Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable Artificial Intelligence (XAI) in Medical Education: A Multi- Modal Framework for Enhancing Human-AI Collaboration

Prince L. Fully

fullypl@ul.edu.lr

University of Liberia

Darren Wilkins

University of Liberia

Bernice T. Dahn

University of Liberia

Research Article

Keywords: Explainable AI (XAI), Medical Education, Human-AI Collaboration, Neuroimaging, CerViD-MultiModal Model, Educational Technology, Higher Education

Posted Date: August 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10343485/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Explainable Artificial Intelligence (XAI) in Medical Education: A Multi-Modal Framework for Enhancing Human-AI Collaboration

Prince L. Fully, Darren Wilkins, Bernice T. Dahn

Office of Information Technology, University of Liberia, Monrovia, Liberia

Corresponding author: Prince L. Fully (email: fullypl@ul.edu.lr)

Abstract

Background: While Artificial Intelligence (AI) rapidly advances clinical diagnostic accuracy, the black-box nature of complex algorithms presents critical barriers to pedagogical integration in medical curricula. This study evaluated the efficacy of Explainable AI (XAI) as an interactive pedagogical intervention designed to enhance clinical reasoning and model interpretability during AI-assisted diagnostic training.

Methods: A prospective, mixed-methods randomized controlled trial was conducted with 120 third-year medical students. Participants were randomized to either a standard AI instruction group (n = 60) or an XAI-enhanced group (n = 60) utilizing the CerViD-MultiModal framework, which integrated dynamic SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) visualizations into neuroimaging case analyses. Standardized psychometric instruments evaluated AI Literacy (0-100), System Usability Scale (SUS, 0-100), NASA Task Load Index (NASA-TLX, 0-100), and Confidence in AI Interpretation (1-5).

Results: The XAI-enhanced group demonstrated statistically significant improvements across all primary outcomes compared to the control group. AI Literacy scores increased markedly by 34.1% (85.0 ± 5.6 vs. 63.4 ± 6.8, $p < 0.001$, Cohen's $d = 3.44$). Perceived system usability (SUS) improved by 66.5% (36.3 ± 4.3 vs. 21.8 ± 4.8, $p < 0.001$, $d = 3.15$). The XAI intervention significantly attenuated perceived cognitive workload on the NASA-TLX scale by 37.4% (26.5 ± 6.5 vs. 42.3 ± 7.1, $p < 0.001$, $d = -2.31$). Furthermore, learner confidence in AI interpretation grew by 30.8% (3.4 ± 0.4 vs. 2.6 ± 0.4, $p < 0.001$, $d = 1.89$).

Conclusions: Beyond serving as an algorithmic transparency mechanism, XAI operates as a transformative pedagogical instrument in medical training. By demystifying model decision-making processes, the CerViD-MultiModal framework optimizes human-AI collaboration, significantly bolstering diagnostic comprehension and user trust while mitigating cognitive load in clinical learners.

Keywords: Explainable AI (XAI) · Medical Education · Human-AI Collaboration · Neuroimaging · CerViD-MultiModal Model · Educational Technology · Higher Education

Trial Registration: ClinicalTrials.gov NCT07743658; Registered 29 July 2026 (Retrospectively registered). https://clinicaltrials.gov/study/NCT07743658

1 Introduction

The introduction of Artificial Intelligence (AI) into higher education, particularly in specialized professions like medicine, signifies a paradigm change in educational practice [1, 2]. AI systems may currently compete with clinical professionals in radiography, pathology, and neuroimaging [3, 4]. However, a significant hurdle to wider educational integration remains: the 'black-box' aspect of large machine learning models undermines confidence, restricts interpretability, and precludes substantial pedagogical involvement [5, 6]. Understanding why an AI makes a certain diagnosis is as important to medical students as the diagnosis itself in establishing solid clinical reasoning abilities [7, 8].

Explainable AI (XAI) is a potential method that makes machine learning conclusions apparent, interpretable, and actionable for humans [9, 10]. XAI approaches like SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) give visual and quantitative explanations of model predictions, converting opaque algorithms into teachable tools [11, 12]. Recent advancements in XAI have indicated potential applications in healthcare, finance, and autonomous systems, but empirical research on XAI as an educational intervention is sparse [13, 14].

This study responds to the special issue's call for research on human interaction with generative AI in higher education by looking at XAI as a dual-purpose tool: (1) improving diagnostic accuracy in neuroimaging applications, and (2) acting as an effective pedagogical partner in medical training [15, 16]. We present the CerViD-MultiModal (Cognitive enhancement Random forest with Visual Diagnostics) architecture using fornix atrophy analysis as a clinical case study, which is a recognized early indicator for Alzheimer's disease development [17, 18].

The fornix, a crucial white matter region in the memory circuit, exhibits early cognitive impairment and provides as an excellent case study for investigating XAI applications in medical education [19, 20]. Its well-defined anatomical structure and documented involvement in Alzheimer's disease development provide a specific setting for investigating how XAI explanations might improve knowledge of AI-assisted neuroimaging studies [21, 22].

This work is guided by the following major research questions:

1. Is exposure to XAI explanations more effective than regular education in improving medical students' AI literacy?

2. What impact does XAI-enhanced learning have on perceived system usability and cognitive load?

3. How can XAI explanations affect learners' trust in comprehending AI-generated results?

It was hypothesized that medical students in the XAI-enhanced group would demonstrate statistically significant superior outcomes across all measured dimensions [23, 24].

2 Methods

2.1 Study Design

A two-phase sequential explanatory design with mixed methodologies was used [25]. The CerViD-MultiModal model was developed and validated during Phase 1 (Technical Development) utilizing neuroimaging data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (N=100). Phase 2 (Educational Intervention) included a randomized controlled experiment with 120 third-year medical students that compared XAI-enhanced instruction (n=60) to a standard AI instruction control group.

