Risks of AI Applications Used in Higher Education

Donna Schaeffer[1], Lori Coombs[1], Jonathan Luckett[1], Marvin Marin[1] and Patrick Olson[2]
[1]Marymount University, School of Technology and Innovation, Arlington, USA
[2]National University, San Diego, USA

Donna.schaeffer@marymount.edu (corresponding author)
ldc72459@marymount.edu
jpl65485@marymount.edu
m0m61652@marymount.edu
polson@nu.edu

https://doi.org/10.34190/ejel.22.6.3457

Abstract: As artificial intelligence (AI) tools become more widely used in higher education, we must pay attention to the risks that can emerge. AI projects, whether applied in classroom learning or used for decision-making regarding admissions, financial aid allocation, or hiring, must include attention to governance and compliance issues, regardless of the project's scope and scale. Concerns highlighted in this work include transparency, user privacy, data confidentiality, data integrity, and system availability, however, we note that this is a non-exhaustive list of risks. In this paper, risk assessment is defined, and two examples of risk management frameworks, namely the United States National Institute of Standards and Technology Artificial Intelligence Risk Management Framework and the non-profit humanitarian effort ForHumanity's Independent Audit of AI, Algorithmic, and Autonomous Systems are briefly described. We identify characteristics of AI applications that need to be assessed for vulnerabilities they may present, such as bias and discrimination. This paper aims to facilitate discussion among stakeholders about the risks that may be encountered from using AI in higher education, as well as to suggest ways developers, decision-makers, and users can mitigate these risks. Much discussion and published literature has focused on risk management frameworks designed for large organizations or enterprises or frameworks that do not consider risks specific to AI. We hope that decision-makers carefully consider the risks, perform due diligence when implementing AI applications, and create a plan for mitigating the risks. This research supports e-learning practice because students and faculty are embracing AI applications. Leaders and decision-makers in higher education need to be proactive in protecting their varied stakeholders. The paper asks what risks may be encountered by institutions of higher education when using AI tools and products in the classroom and for various aspects of decision-making and if published frameworks can mitigate these risks.

Keywords: Artificial intelligence (AI), Risk management framework (RMF), Higher education, Cybersecurity

1. Introduction

This paper describes the risks that higher education institutions face as the use of artificial intelligence (AI) is becoming prevalent. This paper presents definitions, describes risk management frameworks, and discusses implications that decision-makers must consider.

The current discourse perceives AI as a two-edged sword when considering its impact on academia. The first edge of the sword questions ethical use. For example, AI Large Language Model (LLM) systems, such as ChatGPT, empower students with the ability to create content on behalf of prompts, which is known as prompt engineering, and those ideas can be used in whole or in part by students to submit as their work. At the very least, this creates a question of authorship. Does the student get the credit for generating the idea, or does the credit go to the AI for putting the words in the order that may be edited by students? Additionally, what is the demarcation line? If a writer uses a software tool such as Microsoft Word, and that tool has a rudimentary predictive element to it, does the credit belong to the human or the AI component – the predictive tool?

The second edge of the sword is AI's potential as a game-changer for students with disabilities that impact their academic performance. For example, students who suffer from dyscalculia may have an issue with understanding or conceptualizing mathematical concepts. However, a university's accommodations department may work with faculty on a solution where students can access an AI application that can provide additional learning resources. It may have the ability to explain concepts in several ways to reinforce the material, and it can facilitate practicing and reinforcing the lessons learned on the students' schedules when they are most comfortable with learning the material and in a manner that does not fatigue a human educator. In this use, students are not being given an advantage; instead, they are being allowed to be as successful as neurotypical students. This level of equity can be achieved by making AI available to students within this population. What is

Cite this article: Schaeffer, D.et al. 2024. "Risks of AI Applications Used in Higher Education", *Electronic Journal of e-Learning*, 22(6), pp 60-65, https://doi.org/10.34190/ejel.22.6.3457

important to the university and faculty is that students can understand the material presented and, with accommodation(s), convey competence in the material taught following the learning objectives in the course syllabus.

