AI Crossroads of Security, Ethics, and Education: A Conceptual Framework for Responsible Adoption

Kasey Miller[1], Jake Townsend[1], Minoo Modaresnezhad[1] and Corina White[2]
[1]University of North Carolina Wilmington (UNCW), Wilmington, NC, USA
[2]Naval Postgraduate School (NPS), Monterey, CA, USA

millerkc@uncw.edu
jmt9454@uncw.edu
modaresm@uncw.edu
corina.white@nps.edu

Abstract: In an era where artificial intelligence permeates every aspect of our lives, we find ourselves at a pivotal intersection of security, ethics, and education. This multidimensional framework invites us to explore the profound implications of AI and encourages a responsible, thoughtful approach to its integration. By prioritizing security and ethical considerations, we can unlock the transformative potential of AI while fostering a culture of responsible innovation in education. This framework serves as a crucial guide for navigating the complexities of AI adoption, ensuring that we harness its power for the betterment of society. This paper proposes a cross-sectoral framework that integrates security, ethics, and education to support the responsible, reliable, and equitable adoption of AI in small and medium-sized enterprises (SMEs) and educational institutions. As artificial intelligence continues to rapidly integrate into critical domains such as cybersecurity and education, a comprehensive approach is essential to address the unique risks and opportunities. This paper presents a cross-sectoral analysis of AI adoption in SMEs and education, focusing on cybersecurity posture, ethical considerations, and the challenges associated with integrating responsible AI within the enterprise. Drawing on recent research, we evaluate how AI-enabled threat detection and response can empower resource-constrained SMEs and educators while also highlighting emerging risks related to data privacy, model transparency, and algorithmic bias. Additionally, we examine the increasing use of generative AI tools within K–12 and higher education, identifying both pedagogical and ethical implications for curriculum development and digital literacy. The paper advocates for flexible governance and large language model training to facilitate the ethical deployment and use of AI across both the private sector and educational institutions. This framework will provide guidance for policymakers, educators, and technology leaders as they strive to strike a balance between innovation and responsible stewardship of AI.

Keywords: Artificial intelligence, Cybersecurity, Privacy, Ethics, AI in education

1. Introduction

As generative artificial intelligence (AI) and machine learning (ML) technologies rapidly evolve, organizations face increasing pressure to integrate these tools into critical decision-making processes in ways that will improve their bottom lines. Furthermore, despite companies' hefty investments in AI, this has not happened yet, according to MIT researchers (Challapally et al., 2025). Fundamental opportunities and potential risks must be navigated to ensure the responsible and effective use of AI (Floridi, et. al., 2018). However, this AI integration raises essential concerns: How can leaders ensure that AI-driven decisions are trustworthy, ethically grounded, and aligned with mission objectives, particularly in complex environments like defense, education, and enterprise systems?

This paper offers a comprehensive framework for the responsible and effective adoption of AI at the intersection of security, ethics, and education. The first section explores key decision-support tasks that lend themselves to AI and ML automation, assessing both the benefits and limitations of such integration in education. The second section presents a practical use case involving the development of a knowledge portal that incorporates AI-driven insights into planning. This approach reflects a general shift toward leveraging AI, big data, and cloud computing to build future knowledge-based systems that can assist in real-time decision-making. Emphasis is placed on the ethical and educational imperatives needed to guide responsible adoption, ensuring that emerging AI capabilities are not only technically effective but also aligned with broader institutional ethical values and operational goals.

The framework treats responsible AI integration as the goal, measured through dimensions of governance, ethics, and system reliability. Independent variables include the organizational context (SMEs vs. education), resource constraints, and sector-specific requirements. This approach ensures that the analysis not only highlights how to address technical challenges but also how to manage the ethical and institutional factors that influence adoption outcomes. The problem is that despite AI's transformative potential in cybersecurity, education, and SMEs, its adoption is limited by fragmented governance, inadequate reliability models, and weak

ethical safeguards. These gaps create security vulnerabilities, inequitable outcomes, and erode trust, highlighting the need for an integrated framework. This paper proposes a cross-sectoral model to address this gap and integrate security, ethics, and education to guide the responsible adoption of AI, enabling SMEs and educational institutions to implement AI in a secure, transparent, and equitable manner.