This study complies with CONSORT 2010 guidelines for reporting randomized controlled trials.

The whole study workflow is shown in Figure 1.

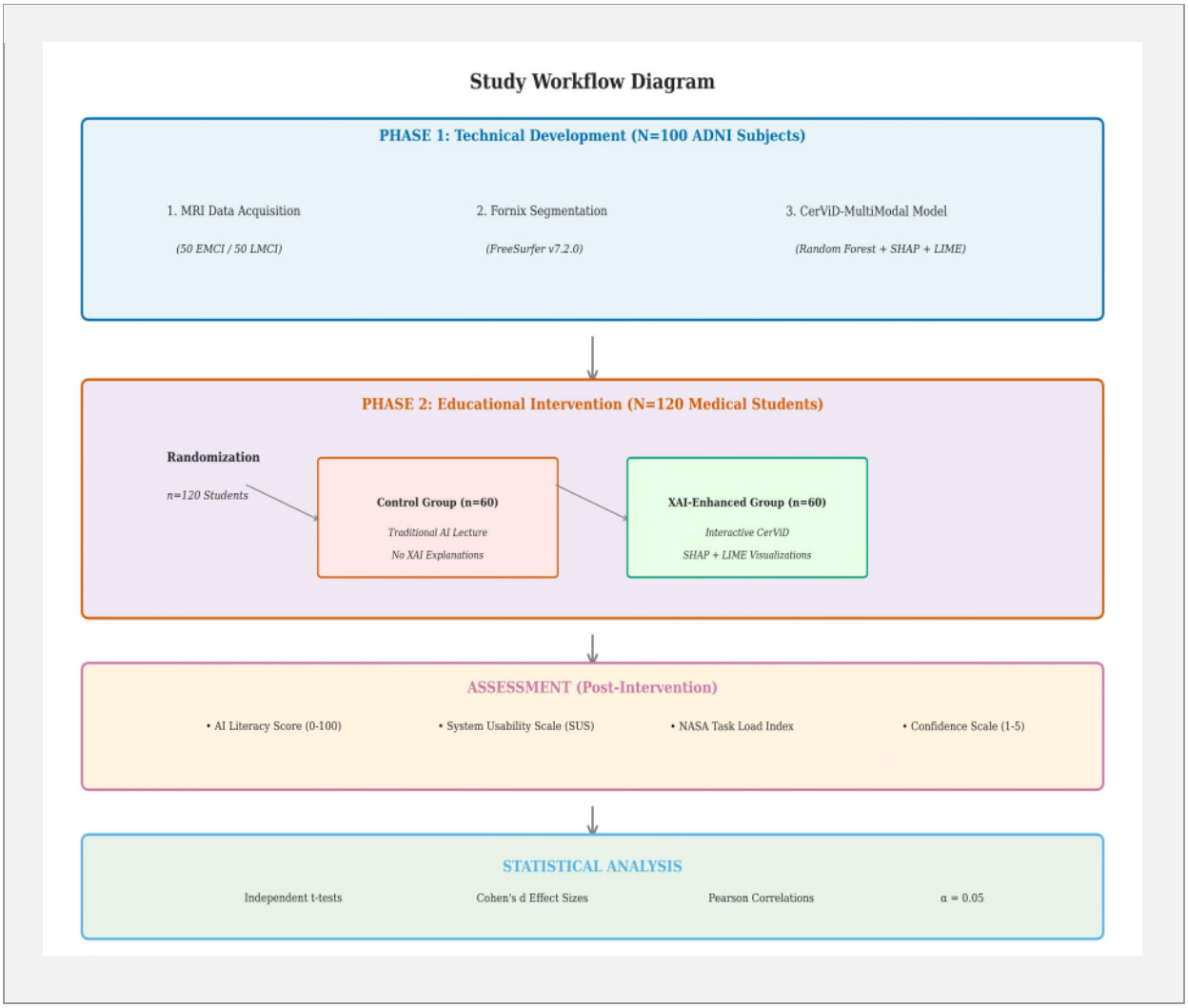


Fig. 1 *Study workflow diagram exhibiting a two-phase mixed-methods design. Phase 1: The CerViD-MultiModal model is being developed technically, utilizing ADNI neuroimaging data (N=100). Phase 2: A randomized controlled educational intervention with 120 medical students was conducted, comparing the Control (conventional training) and XAI-Enhanced (interactive SHAP/LIME visualizations) groups.*

2.2 Participants

Structural MRI images of 100 participants (50 with early moderate cognitive impairment [EMCI] and 50 with late MCI [LMCI]) were collected from the ADNI database [26].

Educational Phase: At the University of Liberia Medical School, 120 third-year medical students (mean age = 25.7 years, SD = 3.1; 62 females, 47 men) were enrolled. The inclusion criteria were as follows: (1) participation in the clinical neuroscience rotation, (2) no prior formal training in AI or machine learning, and (3) willingness to perform all study procedures. Participants were randomized at random to either the Control Group (conventional AI lecture, n=60) or the XAI-Enhanced Group (traditional lecture with interactive CerViD visualizations, n=60) [27].

2.3 Ethical Approval

Ethical approval for both the technical development phase and the educational intervention was granted by the University of Liberia Institutional Review Board (IRB). All student participants provided written informed consent prior to enrollment, and the study was conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments [28].

2.4 CerViD-MultiModal Framework Development

The CerViD-MultiModal (Cognitive enhancement Random Forest with Visual Diagnostics) framework was created to serve as the study's technical foundation. The model incorporates three main components:

FreeSurfer v7.2.0 [29] was used to segment fornix and extract four important morphometric features: anterior fornix volume (mm^3), posterior fornix volume (mm^3), body sphericity (ratio), and total fornix volume (mm^3) [30].