This discussion reveals a unique relationship the education sector has with AI. On the one hand, AI as a field within computer science should be explored for its ability to solve complex problems quickly and redefine how we educate students. On the other hand, it can be a major disruptor in how students' work is created and assessed, how educators approach education and the faculty-student relationship, and how decision-makers allocate resources. The conversation in higher education must shift to AI as a means of learning and teaching the next generation. These are the tools that the current generation of students will invariably use in the future; academics cannot refuse to teach about these valuable tools. Ignoring AI or making AI applications unavailable to students will not put them in a position to excel in the careers of the future and to create new knowledge. However, we must instill within students an ethical understanding and application for the use of AI. In many ways, this is no different from the lessons learned in business schools and the introduction of ethics classes in the curriculum of Master of Business Administration (MBA) degree programs after the collapse of Enron and other economic failures. Embracing AI is practical and prudent. Academia should embrace that students may use AI unethically and develop guidelines, policies, regulations, and educational material to help students understand what is required of them and how they can use the tools available to them. Given that many universities already educate their student population on topics such as academic integrity and academic misconduct, educating them about the ethical use of AI would be a *minimal* expansion of existing efforts.

While the most visible applications of AI may be in curriculum and pedagogy, those applications that touch students directly, much use of AI will happen "behind the scenes" in recruiting and admissions scenarios, allocating resources, and planning. This use will be by university and college administrators and staff and may even be outsourced to third-party vendors. Administrators must have a risk management strategy for outcomes these applications may produce. Risk management strategies typically are built around a risk management framework. After defining key terms and risk scenarios, this paper will describe two risk management frameworks.

2. Definitions

This paper standardizes on definitions provided by the National Institute of Standards and Technology (NIST), an organization within the United States Department of Commerce. Dempsey et al. (2011) define risk as “A measure of the extent to which an entity is threatened by a potential circumstance or event, and typically a function of: (i) the adverse impacts that would arise if the circumstance or event occurs; and (ii) the likelihood of occurrence (Page B-10).” This is often stated as the equation Risk = threats x vulnerabilities.

As NIST becomes involved in the standardization of AI across the US federal government, it has chosen to follow the American Standard Dictionary of Information Technology definition of AI, that is (1) A branch of computer science devoted to developing data processing systems that perform functions commonly associated with human intelligence, such as reasoning, learning, and self-improvement and (2) The capability of a device to perform functions that are normally associated with human intelligence such as reasoning, learning, and self-improvement (National Institute of Standards and Technology, 2019, page 25.). AI is a complex field that encompasses diverse technologies. For example, rule-based AI applications leverage methods that program computers to make decisions based on a set of rules. In contrast, expert systems are programmed to emulate the decision-making abilities of human experts. Machine Language (ML) is a subset of AI that leverages algorithms, data sets, and models to perform specific tasks. Deep learning solves problems in the same manner as the human brain does by using algorithms; however, it requires more extensive data sets than ML. Generative AI is an emerging development that has raised concern in governments and the general public for intellectual property infringement because the AI application draws on patterns that often utilize data lakes (centralized repositories that allow structured and unstructured data *to* be stored) and unlicensed work (Appel et al., 2023).

Several frameworks define risk management in AI. A Risk Management Framework (RMF) is a structured approach used to oversee and manage risk for an enterprise (Nieles, et al., 2017). This paper discusses two RMFs – The National Institute of Science and Technology (NIST) AI Risk Management Framework (AIRMF), which incorporates international standards such as ISO/IEC 23984:2023, and the non-profit organization ForHumanity, which has developed a certification program for auditors around the ethical assessment of AI applications.

This paper's scope of risks includes transparency, privacy, confidentiality, data integrity, system availability, and bias. The reasons these risks are important to consider are explained within the definitions provided in this section of the paper. While the NIST definition of transparency is narrow – its publications define transparency as the amount of information that can be gathered about a supplier, product, or service and how far through the supply chain this information can be obtained (Boyens et al., 2021), it also offers transparency as a synonym for visibility. In AI applications, transparency can be broadened to denote how much can be understood about the AI application, including how it was developed and trained and how it operates when deployed. Users must have access to the data sets that were used to train the model. The characteristic of transparency can illuminate issues of bias. If users understand why the AI application results in the recommendations it provides, they can decide whether to accept or reject the output. To mitigate problems associated with security, trust, and objectivity, models should integrate transparency when designing algorithms.

