Ethical Concerns And Responsible Use Of Generative Artificial Intelligence In Engineering Education

Aditya JOHRI
George Mason University, Fairfax, United States of America, johri@gmu.edu

Euan LINDSAY
Aalborg University, Aalborg, Denmark, edl@plan.aau.dk

Junaid QADIR
Qatar University, Doha, Qatar, jqadir@qu.edu.qa

ETHICAL CONCERNS AND RESPONSIBLE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE IN ENGINEERING EDUCATION

A Johri [1]
George Mason University
Fairfax, USA
https://orcid.org/0000-0001-9018-7574

E Lindsay
Aalborg University
Aalborg, Denmark
https://orcid.org/0000-0003-3266-164X

J Qadir
Qatar University
Doha, Qatar
https://orcid.org/0000-0002-8456-8458

Conference Key Areas: *Education about and education with Artificial Intelligence, Innovative Teaching and Learning Methods, Virtual and Remote Learning*
Keywords: *Technology ethics, Generative AI, Responsible use*

ABSTRACT

The use of educational technologies that use elements of machine learning (ML) and artificial intelligence (AI) are becoming common across the engineering education terrain. With the wide adoption of generative AI based applications, this trend is only going to grow. Not only is the use of these technologies going to impact teaching, but engineering education research practices are as likely to be affected as well. From

---

[1] *A Johri*

johri@gmu.edu

data generation and analysis, to writing and presentation, all aspects of research will potentially be shaped. In this practice paper we discuss the ethical implications of the use of generative AI technologies on engineering teaching and engineering education research. We present a discussion of potential and futuristic concerns raised by the use of these technologies. We bring to the fore larger organizational and institutional issues and the need for a framework for responsible use of technology within engineering education. Finally, we engage with the current literature and popular writing on the topic to build an understanding of the issues with the potential to apply them in teaching and research practices.

1 INTRODUCTION

The contemporary educational sector, including higher education institutions (HEIs), exists in a highly technological state. In addition to traditional applications such as Learning Management Systems (LMS), universities use videoconferencing, automated assessments, and increasingly, Machine Learning (ML) or more generally Artificial Intelligence (AI)-driven applications. From TurnItIn to Grammarly, these new technologies have found broad application including in engineering education. In addition, the data generated by these applications has lead to features that employ Learning Analytics (LA) and Educational Data Mining (EDM) for sensemaking.

The major difference between the newer technologies now in use for education and research and those used earlier, is the generation of data and capabilities that have been developed to analyze analyse and use that data. In recent times, the field of AI has entered a new era marked by remarkable advancements in generative AI applications. One notable example is ChatGPT (Dwivedi et al., 2023), which builds upon the power of Large Language Models (LLMs) (Qadir, 2023). By harnessing the wealth of textual data accessible on the internet, these applications create models with the ability to predict highly probable completions for any given text. As a result, they exhibit language generation and conversational capabilities that closely emulate human-like interactions.

While these technologies have opened up numerous promising applications that align with educational and research objectives of engineering, it is essential for us as a community to address important questions arising from their rapid and uncritical adoption. This collective effort is crucial to ensure their responsible use and mitigate potential concerns. Especially as engineering educators preparing the next generation, we have a moral obligation to think deeply about these issues and reflect on our use of technology across our own practices, as we prepare our students to practice in a world where these technologies exist (Johri et al., 2023; Johri, 2020).

2 ETHICAL CONCERNS

Emerging technologies provide a range of new affordances that we can use in educating future engineers. In deciding how these AI technologies should be employed in our teaching, there are a range of key ethical concerns that we must consider. Scholars in multiple communities, such as LA and EDM, have commented extensively on these issues (Kitto and Knight, 2019; Slade and Prinsloo, 2013; Tzimas and Demetriadis, 2021). Navigating these ethical issues is not always straightforward as often we face dilemmas of conflicting demands of values that we hold. For instance, the use of complex LLMs can improve our ability to make accurate predictions but they also reduce our ability to understand how they work (Whittlestone et al., 2019); the use of generative AI can result in impressive applications but may also result in the loss of jobs and deskilling of humans. In what follows, we discuss some ethical concerns surrounding generative AI and the use of automation in education.