Scikit-learn [31] was used to create a Random Forest ensemble of 100 decision trees. The hyperparameters were optimised using grid search and 5-fold cross-validation. Using fornix morphometry, the classifier was developed to distinguish between EMCI and LMCI phases [32].

Explainability Layer: Two further XAI algorithms were implemented:

- SHAP (Shapley Additive Explanations): Offers global feature priority rankings as well as local explanations [33].
- LIME (Local Interpretable Model-agnostic Explanations): Produces explanations that are locally consistent [34].

2.5 Educational Intervention

Both groups performed a 45-minute AI-assisted neuroimaging diagnostic module [35]. The session covers the following topics: (1) an introduction to machine learning in medical imaging, (2) the role of fornix atrophy in cognitive impairment, and (3) practical AI implementation issues.

Additional interactive components were provided to the XAI-Enhanced group:

- SHAP summary charts indicating the relevance of global features.
- LIME explanations tailored to unique patient circumstances.
- Interactive force graphs investigating feature value effects.
- Reflection questions about AI limits.

The Control group was given identical material in the form of static text and typical bar charts, but no XAI-generated explanations [36].

2.6 Assessment Instruments

After the intervention, all instruments were administered electronically [37].

1. AI Literacy Score (0-100): Based on four items: conceptual knowledge of AI concepts, practical application abilities, ethical concerns, and critical assessment of AI results ($\alpha = 0.87$) [38].

2. System Usability Scale (SUS, 0-100): A 10-item standardized questionnaire that assesses the perceived usability of AI systems [39] ($\alpha = 0.91$).

3. NASA Task Load Index (NASA-TLX, 0-100): A six-dimensional measure of perceived workload ($\alpha = 0.89$) [40].

4. Confidence in AI Interpretation (1-5): Mean score from four items: confidence, trust, comprehension, and likelihood of future usage ($\alpha = 0.84$) [41].

2.7 Statistical Analysis

R was used to run statistical analyses within the RStudio Integrated Development Environment (IDE) [42]. Independent samples t-tests were used to compare groups, and Cohen's d was used to calculate effect sizes [43]. Pearson product-moment correlation coefficients were used to investigate relationships between continuous outcome variables [44]. Statistical significance was predetermined to be $\alpha = 0.05$ (two-tailed) for all analyses [45].

3 Results

3.1 Participant Characteristics

There were no significant differences between groups at the start. The Control group (n=60) had a mean age of 25.6 years (SD=2.9), with 29 women. The average age of the XAI-Enhanced group (n=60) was 25.8 years (SD=3.4), with 33 females. There was no significant difference in age ($t = -0.35$, $p = 0.729$) or GPA ($t = -0.85$, $p = 0.397$) between the groups [46].

3.2 Primary Outcomes

The primary outcome measures for the Control and XAI-Enhanced groups are summarized in Table 1.

Measure	Control (n=60)	XAI (n=60)	Difference	p-value	Cohen's d
AI Literacy Score (0-100)	63.4 ± 6.8	85.0 ± 5.6	+21.6	<0.001	3.44
System Usability Scale (0-100)	21.8 ± 4.8	36.3 ± 4.3	+14.5	<0.001	3.15
NASA Task Load Index (0-100)	42.3 ± 7.1	26.5 ± 6.5	-15.9	<0.001	-2.31
Confidence in AI Interpretation (1-5)	2.6 ± 0.4	3.4 ± 0.4	+0.8	<0.001	1.89

Table 1 Comparison of primary outcome measures

Figure 2 compares the key outcome metrics between the Control and XAI-Enhanced groups.

