ISSN: 2366-1313

Exploring the role of XAI in Academic Performance Prediction: A Bibliographical Approach

Avishek Barman [1], Anal Acharya [2], Soumen Mukherjee [3]

[1] Ramakrishna Mission Vidyamandira, avishekbarman.cs@vidyamandira.ac.in
[2] St. Xavier's College, anal.acharya@sxccal.edu
[3] RCC Institute of Information Technology, soumen.mukherjee@rcciit.org.in

Abstract:

Explainable Artificial Intelligence (XAI) in education represents a transformative approach aimed at improving the transparency, interpretability, and trustworthiness of AI-driven educational technologies. As AI systems are increasingly utilized for personalized learning, assessment, and administrative tasks, the need for comprehensible AI models becomes paramount. This abstract explores the application of XAI in educational contexts, focusing on methods such as analysis of feature importance, model distillation. Also we aim to produces visual explanations to make AI-driven decisions understandable to educators, students, and administrators. XAI facilitates better decision-making by offering perspective into how XAI algorithms determine learning paths, grade assessments, and identify students' strengths and weaknesses. By making AI processes transparent, XAI fosters trust, supports ethical use of AI in education, and empowers educators to tailor instructional strategies more effectively. This discussion underscores the importance of balancing AI complexity with interpretability to create educational AI systems that are both advanced and user-friendly, ultimately enhancing the educational experience and outcomes for all stakeholders.

1. Introduction:

Artificial intelligence (AI) is becoming gradually pivotal in the jurisdiction of teaching and learning. A prime example is personalized teaching systems, which have already gained substantial traction and are backed by mounting evidence demonstrating their efficacy in enhancing learning outcomes. This domain has been termed as AI in Education (AIED) by researchers. The AI in education systems can also leverage varied and refined AI techniques to craft the middleware, a crucial element for enhancing the learning experience [1]. Developments of larger scale AI tools are also in action. The global market of Educational Technology (EdTech) backed by AIED has been predicted to generate approximate $25.7 billion by 2030 [2]. Other studies [3] projected that the application of AI in governance, education and society as a whole would touch $126 billion by 2030. The traditional approach in AIED includes various tools and techniques already in force, like, various machine learning algorithms, feedback assessment systems, learning analytics and so on [4]. The ML systems collect data from various sources which are beyond the classical format of textual and numeric data. Event data from various activities are being collected. Post and during COVID-19, the unstructured data produced by digital media like microphones, cameras, wareable devices are playing a pivotal role in shaping up prediction models for AIED. In educational environments, it is imperative that students, educators,

and administrators grasp not just the outcomes generated by AI systems but also the rationale guiding those outcomes. XAI methodologies enhance the transparency and comprehensibility of AI systems by offering insights into the processes leading to their conclusions or suggestions. Therefore, AIED systems represent a significant domain where the necessity for eXplainable AI is evident.This paper addresses the specific hurdles and methods of eXplainable AI (XAI) within the realm of education.

2. Literature Survey:

Increasing recognition is emerging regarding the significant ethical considerations associated with AIED. AI is perceived as a potential avenue for enhancing employment opportunities, lifelong learning, and democratic engagement, yet it also remains susceptible to egregious misuse. AI has been subjected to numerous criticisms centered on algorithmic bias. The following chart shows the impact of _biased‘ AI for different scenarios which eventually led to an impact on AI in education.

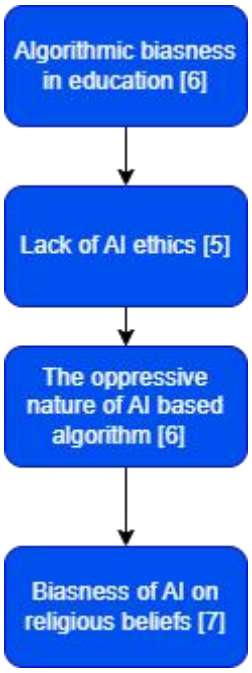

Fig 1: Criticism faced by AI in social perspectives

In the study given by [8], we have seen protection of AI algorithmic groups from further investigation, which is termed as Artificial Immutability. In a research work shown in [9] we have seen the finding that the profound influence of AI on education is undeniable.

