University Governance for Responsible AI

Sang Hoo Oh[a] and Madelyn Sanfilippo[b]

[a]Catholic University of America, USA

[b]University of Illinois at Urbana-Champaign, USA

ohs@cua.edu, madelyns@illinois.edu

ABSTRACT

With the democratization of access to artificial intelligence via large language models, such as ChatGPT, questions about AI ethics in education are at the forefront of educational policymaking. How do we teach students to engage with AI ethically in coursework, research, and careers? What uses of AI constitute violations of academic integrity? What norms need to be established about AI in higher education? This research evaluates official strategy and policy documents from a wide array of research and liberal arts schools in the U.S. via a combination of thematic and structured content analysis, to surface policy and ethics concerns and the institutional structures that shape restrictions, rights, and interventions, following assessment of interrupter reliability. Results highlight the concerns of universities and emerging approaches to intervene with respect to the appropriate use of AI in education. Policy recommendations are made based on the emerging strategies and norms identified via analysis.

ALISE RESEARCH TAXONOMY TOPICS

Information policy; Information ethics; Artificial intelligence; Students; Education.

AUTHOR KEYWORDS

AI ethics; Higher education; Academic integrity; University policy; Responsible use of AI.

DOI: https://doi.org/10.21900/j.alise.2024.1706

INTRODUCTION

Over the past decade, artificial intelligence (AI) has transformed nearly every aspect of our society, including higher education. The rapid deployment of AI into higher education has changed the educational experience for students, faculty, and administrators (Chan, 2023). AI technologies have been applied in various educational settings to increase efficiency in administrative processes, improve personalized teaching and learning, and support research activities. Using AI in these different educational settings indicates its ability to potentially transform education in significant ways. Despite its powerful and profound impact on higher education, there has been a growing concern in the academia about the use of AI, with the emergence of large language models (LLMs), such as ChatGPT, Bing, and Geminia. There are also questions about AI ethics in education as access to AI via large language models has increased significantly in the last few years (Sam & Olbrich, 2023). With the increase in access to AI, universities now have to consider how to teach students to engage with AI ethically in coursework, research, and careers. The universities also need to decide what uses of AI constitute violations of academic integrity and what norms need to be established about AI. These questions sparked debate about the principles and values that should guide the development and use of AI, which led many higher education institutions to draft AI policy and guidelines for responsible use of AI. Hence, in this paper, we examined emerging AI policies and guidelines from the U.S. colleges and universities via a combination of thematic and structured content analysis to identify policy and ethical concerns, as well as the institutional structures that shape restrictions, rights, and interventions, following assessment of interrupter reliability.

BACKGROUND

With the increase in access to AI, governments worldwide have been developing national policies and strategies to offer clearer guidelines on AI usage, aiming to maximize its benefits while minimizing associated risks. These governments have been trying to address key issues such as AI discrimination and bias, privacy breaches, human rights violations, and the malicious use of AI (Greiman, 2021; Hogenhout, 2021; Schiff, 2021). Most national policies on AI prioritized the discussion of ethics to promote responsible and appropriate management of AI technologies, focusing on determining what is morally acceptable or unacceptable (Hogenhout, 2021). Many of these national AI policies have referred to Floridi (2021)'s framework for the ethical use of AI, which emphasized five core principles: beneficence, non-maleficence, autonomy, justice, and explicability. Additionally, Dexe and Franke (2020) reviewed AI strategy documents from Nordic countries and identified various ethical principles as the implicit foundation for policy development. Their study also provides insights into the unique approaches and priorities of Nordic AI policies, which explores their efforts to balance innovation with ethical considerations in the AI landscape.

Scholars have also started examining the content of AI ethics documents in public, private, and non-governmental sectors. For instance, Jobin et al. (2019) identified several themes in a set of 84 documents: transparency, justice, fairness, nonmaleficence, responsibility, and privacy. Similarly, Schiff et al. (2021) assessed more than 100 documents from public, private, and non-governmental sectors through the analysis of 25 ethical topics that they developed through coding. The study highlighted the different ethical considerations that arise in the

development, adoption, and governance of AI technologies across different sectors. In addition, Floridi and Cowls (2019) argued that the 47 AI ethics principles they reviewed align with traditional bioethics principles like beneficence, and justice, along with a new principle of explicability. Overall, previous studies on AI ethics have predominantly focused on assessing the extent to which a global consensus on AI ethics is emerging.