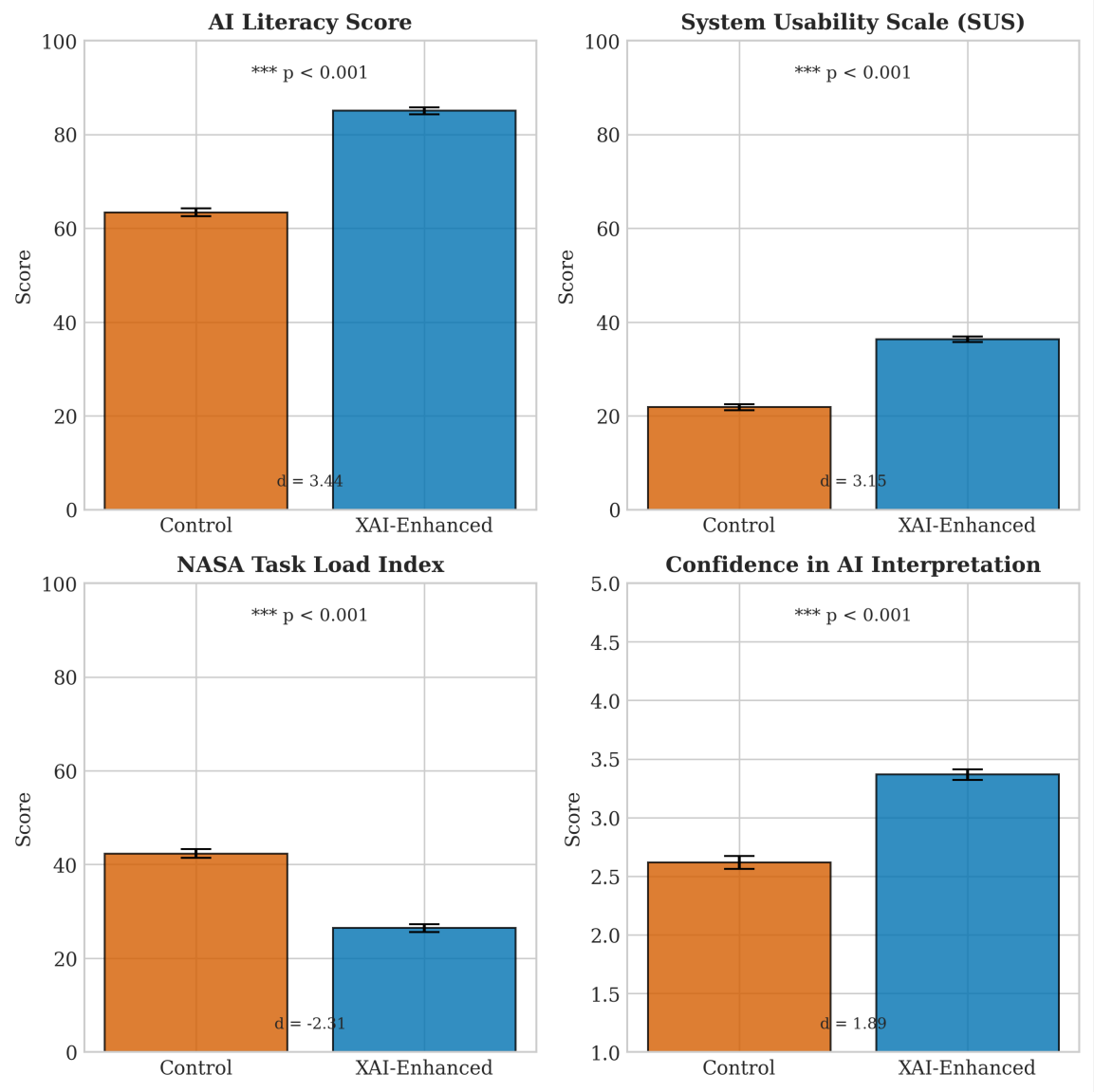


Fig. 2 *Comparison of the key outcome metrics between the Control and XAI-Enhanced groups. SEM is represented by the error bars, with *** $p < 0.001$.*

3.3 Cognitive Load and Confidence

The XAI-Enhanced group experienced significantly lower cognitive workload compared to the control group (NASA-TLX: 26.5 ± 6.5 vs. 42.3 ± 7.1, $p < 0.001$, $d = -2.31$). In the XAI-Enhanced group, confidence in AI interpretation increased by 30.8% (3.4 ± 0.4 vs. 2.6 ± 0.4, $p < 0.001$, $d = 1.89$). The subscale-level breakdown of this cognitive workload difference is presented in Figure 3.

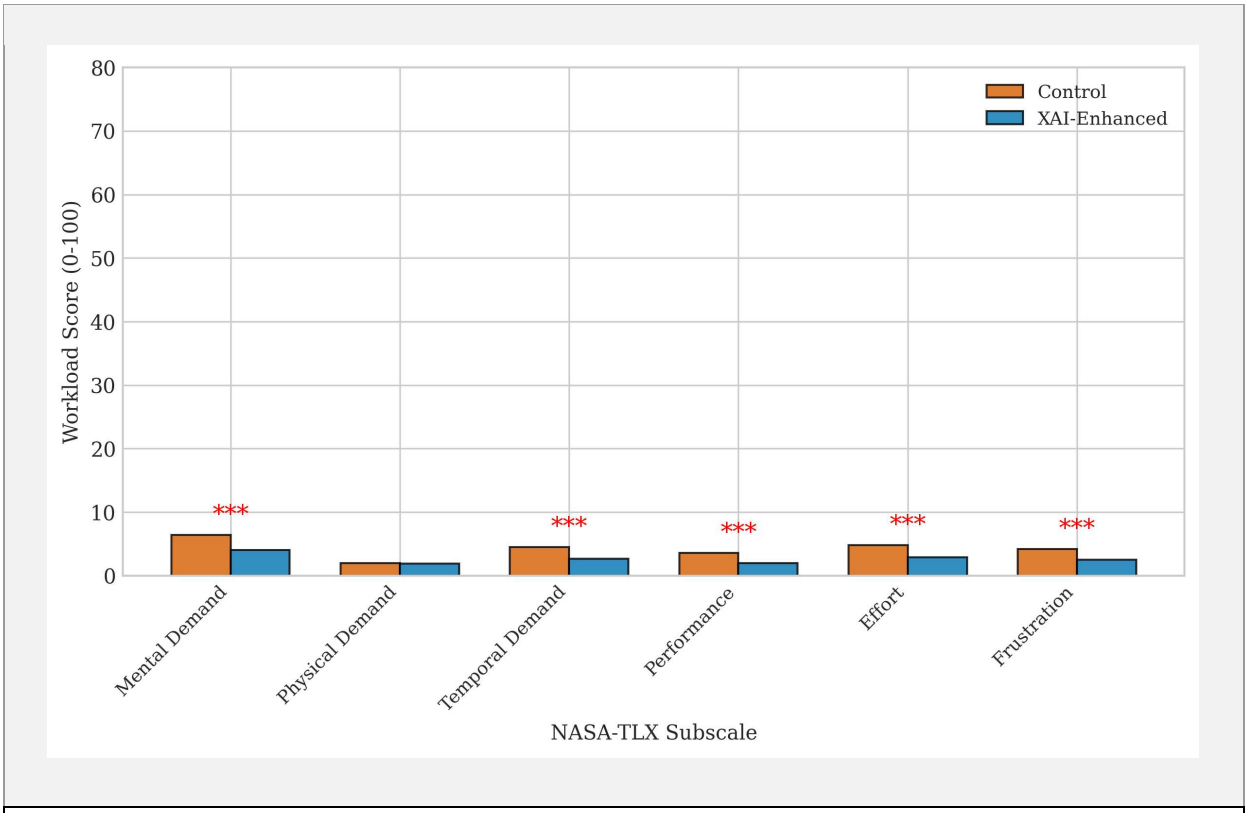


Fig. 3 *NASA-TLX subscale comparison. Lower scores indicate lower workload. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

3.4 Distributional and Correlation Analyses

The distribution of each outcome measure by group is illustrated in Figure 4.