Nonetheless, there is a stark absence of research, consensus on guidelines, development of policies, and enactment of regulations to manage its application in educational settings. Consequently, there is extensive debate regarding the management of risks that may arise concerningaccountability and democracy within the traditional academic systems [10]. The purpose of explicability is to simplify the reconstruction of actions carried out by AI programs and to clarify the potential responsibility for resulting consequences. _Algorithmic transparency; explainer generalizability; and explanation granularity‘ have been in focus as unique attributes of XAI [11]. The objective of this paper, as outlined by the [12], is to comprehend the essence of XAIED, ascertain factors contributing to its effectiveness, and recognize any practical or ethical constraints linked to clarity in teaching-learning processes.

2.1. Explainable AI

XAI advocates for employing methods that ‑enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners‖ [13]. Initially, Explainable AI (XAI) is primarily centered around algorithms. Broadly speaking, the models of machine learning can be classified taking their interpretability into account, which pertains to the extent that a human can understand the rationale behind a decision or replicate precisely what the model accomplishes. Considering the principle of designing, we thrive for finding priority in the models having high interpretability because, in theory, it can detect and rectify the different types of bias thus helping in mitigating biases. It can also enhance defense against adversarial distortions, where noise becomes a part of the input to deceive the AL model while remaining almost subtle to humans. This has the potential to alter predictions and enhance the utilization of relevant variables and induceimpurity in the model's reasoning [14].

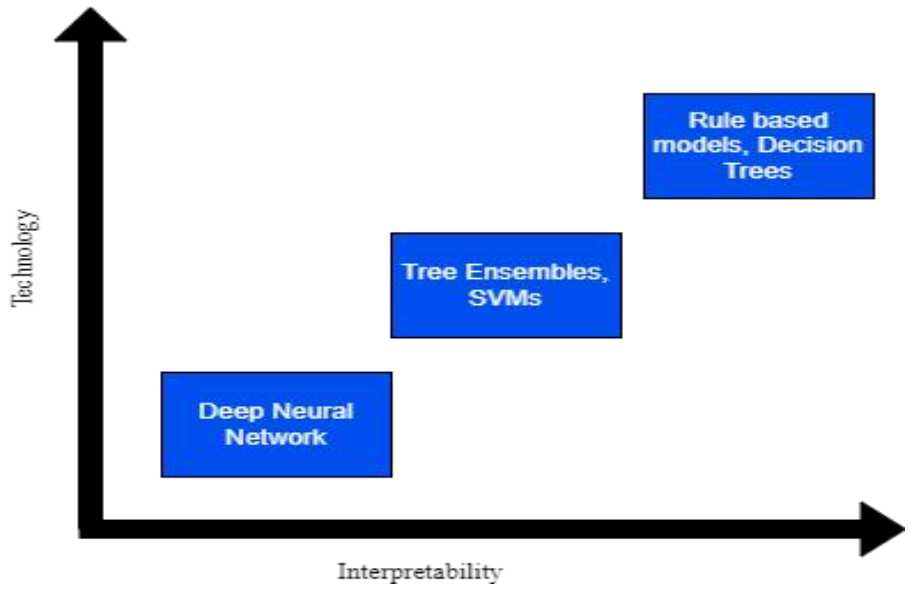

Fig 2: Technology and their interpretability

Certain models like general decision trees, additive models are considered interpretable. Also tools like rule-based models are inherently interpretable because of their relatively simple structure, facilitating an explanation of their functionality. In contrary we have models that are not easily interpretable namely support vector machines, deep neural networks etc. Also tree ensembles are also having such complex structures [15]. In order to render these models comprehensible to humans, the XAI research community has undertaken substantial efforts to develop explainability techniques which are known as post-hoc. These techniques aim to elucidate ideas on generating its predictions but hiding the model's underlying structure [16].Recent advancements in eXplainable Artificial Intelligence (XAI) have experienced a substantial transformation and a progressive evolution towards socially embedded XAI. This entails the introduction and exploration of a _socio-technically‘ [17] informed viewpoint, which integrates the socio-organizational context to elucidate AI-facilitated decision-making processes.

