Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025 ASEE
Paper ID #48859

Empowering Engineering Graduates to Contribute towards Designing Safer Generative AI Tools through an Ethics Course

Sourojit Ghosh, University of Washington

Sourojit Ghosh is a fifth year PhD Candidate at the University of Washington, Seattle in Human Centered Design and Engineering.

Dr. Sarah Marie Coppola, University of Washington

Sarah Coppola is an Assistant Teaching Professor the Department of Human Centered Design & Engineering at the University of Washington. Dr. Coppola is an educator and researcher whose work focuses on how technology and systems design affects people's performance and health. She holds a BS in Mechanical Engineering from Northwestern University, a MS in Human Factors Engineering from Tufts University, and a Doctorate in Ergonomics from Harvard University.

Empowering Engineering Graduates to Contribute towards Designing Safer Generative AI Tools through an Ethics Course

Sourojit Ghosh and Sarah Coppola,

University of Washington Seattle

Introduction

Over the past few years, the world has witnessed the steady proliferation of Generative Artificial Intelligence (GAI) tools in all sectors and industries, being matched by growing levels of public and private investment into creating a more GAI-powered world. However, there remain many unanswered questions about the ethical and moral impact of this emergent technology, both in terms of the harms caused by the outputs of GAI tools towards historically marginalized identities (e.g., [1]–[4]) as well as the ecological impacts of producing and running large GAI systems on a global scale (e.g., [5]–[7]). In such a climate, there arises a strong necessity for training engineering students and future industry professionals in the ethical usage of GAI tools, such that they may champion ethical and harm-informed GAI design and incorporation strategies to their employers.

Towards this end, we developed and taught a 10-week college course on considerations and ethical challenges in using GAI tools within the user-centered design process, which became one of the authors' university's first GAI ethics courses. This course was developed by the lead author, a doctoral student researching this topic at an R1 university, and taught across the department's BS and MS programs in consecutive quarters. The course afforded students the ability to practice the usage of GAI tools at various stages within the user-centered design process, discuss whether and how the use of such tools either benefited or hindered their workflows, consider the ethical considerations and challenges in every decision to engage with GAI, and build safeguards to avoid doing harm through their use of GAI. Students were exposed to various ethical dilemmas within GAI usage, such as identifying when GAI tools might produce disproportionately poor and harmful outcomes towards historically marginalized populations when there might be mismatches between GAI outputs and user needs, and when GAI usage might be inappropriate given specific design conditions and target populations. In particular, the course sought students with little to no coding/CS experience in enacting change within GAI usage policies, countering the popular rhetoric that GAI issues are inherently technical and therefore need technical knowledge to overcome.

In this paper, we present the course syllabus and findings from teaching this course across two cohorts. We conducted semi-structured interviews with students after the conclusion of each quarter, similar to Ghosh and Coppola [8], inquiring into their experiences with course materials and how their own GAI practices and policies were affected by participating in the course. The course is inspired by similar offerings at the University of Chicago Harris School of Public Policy [9], the University of Illinois at Urbana-Champaign [10], the University of Colorado Boulder [11], Brown University [12], and others. We hope this paper and our findings on this course can serve as a starting point for instructors and educators at other universities who are interested in building AI ethics courses within their curricula, which we believe must become a necessity across engineering departments.

Course Structure

Learning Goals and Overview

The course invited students to explore the usage of Generative Artificial Intelligence (GAI) tools and Large Language Models (LLMs) in the User-Centered Design (UCD) process, as they considered the various advantages and limitations they bring. It was established that interested students would need to have completed the departmental Introduction to UCD course (or equivalent) as a prerequisite for enrolling in this class.

The course was set up as a mixture between a seminar-style and project-based structure, with daily readings being due before the start of the class followed by in-class discussions and a short section of class periods being dedicated to group work. The learning goals for the course were as follows:

1. Students should be able to articulate the challenges and harms that GAI tools and LLMs can cause, and acknowledge how these percolate into their usage in the UCD process.
2. Students should be able to incorporate GAI tools and LLMs into the UCD process, and make informed decisions on whether using such tools at any given point is appropriate.
3. Students should gain practical experience working with GAI tools and LLMs within various stages of the UCD process, and be able to reflect upon the efficacy (or lack thereof) of such usage.
4. Students would develop an understanding of the growing body of research on GAI tools and LLMs, and gain insights on the direction of the field.

