Hybridizing Educational Assessment: a Theoretical Model from Judicial Science

Giannangelo Boccuzzi[1,*,†], Alberto Nico[2,†] and Flavio Manganello[1,†]

[1]*National Research Council, Institute of Educational Technologies, Genoa, Italy*
[2]*University of Bari "Aldo Moro", Department of Law, Bari, Italy*

Abstract

Contemporary artificial intelligence (AI) systems are being integrated into decision-making processes across domains including legal systems and educational practices, combining automation with human oversight. This paper examines the potential for adapting hybrid judicial decision-making models to educational assessment, specifically student evaluation processes. The proposed model draws from judicial systems where algorithms generate initial judgment drafts that are subsequently reviewed by judges. Applied to education, this framework involves automated systems performing initial evaluations based on standardized data, with educators reviewing and refining results. This hybrid approach may offer immediate feedback, reduced educator workloads, enhanced objectivity, and personalized assessments while maintaining human authority for final decisions and potentially ensuring ethical outcomes.

1. Introduction

Artificial intelligence integration in decision-making has altered numerous sectors, with education and legal systems representing domains where accurate, fair, and transparent decisions carry significant consequences. As AI technologies develop, the challenge involves creating systems that combine algorithmic efficiency with human insight rather than replacing human judgment [1, 2]. Contemporary research indicates the importance of developing trustworthy AI systems that maintain human agency while benefiting from computational capabilities [3].

In judicial contexts, experimental hybrid models have emerged where AI systems generate initial draft decisions subsequently reviewed, modified, and finalized by human judges. This approach addresses caseload pressures while maintaining human involvement in legal decision-making [4]. Italian legal scholarship has examined the implications of AI in judicial decision-making, with attention to the balance between algorithmic efficiency and human oversight in both civil and criminal contexts [5, 6].

Analyses of AI applications in justice systems have highlighted the need for regulatory frameworks that preserve judicial independence while leveraging technological capabilities [7]. The implementation raises questions about applicability to other domains requiring considered judgment and ethical considerations, particularly where algorithmic accountability and transparency are important [8].

Educational assessment represents a potentially suitable domain for such hybrid approaches. Traditional assessment methods face challenges including scalability issues, potential for human bias, inconsistency across evaluators, and increasing demand for immediate feedback in digital learning environments [9, 10]. Fully automated assessment systems, despite their efficiency, encounter difficulties with contextual understanding, creativity evaluation, and interpretation of student responses that may require human pedagogical expertise [11].

Developments in explainable AI have begun addressing these limitations by providing more transparent and interpretable automated assessment systems [12]. This paper proposes adapting the hybrid decision-making model from judicial contexts to educational assessment, specifically student evaluation processes. The analysis suggests such a model may address limitations of both purely human and fully automated assessment systems while preserving pedagogical values of fairness, transparency, and individualized attention.

2. Background and related work

2.1. Hybrid decision-making in judicial systems

The legal domain has been at the forefront of exploring AI-human collaboration in high-stakes decision-making. Developments in judicial AI systems demonstrate that hybrid models can enhance both efficiency and consistency while maintaining human oversight for final decisions [13]. These systems typically operate through a two-stage process: algorithmic analysis of case data and legal precedents generates initial recommendations, then reviewed and refined by human judges who consider contextual factors, ethical implications, and exceptional circumstances.

Successful judicial hybrid systems operate on key principles encompassing maintaining human authority and accountability for final decisions, ensuring transparency in AI-generated recommendations, providing clear rationales for algorithmic modifications, and incorporating safeguards against algorithmic bias and errors [14]. These principles prove essential for maintaining public trust and legal validity in judicial processes. Research highlights the importance of addressing the "black box" problem in AI systems, particularly where decisions significantly impact individuals' rights and opportunities [15]. AI integration in judicial decision-making requires careful consideration of constitutional principles and procedural guarantees, as demonstrated by Italian scholarship on predictive justice and civil jurisdiction [16]. The broader implications of AI-assisted decision-making in legal contexts have been analysed, emphasizing the need for regulatory frameworks that balance innovation with fundamental rights protection [17].

