Source: https://www.niso.org/publications/z39104-2022-credit

Verified 2026-10-16. Official standard record confirms 14 roles, ANSI approval 14 January 2022, publication 8 February 2022, DOI.

[R9] Committee on Publication Ethics (2023). Authorship and AI tools. COPE position statement.

DOI: https://doi.org/10.24318/cCVRZBms

Source: https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools

Verified 2026-10-16. Official COPE search-indexed statement verified; live open request returned HTTP 403, so the verification is not described as full-page retrieval. Position originated 13 February 2023; later website/PDF updates do not imply new study evidence.

[R10] ORCID (n.d.). Collecting and sharing ORCID iDs. ORCID integration guidance.

Source: https://info.orcid.org/documentation/collecting-and-sharing-orcid-ids/

Verified 2026-10-16. Official page retrieved; OAuth collection, downstream authenticated metadata, limited curated unauthenticated cases, and warnings against name-only matching verified.

[R11] Crossref (n.d.). Relationships. Crossref schema documentation.

Source: https://www.crossref.org/documentation/schema-library/markup-guide-metadata-segments/relationships/

Verified 2026-10-16. Official documentation retrieved; typed relationships and data/software identifier conventions verified. Does not itself confer registration or transfer authority.

[R12] Creative Commons (n.d.). Reusing CC-Licensed Content. Creative Commons reuse guidance.

Source: https://creativecommons.org/reusing-cc-licensed-content/

Verified 2026-10-16. Official guidance retrieved; TASL attribution and source-origin practice verified. License-specific legal code governs actual reuse.

[R13] Fengchun Miao; Wayne Holmes (2023). Guidance for generative AI in education and research. UNESCO.

DOI: https://doi.org/10.54675/EWZM9535

Source: https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

Verified 2026-10-16. Official UNESCO page and digital-library citation verified; page original date 7 September 2023 and update 16 January 2026 distinguish the 2023 publication.

[R14] UNICEF (2025). Guidance on AI and children: Version 3.0. Recommendations for AI policies and systems that uphold child rights. UNICEF, December 2025.

Source: https://www.unicef.org/innocenti/reports/policy-guidance-ai-children

Verified 2026-10-16. Official version 3.0 page retrieved; December 2025 date and child-centered requirements verified. Not an experiment or a compliance certification.

[R15] Elham Tabassi (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology, NIST AI 100-1.

DOI: https://doi.org/10.6028/NIST.AI.100-1

Source: https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10

Verified 2026-10-16. Official NIST publication record confirms author, 26 January 2023 date, report number, DOI and voluntary status.

[R16] Timnit Gebru; Jamie Morgenstern; Briana Vecchione; Jennifer Wortman Vaughan; Hanna Wallach; Hal Daumé III; Kate Crawford (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92.

DOI: https://doi.org/10.1145/3458723

Source: https://dl.acm.org/doi/10.1145/3458723

Verified 2026-10-16. ACM official indexed publication record verifies 2021 venue/DOI; author-posted arXiv 1803.09010 verifies full author list and proposal scope. The 2018 preprint date is distinct from the 2021 journal publication.

Source and authorship notes

Original conceptual architecture and prospective protocol; no primary empirical results, completed systematic review, benchmark execution, ORCID write, DOI registration, or independent human peer review is asserted.

Author intended by commissioning user: Dr. Freedom Cheteni. No Stanford affiliation, endorsement, or historical Designership reputation is inferred.

References were verified on 2026-10-16 against primary proceedings, publisher records, research-group records, or official standards/guidance. Access limitations are recorded in individual evidenceNote fields.

AI assistance was used to research sources, organize the argument, draft prose, and formulate the evaluation protocol. Human author and editors must verify the final claims, approve authorship/contributions, and conduct any claimed independent peer review.

No source content is republished here as a newly authored Cheteni study; original authors and identifiers remain intact. General normative design choices are proposals of this article rather than recommendations attributed wholesale to a cited standard.