# Towards Trustworthy AI in STEM Education: Challenges and Strategies from the Trust-AI platform

Nikolaos Antonios Grammatikos[1,*], Evangelia Anagnostopoulou[2], Dimitris Apostolou[1,2] and Gregoris Mentzas[2]

[1]University of Piraeus, Piraeus, GR
[2]Institute of Communication and Computer Systems, National Technical University of Athens, Athens, GR

**Abstract**

Artificial intelligence (AI) in education has the potential to personalize learning, but trustworthiness is essential for its successful implementation. The present paper provides an overview of the challenges identified during the development and evaluation of the Trust-AI platform. A number of trustworthiness challenges have been identified related to the validity and reliability of AI assessments, explainability of AI models, fairness of AI algorithms that are prone to bias, and overall safety, privacy, and security of our Trust-AI platform. To address these challenges, this paper suggests strategies we have implemented during the design and development of the Trust-AI platform, as well as the mitigation actions we plan to implement during the deployment phase of our platform. The paper concludes that the implementation and deployment of trustworthy AI systems in education needs to be driven by an ongoing, comprehensive dedication to trustworthiness assessment based on ethical values and stakeholder involvement.

**Keywords**

Trustworthy AI, STEM education, Artificial intelligence

## 1. Introduction

The potential of implementing AI in Science, Technology, Engineering and Mathematics (STEM) education is tremendous; AI has the potential to bring revolutionary changes in the learning process. AI-enabled learning tools offer new capabilities in STEM learning, such as personalized learning, immediate feedback, and representation of concepts in exciting and interactive methods, which help to personalize the learning process to individual students [1]. AI technologies have the potential to accommodate both unique paces and forms of learning, and provide support tailored to the individual, which is important to the knowledge gap prevention and consequent mastering of the STEM courses.

As AI tools are growing more complex, it is not only used in basic grade calculation but is now used in smart tutoring [2], educational robots [3], and fully online teaching and learning systems [4], which are radically transforming the future of teaching and learning. However, there have been serious doubts raised about the ethical impacts and bias of AI systems, and thorough studies are needed to make sure that these tools of power are created and implemented in a way which people can trust and believe that their functionalities are sound and objective. Challenges associated with the deployment of AI in schools should be thoroughly examined to develop trust between teachers and students. For example, issues related to algorithmic bias need to be addressed, as the usage of AI might propagate or even increase social disparities based on race, gender, and socioeconomic background [5]. Moreover, privacy and security concerns regarding the amount of student data gathered are of paramount importance, which means that advanced data governance and security measures will be required [6], [7]. The necessity to gain the appropriate skills and knowledge to introduce AI into pedagogical practice and to

*TRUST-AI: The European Workshop on Trustworthy AI. Organized as part of the European Conference of Artificial Intelligence - ECAI 2025. October 2025, Bologna, Italy.*

*Corresponding author.
ngrammatikos@unipi.gr (N. A. Grammatikos); eanagn@mail.ntua.gr (E. Anagnostopoulou); dapost@unipi.gr (D. Apostolou); gmentzas@mail.ntua.gr (G. Mentzas)
0009-0007-9864-6692 (N. A. Grammatikos); 0000-0002-9795-7452 (E. Anagnostopoulou); 0000-0002-5815-8033 (D. Apostolou); 0000-0002-3305-3796 (G. Mentzas)

critically assess the results of these systems also puts pressure on educators to acquire them. Failure to have proper training contributes to the likelihood of wrongful use of AI tools, which can become detrimental to the learning process. Another significant obstacle is the digital divide, where unequal access to AI technology may further harm students of communities with low resources and deepen the current inequality in education [8],[9].

This study aims to help advance the discussion of trustworthy AI in STEM education. The focus of our research is to identify the AI trustworthiness challenges associated with the Trust-AI platform [10],[11], an AI educational system that offers personalized learning and AI-driven student assessment during STEM laboratory and experimental teaching by providing continuous assessment and guidance. To clarify this focus, our research question is to explore the key challenges related to AI trustworthiness in the case of the Trust-AI platform and to investigate the ways in which they may be successfully addressed in STEM education. Further, we propose strategies addressing the identified challenges so that educational institutions can promote an environment of trust, transparency, and accountability in AI-driven education.