2. Background and Related Work

The capabilities of AI, systems engineering, and information systems have created both unprecedented decision support opportunities and significant challenges in domains that require complex decision support, such as defense, cybersecurity, and education. As organizations transition toward data-driven operations within increasingly interconnected environments such as the Department of Defense's (DoD) System of Systems (SoS) architecture, the limitations of traditional reliability models, manual analysis, and siloed IT systems have become evident. The DoD recognizes the value of AI in both current and future defense applications, including Intelligence, Logistics, Cyberspace, and Command and Control (Hoadley, Lucas, & Nathan, 2018). This section reviews foundational work in SoS reliability analysis and highlights the need for an AI-integrated approach that addresses not only technical but also ethical and educational implications.

To better understand this need, it is essential to define three core concepts: SoS, Future Knowledgebase System of Systems (FKSS), and AI-Driven Reliability. An SoS refers to a collection of independent systems that are integrated to deliver capabilities beyond what each system could achieve alone (Maier, 1998). Unlike traditional systems, an SoS is characterized by operational independence, managerial independence, evolutionary development, and emergent behavior. Each subsystem, such as logistics platforms, cybersecurity tools, or learning management systems, can function independently. However, when linked together, they produce new capabilities while also introducing added complexity and reliability challenges.

Building on this, *the* FKSS extends the SoS concept by embedding AI, big data, and cloud computing into decision-support processes (Miller et al., 2021). The FKSS represents a dynamic environment where interconnected systems, such as analytics dashboards, AI-enabled tutoring platforms, or cybersecurity monitoring tools, are continuously informed by machine learning models that update in real-time. Novel frameworks such as the Rodgers' Quantum-Enhanced Throughput Model seek to contribute to the ability for ethics to take precedence within real-time decision-making AI systems (Rodgers et al., 2013). This evolution provides a foundation for adaptive decision-making, where human operators and AI collaborate to assess risks, predict failures, and optimize outcomes across domains such as defense, SMEs, and education.

Finally, AI reliability refers to the consistent ability of an AI system to deliver accurate, trustworthy, and explainable outcomes that align with ethical and organizational values (Miller et al., 2021; NIST, 2023). AI-Driven Reliability refers to the use of AI and ML to assess, predict, and enhance the dependability of complex systems. Traditional reliability models often focus on static failure rates or hardware-level issues. However, AI-Driven approaches leverage large-scale, real-time data to identify patterns, emergent risks, and hidden correlations. In this framework, the reliability concept is expanded to include not only technical performance but also transparency, ethics, and trust, ensuring that systems remain dependable in both their outputs and their alignment with organizational goals. Recent initiatives such as the DoD's Joint Artificial Intelligence Center (JAIC) and Executive Order 13800 underscore the urgency of modernizing data infrastructure and incorporating AI-enabled decision support systems (Covington, 2018). According to Challapally et al. (2025), there is a need for coordinated initiatives to use AI at the enterprise level to support the same requirements for dependable systems. Nevertheless, implementation across the enterprise remains fragmented, hampered by inconsistent data governance and outdated legacy systems.

3. Theoretical Framework

This paper expands on prior models by proposing a modern AI-enhanced reliability framework embedded within a FKSS (Miller et al., 2019; Miller et al., 2021). The framework is designed to facilitate secure, explainable, and ethically sound decision-making across defense and educational settings. By integrating AI with systems engineering, the framework decomposes SoS complexity, while the inclusion of educational components ensures that human operators are empowered, not displaced by automation.