2.2. AI in educational assessment

Educational assessment has experienced advancement in automated evaluation systems, particularly in multiple-choice testing, essay scoring, and skill assessment in programming and mathematics [18]. These systems demonstrate effectiveness at processing large volumes of standardized responses, providing immediate feedback, and maintaining consistent evaluation criteria across contexts. Machine learning developments have expanded automated assessment capabilities, enabling more sophisticated analysis of student performance patterns and learning trajectories [19].

Current automated assessment systems face constraints. They encounter difficulties evaluating creativity, critical thinking, and complex problem-solving that cannot be easily quantified [20]. These systems often fail to account for individual student circumstances, learning disabilities, or cultural contexts that may influence performance but are not reflected in standardized metrics. Research has demonstrated that students' trust in automated assessment systems is affected by the transparency and explainability of algorithms used [21], highlighting the need for more interpretable AI approaches in educational contexts.

2.3. Bias and fairness in AI systems

Research indicates that algorithmic bias can perpetuate or amplify existing inequalities, affecting marginalized groups [22]. In educational contexts, this concern appears acute as biased assessment systems can produce lasting impacts on student opportunities and self-perception. Frameworks for understanding unintended consequences of machine learning highlight the complexity of addressing

bias in AI systems, requiring systematic approaches considering multiple bias sources throughout development and deployment processes [23].

Research indicates that human oversight appears crucial for identifying and mitigating algorithmic bias, as humans can recognize contextual factors and exceptional circumstances that algorithms may miss [24]. This finding supports the argument for hybrid systems combining algorithmic efficiency with human judgment. Studies examining perceptions of algorithmic decision-making reveal that individuals often view purely automated decisions as reducing human dignity and failing to account for personal circumstances [25], reinforcing the importance of maintaining human involvement in assessment processes.

3. Proposed hybrid model for educational assessment

Adapting judicial hybrid systems, this analysis proposes a two-stage educational assessment model combining automated evaluation with human review. The first stage involves AI systems analyzing student responses using standardized criteria, generating preliminary scores with detailed feedback, identifying areas of concern or notable performance, and flagging responses requiring human attention. The second stage encompasses educators reviewing AI-generated assessments, considering contextual factors and student circumstances not captured by algorithmic analysis, and modifying scores based on pedagogical judgment so that final assessments incorporate both algorithmic efficiency and human expertise (see Figure 1).
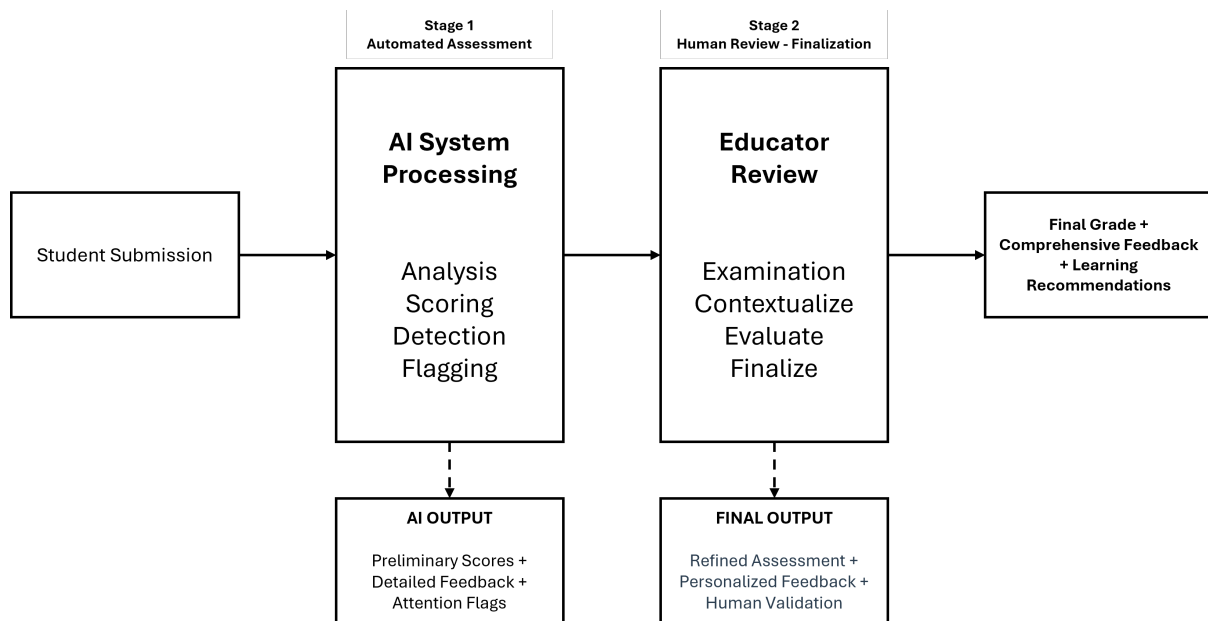

Figure 1: Two-stage hybrid decision-making model for educational assessment.