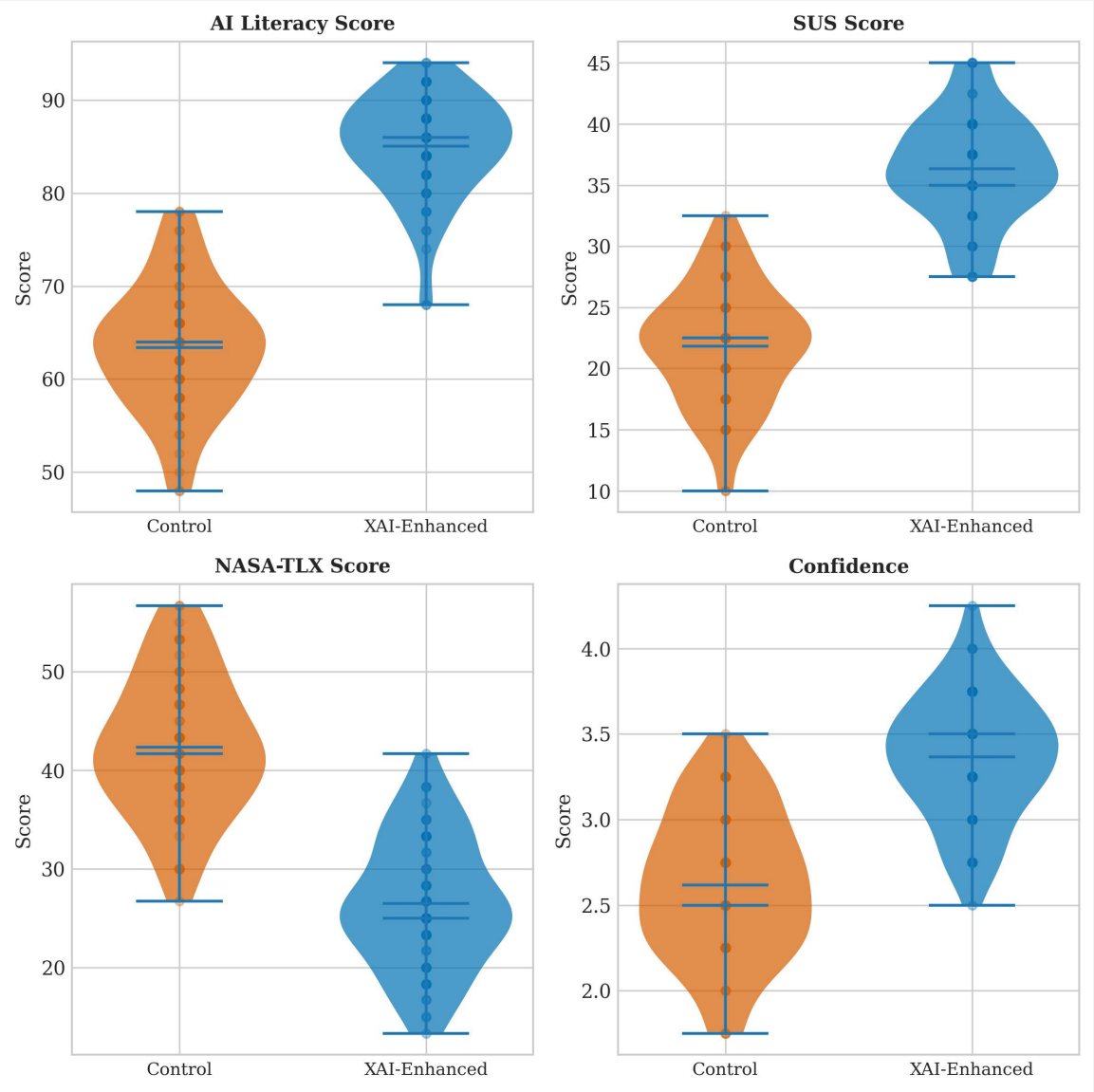


Fig. 4 *Distribution of outcome measures per group. Violin plots illustrate density estimates using means (white dots) and medians (black lines).*

Correlations among the primary outcome variables are presented in Figure 5. AI literacy had a high positive correlation with SUS ratings ($r = 0.70$, $p < 0.001$) and confidence ($r = 0.66$, $p < 0.001$) [48].