Each class period, roughly two hours long, was split into a maximum of three components. The first component, intended to last about an hour, was a student-led group discussion. In this component, self-selected students would be "Discussion Leaders" and briefly summarize the day's readings, and then propose small- and large-group discussion questions and/or short activities to the class. Though the instructor would participate in such discussions, this component would be entirely student-led in the way that the instructor did not have any input into designing the questions and/or activities being proposed. The second component of every class period would be more instructor-led discussions, in which they would bring their expertise to the table and provide their own thoughts and discussion questions/activities about the readings and topics for the day. Finally, the third component would be short periods of time reserved for group work towards the project (discussed in the following section), as appropriate. Not every class period included the final component – this was more prominent towards the second half of the course and typically absent on days when class conversations did not leave enough time for it.

The course was ungraded (see [13]–[15]), where students were asked to submit their own learning goals at the start of the quarter and self-evaluate at the end how they performed with respect to those and the course goals laid out in the syllabus. In this format, course components were still assigned weights as a percentage of the total final grade and were graded complete/incomplete based on meeting the assignment requirements. The use of GAI tools to complete assignments was permitted, since the authors believe that such tools could be important equity measures in a reading-heavy course [16], with the requirement that students attribute their usage of such tools wherever used, such as signing assignments with "proofread by ChatGPT" if done so. Students were also encouraged, in line with some assignment requirements mentioned below, to experiment with various GAI assistants in writing and completing assignments, thus being able to determine which tool could best support which action. The course was offered in a synchronous HyFlex format [8], where students could either participate in person or via Zoom on any given day.

The course was taught by the lead author within their department in two consecutive quarters, once each to the department's BS (taught in Fall 2024) and MS students (taught in Spring 2024), hereafter referred to in this paper as GAIC-B and GAIC-M respectively. The course and assignment structures were identical, and the contents

differed only in the way that GAIC-B contained more readings on account of being taught twice a week whereas GAIC-M was only once a week. The difference in the number of class periods was to accommodate the needs of the respective programs within the department and not the authors' choice.

Assignments

Writing Assignments

These assignments made up 30% of the final grade. They included 3 sets of Reading Responses, where students were required to submit 500-800 word responses to any given day's readings. Such responses required students to begin with short summaries of the readings along with key themes, proceed to critical analysis through their own interpretations of the material, followed by descriptions of how others in the field have received and interacted with the pieces, and conclude with a reflection of how these readings and themes connected to their own interests within the course and in general. Two out of the three Reading Responses could be completed based on any readings within the course list, but the third required students to identify a piece of literature – whether peer-reviewed research or otherwise – and perform the same exercise as above, with the addition of describing why they chose the particular article. A full list of course readings is provided in Table 1.

Week	Readings
Wk1	Eight Things to Know about Large Language Models, by Samuel Bowman. Explained: Generative AI, by Adam Zewe. Prompt Engineering, by Lillian Weng. Introduction to Generative AI, by Google Cloud Tech
Wk2	Artificial intelligence (AI) for user experience (UX) design: a systematic literature review and future research agenda, by Stige et al. User experience design professionals' perceptions of generative artificial intelligence, by Lu et al. StoryDiffusion: How to Support UX Storyboarding With Generative-AI, by Liang et al. Preparing future designers for human-AI collaboration in persona creation, by Goel et al.
Wk3	How Do Analysts Understand and Verify AI-Assisted Data Analyses?, by Gu et al. Generative AI in User Experience Design and Research: How Do UX Practitioners, Teams, and Companies Use GenAI in Industry?, by Takaffoli et al. Bridging the Gap between UX Practitioners' work practices and AI-enabled design support tools, by Lu et al. Enhancing UX Evaluation Through Collaboration with Conversational AI Assistants: Effects of Proactive Dialogue and Timing, by Kuang et al.
Wk4	Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration, by Fui-Hoon Nah et al. Simulating the Human in HCD with ChatGPT: Redesigning Interaction Design with AI, by Schmidt et al. ChatGPT for Learning HCI Techniques: A Case Study on Interviews for Personas, by Barambones et al. Herding AI cats: Lessons from designing a chatbot by prompting GPT-3, by Zamfrescu-Pereira et al.
Wk5	Diffusion Explainer: Visual Explanation for Text-to-image Stable Diffusion, by Lee et al. Creating User Interface Mock-ups from High-Level Text Descriptions with Deep-Learning Models, by Huang et al. DesignAID: Using Generative AI and Semantic Diversity for Design Inspiration, by Cai et al. Text-to-image AI as a tool for the designer's ideation process, by Lamac et al.
Wk6	"They only care to show us the wheelchair": Disability representation in text-to-image AI models, by Mack et al.