2.1 Data privacy and consent

A fundamental concern with the use of new forms of digital technology is how they handle data – what data is collected, how is it stored and retrieved, where is it stored, and what kinds of consent provisions are available to users. The majority of education contexts where AI is deployed are systems and situations that are intrinsically linked to core operations of the university - situations where students are either not asked to consent, or where refusing consent would be impractical. Opting out of automated assessment is impossible; seeking explicit consent for data mining of historical data is similarly impractical.

Increasingly, developers of educational technology products tend to gather more data than what is functionally needed for potential future use and extensions. This has meant that data have been collected en masse with little regards to their actual use, and data can be repurposed for purposes different from their original intended use. Consequently, consent mechanisms are overreaching in what they ask of users; similar to the “I Agree” most of us click on while accessing most digital platforms – a consent that is in conflict with the GDPRs data minimisation principle.

2.2 Algorithmic bias

A related concern to data is how the data are analyzed analysed and used. Increasingly, systems use ML techniques to make sense of the data. To develop these algorithms, they have to be trained on datasets. These datasets are largely developed through data that are readily, conveniently, available and not necessarily represenative of a specific issue. As an example, to develop facial recognition algorithms, a large number of faces have to be fed to the algorithm and then labelled as “face” so that the algorithm knows it is a face. If the data that is used to do this is largely white faces, as has historically been the case, then the algorithm performs poorly on other skin tones. Thus, algorithms inherently develop a bias and the more they are used the more they get “trained” to make a mistake unless it is intentionally corrected. The act of identifying previous implicit biases can itself be problematic. How should we respond to the discovery that an accurate model of our current practice identifies a clear practice of bias? How should we treat a colleague who our algorithm has identified to be biased, but whose bias has only been made visible through their voluntary consent to participate in our modelling study?

2.3 Transparency and opacity

How does one know that the algorithm is biased? Detecting algorithmic bias often occurs unintentionally when it demonstrates flawed behavior during actual implementation. Testing algorithms can be challenging since they are essentially "black boxes" kept proprietary by their developers. Additionally, with the rise of deep learning and LLMs, even the creators may lack full awareness of the inner workings and steps involved in generating outputs. The complexity of training neural network-based algorithms makes understanding their functionality nearly impossible. Balancing the need to protect privacy and intellectual property presents difficulties in disclosing algorithmic workings and the underlying training data. This issue is particularly pronounced when dealing with student data, which is regarded as confidential in Europe and often enjoys federal protection in the United States, making it challenging to access the data used to develop a technology. There are several avenues being pursued in this area, especially the subfield of Explainable AI (XAI), where scholars have developed methods to make the use of AI more transparent across the application lifecycle (Dengel, et al., 2021; Doran, et al., 2018).

2.4 Equity and access

ML based technologies for learning can potentially treat all users as equal, ensuring accessibility that is fair to all. They can also things more equitable by prodiving services to those who need it most. Furthermore, they can support scaling up for services at a faster pace than is possible through purely human resources. Given the differences in learning opportunities, prior knowledge, and different backgrounds of students, this is a high barrier for many technologies to meet. For instance, how do you ensure that everyone understands their rights and consents with full knowledge if their technological literacy is different? There are also students that need accommodation due to different reasons and neurodiverse individuals who should also have equal access. Although technology use seems universal, many nuances that need to be worked out to ensure equity in the use of technologies.

2.5 Individual versus community approach to education and learning

Increasing personalisation of services and information may bring economic and individual benefits, but risks creating or furthering divisions and undermining community solidarity. The attractiveness of AI systems is that they can effectively automate the most common tasks; but this risks introducing a “tyranny of the majority”, where the needs of minorities in the long tail are overlooked because they are difficult to automate. The most effective and accurate algorithms, in terms of their predictive power or accuracy, may be based on complex methods (such as deep learning). The inner workings of such algorithms might not be fully transparent to developers and may result in systemic discrimination against a minority class even if it is on average accurate. As argued by Engelbart (1962), complex problems (commonly referred to as "wicked problems" nowadays, such as addressing hunger, containing terrorism, or fostering rapid economic growth) cannot be solved through technology alone (no matter how advanced it may be). The full potential lies in human-computer symbiosis where technologies like generative AI and algorithms are utilied to augment the collaborative efforts of human communities.