Any discussion of transparency yields decision points. Systems engineers, developers, and systems administrators who develop and deploy AI in educational institutions should consider engaging with lawyers early in the design process. AI and ML algorithms need to be examined for vulnerabilities and liabilities from cybersecurity and human user experience perspectives. Higher education institutions need enhanced transparency for AI models and machine learning algorithms. However, developers may not want to expose their code, patterns, and data sets to users because that level of openness could make applications vulnerable to attacks. While generating more information about the AI application might create tangible benefits for users, it may also create new risks for developers and users (Burt, 2019).

Privacy and confidentiality are related concepts. Indeed, NIST uses the terms in one another's definitions. For example, Powell et al. (2022) define privacy as the assurance that the confidentiality of and access to certain information about an entity is protected. Confidentiality is defined as "Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information (Pub., FIPS, 2006, page 6)." NIST also places privacy in the context of rights – Oldehoeft (1992) states privacy is the right of a party to maintain control over and confidentiality of information about itself. When data is misused, it becomes a source of liability. The data sets that are used to train AI systems must be very large and may include sensitive data. Data sets used for training must not contain personally identifiable information or information that can be aggregated to identify individuals. For example, a recent breach of an AI dataset occurred when developers at Microsoft caused the exposure of 38 terabytes of data, including disk backups of two employees' workstations, confidential corporate information, private keys, passwords, and over 30,000 internal Microsoft Teams messages during a routine update to GitHub (Naraine, 2023). While this example does not involve higher education, it shows the risks of using AI applications**.** Developers must build privacy protection into their systems and applications early in the design phase to protect users. Users should be informed when and how their data is collected and used. When data is collected, users can be offered informed consent agreements, opt-out ability, and the ability to delete their data. AI users must have a method for managing their privacy risks.

Data confidentiality involves protecting information in a system or application so that unauthorized access is prevented. Stakeholders within higher education include administrators, faculty and staff, students, and other related entities. These stakeholders do not want their sensitive and personally identifiable information (PII) exposed when institutions use third-party AI applications. Technical staff must understand their shared responsibility involved with combating commonly encountered threats to information confidentiality, including hackers, unprotected downloaded files, unauthorized user activity, local area networks (LANs), and trojan horses. Confidentiality can be compromised in data, network, end-to-end, application, and disk file scenarios. Third-party AI applications should be listed on an approved list for the institution so that service level agreements (SLAs) are put in place that clearly define the university's and the third-party application developer's shared model of responsibility. A constructive shared responsibility model considers access control, encryption, data masking, secure file transfer protocols (SFTPs), data loss prevention (DLP), and virtual private networks (VPNs) to protect the confidentiality of users (Anonymous, 2023).

Data integrity is a property where data or information has not been altered or destroyed in an unauthorized manner (Scholl et al., 2008). Data integrity is rooted in trusted data. Without trusted data from connected sensors, devices, systems, and applications, they all become vulnerable to improper cyber manipulation, making AI decision-making questionable. Data integrity is one pillar of the widely used concept of the "CIA Triad," which highlights confidentiality, integrity, and availability. Any data integrity breach means that AI and related devices won't be able to operate properly, exposing systems and applications to exploitation and cyber-attacks (Armilis, 2023).

Scholl et al. (2008) define system availability as the assurance that users have timely and reliable access to and use of information. Systems availability risk occurs when decisions are made based on easily and immediately available data without considering further research or additional external perspectives. The danger with availability risks when using AI applications is that generative AI is only sometimes accurate or reliable (Pavlou, 2023). When AI yields inaccurate or incomplete results, broader risk implications are present in higher education.

The final risk in this paper's scope is bias. Barker and Kelsey (2007) state that bias exists if one value from a sample space is more likely to be chosen than another value. Bias is discussed in an upcoming section.