3.1. Key features and safeguards

The proposed model incorporates four key features for effectiveness and ethical operation: (i) transparent AI-generated assessments with clear explanations of scoring rationales, enabling informed human review and building system trust [26], (ii) bias detection mechanisms identifying assessments potentially disadvantaging certain student groups, with human reviewers trained to recognize and address these biases, (iii) customization capabilities allowing adaptation based on individual student needs while maintaining standardized core criteria, and (iv) audit trails maintaining detailed records of both automated and human decisions to support accountability and continuous improvement.

The integration of explainable AI techniques is important for hybrid assessment systems. Research has indicated that post-hoc explanation methods, while useful, may have limitations in educational contexts, suggesting the need for more sophisticated approaches to AI interpretability [27]. The proposed model incorporates multiple explanation strategies, including feature importance rankings, decision trees for specific assessment criteria, and natural language explanations of scoring rationales.

3.2. Implementation framework

Effective implementation requires attention to key components. Technical infrastructure must include functional AI systems capable of processing diverse student responses, secure data management systems, and user-friendly educator interfaces. Educator training programs are necessary, focusing on AI-generated assessment interpretation, algorithmic bias recognition, and integration of automated insights with pedagogical judgment. Quality assurance through regular monitoring and evaluation of system performance, including accuracy metrics, bias detection, and user satisfaction measures, forms the foundation of effective implementation. Clear ethical frameworks governing AI use in assessment, encompassing privacy protection, fairness standards, and transparency requirements, are important for maintaining trust and compliance.

Research has highlighted the importance of involving educators as codesigners in development of AI-powered educational systems to ensure technology aligns with pedagogical goals and practices [28]. The implementation framework must consider regulatory compliance, particularly regarding data protection and algorithmic decision-making in educational contexts [29]. The European Union's General Data Protection Regulation and emerging AI governance frameworks provide guidelines for ethical deployment of AI in educational assessment [30]. Examining legal scholarship on AI governance in judicial contexts, educational institutions can benefit from established frameworks for AI regulation that emphasize transparency, accountability, and human oversight [7]. Regulatory challenges in educational AI resemble those encountered in judicial AI systems, particularly regarding the need to balance technological efficiency with individual rights protection.

4. Advantages and implications

4.1. Benefits for educational practice

The proposed hybrid model offers advantages over traditional assessment methods and fully automated systems. By automating initial assessment tasks, educators can focus time and expertise on cases requiring considered judgment and personalized attention, with this efficiency gain proving valuable in contexts with high student-to-teacher ratios or resource constraints. Algorithmic components provide consistent application of evaluation criteria across students and contexts, reducing variability that can arise from human factors such as fatigue, mood, or unconscious bias. Students receive rapid initial feedback on performance, supporting timely learning interventions and maintaining engagement in the learning process. The combination of algorithmic analysis and human insight enables more personalized assessment that considers individual student circumstances, learning styles, and developmental needs. While maintaining human judgment for final decisions, the model reduces the impact of subjective biases through systematic algorithmic analysis and transparent decision-making processes.

4.2. Ethical and pedagogical considerations

Implementation of hybrid AI-human assessment systems raises ethical and pedagogical questions that must be addressed. Student assessment data requires careful handling to protect privacy while enabling system functionality, necessitating clear policies and technical safeguards to govern data collection, storage, and use. Students and educators must understand how AI components contribute to assessment decisions, requiring systems that can provide clear explanations of reasoning processes. Educators need ongoing support and training to effectively work with AI systems, interpret algorithmic insights, and

maintain their role in assessment processes. Benefits of hybrid assessment systems must be accessible to all students and educational contexts, requiring attention to digital divides and resource disparities that might otherwise limit equitable access to these technological advantages.