Recently, AI has also emerged as a transformative force in higher education, prompting universities around the world to establish policies, ethics frameworks, and guidelines to govern its development and deployment. In recent years, scholarly discourse has extensively addressed key issues and trends in the evolving landscape of AI policy and ethics in higher education. To address the ethical questions posed by AI in higher education, previous studies recommended the incorporation of AI ethics into curriculum. Borenstein and Howard (2021) emphasized the growing significance of AI ethics education in response to emerging challenges in AI technologies. This study highlighted the significance of integrating AI ethics more comprehensively and systematically into the curriculum and provided a series of suggestions regarding AI ethics pedagogy. The authors also suggest the need to address ethical concerns such as bias, privacy, accountability, and transparency in AI systems. Another research conducted by Chan (2023) introduced an AI Ecological Education Policy Framework for university teaching and learning. This framework was developed by exploring the perceptions and implications of text generative AI technologies. By proposing the AI Ecological Education Policy Framework, this study promotes a detailed comprehension of the implications of AI deployment within academic environments and ensures that stakeholders in academic settings understand the implications of integrating AI, enabling them to fulfill their responsibilities and respond accordingly. Moreover, UNESCO (United Nations Educational, Scientific and Cultural Organization), a respected international organization in education, developed its guidelines and recommendations grounded in extensive research and global expertise, ensuring relevance and adaptability across various educational systems and cultural contexts (UNESCO 2021a; UNESCO, 2021b). These guidelines and recommendations have offered a comprehensive framework for integrating AI into education, addressing ethical, social, economic, and technological dimensions crucial for developing effective policies. They have also provided a structured foundation for addressing specific AI policy and ethics issues in university teaching and learning.

Other studies related to AI policy and ethics in higher education have discussed the relationship between academic integrity and AI Large Language Models (LLMSs). Perkins (2023) discussed the academic integrity considerations of using AI LLMs in formal assessment. This study explored the potential benefits of AI LLMs in education to support various aspects of student education, including writing instruction, co-creation with AI, aiding English as a Foreign Language (EFL) learners, and enhancing Automated Writing Evaluations (AWE). However, the paper also highlights concerns regarding the ability of AI LLMs to generate original, coherent text that can evade detection by current technological methods and trained academic staff, which poses significant academic integrity challenges. This research concluded that whether student use of AI tools constitutes plagiarism or academic misconduct depends on whether the student discloses their use and aligns with the academic integrity policies of universities, which must adapt to accommodate the evolving role of these tools in education. In addition, Kumar et al. (2023) highlighted the ongoing evolution of LLMs and algorithmic writing, which is expected to shape the understanding and conceptualization of academic integrity. This study specifically explored how LLMs like ChatGPT, affects academic integrity. Authors suggests that educators

will need to adapt their approach in this new environment in three stages: awareness, detection and support, and integration of technologically assisted writing into pedagogy, with recommendations provided for pedagogy and policy adaptation accordingly.

Furthermore, the use of AI technologies has raised different issues of concern in higher education (Chan & Tsi, 2023; Chan & Zhou, 2023). The primary concerns revolve around questions such as how AI might reshape the design of assessment and curriculum, ensuring that everyone has fair access to these new technologies, redefining the role of teachers, and addressing the lack of technological support in developing countries (Popenici & Kerr, 2017; Swiecki et al., 2022; UNESCO, 2021a). With these concerns, AI policies in education are directed towards addressing several key issues: maintaining the fundamental values of traditional teaching methods, including teacher-student and student-student relationships (Luan et al., 2020; UNESCO, 2021b); ensuring inclusivity and equity in the adoption of AI technologies (Tanveer et al., 2020; UNESCO, 2021a); supporting the professional development of educators to enhance their skills and adapt their roles (Ocaña-Fernández et al., 2019; Wang et al., 2021).