These baseline definitions allow us to recognize risks inherent to and introduced through AI application development and deployment. Thus, multiple stakeholders, including developers, administrators in higher education institutions, faculty, students, and other persons who interact with the application, are involved. The definitions also facilitate the identification of risks that may arise in the development and use of AI applications.

3. Underlying Factors That can Cause Risks

Each of the risks defined in the previous section of the paper can be traced back to one of several occurrences that may be inherent in AI and machine learning (ML). These occurrences are data persistence, data repurposing, and data spillover. Pearce (2021) states that data persistence occurs when data exists longer than the developers intend. Data repurposing is when data is used beyond its originally designated purpose (Pearce, 2021). Data spillover occurs when data is collected on entities that are not the intended target of data collection (Pearce, 2021). It is easy to see how any of these occurrences can be present in AI applications because the training sets are often supplemented with additional data as time progresses to fine-tune the system.

Mitigating risk is essential when developing and deploying AI. The Definitions section of this paper includes inherent dangers that become exposed when managing different types of risks. This section briefly describes two Risk Management Frameworks: ForHumanity is designed for developers, and the NIST AI RMF is appropriate for large-scale government enterprises. At the time of this writing, no frameworks consider specific risks for institutions of higher education. The section concludes with a short discussion about how each of these frameworks handles one risk—that of bias.

ForHumanity is a non-profit organization founded in 2016 to develop an independent certification for those who audit AI systems. It is a volunteer-based organization with a small board and close to 1400 volunteers from 89 countries. ForHumanity offers several certifications, such as the ForHumanity Certified Auditor (FHCA), which includes a Code of Ethics. This group has codified a body of knowledge on various areas of compliance that auditors need to consider. Some areas include accuracy, validity, reliability, resilience and robustness, anti-discrimination, and data security. The non-profit also offers introductory courses in risk management for AI and autonomous systems and houses over 50 fellows involved in international projects, such as adapting audit standards for the European Union and establishing standards for emerging technologies like biometrics.

Training and assessment of auditors include a focus on data inputs and outcomes. Data integrity is apparent in the ForHumanity framework, with attention to the inputs, outcomes, and pipeline. A data pipeline contains the flow of data from its ingestion to a data set used by the AI through its transformation and storage. Auditors use checks and balances to ensure the data meets established metrics and measurements. The hope is that compliance with ForHumanity's' audit will become a "a seal of approval" that will influence buyer behavior over time. To quote founder Ryan Carrier's (2019) philosophy: If we can make good, safe, and responsible AI profitable, whilst making dangerous and irresponsible AIs costly, then we achieve the best possible result for humanity.

This philosophy is rooted in deterrence theory and is modeled on programs like the U.S. Department of Energy's (DOE) Energy Star and Intel Corporation's branding "Intel Inside." Since 1992, the voluntary Energy Star program has become an international standard, with more than 40% of Fortune 500 companies purchasing equipment and 45% of American households knowingly purchasing an ENERGY STAR certified product in 2022 (Environmental Protection Agency, 2023). The "Intel Inside" branding campaign started in 1991, intending to assure non-technical electronics buyers that their choice had quality components. It is easy to see similarities that a ForHumanity certification can make AI tools and applications more acceptable to users.

Auditors certified by ForHumanity learn about the relevant legal frameworks that regulate and legislate bias. For example, AI applications that are used in admissions decisions in US colleges and universities can no longer consider race as a data point. In June 2023, the US Supreme Court ruled that Harvard University and the

University of North Carolina violated the 14th Amendment of the US Constitution and Title VI of the Civil Rights Act of 1964. ForHumanity's training and assessments are also designed with European Union regulations.

The NIST AI Framework was introduced in January 2023. It is prescriptive around four pillars. The first pillar is governance, where the organization sets the overall direction and policy for its use of AI. The framework calls for organizations to identify roles, responsibilities, and resources for mitigating risks. The Mapping pillar is where the potential risks are identified, and their likelihood and impacts are quantified. The framework also includes Measure, a pillar in which data about AI risks are monitored and tracked over time. The fourth pillar is Act, which prescribes the management of risks via controls, monitoring, and adjustments. The NIST AI RMF identifies three categories of bias in AI: systemic, statistical, and human. It identifies three broad challenges: datasets, testing and evaluation, and human factors. It also introduces preliminary guidance for addressing bias in AI applications.