	The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT, by Wach et al. In-Between Visuals and Visible: The Impacts of Text-to-Image Generative AI Tools on Digital Image-making Practices in the Global South, by Mim et al. AI's Regimes of Representation: A Community-centered Study of Text-to-Image Models in South Asia, by Qadri et al.
Wk7	ChatGPT for good? On opportunities and challenges of large language models for education, by Kasneci et al. Investing in AI for social good: an analysis of European national strategies, by Foffano et al. "AI enhances our performance, I have no doubt this one will do the same": The Placebo effect is robust to negative descriptions of AI, by Kloft et al. Consumer reactions to AI design: Exploring consumer willingness to pay for AI-designed products, by Zhang et al.
Wk8	Design Principles for Generative AI Applications, by Wiesz et al. Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare, by Oniani et al. Designing Responsible AI: Adaptations of UX Practice to Meet Responsible AI Challenges, by Wang et al. AI ethics principles in practice: Perspectives of designers and developers, by Sanderson et al.
Wk9	Canvil: Designerly Adaptation for LLM-Powered User Experiences, by Feng et al. Conjure AI: DALL-E Icon Generator, by Robert Nowell
Wk10	Google Cloud: Introduction to Generative AI AWS: Generative AI with Large Language Models Microsoft: Fundamentals of Generative AI

Table 1: Readings List for GAIC-B

In addition to the three Reading Responses, students were also required to write a Final Paper, up to 1500 words in length. This assignment invited them to synthesize their readings and conversations across course topics and accommodated various formats. Students could write traditional 5-paragraph essays on topics of their choice [17], prototype an artefact that uses GAI and write a report on the design and user testing, or take a creative writing approach to imagine their futures in a GAI-powered world. They were required to submit a short Proposal midway through the quarter detailing their plans and received instructor feedback before they went on to craft their papers.

Group Project

This set of assignments provided students with the opportunity to experiment with GAI tools within various stages of the UCD process in a sandboxed course environment. Since one of the largest course goals was to determine the possibilities of using GAI tools within the UCD process, but also recognizing *where* and *how* it can be successful while also accounting for potential drawbacks and harmful consequences, the Final Project accounted for 45% of the overall course grade and was the largest deliverable for the course. Students self-sorted into groups of 3-4 and were required to first identify which sites/stages within the UCD process they wanted to try to use GAI tools in. They were required to submit a Site Selection assignment detailing their plans for using GAI tools in their chosen UCD stages, with short descriptions (200-300 words each) on why they chose these stages, what specific GAI tools they were interested in using, and why, how the application of GAI tools could be successful in those stages, as well as potential pitfalls they would expect.

Having submitted this report and received instructor feedback, student groups then put their theories on success and failure to the test. They were required to perform the chosen stages of the UCD process twice, once each with and without using GAI tools, and compare the two processes. In the second component, groups were required

to submit detailed reports of their processes with replicable detail and reflect upon them based on the hypotheses presented during Site Selection. Students were asked to include individual reflections about their future plans for GAI usage within the UCD process, given their observations and experiences in this assignment. Finally, on the last day of class, each group was required to present their work in this assignment to the class.

Class Participation

This component of the assignments evaluated students' participation in daily classes, including being a Discussion Leader (at least once in GAIC-M and twice in GAIC-B), posting short comments to readings on Canvas boards and interacting with each others' comments, and otherwise being an active participant in class discussions. Students were not graded on attendance, as that is against University and departmental policy.

Methods

To study the effectiveness of GAIC-M and GAIC-B, we decided to interview students to discuss their course experiences. Similar to Ghosh and Coppola [8], we waited until after grades for each course had been posted to recruit students for the interview series. We began with emailing respective class listservs soliciting their participation in these interviews, mentioning the purpose as stated above and that they would be compensated $20 for their participation. We recruited participants for 30-45 minute Zoom interviews, asking interested interviewees to indicate their preferred meeting times over a provided period of time. We interviewed a total of 12 participants across the two courses – 7 from GAIC-M (out of 23 total students) and 5 from GAIC-B (out of 13 total students). Participant information is provided in Table 2 along with participant pseudonyms similar to Ghosh and Coppola [8], a practice which avoids dehumanizing participants by boiling their identities down to numbers ([1], [18]). Pseudonyms are assigned based on famous philosophers sharing last initials with participants.