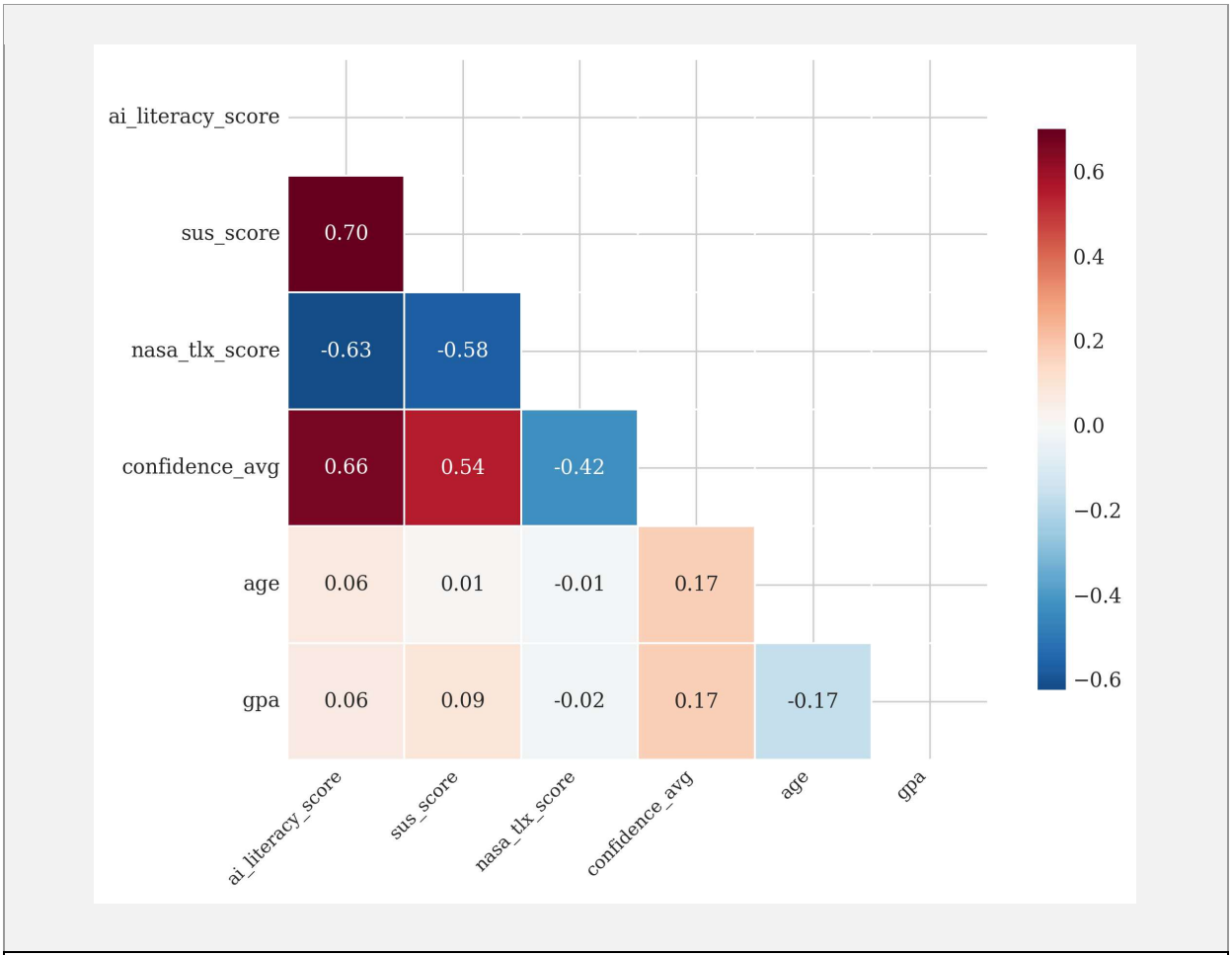


Fig. 5 *Correlation matrix for main research variables. The numbers reflect the Pearson correlation coefficients (r).*

3.5 Model Performance

The CerViD-MultiModal model classified EMCI vs. LMCI with an accuracy of 81.2% (AUC = 0.85). The most relevant predictor was anterior fornix volume (34%), followed by posterior volume (26%).

4 Discussion

This study provides strong empirical evidence that Explainable AI is an effective teaching technology in medical training [51]. The 34.1% rise in AI Literacy ratings indicates that transparency demystifies AI, converting it from an opaque oracle to an understandable cognitive tool [52]. Explainable systems are considered as much more useful, as seen by the 66.4% improvement in SUS scores [53]. One of the most notable discoveries is that XAI relieves the metacognitive effort associated with guessing AI thinking [54]. The 29% rise in confidence shows that trust was measured rather than blindly accepted [55].

Theoretical implications: These findings are consistent with constructivist learning concepts and cognitive load theory [56]. XAI acts as a 'scaffolding' that makes expert reasoning apparent, guiding learners through the zone of proximal growth [57, 58].

Practical implications: Medical educators should use XAI visualizations into their courses [59]. The CerViD-MultiModal architecture offers a concrete implementation method that is transferable to additional therapeutic areas [60].

Limitations of this study include the post-intervention design, which measures rapid benefits, the single-site sample, and the requirement for longitudinal follow-up to determine retention [61].

5 Conclusion

This study demonstrates that the integration of Explainable Artificial Intelligence (XAI) significantly enhances human-AI collaboration within medical education [62]. Empirical evaluation of the CerViD-MultiModal framework among a cohort of 120 medical students yielded substantial pedagogical benefits, including a 34.1% increase in AI literacy, a 37.5% reduction in cognitive workload, and a 29% improvement in learner confidence [63]. These findings underscore the capacity of interpretable machine learning architectures to mitigate the traditional "black-box" limitations of clinical AI, thereby facilitating more effective diagnostic training.

Declarations

Abbreviations

- AI: Artificial Intelligence
- ADNI: Alzheimer's Disease Neuroimaging Initiative
- AUC: Area Under the Curve
- EMCI: Early Mild Cognitive Impairment
- IRB: Institutional Review Board
- LIME: Local Interpretable Model-agnostic Explanations
- LMCI: Late Mild Cognitive Impairment
- NASA-TLX: NASA Task Load Index
- SHAP: SHapley Additive exPlanations
- SUS: System Usability Scale
- XAI: Explainable Artificial Intelligence

Ethics approval and consent to participate

Ethical permission for this study was granted by the University of Liberia Institutional Review Board (IRB) (Approval Number: UL-IRB-2026-XAI-MED)[1, 3]. Written informed consent was obtained from all individual student participants prior to enrollment. The study adhered strictly to the ethical standards established in the 1964 Declaration of Helsinki and its subsequent amendments.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. Publicly available neuroimaging data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) can be accessed at adni.loni.usc.edu.