In the popular Harry Potter series of books, students at Hogwarts School of Witchcraft and Wizardry are sorted into their "best-fit" houses by a sorting hat that can sing, talk, and look into the students' minds (Rowling, 1997). This can be likened to an AI application assessing students and assigning them to classes that best align with their capabilities. Having the ability to screen and sort candidates to a pathway that best aligns with their needs may positively impact student retention and success rates while maintaining academic standards at a lower administrative overhead cost than traditional methods, which rely on student transcripts, essays, assessment testing, and the evaluation of those artifacts by staff members. Based on results provided by the AI application, institutions may be able to tailor programs to advanced students or to filter students to classes where they may have additional opportunities or alternative pathways to success (Pallathadka et al., 2023).

In the example of using AI for candidate acceptance or placement within programs, depending on how the AI model is trained, human bias may either be removed from the screening process, and the resulting pool could be a more diversified mixture of candidates or learning bias could negatively influence the results. An AI algorithm will yield results that are systemically prejudiced due to inaccurate assumptions in the ML process. This bias can be injected into algorithms unconsciously or consciously by the systems developers and administrators that develop the ML systems or due to flawed data sets used to sequence the systems (Verma, 2023).

Data persistence can cause AI applications to be biased in their recommendations. Data sets that were created before the afore-mentioned United States Supreme Court decision on college admissions may involve rules that include race as an attribute. This brings into question the longevity of an application's predictive power. Data sets will have to be regularly culled and updated.

AI application developers may repurpose data sets and use the same data in training sets for multiple applications to capture economies of scale. This may cause the system to make Incorrect correlations, resulting in bad decisions or recommendations. This same risk is present in the case of data spillover among training sets used in different applications. Taken out of context, the data that is spilled over may result in bad decisions or recommendations.

4. Conclusion

The biggest challenge for the field of AI may be finding the balance between protecting privacy and restraining advancement. Higher education institutions will benefit from allocating costs towards incorporating measures that consider how to properly engage allowable third-party AI applications securely for student, staff, and administrator use. Faculty need to assess the risks from academia/industry partnerships that are prevalent in the AI field. These partnerships may take the form of subscribing to datasets from third-party providers, which goes back to the prior discussion on transparency and data repurposing. While this paper addresses inherent risks, there are also opportunities, such as cooperation on or funding the research and development of AI applications for higher education. As noted, this discussion shows that higher education has a unique relationship with AI. This is due to the dual uses of AI in higher education, academic and administrative.

Developing a Risk Management Framework per published standards may not seem feasible given the staff and budget constraints that many higher education institutions face. However, it is possible to incorporate aspects of an RMF into institutions' technology policies and decision-making. As of the mid-year 2024, 200 information technology specialists held at least one ForHumanity certification. University and college administrators could engage with a consultant to audit the implementation of AI applications.

Another example is the Govern pillar of the NIST AI RMF. This pillar calls for identifying ownership, writing, and enforcing policies for acceptable use. Schaeffer, Dehghanpour, and Olson (2024) analyze university and college policies about using generative AI in the classroom. They found that many policies give ownership to individual faculty members, who may sanction or prohibit the use. Misuse by students was viewed as a violation of academic integrity in most policies. Managing the risks can build on already defined measures, such as the consequences of academic dishonesty.

The risks of utilizing AI applications in higher education settings are profound, especially risks of bias. Therefore, we recommend that administrators decide on the AI applications and tools based on desired outcomes. Typical applications such as AI as a supplement to classroom activities or in decision-making for admissions, financial aid allocation, or placement, or those AI applications used for recruiting and hiring decisions, require considering the risks and benefits. The benefits that AI applications offer higher education will only be accrued when institutions mandate governance and compliance for the AI applications their stakeholders choose to use.