3. The present scenario in AIED

Numerous envisaged applications of AIED hinge on the presumption that extensive data collection and examination will occur. This encompasses contentious methods of collecting data regarding learner progress within a virtual learning setting like collecting biometric data, sampling voices, and employing eye-tracking techniques [18]. There is already significant dependence on contentious monitoring technologies in proctoring and evaluation. The notable aspects of surveillance have been presented in [19]. These can be encapsulated as power imbalances existing between owners/farmsof data and individuals of which the data is being collected, analyzed, shared and managed; as well as the integration and amplification of data-driven digital platforms into the fundamental operations of educational institutions.It's impossible to disentangle the utilization of analytics from surveillance. Nevertheless, the volume of machine learning data collection may be huge.With the rising prominence of AIED, focus is transitioning from technical aspects to the socio-technical viewpoint. Much of the existing AIED literature stems from quantitative computer science, leaving a scarcity of expertise in AI within the humanities [20]. Few literatures suggested multilayer categorization of the challenges AIED is facing [21]. Thenext section represents the methodologies and approaches that are emerging by the introduction of XAI in AIED and how it is going to have disruptive impacts on the conventional roles and tasks of both learners and educators.

4. Methods and approaches

The assertions presented in this paper stem from a thematic review of literature across various disciplines pertinent to eXplainable AI for education. The sample population has unique characteristic and still hard to find. Resources are compiled from Google keyword search. Papers published till December 2023 was consulted for the compilation.

4.1. The 'Black Box' approach

The opaque characteristics of deep learning remain unresolved, and numerous machine-generated decisions remain inadequately comprehended [22].The primary structural characteristic of the computational model'black box' lies in the lack of transparency regarding the processes and mechanisms that translate input into output. Machine learning has made limited headway in capturing higher-order thoughts, abstract concepts, creative language use, or 'common sense'. Significant strides have been achieved in functional or "weak" applications, particularly in NLP programming, often promoted in the exuberant language of AI marketing [23].The research work presented in [24] introduced a comprehensive typology aimed at elucidating matters concerning 'black box' computation. These differences depend on the particular problem of explanation under consideration, the chosen type of explainer, the black box model being examined, and the nature of the data utilized as input by the operating model. Similar model based explainer has been seen in [25]. The objective facing Explainable AI (XAI) is to achieve a dual goal: establishing a suitable alignment between interpretive techniques and opaque models, while also crafting interpretive models and tools that are readily comprehensible to their target users.