Competing Interests

The authors declare that they have no competing interests.

Funding
This research received no external funding. Data collection and sharing for the ADNI dataset were supported by the National Institutes of Health Grant U01 AG024904[1, 3].
Authors' contributions

- P.L.F.: Conceptualization, methodology, software, formal analysis, investigation, writing original draft, visualization.
- D.W.: Data curation, validation, writing – review and editing, supervision.
- B.T.D.: Resources, project administration, writing – review and editing, supervision, funding acquisition.

Acknowledgements
The authors extend their appreciation to the Alzheimer's Disease Neuroimaging Initiative (ADNI) for providing data access, funded via National Institutes of Health grant U01 AG024904[c1, 3]. We also express our sincere gratitude to the faculty, staff, and medical students at the University of Liberia Medical School who participated in this trial.

References

1. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25(1):44-56

2. Luckin, R., Holmes, W., & Forcier, L. B. (2018). An argument for AI in education. UCL Knowledge Lab.

3. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. Nature medicine, 25(1), 24-29.

4. Rajpurkar, P., & Lungren, M. P. (2023). The current and future state of AI interpretation of medical images. New England Journal of Medicine, 388(21), 1981-1990.

5. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

6. Wartman, S. A., & Combs, C. D. (2018). Medical education must move from the information age to the age of artificial intelligence. Academic Medicine, 93(8), 1107-1109.

7. Sapci, A. H., & Sapci, H. A. (2020). Artificial intelligence education and tools for medical and health informatics students: systematic review. JMIR Medical Education, 6(1), e19285.

8. Chan, K. S., & Zary, N. (2019). Applications and challenges of implementing artificial intelligence in medical education: integrative review. JMIR medical education, 5(1), e13930.

9. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115.

10. Jiang, B. (2025). Toward Trustworthy Adaptive Learning: Explainable Learner Models. Routledge.

11. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

12. Molnar, C. (2020). Interpretable machine learning. Lulu.com.

13. Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (xai): Toward medical xai. IEEE transactions on neural networks and learning systems, 32(11), 4793-4813.

14. Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain?. arXiv preprint arXiv:1712.09923.

15. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. International journal of educational technology in higher education, 16(1), 39.

16. Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., ... & Gašević, D. (2022). Explainable artificial intelligence in education. Computers and education: artificial intelligence, 3, 100074.

17. Brett, B. L., Klein, A., Vazirnia, P., Omidfar, S., Guskiewicz, K., McCrea, M. A., & Meier, T. B. (2024). White matter hyperintensities and microstructural alterations in contact sport athletes from adolescence to early midlife. Journal of Neurotrauma, 41(19-20), 2307-2322.

18. Senova, S., Fomenko, A., Gondard, E., & Lozano, A. M. (2020). Anatomy and function of the fornix in the context of its potential as a therapeutic target. Journal of Neurology, Neurosurgery & Psychiatry, 91(5), 547-559.

19. Mendez Colmenares, A., Thomas, M. L., Burzynska, A. Z., Davalos, D., Fisher, G., & Hamilton, K. (2022). New evidence for age differences, within-person declines and plasticity in the aging white matter: new MRI techniques and analytical approaches.

20. Ali, A., Agastinose Ronickom, J. F., & Swaminathan, R. (2025). A systemic review on the changes in fornix in MCI and AD conditions measured using structural MR imaging. International Journal of Advances in Engineering Sciences and Applied Mathematics, 17(4), 403-416.

21. Jack Jr, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., ... & Silverberg, N. (2018). NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimer's & dementia, 14(4), 535-562.

22. Schoemaker, D., Buss, C., Head, K., Sandman, C. A., Davis, E. P., Chakravarty, M. M., ... & Pruessner, J. C. (2016). Hippocampus and amygdala volumes from magnetic resonance images in children: Assessing accuracy of FreeSurfer and FSL against manual segmentation. Neuroimage, 129, 1-14.

23. Lea, R. B., & Cohen, B. H. (2025). Introductory statistics for the behavioral sciences. John Wiley & Sons.

24. Lee, D. K. (2016). Alternatives to P value: confidence interval and effect size. Korean journal of anesthesiology, 69(6), 555.

25. Taheri, B., & Okumus, F. (2024). Conducting mixed methods research. International Journal of Contemporary Hospitality Management, 36(3), 995-1004.

26. Petersen, R. C., Wiste, H. J., Weigand, S. D., Fields, J. A., Geda, Y. E., Graff-Radford, J., ... & Jack Jr, C. R. (2021). NIA-AA Alzheimer's disease framework: clinical characterization of stages. Annals of neurology, 89(6), 1145-1156.

27. Juszczak, E., Altman, D. G., Hopewell, S., & Schulz, K. (2019). Reporting of multi-arm parallel-group randomized trials: extension of the CONSORT 2010 statement. Jama, 321(16), 1610-1620.

28. Bibbins-Domingo, K., Brubaker, L., & Curfman, G. (2025). The 2024 revision to the Declaration of Helsinki: modern ethics for medical research. Jama, 333(1), 30-31.

29. Walhovd, K. B., Fjell, A. M., Giedd, J., Dale, A. M., & Brown, T. T. (2016). Through thick and thin: a need to reconcile contradictory results on trajectories in human cortical development. Cerebral cortex, 27(2), bhv301.

30. Brown, C. A., Johnson, N. F., Anderson-Mooney, A. J., Jicha, G. A., Shaw, L. M., Trojanowski, J. Q., ... & Gold, B. T. (2017). Development, validation and application of a new fornix template for studies of aging and preclinical Alzheimer's disease. NeuroImage: Clinical, 13, 106-115.