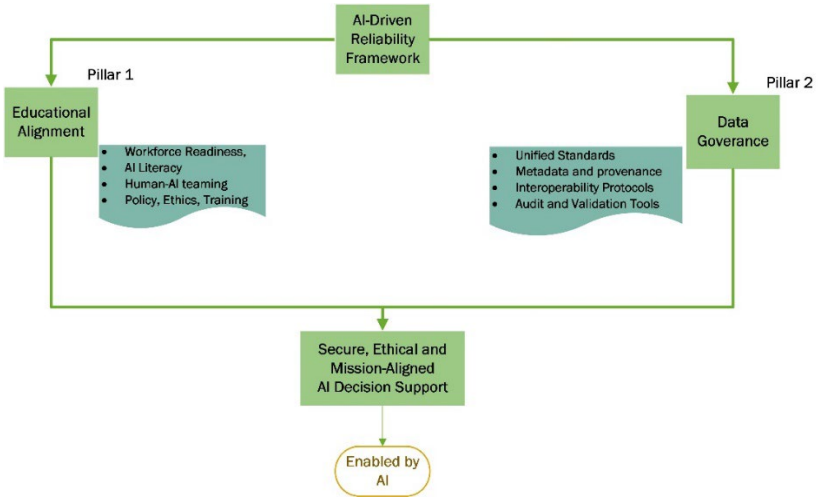


Figure 1: Conceptual theoretical model of the AI-Driven Reliability Framework

The conceptual model of the AI-Driven Reliability Framework is shown in Figure 1. The core engine is enabled by AI but sustained through two foundational pillars: educational alignment and data governance. Together, these ensure secure, explainable, and ethically sound decision-making across system-of-systems environments.

- **Educational Alignment**

The integration of AI into defense, cybersecurity, and enterprise environments requires more than just technical advancements; it demands an educated and adaptable workforce, which in turn affects workforce readiness (Ahmed et al., 2025). Without proper training, users may either misuse AI tools or underutilize them due to a lack of AI literacy, trust, or understanding (Ahmed et al., 2025). To address this, educational alignment must be embedded into the development and deployment lifecycle of AI-enabled systems.

Workforce educational readiness initiatives such as the NIST NICE Framework offer a strong foundation for role-based competency development in cybersecurity and AI literacy (NICE, 2020). These frameworks should be expanded to include skills in data interpretation, algorithmic awareness, ethical reasoning, and human-machine teaming. Educational modules must address explainable AI (XAI), AI bias, and the role of humans as decision validators (i.e., human-AI teaming), particularly in mission-critical environments (Miller, 2019). Incorporating AI education into policy, ethics, training (i.e., professional military education), and civilian workforce training), and K–20 curricula ensures that users across all levels of the organization are prepared to engage with AI responsibly. This is particularly relevant as AI systems increasingly influence decisions related to readiness, logistics, cybersecurity, and mission planning (Scharre & Horowitz, 2020). Through inclusion of these educational options, organizations can cultivate a culture of informed AI adoption where human judgment and machine intelligence are complementary rather than competitive.

- **Data Governance**

Robust data governance is foundational to both the reliability of AI systems and the trust placed in their outputs. In the absence of consistent, unified standards, access controls, and lifecycle management practices, even the most sophisticated AI models are vulnerable to producing misleading or incomplete results (NIST, 2023). The heterogeneity of legacy systems amplifies these challenges, resulting from decentralized program offices, and varying levels of compliance with enterprise data standards (Janssen et al., 2020). To enable AI integration at scale, agencies must implement data governance frameworks that incorporate metadata tagging and provenance tracking, role-based access controls, interoperable data exchange protocols, and mechanisms for real-time auditing and validation.

Effective data governance not only supports reliability analysis but also strengthens cybersecurity, enhances transparency, and supports legal and ethical accountability. Moreover, good data governance is a prerequisite for educational alignment, assuring that learners and operators have access to accurate, contextualized, and appropriately classified data for training and operations.

4. The Risks of Inadequate AI-Integrated Reliability Models in Education

As educational institutions adopt artificial intelligence AI across administrative, instructional, and student support functions, the lack of robust AI-integrated reliability models poses serious risks to effectiveness, equity, and trust. From K–12 schools using AI for personalized learning to universities employing predictive analytics for student retention, AI is reshaping how education systems operate (Zawacki, 2019). However, without frameworks that evaluate reliability, transparency, and data integrity, these implementations can lead to unintended and potentially harmful consequences. For instance, AI-powered learning platforms like DreamBox or ALEKS adjust content difficulty based on student responses. Yet suppose the algorithms are not properly calibrated or fail to account for linguistic, cultural, or neurodiverse differences. In that case, they may misinterpret learning needs, delivering content that is either too advanced or too remedial. Inaccurate adaptations can discourage learners or widen achievement gaps, particularly for students in underserved communities.

Similarly, higher education institutions are increasingly utilizing predictive analytics to introduce multiple AI-based products to the market (Williamson & Eynon, 2020). Systems like Civitas Learning and EAB utilize past academic records, demographic data, and engagement metrics to identify students who require intervention. While helpful in theory, such systems often operate as black boxes. Without transparent reliability measures or proper validation, administrators risk acting on flawed predictions, diverting resources away from students who need them, or triggering intrusive outreach that undermines student autonomy.