## 2. Methodology

This section presents the methodology (Figure 1) that we followed to identify the challenges related to trustworthiness of our Trust-AI platform. The approach follows a multi-phase process: first, core principles were defined based on a literature review; next, stakeholder workshops captured trust-related challenges; and finally, system demonstration and evaluation involved teacher interaction, feedback collection, and validation of AI features.

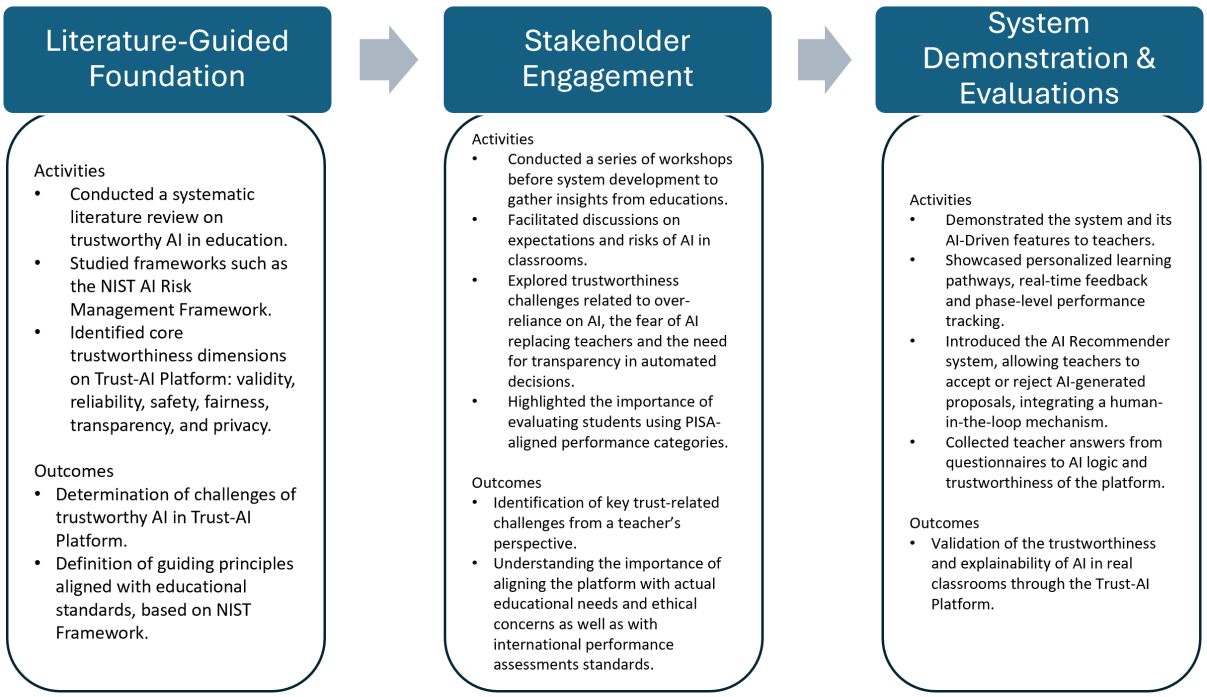


**Figure 1:** Overview of the multi-phase methodology followed to assess and enhance the trustworthiness of the Trust-AI platform. The process includes (1) a literature-guided foundation to define core principles, (2) stakeholder engagement through workshops to identify real-world concerns, and (3) system demonstration and evaluation involving hands-on teacher interaction, feedback collection, and validation of AI features.

Our research started by conducting a systematic literature review to determine the current challenges of trustworthy Artificial Intelligence in education [12]. We investigated how AI trustworthiness can be enhanced in real-world educational settings and surveyed state-of-the-art (SOTA) approaches, frameworks, and guidelines, with particular attention to their applicability in STEM education contexts.

We then established the guiding principles which were primarily based on the NIST AI Risk Management Framework [13] and its core trustworthiness characteristics of validity, reliability, safety, fairness, transparency, and privacy, as it offers a systematic, risk-driven framework that goes beyond the ethical self-assessment frameworks like ALTAI. These principles served as foundational pillars in the design of our platform and the structure of our evaluation activities. In the second phase of our methodology, we had direct interaction with relevant stakeholders, and we held a series of workshops with teachers to discuss the real possibilities and difficulties of the use of AI in the classroom. Among the key issues raised were fears of teacher replacement, over-reliance on AI, the opacity of automated decisions, and the need to align AI evaluation with educational standards such as PISA's performance level classification [14].