2.6 Human-Centered Learning and Human-in-the-Loop Learning

Human learning is not merely a technological challenge. It is important to temper our expectations and recall our past underwhelming experiences with supposedly revolutionary technologies, as emphasised by Langdon Winner (2009). Education is about the humans involved and should remain human-centered. It must also be remembered that education involves attaining proficiency through practice and mastery through understanding, which cannot be automatized automatised or rushed and attained instantaneously. We will do well to benefit from previous systematic thinking on human augmentation so that human capacities and capabilities are effectively augmented to solve the problems humanity face (Johri, 2022).

The operational intention of many ML approaches is that they should be decision support systems, making recommendations to humans who actually make the decisions. Over time, however, there is the risk that this will drift – particularly if the models turn out to be very effective. If a model only rarely turns out false negatives (for instance, failing an assignment it should have passed), for how long will we commit the resources to check all of the negatives for the false one? This question is particularly relevant if false negatives result from implicit bias in our models.

2.7 Speed of innovation versus equality, safety, credibility and sustainability

The rise of AI, and generative AI specifically, has given rise to an influx of funding support for innovation from both the private sector and governments. Across the world, new companies and industries are being formed, leading to new products. Although this innovation is necessary for using AI beneficially, this arms race of sorts is also likely to lead to sustainability and climate change challenges, as well as issues of inequality if this continues to be a winner-takes-all battle. More resources are likely to be put into technologies that will benefit a few as opposed to those with lower profitability but broader impact.

Pursuing technological progress at breakneck speed may compromise the safety, robustness, and reliability of these developments. While adapting to changing times is desirable, universities have historically been slow to evolve. The credibility of universities rests on thorough quality assurance processes, which often struggle to keep up with the latest technological advancements. While an ill-advised response could be to outright ban such technologies, this approach merely introduces enforcement and compliance issues, while further distancing academic practices from the eventual professional practices that are inevitably on the horizon.

2.8 Efficiency vs. effectiveness

AI offers great promise in automating and streamlining the common and recurring aspects of the learning experience. However, in the pursuit of efficiency, we risk neglecting the less common but equally important elements. Although AI is proficient in addressing the majority of learning features, it may encounter difficulties when confronted with exceptional cases and outliers. While we can identify the primary feedback provided by humans to students and create automated systems to deliver it on a large scale, there is a danger of disregarding the valuable feedback that falls outside the common patterns. We must consider how human-machine augmentation can be best practiced so that it does not sacrifice human ingenuity in search for efficiency (Dengel, Devillers, and Schaal, 2021). Having invested in automating this feedback, we may be tempted to reuse the same model in subsequent years without updating it to account for contextual changes, evolving theories, and the distinct learning profiles of each year's student cohort. While there is immense potential for more efficient resource allocation, we must ask ourselves if we can still uphold our graduate standards if we solely focus on automatable outcomes. Moreover, it is essential to consider whether the resources freed up by automation will be redirected to address non-automatable elements in teaching or instead diverted to research, central administration, or budget cuts.

2.9 The dignity of academic work

Technological determinism often dictates that we machines to the extent possible; however, much of higher education has always been experienced, and valued, as an artisanal human process (Crawford, 2009). Increasing automation and quantification could make lives more convenient, but risks undermining those unquantifiable values and skills that constitute human dignity and individuality. This is especially applicable to teaching, which is a personal profession and in most cases a respected profession. The use of technology to automate educational practices risks making the work less dignified devoid of purpose.

3 FRAMEWORKS AND CHECKLISTS

The range of issues highlighted in the previous section motivates systemic approaches to the design and development of AI systems in engineering education. In this section, we will present two existing frameworks, DELICATE and RESPACT, which are relevant for operationalizing AI in engineering education. We then discuss unique challenges posed by generative AI which future frameworks should consider.

3.1 DELICATE

The DELICATE checklist (Drachsler and Greller 2016) was developed as an instrument for educational institutions to engage in ethics and privacy discussions around the use of educational technologies that use Learning Analytics (LA). The authors argue that there are ways to design privacy protections and consent mechanisms so that all stakeholders are benefitted.The checklist consists of the following elements to guide the use of LA applications:

1) ***D****etermination*: Why do you want to apply Learning Analytics;
2) ***E****xplain*: Be open about your intentions and objectives;
3) ***L****egitimate*: Why you are allowed to have the data;
4) ***I****nvolve*: Involve all stakeholders and the data subjects;
5) ***C****onsent*: Make a contract with the data subjects;

6) ***A**nonymise*: Make the individual nonretrievable;
7) ***T**echnical*: Procedures to guarantee privacy; and,
8) ***E**xternal*: If you work with external providers.