Despite recognizing numerous concerns within educational environment, policies regarding AI in education tend to be broad and implicit due to the lack of concrete evidence of AI technology implementation (UNESCO, 2021a; Schiff, 2022; Chan, 2023). Schiff's (2022) analysis of 24 AI policy strategies focusing on education's role in global AI policy discussions, policymakers predominantly perceived education as a means to facilitate workforce development and train AI specialists. The study highlighted a notable absence of discussions on AI's actual integration into education policy, emphasizing instead its instrumental role in preparing a workforce adept in AI. Schiff concluded that if this policy trend continues, policymakers might overlook the transformative capabilities of AI in education and inadequately allocate resources, oversight, and attention to address the ethical implications of AI integration in educational settings. The current literature also lacks adequate attention to AI scholarship and education governance, and there's limited public understanding of AI policy implications (Feldstein, 2019; Gellai, 2022). While efforts are needed to develop more comprehensive and targeted policy frameworks for AI in education, ethics emerges as a crucial starting point for further discourse, with researchers urged to engage policymakers by focusing on ethical considerations in AI education (Sam & Olbrich, 2023; Schiff, 2022).

METHODS

To investigate how the U.S. universities are responding to emerging challenges around AI in higher education, official documentation regarding college and university strategies and policies were collected from all universities in the 2024 US News ranking of top 100 research institutions and top 100 universities for undergraduate teaching. Note that 24 universities appeared on both lists; thus, a total of 176 universities were considered.

After assembling this data set, policies and strategy documents were thematically coded. This iterative process began with a subset of 10 random policies to surface concepts. Both investigators worked together to compare and consolidate codes, then cluster them categorically. When definitions were agreed upon, interrater reliability testing was conducted. This process took two rounds to reach good or excellent agreement for all codes, after refining the codebook; the final Krippendorf scores ranged from 0.765 to 1.0. The lowest agreement was present for sparse codes. Table 1 presents the thematic codes, organized categorically, employed to assess

ethical and policy concerns surfaced in this research. Note that the category of institutions, including codes differentiating between strategies, norms, and rules, were applied in a structural sense, drawing on the validated institutional grammar (Crawford & Ostrom, 1995) associated with the knowledge commons (e.g., Sanfilippo, Frischmann, & Strandburg, 2021) and institutional analysis and development (e.g., Ostrom, 1990) frameworks.

Table 1
Codebook

	Thematic Codes		Thematic Codes
Actors	Administrator Instructor Student Staff User	*Learning Objectives*	AI literacy Critical thinking Understanding LLMs
Attributes	Role Responsibility	*Decision-Making*	Advantage AI adoption challenges AI adoption opportunities Communication Compliance Limitation
Academic Integrity	Academic integrity Citation and attribution Copyright Plagiarism	*Concerns*	Data privacy Data security Harms Risks Threats
Values and Ethics	Accountability Equity Ethical considerations Fairness Reliability Responsible Use Transparency	*Types of AI*	AI detection tools Generative AI Predictive AI
Instruction	AI integration with teaching Assessment methods, strategies	*Institutions*	Norms Rules Strategies

Instructional design

CONCLUSION

This research presents a comprehensive analysis of the emerging policies and strategies for governing AI at the U.S. universities. To examine how universities are addressing emerging challenges related to AI in higher education, we evaluate official university strategy and policy documents from a wide array of research and liberal arts schools across the U.S. The combination of thematic and structured content analysis has been conducted to surface policy and ethics concerns and the institutional structures that shape restrictions, rights, and interventions, following assessment of interrupter reliability.

This study provides a foundation for understanding the diverse approaches to AI governance within US higher education. By identifying key similarities and differences between R1, undergraduate teaching, and universities that are classified as both R1 and undergraduate teaching institutions, this research offers insights into policy and ethics concerns in the U.S. universities. The study also contributes to academic literature by offering empirical evidence and analysis regarding the emerging strategies and enforcement mechanisms for governing AI at the U.S. universities. In addition, this research aims to inform policymakers, educators, and stakeholders about effective practices and potential areas for improvement in AI governance within higher education institutions. Future research should delve deeper into the specific content of AI policies, examining how these policies are implemented and enforced. Additionally, exploring the impact of AI policies on students, faculty, and staff would provide valuable insights into the effectiveness of these frameworks.

REFERENCES

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, *1*, 61-65.

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International journal of educational technology in higher education*, *20*(1), 38.

Chan, C. K. Y., & Tsi, L. H. Y. (2023). The AI Revolution in Education: Will AI Replace or Assist Teachers in Higher Education? [Preprint]. arxiv:2305.01185

Chan, C. K. Y., & Zhou, W. (2023). Deconstructing Student Perceptions of Generative AI (GenAI) through an Expectancy Value Theory (EVT)-based Instrument [Preprint]. arxiv:2305.01186.