In the third phase of our methodology, the Trust-AI platform was demonstrated and explained during three workshops, involving 60 teachers, allowing them to have a first-hand understanding and evaluation of the system and its contribution to typical STEM teaching activities. With the Trust-AI platform, each student follows a customized learning pathway suggested by the AI assistant, which also corrects misconceptions and offers immediate feedback. This personalized interaction supports differentiated learning and encourages student autonomy while maintaining teacher oversight. By facilitating one-on-one interactions between students and tutors, this setup aims to improve the usual student-teacher dynamic and foster deeper learning. Additionally, teachers can track students' progress in real time, which enables them to offer tailored advice and step in when needed.

Additionally, the Trust-AI platform includes an AI learning pathways recommender system that flags underperforming activities using student performance data. Teachers can review these recommendations, so that they can either accept or reject them, assessing their alignment with educational standards and pedagogical goals. Their feedback is recorded and used to continuously refine the AI system through reinforcement learning, forming a human-in-the-loop architecture where educators remain central in decision-making.

Following the workshops and in order to uncover more information, we held interviews with teachers to elicit their opinion and evaluate trustworthiness characteristics of the Trust-AI platform. To quantify teachers' feedback, we also asked them to complete a structured questionnaire to gauge the perceived trustworthiness of AI effectiveness as well as perceived risks, such as over reliance, academic integrity, and data privacy. The combination of literature-grounded design, stakeholder consultation, live system evaluation, and teacher-centered feedback mechanisms enabled a comprehensive exploration of the challenges and opportunities in building a trustworthy AI system for STEM education.

## 3. AI Trustworthiness Challenges in our Educational AI Solution

This section discusses the challenges related to AI trustworthiness as identified in the Trust AI platform. We group the identified challenges under the seven characteristics of trustworthy AI systems defined by the NIST Framework [13].

### 3.1. Valid and Reliable

Validity and reliability were prioritized as the most important characteristics to ensure the trustworthiness of our platform. Failure in this area can lead to misleading guidance among students. Specifically, a major risk for the Trust-AI platform is its potential inability to accurately assess the level of a student's problem-solving skills. In case the system provides assessments that are not correct, teachers will not trust it. Teachers will not use the system if they lack confidence in it. Furthermore, inaccurate evaluation of student performance may result in recommendations for personalized learning that are not reliable. These problems have the potential to interfere with the learning process and produce unreliable educational results.

This risk of inaccuracy extends to all functions of the Trust-AI platform. The ability to suggest personalized learning paths depends entirely on the accuracy of problem-solving skills assessment. An inaccurate assessment can lead to learning recommendations that are not relevant to the students'

needs. Similarly, feedback from the system should be consistently accurate and personalized for student learning needs. Unreliable or inaccurate feedback will result in students not trusting the system.

Equally important is the robustness of the Trust-AI platform. Robustness is the feature of the system to work with the same and reliable results in different classroom conditions, incomplete student inputs, and other unpredicted situations. Without a robust platform, there will be a possibility of having unstable or inconsistent outputs given unexpected inputs and variations in classroom conditions. These failures may disrupt the process of learning by providing incorrect suggestions, lowering the accuracy of feedback, and eventually making teachers and students unwilling to use the platform.

### 3.2. Explainability

When an AI's decision-making process is a 'black box', it creates a fundamental barrier to trust. Educators cannot be expected to implement recommendations without understanding their rationale. A significant challenge is the erosion of teacher trust. If a teacher cannot understand why the AI recommended a specific learning path, they cannot professionally endorse it or integrate it into their teaching. An AI learning pathway recommendation without justification is not useful. If the system flags an activity as 'at-risk' without explaining which concepts they are struggling with, the teacher has no basis for intervention. This lack of justifying a specific decision and clarity about the overall system design and data creates an 'accountability vacuum' where no one can be held responsible for the AI's outcomes.

### 3.3. Safety

In education, safety is related to the well-being of students and teachers. One relevant identified challenge is the fear of educators that their role may be replaced by the system. The feeling of potential loss of jobs due to AI performing and taking control of the essential teaching aspects encourages professional insecurity, causing teachers to distrust and resist the emerging technology in the educational field.