Participant #	Pseudonym	Course Attended
P1	Adorno	GAIC-M
P2	Irigaray	GAIC-B
P3	Camus	GAIC-B
P4	Aspasia	GAIC-B
P5	Kant	GAIC-M
P6	Rousseau	GAIC-M
P7	Elisabeth	GAIC-B
P8	Hegel	GAIC-M
P9	Hypatia	GAIC-M
P10	Chomsky	GAIC-B
P11	Rawls	GAIC-M
P12	Lovelace	GAIC-M

Table 2: List of Participants

In these semi-structured interviews, participants were first asked about their rationale for signing up for the course and their initial assessments, based on syllabi and first day's instructor presentation overviewing the class. They were then asked to detail the contributions (or lack thereof) of each of the class components – Discussion Leading and in-class discussions, readings, Reading Responses, Final Project, and Final Paper – on their learning of the topics. They were then asked to reflect on their overall experience with the class, in terms of achieving their own learning goals and those set out in the syllabus, as well as pointing out specific things they liked and disliked about the course. Finally, participants were asked if conversations within or topics studied in the course affected their own usage or understanding of GAI tools, and if they have been able to apply anything they gained from the course to their outside of it, such as in jobs/internships, research, other classes they have taken or are currently taking, or personal usage of GAI tools. Interviews were recorded with participant consent. The entire study was approved by the University's Institutional Review Board.

Interviews were thematically analyzed [19] using qualitative coding, where interview transcripts were initially open-coded as authors read them in conjunction with recorded interviews. These codes were then organized into themes and patterns of findings, as reported below.

Findings

One of the primary findings from the conducted interviews was that the course was incredibly successful, as documented through various patterns of comments.

All of the participants, across both courses, noted their satisfaction with the course in terms of achieving their own learning goals as well as those stated in the syllabus. A few salient examples are shown below.

> *"I signed up for this course to gain a stronger understanding of how GAI tools work and how I could use them in my job interviews, and I was able to do that in the class."* - Adorno (GAIC-M)
> *"I wanted to learn more about how AI causes harm and we talked about that a lot across different readings, and I now have a much better idea about it."* - Camus (GAIC-B)

However, there were some students – 2 from GAIC-M and 3 from GAIC-B – who expressed their dissatisfaction with one common personal learning goal not being met as much as they would have liked to.

> *"I initially thought the class would give me much more hands-on experience with different AI tools and while we did get to play around with a few different things in our project, I would have liked to learn a lot more about specific AIs and how/when to use them."* - Hegel (GAIC-M)
> *"When I signed up for the class, I initially thought we'd get to try out a lot of different AI tools and almost build pros/cons lists, and we didn't get to do that, so that was something I was a little disappointed by."* - Chomsky (GAIC-B)

The sentiment that the course could have incorporated more content around the technical details and underlying mechanisms within GAI tools was particularly prevalent among GAIC-M interviewees. 6 out of 7 GAIC-M interviewees expressed such a sentiment, as shown below.

> *"I liked that we spent some time talking about details about how AI tools work, but feel like we could have done a lot more of that."* - Hegel (GAIC-M)
> *"In general, I feel like we could have incorporated more technical readings into the syllabus. I know not everyone has the background to understand all of the details, but it would have been helpful even as optional readings for people who wanted to learn more. One of the strongest things*

I took away from the class was (the instructor's) explanation of how Stable Diffusion works, like adding in noise and removing slowly, that was very useful to understand." - Lovelace (GAIC-M)

Though the same sentiment of dissatisfaction with the low volume of technical content was not prevalent among interviewees from GAIC-B, there was still a strong appreciation for such content. In particular:

"One of the strongest things I took away from the class was understanding how ChatGPT/OpenAI does tokenization, and the example of counting r's in strawberry. It gave me a lot more clarity on why AI seems to get some easy things wrong and ends up looking silly." - Irigaray (GAIC-B)
"I liked a lot of the technical content about how AI works, that actually pushed me towards even thinking about a Master's program in the area." - Aspasia (GAIC-B)

The course content focused on the technical details of GAI tools was one of several topics that students appreciated within the syllabus. Overall, the readings and the themes were well-received across all of our interviewees. A few examples are highlighted below.