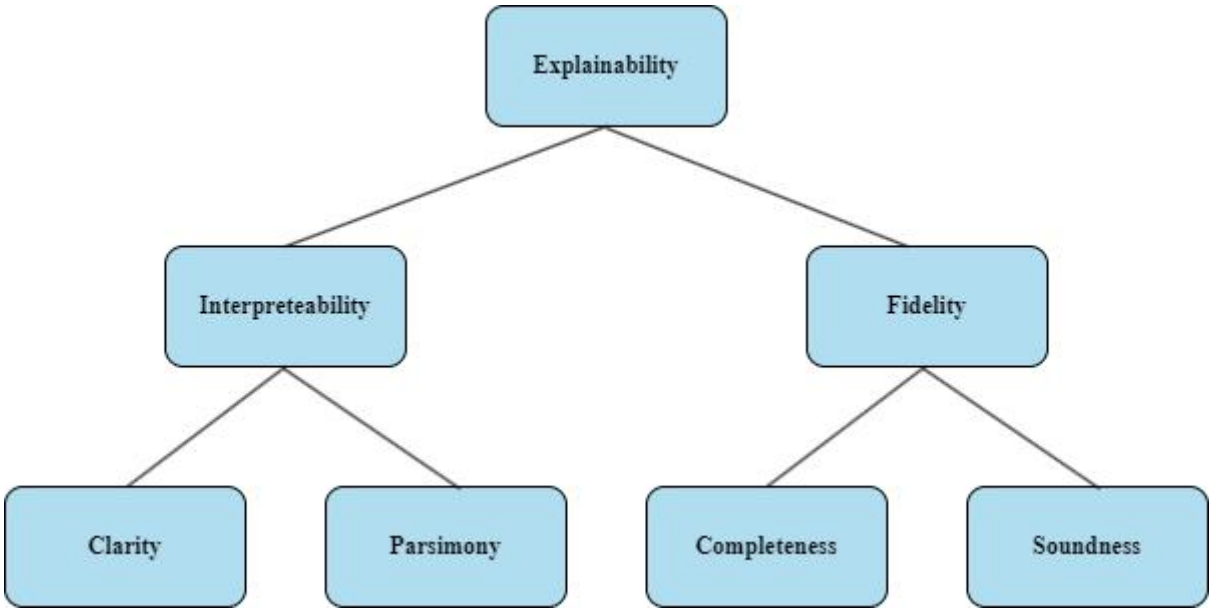

Fig 3: Summery of definitions of explainabilty [26]

4.2. Perspectives blending social and technical aspects of XAIED

Grasping the complete socio-technical scope of AI implementations poses a challenge. In a proposal for social transparency in AIED [27] we came across the social technical risks and their probable improvements. We may summarize the key point in the following table:

ISSN: 2366-1313

Socio Technical AI Risk	Description	Improvement
The inability to comprehend how adapting algorithm based solutions originally developed taking social context into consideration might lead to misinterpretation, inaccuracies, or potential harm when implemented in a different context.	Risk embedded within machine learning. It talks about the notion that algorithms can be utilized across various contexts. Used for promoting their portability; however, this can result in a lack of sensitivity to specific social contexts within particular domains.	Acknowledge the need to adapt scripts for new application contexts; acknowledge that normative concepts are linked not restricted to specific objects but to particular social contexts; and understand that all algorithms require contextualization.
Inability to accurately represent the entire system in which social criterions, like fairness, will be applied.	Algorithmic decisions are based on specific data points, and abstraction alone cannot fully capture socio-technical intricacies.	Concurrent examination of machine activities and human together within the system.
Failure to grasp how the introduction of technology into an established social system alters the behaviors and ingrained values involved in the existing system.	Technologies have the potential to induce changes in social customs and values through their continual use, resulting in both deliberate and unintended outcomes.	Develop a thorough understanding of the current ripple effects to foresee potential 'what if?' scenarios; utilize domain expertise to access risks.
Failure to acknowledge fact that technology may not always be the only way to achieve the optimal solution.	Machine learning is limited to anticipating and devising technological solutions, such as algorithmic adjustments.	Exercise caution in determining how and when to devise technological systems, recognizing that sticking to a platform is not the solution for every scenario.

Table 1: Social challenges and remedies

The framework [28] comprises essential inquiries concerning stakeholders, the advantages of Explainable AI (XAI) and user experience, the approach to explaining AI, and potential pitfalls. The suggestion here is to differentiate between the various scopes of explanations, namely global and local forms of explanation, utilizing proxies that are simpler to comprehend.The XAI-ED model offers a versatile perspective on XAI matters and proposes practical approaches to aligning various beneficiaries with suitable communication strategies along with the proxies.

5. The Strengths, Opportunities and Challenges in XAIED

As its prevalence grows, learners also have the potential to gain from Explainable AI (XAI) when it aids their understanding of decision-making processes that impact their learning within AIED.

Additionally, various stakeholders engaged in educational processes arguably experience empowerment. By going by convention, the stakeholders include administrators, managers, librarians, technicians, designers and few more statutory. An explanation of the same AIED system from these different viewpoints might vary significantly.

- Educators could utilize automated assessment tools, plagiarism detectors, administrative aids, and also assess predictive analytics dashboards.
- Learners could engage with learning management systems which are adaptive in nature and receive assistance from interactive bots or AI teaching agents.
- Institutions can be benefited of the summarized data to manage and plan operations.

Each of the viewpoints can be considered as _algorithmic transparency; explainer generalizability; and explanation granularity‘ [29].A more detailed explanation may be necessary when AIED assumes a central role, though educators should possess the ability to elucidate AIED processes. Conversely, learners may only require a straightforward model to comprehend the contribution of AI in academics at a large scale, but this might restrict transparency in the algorithmic approach.Giving a detailed explanation of how the algorithm affects the learning process could also influence learner behavior. This might divert attention from genuine learning or potentially be perceived as efforts to manipulate the algorithm.