Overall, the checklist contains guidance on paying attention to the value of LA and the rights of participants, ensuring that there is transparency about the use of LA and that users give consent openly and willingly. There is also an emphasis on data anonymization and institutional guidance for adopting clear and transparent obligations with any external agencies involved (Drachsler and Greller 2016).

3.2 RESPACT

Another related framework developed on the basis of review of the literature and empirical work (Johri and Hingle, *forthcoming*; Johri and Hingle, 2023) is RESPACT, with applications specifically for educational technologies that use ML/AI and are implemented in an HEI context. The framework consists of the following elements:

1) ***R**esponsive*: The technology needs to be *responsive* to user needs and work *responsibly*. Often the implementation of technologies is done without consideration of whether it fills a need and is usable.
2) ***E**thical*. It is imperative that organizations use some set of *ethical guidelines* for technology procurement and implementation. Institutions may already have access to guidelines for protecting student data that can be expanded for this.
3) ***S**ecure*. The *security* of data is paramount in any technology implementation. With increasing attacks on systems and stolen data becoming common it is vital that insitutions work sincerely towards securing their infrastructure.
4) ***P**rivate*. Privacy is one of the most contentious aspects of technology use and implementation. For educational institutions, it is essential that they view privacy contextually and are guided not just by the law but also by their ethos.
5) ***A**ccountable*. Accountability is another consideration as the misuse of data or of technology has to be righted, and the insitution needs to ensure compliance as well as working within appropriate frameworks.
6) ***C**onsent-Driven*. Consent needs to go beyond simply informed consent to all the involved parties. Consent should extend toward differential schema, so that diverse users can agree to the terms based on their preference.
7) ***T**ransparent*. As technologies get more complex, it is hard for a user to understand all the integrated functions and services and the flow of information or data through the overall system. Transparency is essential for explainability and, consequently, for trust; a user will trust more what they understands better.

Overall, the RESPACT framework, which comes out of empirical work on video-based monitoring of exams (Johri and Hingle, 2023), provides a institutional level set of guidelines for technology use and implementation. While it recognizes that user protection is essential, it also emphasizes institutional imperative of data security and responsiveness to user needs.

Through the exploration of the DELICATE and RESPACT frameworks, we have identified key ethical considerations for the use of AI in engineering education. Both these frameworks stress the importance of transparency, accountability, and the involvement of all stakeholders. They also emphasize the need for consent-driven approaches, addressing privacy concerns, and ensuring the security of data.

3.3 Applicability of these frameworks for Generative AI

While the frameworks listed above are generally applicable to a range of edtech and AI scenarios, generative AI poses some unique challenges, which requires special attention, demanding extension of these frameworks in future work (Kasneci et al., 2023; Weidinger et al., 2021). Some of these new challenges are outlined next:

a) ***Avoiding AI-aided plagiarism***: How to cope with situations where ready-made answers become available to students, who can simply copy-paste them and not disclose such AI-generated plagiarism?
b) ***Avoiding deskilling***: How to cope with deskilling that emerges as students are not required to engage in learning and provided directed answers?
c) ***Enhancement of human capacity***: How can we ensure that the use of generative AI augments human capabilities and does not atrophy them?
d) ***Fair use of AI generated content (AIGC) and AI agents***: Is it ever fair to use AIGC without full disclosure? Is it ethical to leave decision making in the hands of algorithms? Who is liable in case of an inappropriate response?
e) ***Equity and accessibility***: How to ensure that everyone can equitably access the benefits of generative AI and tools do not exacerbate inequality?

Future frameworks should look at the dangers of automation in the field of engineering education and attempt to keep engineering education human-centered and avoid what Brynjolfsson (2022) calls the “Turing Trap” (the focus on automation and artificial intelligence rather than on human intelligence augmentation).