"I really liked the readings around AI bias against South Asian people, it was something new because I've read a lot about how AI is racist and sexist within US contexts, but not much outside that." - Irigaray (GAIC-B)
"I enjoyed all of the readings, it made me want to read more about specific topics and I did that pretty much every day after class, Googling on my own about the topics we read and talked about in class." - Rousseau (GAIC-M)

However, there emerged one pattern of dissatisfaction with the readings across both cohorts from GAIC-M and GAIC-B: some topics felt overdone by the end of the course.

"Some of the topics got a bit stale by the last few weeks. For instance, topics around bias. We pretty much circled back to 'AI is biased' across like half of the readings, and there was nothing more to add that we hadn't already discussed." - Chomsky (GAIC-B)
"A few topics were repetitive, like we spent a lot of time talking about UX practitioners' use of AI tools and it was cool to learn about the different ways but at the same time we quickly figured out how there were a few common underlying principles." - Rawls (GAIC-M)

Despite the sentiment that some readings and resultant class conversations got repetitive at times, most students (8/12 across both groups) rated class discussions as the most influential component of the course towards their own learning. A few examples are shown below.

"The most helpful part of the course were the conversations we had in class. It was good to hear a lot of different perspectives based on everyone's backgrounds, and get some expert comments from the instructor, and that has stayed with me the most." - Adorno (GAIC-M)
"I loved the conversations we had in class. Normally in readings-based classes, people tend to not show up or stay quiet because those really don't count towards the grade, but in this class we had everyone engaging with the readings and each other and that's probably down to how popular the topic is but also the readings were great and thought-provoking and discussion questions were interesting most times." - Elisabeth (GAIC-B)

A lot of the success of the in-class conversations was down to the quality work done by daily Discussion Leaders, something which was also recognized by our interviewees:

"I think a lot of the good conversations we had in class came down to the Discussion Leaders doing a good job with the questions and prompts they brought to class, it was a lot of different perspectives where everyone was coming from." - Rousseau (GAIC-M)
"I really liked the Discussion Leading, because everyone came up with interesting questions and we had good conversations everytime."- Aspasia (GAIC-B)

In fact, the Discussion Leader part of the course ended up becoming the most popular component across both sets of interviewees, to the point where 7/12 interviewees felt that it was the most helpful aspect of the course in terms of their learning. This is perhaps best summarized by the two comments below:

"Being a Discussion Leader was my favorite part of the class. I've never felt as trusted by an instructor to lead a class as this, and that made me a lot more focused at what I was doing. I imagine a lot of people felt this way when it was their turns, and that is probably why the conversations were so good." - Chomsky (GAIC-B)
"Discussion Leading initially seemed very intimidating but I thought it was the most helpful part of the course because it made us very very closely read some papers we were really interested in. I remember I did mine on a topic I was really curious about, and that had a direct effect on what I was doing in my project and even later on after the class." - Kant (GAIC-M)

However, the success of the Discussion Leading was not absolute, as interviewees noted that this attribute of the course took away from time that the instructor could lead or "lecture" the class, which some took issue with.

"Yeah Discussion Leading was cool but I would've liked a lot more instructor-led lecturing or conversations, because I feel like we missed out a little bit on the 'expert' perspective from the instructor." - Camus (GAIC-B)
"My biggest complaint is how little we got to hear from the instructor in the course. One of the big draws of the class for me was the instructor because I knew his research and wanted to learn from him, so getting to hear such little from the instructor and relying a lot on student conversations felt a bit of a letdown." - Adorno (GAIC-M)

After Discussion Leading, the most impactful aspect of the course for interviewees' learning was the Final Project, as the remaining 5/12 students agreed. They said:

"I really liked the Project component of the class, it allowed me to try out some GAI tools and experiment with stuff in the safety of the class, and I didn't have to worry too much about actually causing harm to people." - Irigaray (GAIC-B)
"The Project was my favorite part of the course, it was the closest to what I might actually be doing using GAI as a UX Designer." - Hypatia (GAIC-M)

In summation, the course was a strong success in terms of students achieving course goals as well as their own learning objectives. Over and above the aforementioned comments about student appreciation for course contents, perhaps the most glowing endorsement of the course is that most interviewees (9/12) mentioned that they took away lessons from the course they implemented in their personal usage patterns of GAI tools.

"This class definitely affected my use of GAI tools – the information about how ChatGPT tokenizing works definitely helped me ask better questions of ChatGPT." - Rawls (GAIC-M)
"I've definitely stopped using ChatGPT for like anything and everything after I learned about a lot of the ways it can be bad, based on what we talked about in class." - Camus (GAIC-M)

Furthermore, interviewees mentioned being able to take learnings from the class into their workplaces, which was especially the case for Master's students.