Crawford, S. E., & Ostrom, E. (1995). A grammar of institutions. *American political science review*, *89*(3), 582-600.

Cronan, T. P., McHaney, R., Douglas, D. E., & Mullins, J. K. (2017). Changing the academic integrity climate on campus using a technology-based intervention. *Ethics & Behavior*, *27*(2), 89-105.

Dexe, J., & Franke, U. (2020). Nordic lights? National AI policies for doing well by doing good. *Journal of Cyber Policy, 5*(3), 332–349.

Feldstein, S. (2019). The road to digital unfreedom: How artifcial intelligence is reshaping repression. *Journal of Democracy, 30*(1), 40–52.

Floridi, L. (2021). A Unified Framework of Five Principles for AI in Society. *In Ethics, Governance, and Policies in Artifcial Intelligence* (Vol. 144, pp. 5–17). Springer International Publishing AG.

Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, 535-545.

Garrett, N., Beard, N., & Fiesler, C. (2020, February). More than" If Time Allows" the role of ethics in AI education. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 272-278).

Gellai, D. B. (2022). Enterprising academics: Heterarchical policy networks for artifcial intelligence in British higher education. *ECNU Review of Education*.

Greiman, V. A. (2021). Human rights and artifcial intelligence: A universal challenge. *Journal of Information Warfare, 20*(1), 50–62.

Hogenhout, L. (2021). A Framework for Ethical AI at the United Nations. https://doi.org/10.48550/arxiv.2104.12547 IMDA & PDPC (2020). *Model Artifcial Intelligence Governance Framework.* Retrieved from https://www.pdpc.gov.sg/-/media/ fles/pdpc/pdf-fles/resource-for-organisation/ai/sgmodelaigovframework2.pdf.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9), 389-399.

Kumar, R., Eaton, S. E., Mindzak, M., & Morrison, R. (2023). Academic integrity and artificial intelligence: An overview. *Handbook of academic integrity*, 1583-1596.

Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C.-C. (2020). Challenges and future directions of big data and artifcial intelligence in education. *Frontiers in Psychology, 11*, 580820.

Ocaña-Fernández, Y., Valenzuela- Fernández, L. A., & Garro-Aburto, L. L. (2019). Artifcial intelligence and its implications in higher education. *Journal of Educational Psychology, 7*(2), 553–568

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of university teaching & learning practice*, *20*(2), 07.

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artifcial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12, 22.

Raji, I. D., Scheuerman, M. K., & Amironesei, R. (2021, March). You can't sit with us: Exclusionary pedagogy in ai ethics education. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 515-525).

Sam, A. K., & Olbrich, P. (2023). The need for AI ethics in higher education. In C. C. Corrigan, S. A. Asakipaam, J. J. Kponyo, & C. Luetge (Eds.), *AI ethics in higher education: Insights from Africa and beyond* (pp. 3–10). Springer. https://doi.org/10.1007/978-3-031-23035-6_1.

Sanfilippo, M. R., Frischmann, B. M., & Strandburg, K. J. (Eds.). (2021). *Governing privacy in knowledge commons*. Cambridge University Press.

Schiff, D., Borenstein, J., Laas, K., & Biddle, J. (2021). AI Ethics in the Public, Private, and NGO Sectors: A Review of a Global Document Collection. *IEEE Transactions on Technology and Society*, *2*(1), 31-42.

Schiff, D. (2022). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education*, *32*(3), 527-563.

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artifcial intelligence. *Computers and Education: Artificial Intelligence, 3*, 100075.

Tanveer, M., Hassan, S., & Bhaumik, A. (2020). Academic policy regarding sustainability and artifcial intelligence (AI). *Sustainability, 12*(22), 9435.

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W. G. (2021). Framing governance for a contested emerging technology: insights from AI policy. *Policy and Society*, *40*(2), 158-177.

UNESCO. (2021a). AI and education: Guidance for policy-makers. UNESCO.

UNESCO. (2021b). Recommendations on the Ethics of Artifcial Intelligence. UNESCO.

Wang, S., Wang, G., Chen, X., Wang, W., & Ding, X. (2021). A review of content analysis on China artifcial intelligence (AI) education policies. *Artifcial intelligence in education and teaching assessment*, 1-8.