4 CONCLUSIONS

The widespread adoption of generative AI and automation within engineering education raises significant ethical concerns that cannot be ignored. We have highlighted some new unique challenges posed by generative AI such as AI-aided plagiarism, deskilling, the imperative for human augmentation and capacity enhancement, fair use of AI-generated content, and equity and accessibility. These concerns highlight the need for the development of comprehensive ethical AI frameworks that address the unique challenges posed by generative AI. As generative AI technologies become more integrated into education, it is crucial to assess both their short-term benefits and long-term consequences for teaching and learning. The scalability and complexity of these applications require a deliberate effort within the educational community to minimize harm and promote responsible use. Extending existing ethical guidelines to meet new challenges posed by generative AI and emphasizing transparency, accountability, privacy, and human-centeredness, we can mitigate the risks and maximize the positive impact of generative AI in engineering education.

5 ACKNOWLEDGMENTS

Johri's work was partly supported by U.S. NSF Awards#204863, 1937950, 1939105; USDA/NIFA Award#2021-67021-35329. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

Brynjolfsson Erik. "The Turing trap: The promise & peril of human-like artificial intelligence." *Daedalus*. 2022 May 1;151(2):272-87.

Crawford, Matthew B. *Shop class as soulcraft: An inquiry into the value of work*. Penguin, 2009.

Dengel, Andreas, Laurence Devillers, and Laura Maria Schaal. "Augmented human and human-machine co-evolution: efficiency and ethics." *Reflections on Artificial Intelligence for Humanity* (2021): 203-227.

Doran, Derek, Sarah Schulz, and Tarek R. Besold. "What Does Explainable AI Really Mean? A New Conceptualization of Perspectives." (2018). *CEUR Workshop Proceedings*.

Drachsler, Hendrik, and Wolfgang Greller. "Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics." In *Proceedings of the sixth international conference on learning analytics & knowledge*, pp. 89-98. 2016.

Dwivedi, Yogesh K., Nir Kshetri, Laurie Hughes, Emma Louise Slade, Anand Jeyaraj, Arpan Kumar Kar, Abdullah M. Baabdullah et al. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy." *International Journal of Information Management* 71 (2023): 102642.

Engelbart, Douglas, C., 1962. *Augmenting Human Intellect: A Conceptual Framework*. dougengelbart.org.

Johri, Aditya, and Ashish Hingle. "Technological Paradox as Occasion for Restructuring Educational Practices and Igniting Moral Imagination." D.A. Thomas & V. Laterza (eds.) *Critical Perspectives on EdTech in Higher Education - Varieties of Platformisation*. Palgrave Macmillan. (forthcoming)

Johri, Aditya, Andrew S. Katz, Junaid Qadir, and Ashish Hingle. "Generative artificial intelligence and engineering education." *Journal of Engineering Education* (2023).

Johri, Aditya, and Ashish Hingle. "Students' technological ambivalence toward online proctoring and the need for responsible use of educational technologies." *Journal of Engineering Education* 112, no. 1 (2023): 221-242.

Johri, Aditya. "Augmented sociomateriality: implications of artificial intelligence for the field of learning technology." *Research in Learning Technology* 30 (2022).

Johri, Aditya. "Artificial intelligence and engineering education." *Journal of Engineering Education* 3 (2020): 358-361.

Kasneci, Enkelejda, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser et al. "ChatGPT for good? On opportunities and challenges of large language models for education." *Learning and Individual Differences* 103 (2023): 102274.

Kitto, Kirsty, and Simon Knight. "Practical ethics for building learning analytics." *British Journal of Educational Technology* 50, no. 6 (2019): 2855-2870.

Qadir, Junaid. "Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education." In *2023 IEEE Global Engineering Education Conference* (EDUCON), pp. 1-9. IEEE, 2023.

Slade, Sharon, and Paul Prinsloo. "Learning analytics: Ethical issues and dilemmas." *American Behavioral Scientist* 57, no. 10 (2013): 1510-1529.

Tzimas, Dimitrios, and Stavros Demetriadis. "Ethical issues in learning analytics: a review of the field." *Educational Technology Research and Development* 69 (2021): 1101-1133.

Whittlestone, Jess, Rune Nyrup, Anna Alexandrova, Kanta Dihal and Steve Cave. *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*. London: Nuffield Foundation. (2019)

Winner, Langdon. "Information technology and educational amnesia." *Policy Futures in Education*. 2009 7(6):587-91.

Weidinger, Laura, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng et al. "Ethical and social risks of harm from language models." arXiv preprint arXiv:2112.04359 (2021).