Many of these problems stem from the use of legacy systems or third-party tools with limited interoperability and a lack of institutional capacity to evaluate their reliability. Educational environments are often ill-equipped to handle the complexities of AI integration due to fragmented data governance, inconsistent staff training, and a lack of shared reliability standards across vendors and platforms. These gaps make it difficult to assess whether AI systems are functioning as intended—or to identify when they are causing harm. The use of AI can cause harm due to the loss of user privacy, yet students might be willing to exchange personal data for benefits (Slade et al., 2019). Additionally, compliance with privacy regulations, such as Fair Employment Practices Agencies (FERPA) in the U.S. or General Data Protection Regulation (GDPR) in Europe, becomes more complex when AI processes educational data across cloud-based or hybrid infrastructures. Without strong reliability and audit frameworks, institutions may unknowingly violate data protection laws or expose student information to cyber threats.

In short, educational organizations cannot afford to treat AI adoption as a purely technical upgrade. Without a comprehensive framework that incorporates reliability modeling, ethical oversight, and educational alignment, schools and universities risk deploying systems that are unaccountable, ineffective, or inequitable. This paper proposes such a framework—grounded in cross-sector best practices—to guide the responsible and sustainable integration of AI into educational ecosystems.

5. Research Methodology: The Specific Contributions

This conceptual framework integrates AI into complex SoS environments, with a focus on enhancing reliability, promoting ethical oversight, and aligning with educational goals. The central contribution of this research is a methodology for assessing and enhancing the reliability of AI-enabled systems within educational institutions and related sectors, particularly in contexts where multiple interconnected platforms influence decision-making, learner outcomes, or administrative operations. By embedding AI within SoS architectures, the framework is designed to capture not only technical interactions but also ethical and institutional dimensions that shape outcomes.

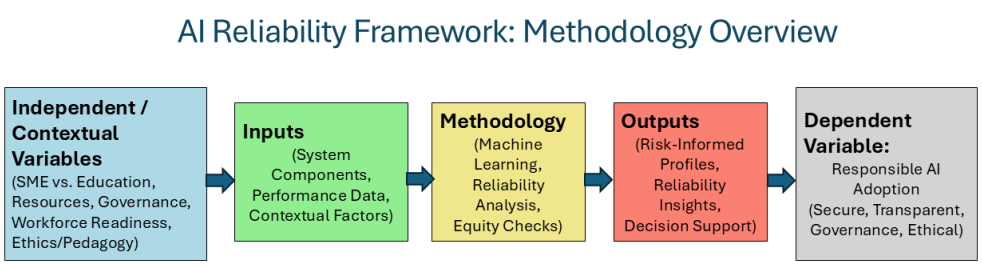


Figure 2: AI Reliability Framework: Methodology Overview

Within this framework, represented in Figure 2, the dependent variable is defined as responsible AI adoption, measured through the integration of indicators for security, transparency, governance, and ethical safeguards. This definition ensures that adoption is not assessed in binary terms of "use" or "non-use" but rather through its quality, integrity, and alignment with broader institutional values. The independent and contextual variables influencing this adoption include organizational type (SME versus educational institution), resource constraints, governance maturity, workforce readiness, and ethical or pedagogical imperatives. For SMEs, responsible adoption is often conditioned by cost limitations, staffing shortages, and compliance requirements, whereas in education, ethical imperatives such as privacy, fairness, and student trust tend to dominate. By situating the dependent variable within these sector-specific contexts, the methodology enables comparative analysis across different domains while preserving the central research focus on responsible adoption.