Raising concerns about the utilization of AI tools in socio-governance domains like border management, law enforcement, criminal justice, national security, andthey advocate for comprehensive regulation spanning multiple sectors. They also advocate for significantly enhanced transparency to address the 'black box' issue associated with AI-driven decision-making, where algorithmic suggestionsare provided without the ability to reconstruct or explain the underlying process of generating recommendations. Statutory bodies of US, The White House [30] and Council of the European Union [31] have come up with bills Prohibit any applications that fail to operate in complete accordance with human rights laws. These bills and regulations propose reporting in clear, technically accurate language that is both meaningful and should be made publicly available whenever feasible.Moving forward, the primary regulatory obstacle lies in discovering non-reductive approaches to render socio-technical processes involving AI which are transparent and also comprehensible.

6. Conclusion

Enhanced transparency and explicability offer a pathway for critical examination of algorithmic applications in education and broader societal contexts involving AI. This critical assessment, as evidenced in the literature, underscores the importance of adopting a socio-technical perspective for eXplainable AI in Education (XAIED). To engage educators and learners effectively in AIED, it is imperative that they comprehend and consent meaningfully to AI interventions, while trust is cultivated through maximum transparency. As the risks and impacts of AIED continue to evolve, XAIED emerges as indispensable, especially considering the alternatives of opaque AI or no AI at all. Given its potential to foster trust, mitigate risks, and facilitate the exchange of participants

viewpoints, XAIED could be seen as a default stance for Universities aiming to promote transparency and accountability.For learners to engage effectively in AIED, they must comprehend and give meaningful consent to the procedures and consequences of algorithmic involvement. However, achieving this is challenging unless those supporting learners also grasp the processes and ethical implications involved. Even if algorithms were made with utmost transparency and it is fully explicable, the intricate socio-technical ecosystems encompassing the development, assembly, programming, training, deployment, and maintenance of AI systems are so dispersed that they remain largely obscured from any individual perspective. By scrutinizing the scheme of XAIED using a socio-technical standpoint, it becomes evident that eXplainable AI (XAI) alone cannot fully address the problemsshown by AI; rather, it represents both a starting point and a crucial prerequisite for meaningful discussions about potential futures.

7. References

1. Adams, R. 2021. ‐Can Artificial Intelligence be Decolonized?‖ Interdisciplinary Science Reviews 46 (1-2): 176–197. doi:10.1080/03080188.2020.1840225.

2. Byrne, R. M. J. 2019. ‐Counterfactuals in Explainable Artificial Intelligence (XAI): Evidence from Human Reasoning.‖ In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence Survey track. 6276–6282. doi:10.24963/ijcai.2019/876.

3. Dignum, V. 2021. ‐The Role and Challenges of Education for Responsible AI.‖ London Review of Education 19 (1): 1– 11. doi:10.14324/LRE.19.1.01.

4. Elicit. n.d. https://elicit.org/search.

5. Birhane, A., E. Ruane, T. Laurent, M. S. Brown, J. Flowers, A. Ventresque, and C. L. Dancy. 2022. ‐The Forgotten Margins of AI Ethics.‖ FAccT ‗22: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (Forthcoming). doi:10.1145/3531146.3533157

6. Noble, S. U. 2018. Algorithms of Oppression. NYU Press

7. Samuel, S. 2021. ‐AI‘s Islamophobia problem.‖Vox. https://www.vox.com/future-perfect/22672414/ai-artificialintelligence-gpt-3-bias-muslim

8. Zuboff, S. 2019. The Age of Surveillance Capitalism.Public Affairs Books.

9. Bulathwela, S., M. Pérez-Ortiz, C. Holloway, and J. Shawe-Taylor. 2021. ‐Could AI Democratise Education? SocioTechnical Imaginaries of an EdTech Revolution.‖35th Conference on Neural Information Processing Systems (NeurIPS 2021).ArXiv, abs/2112.02034. doi:10.48550/arXiv.2112.02034

10. Floridi, L., J. Cowls, M. Beltrametti, et al. 2018. ‒AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.‖ Minds & Machines 28: 689–707. doi:10.1007/s11023- 018-9482-5.