References

Anonymous 007. (2023). Information Security Confidentiality Geeks for Geeks. https://www.geeksforgeeks.org/information-security-confidentiality/

Appel, G., Neelbauer, J. and Schweidel, D.A., 2023. Generative AI has an intellectual property problem. *Harvard Business Review*, 7.

Armilis. (2023). Data integrity in industry 4.0. https://armilis.com/blog/data-integrity-in-industry-4-0/

Barker, E.B. and Kelsey, J.M., 2007. Recommendation for random number generation using deterministic random bit generators (revised) (pp. 800-900). Washington, DC, USA: US Department of Commerce, Technology Administration, National Institute of Standards and Technology, Computer Security Division, Information Technology Laboratory.

Boyens, J., Smith, A., Bartol, N., Winkler, K., Holbrook, A. and Fallon, M., 2021. Cyber Supply Chain Risk Management Practices for Systems and Organizations (No. NIST Special Publication (SP) 800-161 Rev. 1 (Draft)). National Institute of Standards and Technology.

Burt, A., 2019. The AI transparency paradox. *Harvard Business Review*, 13.

Carrier, R., 2019. Implementing guidelines for governance, oversight of AI, and automation. *Communications of the ACM*, 62(5), pp.12-13.

Dempsey, K.L., Johnson, L.A., Scholl, M.A., Stine, K.M., Jones, A.C., Orebaugh, A., Chawla, N.S. and Johnston, R., 2011. Information security continuous monitoring (ISCM) for federal information systems and organizations.

Environmental Protection Agency Office of Air and Radiation, Climate Protection Partnerships Division. (2023). *National Awareness of ENERGY STAR® for 2022: Analysis of 2022 CEE Household Survey*

Naraine, R. 2023. Microsoft AI Researchers Expose 38TB of Data, Including Keys, Passwords and Internal Messages, *Security Week.* https://www.securityweek.com/microsoft-ai-researchers-expose-38tb-of-data-including-keys-passwords-and-internal-messages/

National Science and Technology Council (US). Select Committee on Artificial Intelligence, 2019. The national artificial intelligence research and development strategic plan: 2019 update (p. 0050). National Science and Technology Council (US), Select Committee on Artificial Intelligence.

Nieles, M., Dempsey, K. and Pillitteri, V.Y., 2017. An introduction to information security. NIST special publication, 800(12).

Oldehoeft, A.E., 1992. Foundations of a security policy for use of the National Research and Educational Network. US Department of Commerce, National Institute of Standards and Technology.

Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J. and Phasinam, K., 2023. Classification and prediction of student performance data using various machine learning algorithms. *Materials today: proceedings*, 80, pp.3782-3785.

Pavlou, C. (2023). Availability bias in the AI era: Understanding its impact on decision-making. https://www.talentlms.com/blog/availability-bias-ai-era/

Pearce, G., 2021. Beware the privacy violations in artificial intelligence applications. ISACA Foundation.

Powell, M., Brule, J., Pease, M., Stouffer, K., Tang, C., Zimmerman, T., Deane, C., Hoyt, J., Raguso, M., Sherule, A. and Zheng, K., 2022. Protecting Information and System Integrity in Industrial Control System Environments.

Pub, F.I.P.S., 2006. Minimum security requirements for federal information and information systems.

Scholl, M.A., Stine, K., Hash, J., Bowen, P., Johnson, L.A., Smith, C.D. and Steinberg, D., 2008. An introductory resource guide for implementing the health insurance portability and accountability act (HIPAA) security rule

Rowling, J. K. 1997*. Harry Potter and the Philosopher's Stone.* Bloomsbury Pub Limited.

Schaeffer, D. M., M. Dehghanpour, and P.C. Olson. 2024. Risk management framework for artificial intelligence in the classroom. *Proceedings of the International Association of Computer Information Systems.*

Verma, R. (2023). Mitigating AI Bias: The Key To A Better Future For Businesses And Society. *Forbes.* https://www.forbes.com/sites/forbesbusinesscouncil/2023/09/13/mitigating-ai-bias-the-key-to-a-better-future-for-businesses-and-society/?sh=2e904a895a93