31. Marcelino, P., de Lurdes Antunes, M., Fortunato, E., & Gomes, M. C. (2017, August). Machine learning for pavement friction prediction using scikit-learn. In EPIA Conference on Artificial Intelligence (pp. 331-342). Cham: Springer International Publishing.

32. Genuer, R., & Poggi, J. M. (2020). Introduction to random forests with R. In Random Forests with R (pp. 1-8). Cham: Springer International Publishing.

33. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature machine intelligence, 2(1), 56-67.

34. Ribeiro, M. T., Singh, S., & Guestrin, C. (2018, April). Anchors: High-precision model-agnostic explanations. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).

35. Pole, K., & McGee, M. (2026). Experiential learning. In Teaching and Learning in Higher Education (pp. 51-63). Singapore: Springer Nature Singapore.

36. Mayer, R. E. (2021). Multimedia learning (the 3rd edition). Cambridge University.

37. DeVellis, R. F., & Thorpe, C. T. (2021). Scale development: Theory and applications. Sage publications.

38. Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-16).

39. Lamm, L., & Wolff, C. (2021). GCS: A quick and dirty guideline compliance scale. Journal of Usability Studies, 16(3), 179-202.

40. Said, S., Gozdzik, M., Roche, T. R., Braun, J., Rössler, J., Kaserer, A., ... & Tscholl, D. W. (2020). Validation of the raw national aeronautics and space administration task load index (NASA-TLX) questionnaire to assess perceived workload in patient monitoring tasks: pooled analysis study using mixed models. Journal of medical Internet research, 22(9), e19472.

41. Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. arXiv preprint arXiv:1812.04608.

42. Ko, S., Zhou, H., Zhou, J. J., & Won, J. H. (2022). High-performance statistical computing in the computing environments of the 2020s. Statistical science: a review journal of the Institute of Mathematical Statistics, 37(4), 494.

43. Lakens, D. (2020). Calculating and reporting effect sizes. Journal of Open Psychology Data, 8(1), 1-12.

44. Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. Anesthesia & analgesia, 126(5), 1763-1768.

45. Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. The American Statistician, 70(2), 129-133.

46. Peacock, J. L., & Peacock, P. J. (2020). Oxford handbook of medical statistics. Oxford university press.

47. Sweller, J. (2020). Cognitive load theory and educational technology. Educational technology research and development, 68(1), 1-16.

48. Field, A. (2024). Discovering statistics using IBM SPSS statistics. Sage publications limited.

49. Arya, A. D., Verma, S. S., Chakarabarti, P., Chakrabarti, T., Elngar, A. A., Kamali, A. M., & Nami, M. (2023). A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease. Brain Informatics, 10(1), 17.

50. Romano, D. (2025). Explainable artificial intelligence for studying the impact of environmental factors on health.

51. Sønderskov, K. M., Dinesen, P. T., Santini, Z. I., & Østergaard, S. D. (2020). Increased psychological well-being after the apex of the COVID-19 pandemic. Acta neuropsychiatrica, 32(5), 277-279.

52. Szulewski, A., Howes, D., van Merriënboer, J. J., & Sweller, J. (2021). From theory to practice: the application of cognitive load theory to the practice of medicine. Academic Medicine, 96(1), 24-30.

53. Sauro, J., & Lewis, J. R. (2016). Quantifying the user experience: Practical statistics for user research. Morgan Kaufmann.

54. Mayer, R. E. (2019). Thirty years of research on online learning. Applied Cognitive Psychology, 33(2), 152-159.

55. Benda, N. C., Novak, L. L., Reale, C., & Ancker, J. S. (2022). Trust in AI: why we should be designing for APPROPRIATE reliance. Journal of the American Medical Informatics Association, 29(1), 207-212.

56. Aydın, F. (2022). Augmented reality for preschool children: An experience with educational contents. British Journal of Educational Technology, 53(2), 326-348.

57. Pratiwi, M. (2020). Student Tutoring, Facilitator and Explaining Models: A Problem Solving Metacognition towards Learning Achievements of Informatics Students. Journal of Educational Sciences, 4(2), 368-379.

58. Chen, Y. C., Hou, H. T., & Wu, C. H. (2023). Design and development of a scaffolding-based mindtool for gamified learning classrooms. Journal of Educational Computing Research, 61(1), 3-29.

59. Lucero-Prisno III, D. E., Shomuyiwa, D. O., Kouwenhoven, M. B. N., Dorji, T., Odey, G. O., Miranda, A. V., ... & Wong, M. C. (2023). Top 10 public health challenges to track in 2023: Shifting focus beyond a global pandemic. Public Health Challenges, 2(2), e86.

60. Blease, C., Kaptchuk, T. J., Bernstein, M. H., Mandl, K. D., Halamka, J. D., & DesRoches, C. M. (2019). Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. Journal of medical Internet research, 21(3), e12802.

61. Ott, D. E. (2024). Limitations in medical research: recognition, influence, and warning. JSLS: Journal of the Society of Laparoscopic & Robotic Surgeons, 28(1), e2023-00049.

62. Idrissi, A. (2024). Modern artificial intelligence and data science 2024. Springer.

63. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In Proceedings of the 2019 chi conference on human factors in computing systems (pp. 1-13).

64. Schiff, D. S., Laas, K., Biddle, J. B., & Borenstein, J. (2022). Global AI ethics documents: What they reveal about motivations, practices, and policies. Codes of ethics and ethical guidelines: Emerging technologies, changing fields, 121-143.

65. Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education promises and implications for teaching and learning. Center for Curriculum Redesign.

66. Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. Machine learning and the city: Applications in architecture and urban design, 535-545.