4.3. Challenges and limitations

The proposed hybrid model faces challenges that must be acknowledged and addressed. Developing and maintaining sophisticated AI systems requires technical expertise and resources, potentially creating barriers for smaller educational institutions that may lack necessary infrastructure or funding. Implementing hybrid systems requires changes to existing assessment practices, necessitating careful change management strategies and stakeholder engagement to ensure successful adoption. Both educators and students must develop trust in AI-assisted assessment systems, requiring transparent communication about system capabilities and limitations to build confidence in technology. Ensuring consistent quality across AI systems and human reviewers requires ongoing monitoring and calibration efforts to maintain reliability and fairness in assessment outcomes.

5. Conclusion and future work

This paper has proposed adapting hybrid judicial decision-making models to educational assessment, suggesting that such an approach may address the limitations of both purely human and fully automated assessment systems. The proposed model maintains human authority and accountability while leveraging AI capabilities for efficiency and consistency. Figure 2 illustrates the comparative framework of hybrid decision-making processes across judicial and educational domains, highlighting their shared reliance on algorithmic support combined with human oversight.
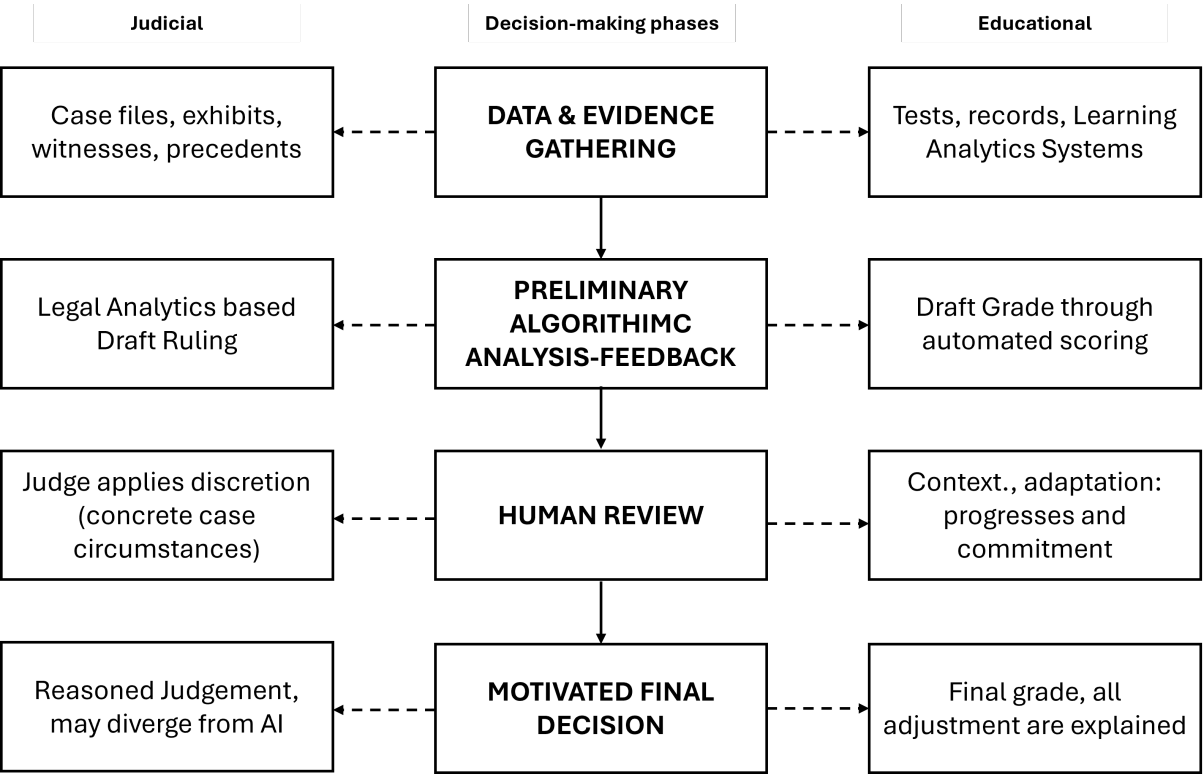
Judicial	Decision-making phases	Educational
Case files, exhibits, witnesses, precedents	**DATA & EVIDENCE GATHERING**	Tests, records, Learning Analytics Systems
Legal Analytics based Draft Ruling	**PRELIMINARY ALGORITHIMC ANALYSIS-FEEDBACK**	Draft Grade through automated scoring
Judge applies discretion (concrete case circumstances)	**HUMAN REVIEW**	Context., adaptation: progresses and commitment
Reasoned Judgement, may diverge from AI	**MOTIVATED FINAL DECISION**	Final grade, all adjustment are explained

Figure 2: Comparative framework of hybrid decision-making processes: Judicial vs. educational assessment systems.