11. Antoniadi, A. M., Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney. 2021. ‒Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review.‖ Applied Sciences 11 (11): 5088. doi:10.3390/app11115088. MDPI AG.

12. Hu, B., P. Tunison, B. Vasu, N. Menon, R. Collins, and A. Hoogs. 2021. ‒XAITK: The Explainable AI Toolkit.‖ Applied AI Letters 2 (4), doi:10.1002/ail2.40.

13. Gunning 2017 XAI—Explainable artificial intelligenceD Gunning, M Stefik, J Choi, T Miller, S Stumpf, GZ Yang

14. Moosavi-Dezfooli, S. M., Fawzi, A., Fawzi, O., &Frossard, P. (2017). Universal adversarial perturbations. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1765–1773)

15. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.

16. Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. ACM Queue, 16, 31–57.

17. Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., &Weisz, J. D. (2021). Expanding explainability: Towards social transparency in AI systems. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445188.

18. Luckin, R., W. Holmes, M. Griffiths, and L. B. Forcier. 2016. Intelligence Unleashed. An Argument for AI in Education. London: Pearson. https://discovery.ucl.ac.uk/id/eprint/1475756.

19. Beetham, H., A. Collier, L. Czerniewicz, B. Lamb, Y. Lin, J. Ross, A.-M. Scott, and A. Wilson. 2022. ‒Surveillance Practices, Risks and Responses in the Post Pandemic University.‖ Digital Culture & Education 14 (1): 16–37. https://www.digitalcultureandeducation.com/volume-14-1.

20. Zawacki-Richter, O., V. I. Marín, M. Bond, and F. Gouveneur. 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: 39. doi:10.1186/s41239-019-0171-0.

21. Selbst, A. D., D. Boyd, F. Sorelle, V. Suresh, and J. Vertesi. 2018. –Fairness and Abstraction in Sociotechnical Systems (August 23, 2018).‖ In 2019 ACM Conference on Fairness, Accountability, and Transparency (FAT*), 59-68.29. (Khosravi et al. 2022)

22. Tjoa, E., and C. Guan. 2021. –A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI.‖ IEEE Transactions on Neural Networks and Learning Systems 32 (11): 4793–4813. DOI:10.1109/tnnls.2020.3027314. PMID: 33079674.

23. Russell, S. J., and P. Norvig. 2021. Artificial Intelligence: A Modern Approach. 4th ed. Prentice Hall.

24. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. and Pedreschi, D. (2018, August). –A Survey of Methods for Explaining Black Box Models.‖ ACM Computing Surveys 51 (5): Article 93. doi:10.1145/3236009.

25. Markus, A. F., J. A. Kors, and P. R. Rijnbeek. 2021. –The Role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care: A Comprehensive Survey of the Terminology, Design Choices, and Evaluation Strategies.‖ Journal of Biomedical Informatics 113), doi:10.1016/j.jbi.2020.103655.

26. Markus, A. F., J. A. Kors, and P. R. Rijnbeek. 2021. –The Role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care: A Comprehensive Survey of the Terminology, Design Choices, and Evaluation Strategies.‖ Journal of Biomedical Informatics 113), doi:10.1016/j.jbi.2020.103655

27. Baker, R. S., and A. Hawn. 2021. –Algorithmic Bias in Education.‖ doi:10.35542/osf.io/pbmvz.

28. Birhane, A., E. Ruane, T. Laurent, M. S. Brown, J. Flowers, A. Ventresque, and C. L. Dancy. 2022. –The Forgotten Margins of AI Ethics.‖ FAccT _22: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (Forthcoming). doi:10.1145/3531146.3533157

29. Antoniadi, A. M., Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney. 2021. –Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review.‖ Applied Sciences 11 (11): 5088. doi:10.3390/app11115088. MDPI AG.

30. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

31. https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/#:~:text=Background,AI%20act%20in%20April%202021.