Another challenge related to safety is that over-reliance on AI for basic tasks such as lesson planning and assessment could cause reduced teacher involvement in the educational process. The blindly outsourcing of the basic activities to the AI system threatens the phenomenon of de-skilling. This may mean that pedagogical competence and creativity may be impaired in educators, and this compromises the teaching process and reduces the outcome of learning in the students.

### 3.4. Accountability and Transparency

One of the major challenges regarding the implementation of AI into education concerns transparency and accountability. One of the most important elements of the educational setting is transparency, which encourages trust needed among students, educators, and institutions. Yet, this contradicts this principle directly since the AI algorithm is usually opaque, as is the case with highly advanced AI algorithms. With our Trust-AI platform, it is important for teachers to understand the logic behind the algorithm. There is a strong need that teachers to understand how AI models are assessing the problem-solving abilities of students and what data have been used to train the AI model so as not to reproduce the biases built into the system. Moreover, at the central level, there must be transparency in order to allow a clear chain of accountability. When an algorithm fails or produces a biased result, the black box characteristic of a system is not known, and it is unclear who should be held liable between a teacher or the AI developer. This confusion is not only an undermining factor of trust, but it is also a violation of procedural fairness.

### 3.5. Privacy

Privacy protection of students' behavior data, skills, and performance was raised as a major concern by teachers. The NIST Privacy Framework [13] and General Data Protection Regulation (GDPR) [15] require systems to be designed based on the principles of privacy by design with respect to individual

rights and without the possibility of unauthorized processing or identification. The possibility of reidentification, forming extensive profile data of users, and being able to misuse sensitive information poses a significant challenge. In case they are not handled properly, these security holes in privacy may result in misuse of personal data by unauthorized third parties. The loss of privacy that results due to this may in a sense be extremely damaging to the confidence of the educators in the system. This has a high potential of making teachers reluctant to stop using our platform. Within the framework of our system, the process of assessing students according to their problem-solving ability is performed utilizing entirely anonymized data. However, adherence to the principles established by GDPR is critical to guarantee that data anonymity is maintained at every level of data collection, analysis, and storage and that the utilization of such data has purely and merely educational intentions.

### 3.6. Security

Security is one of the pillars of trust in any education technology infrastructure, and the challenges it poses must be dealt with in a multidimensional manner. Although our platform is specifically built not to collect any personally identifiable student data, and hence addresses some of the most serious challenges, there are still some security concerns, particularly as far as application of an effective access control model is concerned. Since our platform works with anonymized data of interaction and assessment, it is necessary to prioritize data integrity, as well as the right to access information.

In a multi-user environment with different roles of users, like in our case where we have students, teachers and administrators access rights, incorrect access rights may create vulnerabilities that are cause of great concern to security. The possibility that the assessment results will be seen by the students is a serious one. Even anonymization of data may result in performance anxiety when students are afraid to make errors because of this perception. Also, presentation of assessment data to the students may result in inadvertent peer comparison or the sense of exposure, which interferes with psychological safety among students. Thus, a well-defined access control mechanism cannot be seen as a system feature, but rather a basic requirement to address the confidentiality of data, maintain the integrity of a system, and achieve user confidence.

Moreover, most systems containing student data are vulnerable to attacks. These risks include overt threats such as data breaches, which can seriously damage user trust, as well as more subtle forms like data poisoning, where malicious inputs manipulate the behavior of AI systems. However, our platform is less likely to face such severe consequences. Since the Trust-AI platform is designed to support the educational process, assisting students and teachers without storing sensitive personal data, it presents a lower-value target for potential attackers.

### 3.7. Fairness

In educational AI systems, fairness is crucial as any bias can reinforce or create inequalities, disproportionately affecting students based on different factors. Since our platform is not intended to hold demographic or any other sensitive personal data, direct discrimination by protected characteristics is automatically eliminated. Nonetheless, the issues of fairness remain on a more insidious, algorithmic level. One possible issue is bias in algorithms. This means that the AI models may be trained on biased data in the background. These may cause the system to be biased towards certain forms of expression or approaches to solving problems at the expense of the students who have alternative forms.