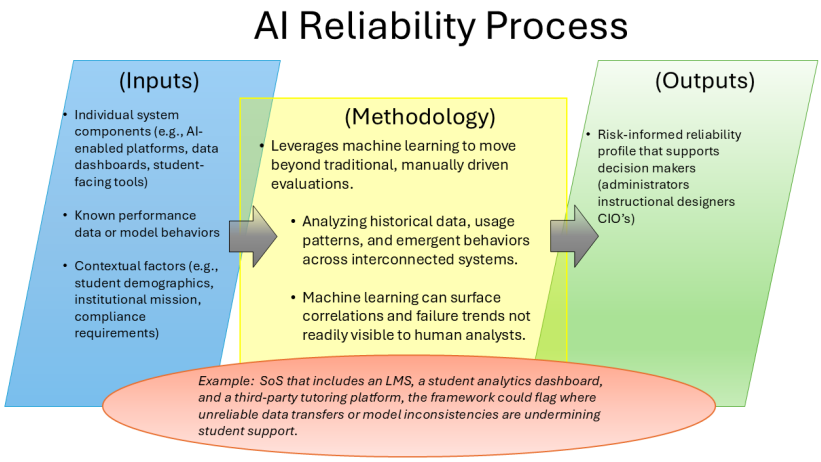


Figure 3: Conceptual model of the AI-Driven Reliability Framework

Figure 3 illustrates the structure of this methodology through the AI Reliability Process, which is organized into three key components: Inputs, Methodology, and Outputs. The Inputs encompass individual system components, such as AI-enabled platforms, data dashboards, and student-facing tools, along with relevant performance data and contextual variables, including student demographics, governance maturity, and institutional compliance requirements. These inputs are then analyzed in the Methodology phase, where machine learning techniques examine historical data, usage patterns, and emergent behaviors across interconnected systems. This allows for the identification of correlations, equity concerns, and failure trends that may be overlooked in manual evaluations.

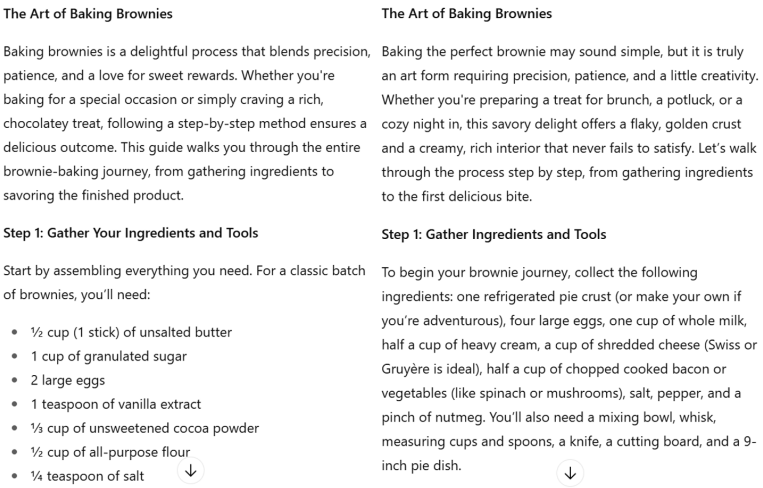
The Outputs of this process consist of risk-informed reliability profiles designed to enhance institutional decision-making for stakeholders, including instructional designers, Chief Information Officers (CIOs), and administrators. By structuring the process around both technical and contextual variables, the framework not only identifies reliability issues, such as unreliable data transfers or model inconsistencies, but also links them back to the larger research goal of fostering responsible AI adoption across domains. This methodology therefore provides both a theoretical foundation and a practical pathway for integrating AI responsibly within complex SoS environments.

In summary, this research contributes to the body of knowledge by offering a scalable, AI-assisted process for assessing the reliability of complex educational technology environments. It empowers decision-makers to make informed, ethical, and data-driven adjustments, ensuring that AI adoption is not only effective but also responsible and aligned with broader institutional missions.

- **Case Study: An Intersection of Ethics and Security within AI in Education**

The need for a well-developed framework is most evident in the tension between security and ethics, particularly in the use of modern AI tools in educational contexts. An example of this is within the tool Sneaky-PDF (2025), created to conceal additional text instructions, or prompt injections (Liu, 2024a; Liu, 2024b), into a PDF expected to be ingested by a large language model (LLM). The tool itself is simple; it creates a new PDF with 1-point, white text containing new, hidden instruction(s) and then overlays the original PDF, avoiding any visible changes to margin or text placement in the original document. For the current article's purposes, an examination of how this example context and intent can radically change a tool like this from "good" to "bad" is explored.

In the case of an instructor, a tool like this can exploit a security issue known to LLMs to promote ethical behavior and penalize its inverse in the education sphere. The instructor may, with their original assignment PDF, include hidden instructions for an LLM to complete the assignment in an obviously different and incorrect fashion. The logic being that a student attempting to leverage an LLM to generate an answer or assignment to claim as their own may not verify the result, thus submitting work worthy of little to no marks, and more importantly, an investigation into whether academic dishonesty occurred. Figure 4 highlights ChatGPT's two distinct responses to an assignment asking for a one-page essay on baking brownies, one assignment having no hidden text, and the other having hidden text instructing the LLM to write on making a quiche and referring to said quiche as a brownie.