The parallels between judicial and educational contexts, both requiring fair, transparent, and ethically sound decisions, suggest that hybrid models from legal domains may inform educational practice. How-

ever, characteristics of educational assessment, including its developmental and pedagogical purposes, require careful adaptation of these models. Future work should focus on key areas, including developing technical prototypes of hybrid assessment systems, conducting empirical studies of effectiveness compared to traditional methods, investigating training needs of educators working with AI-assisted assessment, and exploring broader implications of hybrid systems for educational equity and access. Research should examine long-term impacts of hybrid assessment on student learning outcomes, teacher professional development, and institutional assessment practices.

Technologies such as large language models present both opportunities and challenges for hybrid assessment systems, requiring careful consideration of integration into educational contexts [31]. Development of prescriptive analytics approaches that go beyond prediction to provide actionable recommendations for educational interventions represents a direction for future research. Emphasis on interpretable machine learning in educational contexts suggests the need for continued innovation in explainable AI techniques specifically designed for assessment applications [32].

The potential of hybrid AI-human assessment systems to influence educational practice while preserving human values represents a direction for educational technology research and development. As AI capabilities continue to advance, the challenge will be not whether to integrate these technologies into educational assessment, but how to do so in ways that enhance rather than diminish human aspects of teaching and learning.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] B. Williamson, Algorithmic skin: Health-tracking technologies, personal analytics and the biopedagogies of digitized health and physical education, Sport, Education and Society 20 (2015) 133–151. doi:10.1080/13573322.2014.962494.

[2] G.-J. Hwang, H. Xie, B. W. Wah, D. Gašević, Vision, challenges, roles and research issues of artificial intelligence in education, Computers and Education: Artificial Intelligence 1 (2020) 100001. doi:10.1016/j.caeai.2020.100001.

[3] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. Vayena, Ai4people—an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations, Minds and Machines 28 (2018) 689–707. doi:10.1007/s11023-018-9482-5.

[4] T. Sourdin, Judge v robot?: Artificial intelligence and judicial decision-making, UNSW Law Journal 41 (2018) 1114–1133. URL: https://search.informit.org/doi/10.3316/informit.040979608613368.

[5] F. L. Donati, Intelligenza artificiale e giustizia, Rivista AIC (2020) 415–436.

[6] G. Sartor, Quando a decidere in materia penale sono (anche) algoritmi e ia, Diritto di Internet 2 (2019) 34–49.

[7] U. Ruffolo, A. Amidei, Diritto dell'intelligenza artificiale. Vol. 2, Luiss University Press, Rome, 2024.

[8] M. E. Kaminski, Binary governance: Lessons from the gdpr's approach to algorithmic accountability, Southern California Law Review 92 (2018) 1529–1616.

[9] W. Holmes, M. Bialik, C. Fadel, Artificial intelligence in education: Promises and implications for teaching and learning, 1 ed., Center for Curriculum Redesign, Boston, MA, 2019.

[10] W. Holmes, K. Porayska-Pomsta, K. Holstein, E. Sutherland, R. T. Baker, S. Buckingham Shum, K. R. Koedinger, Ethics of ai in education: Towards a community-wide framework, International Journal of Artificial Intelligence in Education 32 (2022) 504–526. doi:10.1007/s40593-021-00239-1.

[11] N. Selwyn, On the limits of artificial intelligence (ai) in education, Nordisk tidsskrift for pedagogikk og kritikk 10 (2024) 3–14. doi:10.23865/ntpk.v10.6062.

[12] S. Gunasekara, M. Saarela, Explainable ai in education: Techniques and qualitative assessment, Applied Sciences 15 (2025) 1239. doi:10.3390/app15031239.

[13] S. Chesterman, L. B. Moses, U. Pagallo, All rise for the honourable robot judge? using artificial intelligence to regulate ai: A debate, Technology and Regulation 2023 (2023) 45–57. doi:10.71265/0p137y60.