## 4. Strategies for Addressing AI Trustworthiness Challenges

To build a trustworthy AI system in the education domain, a wide range of mitigation actions must be taken across the different phases of the AI lifecycle. In this section, we present the strategies we have implemented during the design and development of the Trust AI platform, as well as the mitigation actions we plan to implement during the deployment phase of our platform to address the challenges described in the previous section.

**Table 1**
Summary of key trustworthiness challenges identified in the Trust-AI platform, along with corresponding strategic responses, mitigation actions, and potential risks if left unaddressed.

| Challenge | Strategy | Mitigating Action | Risk if Unaddressed |
|---|---|---|---|
| Invalid assessments | Accuracy of assessment and benchmarking of results; Integration of Human-In-The-Loop approaches | Correlate outcomes with PISA; continuously monitor student performance; teacher review of AI outcomes; use accurate interaction data | Unreliable educational results; misleading feedback; loss of teacher trust; withdrawal of using the platform |
| Algorithm opacity | Promote transparency and implement explainability features | Display performance metrics and explanations linked to proposals; allow human oversight | Lack of explainability and accountability; teachers may not trust system decisions |
| Privacy | Apply a privacy-by-design approach | GDPR-aligned data practices; secure anonymization of student data | Legal noncompliance; misuse of data; trust erosion |
| Fairness bias | Ensure fairness across student performance categories | Manual oversight of flagged issues; reinforcement learning with human feedback | Systematic exclusion of learners; amplifying inequities in outcomes |
| Security weaknesses | Design for minimal risk of data exposure | Clear access rules; data integrity enforcement; limited attack surface; adopt proactive security measures | Data breaches; unauthorized access; loss of stakeholder confidence |
| Teacher overreliance on AI | Maintain human oversight and pedagogical control | Keep teacher in-the-loop; promote critical assessment of AI; offer AI training | Reduced teacher agency and resistance to system adoption; poor educational outcomes |

Table 1 presents an overview of the main trustworthiness challenges identified, paired with a corresponding strategy, concrete mitigation measures implemented within the platform, and the potential risks if these issues are not addressed.

Validity in educational AI systems refers to the degree to which the system's outputs, such as student assessments, feedback and content adaptation, accurately represent the intended learning outcomes. Ensuring validity is crucial when AI is used to estimate complex traits like problem-solving skills or conceptual understanding, as flawed evaluations might mislead both educators and students.

One prominent challenge related to the validity and reliability of our platform is whether the classification of students who used the AI system into high, low, and moderate performers has been done correctly. To confirm the validity of our student performance categorization, we will constantly monitor and assess students' progress in consecutive STEM lab courses. Additionally, the classification of the students will be compared to world-renowned models. In particular, our assessments of problem-solving abilities will be correlated with the performance indicators identified in the PISA framework of the OECD Program for International Student Assessment (PISA)[15], which categorizes students as high, moderate, or low performers. Our platform adopts several mechanisms to ensure valid assessment and be consistent and aligned with international education standards.

In addition, human-in-the-loop approaches for teacher oversight should be established. That way, the educators can revise and certify AI results and take over control of the process when they believe their professional discretion overrides the technology output in order to control quality and align the technology with their professional judgment. To address this, the platform allows AI suggestions to be monitored in real-time, and the educators are offered the analytics dashboard that displays the

visualization of the student's performance and thus promotes the process of formative validation. Lastly, through providing educators with a choice to accept AI recommendations or reject them, the validity is once again strengthened so as to coordinate the automated products with the pedagogical intentions and the practical observations in a classroom environment.

The system should be highly explainable and transparent as a way of demystifying the black box and gaining confidence. The important step is to integrate explainable AI techniques, which give clear and human understanding explanations of the results of the system. This enables the teachers and students to know the reasons that a certain learning path was suggested, which is central to establishing trust. In our system, we have implemented explainability mechanisms that provide insights into the reasoning behind recommended learning pathway changes. By analyzing and displaying student performance data from prior implementations, our platform provides an explainability mechanism (Figure 2). The system presents aggregated performance metrics for each activity and highlights risk indicators when an activity either poses consistent challenges or lacks sufficient challenge for learners. This makes the rationale behind each recommendation explicit and provides teachers with transparent evidence supporting AI-generated suggestions. Specifically, the system offers teachers explanations as to why it identifies an activity as difficult for students and suggests that it needs to be modified. For every AI-generated activity suggestion, specific performance metrics are linked and made fully visible to the reviewing teacher. Educators can accept or reject these proposals, with each decision logged alongside the reviewer's identity and a timestamp, supporting human oversight and auditability.