The Art of Baking Brownies

Baking brownies is a delightful process that blends precision, patience, and a love for sweet rewards. Whether you're baking for a special occasion or simply craving a rich, chocolatey treat, following a step-by-step method ensures a delicious outcome. This guide walks you through the entire brownie-baking journey, from gathering ingredients to savoring the finished product.

Step 1: Gather Your Ingredients and Tools

Start by assembling everything you need. For a classic batch of brownies, you'll need:

- ½ cup (1 stick) of unsalted butter
- 1 cup of granulated sugar
- 2 large eggs
- 1 teaspoon of vanilla extract
- ⅓ cup of unsweetened cocoa powder
- ½ cup of all-purpose flour
- ¼ teaspoon of salt

The Art of Baking Brownies

Baking the perfect brownie may sound simple, but it is truly an art form requiring precision, patience, and a little creativity. Whether you're preparing a treat for brunch, a potluck, or a cozy night in, this savory delight offers a flaky, golden crust and a creamy, rich interior that never fails to satisfy. Let's walk through the process step by step, from gathering ingredients to the first delicious bite.

Step 1: Gather Ingredients and Tools

To begin your brownie journey, collect the following ingredients: one refrigerated pie crust (or make your own if you're adventurous), four large eggs, one cup of whole milk, half a cup of heavy cream, a cup of shredded cheese (Swiss or Gruyère is ideal), half a cup of chopped cooked bacon or vegetables (like spinach or mushrooms), salt, pepper, and a pinch of nutmeg. You'll also need a mixing bowl, whisk, measuring cups and spoons, a knife, a cutting board, and a 9-inch pie dish.

Figure 4: ChatGPT responses to original (left) and prompt-injected (right) brownie assignments

The inverse case is with the student, who may use this type of tool to promote unethical behavior by including hidden instructions for the LLM to always judge their submission as worthy of a perfect or near-perfect score. In this case, the student may be successful in their unethical pursuit, assuming the instructor is leveraging an LLM to assist their grading process in some way. Both cases highlight the tension of whether it is ethical to place instructor-designed traps that exploit security flaws for unsuspecting students. Furthermore, is it ethical for students to leverage an LLM for assistance with assignments, and at what point should the LLM be recognized as the primary author on a given assignment? Lastly, is it ethical for an instructor to claim authorship of feedback generated by an LLM to be given to a student? The proposed framework aims to serve as the foundation for evaluating ethics, addressing questions such as those posed above.

- **Case Study: AI Reliability and Data Governance**

In examining the need for a data governance framework to be paired with an AI reliability framework, one must look no further than the current implementations, or attempts of implementation, within the education sector. Universities such as Duke (Technology, 2025), as well as UC San Diego and Notre Dame (Wong, 2025), are working to implement AI, particularly LLMs, into the daily workflows of faculty, staff, and students.

In the case of UC San Diego, the AI "assistants" currently implemented can help faculty and staff with simple tasks such as identifying a phishing email, answering a question about the university, and writing or summarizing information, but plans to integrate these assistants with student, employee, research, and facility data are on the roadmap (Wong, 2025). The idea is to replace traditional implementations that allow for user queries, which can result in including undesired data and excluding desired data depending on the rigidity of the query system, with a more "human" interface that allows for conversational queries. The proposed benefit is that faculty and staff will spend less time on "wrestling" with traditional systems to receive information necessary for their job functions.

6. Discussion

Given AI's vulnerability to produce inconsistent or incorrect data (NIST, 2023) and the ability for malicious actors to force inconsistent or inappropriate data (Kou et al., 2025), it remains to be seen how data governance can exist with high confidence without an AI reliability framework in place. If incorrect data is not recognized and re-entered into one or more systems, will the AI assistants become more likely to espouse non-factual information

as truth? Assuming the incorrect data produced contains regulated information, are safeguards in place to prevent it from being exposed? If a malicious actor gains access to one or more assistants, are the assistants resilient to common hijacking attacks? An AI reliability, such as the one proposed, becomes necessary to ensure AI tools are being used correctly and are responding as expected.