[14] C. G. S. Okoh, Robotic judges: A future to desire or not?, SSRN Electronic Journal (2023) 5–14. doi:10.2139/ssrn.4387301.

[15] L. Edwards, M. Veale, Slave to the algorithm? why a 'right to an explanation' is probably not the remedy you are looking for, Duke Law & Technology Review 16 (2017) 18–84.

[16] M. Libertini, M. R. Maugeri, E. Vincenti, Giustizia predittiva e giurisdizione civile. primi appunti, in: A. Pajno, F. Donati, A. Perrucci (Eds.), Intelligenza artificiale e diritto: una rivoluzione?, volume II, Il Mulino, Bologna, 2022, pp. 496–529.

[17] A. Santosuosso, G. Sartor, Decidere con l'IA. Intelligenze artificiali e naturali nel diritto, Società Editrice il Mulino, Bologna, 2024.

[18] T. Wongvorachan, K. W. Lai, O. Bulut, Y.-S. Tsai, G.-D. Chen, Artificial intelligence: Transforming the future of feedback in education, Journal of Applied Testing Technology 23 (2022) 95–116. URL: http://jattjournal.net/index.php/atp/article/view/170387.

[19] F. Kieser, P. Tschisgale, S. Rauh, X. Bai, H. Maus, S. Petersen, M. Stede, K. Neumann, P. Wulff, David vs goliath: Comparing conventional machine learning and a large language model for assessing students' concept use in a physics problem, Frontiers in Artificial Intelligence 7 (2024) 1–15. doi:10.3389/frai.2024.1408817.

[20] R. Luckin, W. Holmes, M. Griffiths, L. B. Forcier, Intelligence unleashed: An argument for AI in education, Pearson Education, London, 2016.

[21] R. Conijn, P. Kahr, C. Snijders, The effects of explanations in automated essay scoring systems on student trust and motivation, Journal of Learning Analytics 10 (2023) 37–53. doi:10.18608/jla.2023.7801.

[22] K. Holstein, J. Wortman Vaughan, H. Daumé III, M. Dudik, H. Wallach, Improving fairness in machine learning systems, Communications of the ACM 62 (2019) 60–71. doi:10.1145/3290605.3300830.

[23] H. Suresh, J. V. Guttag, A framework for understanding unintended consequences of machine learning, 2019. arXiv:1901.10002, arXiv preprint arXiv:1901.10002.

[24] F. Pasquale, The black box society: The secret algorithms that control money and information, Harvard University Press, Cambridge, MA, 2015.

[25] R. Binns, M. Veale, M. Van Kleek, N. Shadbolt, 'it's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions, in: CHI Conference on Human Factors in Computing Systems, ACM, Montreal, Canada, 2018, pp. 1–14. doi:10.1145/3173574.3173951.

[26] B. Memarian, T. Doleck, Fairness, accountability, transparency, and ethics (fate) in artificial intelligence (ai) and higher education: A systematic review, Computers and Education: Artificial Intelligence 5 (2023) 100152. doi:10.1016/j.caeai.2023.100152.

[27] D. Hooshyar, Y. Yang, Problems with shap and lime in interpretable ai for education: A comparative study of post-hoc explanations and neural-symbolic rule extraction, IEEE Access 12 (2024) 137472–137490. doi:10.1109/ACCESS.2024.3463948.

[28] X. Duan, B. Pei, G. A. Ambrose, et al., Towards transparent and trustworthy prediction of student learning achievement by including instructors as co-designers: A case study, Education and Information Technologies 29 (2024) 3075–3096. doi:10.1007/s10639-023-11954-8.

[29] L. Colonna, Teachers in the loop? an analysis of automatic assessment systems under article 22 gdpr, International Data Privacy Law 14 (2024) 3–18. doi:10.1093/idpl/ipad024.

[30] M. E. G. David, Government by algorithms at the light of freedom of information regimes: A case-by-case approach on adm systems within public education sector, Indiana Journal of Global Legal Studies 30 (2023) 105–171.

[31] T. Susnjak, Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and chatgpt, International Journal of Artificial Intelligence in

Education 34 (2024) 452–482. doi:`10.1007/s40593-023-00336-3`.

[32] M. Nagy, R. Molontay, Interpretable dropout prediction: Towards xai-based personalized intervention, International Journal of Artificial Intelligence in Education 34 (2024) 274–300. doi:`10.1007/s40593-023-00331-8`.