Q1 - maximum kinetic energy & cutoff voltage

| Category At Risk | Flag Type | Reason |
|---|---|---|
| Moderate | Extreme correctness pattern | Extreme correctness pattern: High (20.8%) or Low (92.3%) group not at extremes, Moderate wrong is high (76.9%). |
| Moderate | Engagement risk (too fast, question) | Engagement risk: Moderate group's Avg Time (58.7s) is > 1 std below mean (84.23 ± 24.45). |
| Low | Engagement risk (too slow, question) | Engagement risk: Low group's Avg Time (117.2s) is > 1 std above mean (84.23 ± 24.45). |

**Figure 2:** Explainability view of the recommendation system. For each activity, the platform surfaces risk flags by learner category (High/Moderate/Low) when dynamic correctness thresholds are exceeded or engagement timing deviates from the cohort mean (e.g., >1 SD)

To make sure the AI system does not become a crutch instead of the helping tool, a culture of responsible use has to be established. Both students and teachers should have clear and explicit policies outlining the effective use of AI and establishing acceptable limits between the use of AI and another form of academic dishonesty. This needs to be followed by AI training for teachers, which goes beyond technical knowledge to encourage the ability to critically assess the AI results and the limitations of the technology. The system itself is created in such a manner that it enhances and does not substitute for a person and his abilities. The AI can take certain administrative duties conventionally assigned to educators so that these professionals can work on the duties that only humans can perform. The concern with the integrity of the educational process also lies in the fact that it is essential that a human be in a loop during critical decision-making.

The security of student data is not negotiable and needs privacy by design strategy. This implies incorporating strong privacy-enhancing technologies at the beginning such as end-to-end encryption and strict regimes of data minimization to collect only the data that is strictly necessary. The system should be fully compliant with data protection laws such as GDPR and should have clear policies per user that allow him or her full control over the data. The school districts are supposed to maintain ultimate ownership of all the data of the students, and the contracts may not allow the vendor to use the data for other unintended purposes. There is also the need to adopt proactive measures such as giving the system frequent and stringent security scrutinizes, penetration, and malicious resistance testing to guard against external risks to security, such as the ability to deport venerability and counter malicious attacks that may occur.

A privacy-by-design approach has been developed, minimizing data collection to only what is necessary for its operation. No student data are collected or stored, and student interactions are linked to system-generated identifiers. Teacher accounts require names and emails for authentication on the

platform. This data is collected with explicit consent during account creation, under clearly defined terms of use and data handling policies. All teacher data are securely stored, access-controlled and never shared with third parties.

The solution was designed with security in mind and thus no identifiable student data is stored or processed. All data of the interaction is anonymized and can only be linked with machine-generated identifiers, keeping no sensitive information revealed and stored. With the implementation of this architecture, the threat of being victimized by the misuse of data or use of data to expose their identity as a result of breach of security is extremely minimized.

The willingness to implement fairness demands the initiative to investigate and detect bias in all phases of the AI lifecycle. It is important to integrate bias detection algorithms and conduct regular bias audits that can identify and correct discriminatory patterns. Using diverse, and representative data to train the models is necessary to avoid a possible introduction of bias into the system in the first place. Last but not least, a multi-stakeholder AI ethics review board can offer essential guidance to keep the development of fair AI practices.

Our solution, without relying on demographic or personal identifiers, supports equitable learning experiences by dynamically analyzing student performance. Interactional data such as correctness, and timing are the only factors on which all AI recommendations and analytics are made, which lowers the probability of bias caused by background knowledge and knowledge. Moreover, the system combines category-sensitive performance analysis and identifies activities that discriminate against any performance groups, actively managing equity. Educators maintain absolute control over identified problems and suggested solutions and play the role of a human-in-the-loop to approve or ignore AI recommendations. Reinforcement learning in taking a decision to either approve a proposed change or even reject it encourages fairness as well as reduces algorithmic biases.