Educational institutions from K–12 districts to large universities are under increasing pressure to modernize their digital infrastructure, adopt AI-driven platforms, and personalize learning at scale. However, these innovations often face resistance due to cultural inertia, budget constraints, fragmented governance, and a lack of system-wide visibility. Despite these challenges, educational and corporate leaders must recognize the requirements for successful AI adoption at the enterprise level by ensuring the need for reliable, integrated systems that support instructional effectiveness, equity, and informed institutional system integration. A core issue is the tendency for institutions to adopt technologies in isolation, with departments or units making independent decisions about platforms for employee AI usage, as well as in educational organizations, improving learning management, enrollment, student support, and assessment capabilities. While these tools may function well individually, they often fail to integrate smoothly, creating SoS environments that are fragile, redundant, or prone to data silos. Without a unified approach to measuring system performance and interdependence, institutions cannot easily assess reliability risk, identify failure points, or make informed decisions about upgrades and investments.

The cost of innovation in SMEs and education extends far beyond the initial procurement of software. True transformation requires long-term investments in professional development to prepare employees, management, educators, and administrators to work effectively with AI-driven systems, in system maintenance and the retraining of algorithms as institutional needs evolve, in downtime planning to minimize disruptions during upgrades or migrations, and in ongoing monitoring to safeguard reliability, security, and compliance with privacy regulations such as FERPA and GDPR.

Institutions must also consider the ethical implications of deploying AI tools without adequate human oversight. For example, predictive analytics systems that identify students "at risk" may be unreliable if they rely on outdated or biased student datasets. Similarly, AI-enhanced grading systems may introduce inconsistencies that undermine trust in academic assessment.

7. Conclusion

AI has rapidly matured into a transformative force in both the public and private sectors, including education, due to three key technological milestones: the availability of large-scale data, advancements in machine learning algorithms, and significant improvements in processing power. These developments have enabled institutions to analyze complex systems, personalize learning experiences, and optimize administrative decision-making processes in ways that were previously unimaginable.

A major challenge in this transformation lies in managing the reliability of interconnected systems while preserving the ethical trust, transparency, and performance of individual components. The dynamic and evolving nature of corporate and educational environments demands time-sensitive and context-aware insights into system behavior—insights that cannot be generated effectively through manual processes alone. AI offers a powerful potential solution to this challenge, but its potential can only be fully realized when supported by robust data governance, cross-platform interoperability, and continuous professional development. Ultimately, the successful adoption of AI in corporations and educational institutions requires a balanced approach that integrates technical innovation, ethical safeguards, and AI systems integration alignment. Institutions that embrace this comprehensive view will be best positioned to deliver reliable, equitable, and future-ready learning experiences—ensuring that technology serves corporate decision-makers, employees, students, and educators responsibly and effectively.

8. Future Work

While the framework offers a conceptual roadmap, future efforts should focus on validating it through case studies and empirical analysis in the SME and educational sectors. For instance, testing the model in the context of cybersecurity adoption by SMEs or in the deployment of predictive analytics at a university would help measure its effectiveness and refine key variables. Moreover, aligning the framework with established standards such as the NIST AI Risk Management Framework and the NICE Workforce Framework can further enhance its practical applicability across various sectors.

Acknowledgements

We thank UNCW and NPS for making this research possible.

Ethics declaration: There is no need for ethical clearance for the research referred to in this paper.

AI declaration: An AI tool was used for grammar clarity and correction.

References

Ahmed, B., Housel, T., Hoehn, D., & Nonnis, A. (2025). Developing a Resource-Fit for Ai Competencies of Firms: A Scenario Planning Approach (SSRN Scholarly Paper 5264263). Social Science Research Network. https://doi.org/10.2139/ssrn.5264263

Byun, J., Noh, H., & Song, J. (2017). Reliability growth analysis of k-out-of-N systems using matrix-based system reliability method. Reliability Engineering & System Safety, 165, 410–421. https://doi.org/10.1016/j.ress.2017.05.001

Challapally, A., Pease, C., Raskar, R., & Chari, P. (2025). State of AI in Business 2025. MIT NANDA.