## 5. Conclusion

In this paper we presented a comprehensive analysis of the AI trustworthiness challenges associated with the Trust-AI platform identifying significant challenges across validity, reliability, explainability, fairness, safety, and privacy. To address these issues, a wide range of mitigation actions must be taken during the design and development of our platform. We also presented several strategies for addressing the identified challenges, as well as the mitigation actions we plan to implement during the deployment phase of our system to address the identified challenges.

Our study is limited in scope by focusing on one specific AI system; still, it helped us identify several empirical findings and draw the following conclusions that may help educational institutions to promote an environment of trust, transparency, and accountability in AI-driven education. Building and sustaining a trustworthy, educational AI system requires a framework that incorporates validity assessment, explainable AI methodologies, bias detection algorithms, human-in-the-loop approaches for teacher oversight, and teacher training. Moreover, the successful integration of AI in education is not merely a technical achievement but an ongoing commitment to ethical principles and stakeholder collaboration, ensuring that technology serves to empower, not undermine, the fundamental goals of learning.

## Acknowledgments

The work presented here is funded by the European Union's Horizon FAITH project (Fostering Artificial Intelligence Trust for Humans towards the optimization of trustworthiness through large-scale pilots in critical domains), Grant agreement No: 101135932. The work presented here reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.

## Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

## References


[1] M. Chassignol, A. Khoroshavin, A. Klimova, A. Bilyatdinova, Artificial intelligence trends in education: a narrative overview, Procedia Computer Science 136 (2018) 16–24.

[2] K. Meenakshi, R. Sunder, A. Kumar, N. Sharma, An intelligent smart tutor system based on emotion analysis and recommendation engine, in: 2017 International Conference on IoT and Application (ICIOT), IEEE, 2017, pp. 1–4.

[3] N. Wang, W. L. Johnson, Pilot study with rall-e: Robot-assisted language learning in education, in: Intelligent Tutoring Systems, Springer, Croatia, 2016, p. 514.

[4] R. Yilmaz, H. Yurdugül, F. G. K. Yilmaz, M. Şahin, S. Sulak, F. Aydin, Oral, Smart mooc integrated with intelligent tutoring: A system architecture and framework model proposal, Computers and Education: Artificial Intelligence 3 (2022) 100092.

[5] O. Boateng, B. Boateng, Algorithmic bias in educational systems: Examining the impact of ai-driven decision making in modern education, World Journal of Advanced Research and Reviews 25 (2025) 2012–2017.

[6] I. A. Ismail, J. M. R. Aloshi, Data privacy in ai-driven education: An in-depth exploration into the data privacy concerns and potential solutions, in: AI Applications and Strategies in Teacher Education, IGI Global, 2025, pp. 223–252.

[7] I. A. Ismail, Protecting privacy in ai-enhanced education: A comprehensive examination of data privacy concerns and solutions in ai-based learning, in: Impacts of Generative AI on the Future of Research and Education, 2025, pp. 117–142.

[8] Y. Waqar, S. Rashid, F. Anis, Y. Muhammad, Digital divide inclusive education: Examining how unequal access to technology affects educational inclusivity in urban versus rural pakistan, Journal of Social & Organizational Matters 3 (2024) 1–13.

[9] A. A. Tawfik, T. D. Reeves, A. Stich, Intended and unintended consequences of educational technology on social inequality, TechTrends 60 (2016) 598–605.

[10] N. A. Grammatikos, E. Anagnostopoulou, D. Apostolou, G. Mentzas, An ai-powered learning companion for adaptive and personalized stem education, in: International Symposium on Chatbots and Human-Centered AI, Springer Nature Switzerland, Cham, 2024, pp. 87–95.

[11] N. A. Grammatikos, E. Anagnostopoulou, D. Apostolou, G. Mentzas, A conversational digital assistant for stem education, in: 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA), IEEE, 2023, pp. 1–7.

[12] E. Anagnostopoulou, N. A. Grammatikos, D. Apostolou, G. Mentzas, Trustworthy ai in education: A roadmap for ethical and effective implementation, in: Proceedings of the 13th Hellenic Conference on Artificial Intelligence, 2024, pp. 1–7.

[13] E. Tabassi, Nist ai rmf playbook, https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook, 2023. Accessed: 2025-07-21.

[14] A. Schleicher, PISA 2018: Insights and interpretations, OECD Publishing, 2019.

[15] P. Regulation, General data protection regulation, Intouch 25 (2018) 1–5.