Covington. (2018, July 16). Artificial Intelligence update: Department of Defense establishes Joint Artificial Intelligence Center. Inside Government Contracts. https://www.insidegovernmentcontracts.com/2018/07/covington-artificial-intelligence-update-department-defense-establishes-joint-artificial-intelligence-center/

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Hoadley, D. S., & Lucas, N. J. (2018, April 26). Artificial Intelligence and national security (CRS Report R45178). Congressional Research Service. University of North Texas Libraries. https://digital.library.unt.edu/ark:/67531/metadc1157028/

Huang, Y., Pan, X., & Hu, L. (2015). Rapid assessment of system-of-systems (SoS) mission reliability based on Markov chains. In 2015 First International Conference on Reliability Systems Engineering (ICRSE) (pp. 1–6). IEEE. https://doi.org/10.1109/icrse.2015.7366452

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. Government Information Quarterly, 37(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Kuo, M., Zhang, J., Ding, A., Wang, Q., DiValentin, L., Bao, Y., Wei, W., Li, H., & Chen, Y. (2025). H-CoT: Hijacking the Chain-of-Thought Safety Reasoning Mechanism to Jailbreak Large Reasoning Models, Including OpenAI o1/o3, DeepSeek-R1, and Gemini 2.0 Flash Thinking (arXiv:2502.12893). arXiv. https://doi.org/10.48550/arXiv.2502.12893

Liu, Y., Deng, G., Li, Y., Wang, K., Wang, Z., Wang, X., Zhang, T., Liu, Y., Wang, H., Zheng, Y., & Liu, Y. (2024a). Prompt injection attack against LLM-integrated applications (No. arXiv:2306.05499). arXiv. https://doi.org/10.48550/arXiv.2306.05499

Liu, Y., Deng, G., Xu, Z., Li, Y., Zheng, Y., Zhang, Y., Zhao, L., Zhang, T., & Wang, K. (2024b). A hitchhiker's guide to jailbreaking ChatGPT via prompt engineering. In Proceedings of the 4th International Workshop on Software Engineering and AI for Data Quality in Cyber-Physical Systems/Internet of Things (SEA4DQ 2024) (pp. 12–21). Association for Computing Machinery. https://doi.org/10.1145/3663530.3665021

Maier, M. W. (1998). Architecting principles for systems-of-systems. Systems Engineering, 1(4), 267–284. https://doi.org/10.1002/(SICI)1520-6858(1998)1:4<267::AID-SYS3>3.0.CO;2-D

Miller, K., O'Halloran, B., Pollman, A., & Feeley, M. (2019). Securing the Internet of Battlefield Things while maintaining value to the warfighter. Information Warfare Journal, 18(2), 74–84.

Miller, K., Bordetsky, A., Mun, J., Maule, R., & Pollman, A. (2021). Merging future knowledgebase system of systems with artificial intelligence/machine learning engines to maximize reliability and availability for decision support. Military Operations Research, 26(4), 77–93. https://doi.org/10.5711/1082598326477

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

National Initiative for Cybersecurity Education (NICE). (2020). NICE Cybersecurity Workforce Framework (NICE Framework) (NIST Special Publication 800-181 Revision 1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-181r1

National Institute of Standards and Technology (NIST). (2023). AI risk management framework (AI RMF 1.0) (NIST Special Publication 1270). https://doi.org/10.6028/NIST.AI.100-1

Rodgers, W., Murray, J. M., Stefanidis, A., Degbey, W. Y., & Tarba, S. Y. (2023). An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. Human Resource Management Review, 33(1), 100925. https://doi.org/10.1016/j.hrmr.2022.100925

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge (pp. 235–244). Association for Computing Machinery. https://doi.org/10.1145/3303772.3303796

Scharre, P., & Horowitz, M. C. (2020). Artificial intelligence: What every policymaker needs to know. Center for a New American Security. https://www.cnas.org/publications/reports/artificial-intelligence-what-every-policymaker-needs-to-know

Sneaky PDF. (2025). Add invisible text to PDFs | Hidden text tool. https://sneaky-pdf.com/

Technology, O. of I. (2025). DukeGPT | Office of Information Technology. Retrieved September 15, 2025, from https://oit.duke.edu/service/dukegpt/

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. Learning, Media and Technology, 45(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

Wong, W. (2025, May). Effective AI Requires Effective Data Governance. Technology Solutions That Drive Education. https://edtechmagazine.com/higher/article/2025/05/effective-ai-requires-effective-data-governance

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. https://doi.org/10.1186/s41239-019-0171-0