> *"It was really cool to be able to apply things we talked about in class directly to my work and internships, because now I have a better understanding of how it all works, both in terms of technical details and impact. So I can advocate for better and safer AI tools, especially for people who look like me."* - Lovelace (GAIC-M)
> *"I worked on an internship in the summer at a startup and I was able to influence a lot of their AI usage decisions, and I could use things I learned in class to talk about how AI can be harmful, that was very good to be able to do – felt like I was actually making a difference.*" - Hypatia (GAIC-M)

Discussion and Reflections for Future Iterations

Through the interviews with students from GAIC-B and GAIC-M, we can proudly say that the course was successful across both quarters. Students spoke highly of course content and experiences, appreciated every structural aspect of course, and found application of the course material into their personal and professional usage patterns of GAI tools. The success of the course and its students, both those interviewed for this paper and beyond, are also evidenced by their strong performances in terms of course assignments and projects. Particularly successful were the Discussion Leader and Final Project components of the course, where students gained both theoretical knowledge from different perspectives within the classroom as well as strong hands-on experience working with GAI tools that they could leverage in their personal and professional usage practices.

One of the most salient outcomes of both GAIC-M and GAIC-B was how students were able to recognize the multiple facets of AI-perpetrated harm upon historically marginalized communities, in achievement of Learning Goal 1 for the course. Having chosen to weave in AI ethics discourse across every set of readings, we are grateful that these topics resonated with students during the class and stayed with them after they graduated, informing their usage patterns. This is particularly evident in above-mentioned comments from Camus (GAIC-B), Irigaray (GAIC-B), Lovelace (GAIC-M) and Hypatia (GAIC-M), but also overall from class conversations across both courses, leaving little doubt about the fact that GAIC succeeded as an AI ethics course.

However, though the course was marketed as an AI "ethics" course, the content did not include traditional ethics literature such as theories of ethics or moral philosophy. While conversations in class, both from the instructor as well as students, referenced and evoked ethical theories such as Consequentialism [20] and the works of prominent ethicists such as Kant to describe phenomena or patterns of GAI outputs, there was not a course unit dedicated to such content. While such a unit would undoubtedly have made the class more of an AI ethics course, the lead author chose not to include it outright because they felt that their students, in an Engineering major at a predominantly-Engineering school, might be less than appreciative of having extensive readings focused on ethics theory. Upon reflection, we believe that though the class was successful without a dedicated Ethics unit, its presence would not have detracted from this success. This would be especially true if the Ethics unit incorporated more contemporary research from AI or tech ethicists, such as Timnit Gebru [21], Emily Bender, Abeba Birhane [22], Kate Crawford [23], and others, as opposed to foundational ethicists. Future iterations of this course could take inspiration from the work of Deborah Johnson [24] into incorporating ethics units into engineering courses.

It is important to reflect upon the differences between the lines of feedback from GAIC-B and GAIC-M students. Most notably, GAIC-M students expressed a stronger affinity towards course content on the technical details within GAI tools and were disappointed there was not more of it. Our department at the University of Washington is one with a part-time Master's program where Master's students often have full-time jobs and attend classes during evenings, and students in such programs are commonly "certain and narrow" [25] in the way that they

seek actionable information that they apply at their jobs in the very near future or to leverage for a new/higher job ([26], [27]). Therefore, such an appreciation for technical content among GAIC-M students is understandable and the yearning for more of it should be met in future iterations of this course, especially in light of interviewee comments about leveraging course knowledge in their workplaces. Additionally, this finding is part of a larger point of reflection that the contents of the two courses at the Bachelor's and Master's levels should be different in some way, if only to reflect the different attitudes of the students and learning goals of the programs. Future iterations of these courses could therefore focus more on providing Master's students with a stronger technical understanding of GAI tools and tie in ethical challenges to such inner workings, whereas the Bachelor's version could remain as-is.

Additionally, another aspect of course design that warrants strong reflection is the Discussion Leader system. Given that students in both GAIC-B and GAIC-M had taken at least two quarters' worth of classes (if not more) at our department, it is fair to assume their familiarity with the department's primary mode of learning – the flipped classroom [28]. In the flipped classroom model, students are expected to complete readings and view pre-recorded lectures on their own time before attending class, and use class time to ask questions of the instructor and otherwise work on their projects. The lecture-style mode of learning [29], where instructors deliver new content to students in class as a major section of class time, is not commonly practiced in our department. Under these conditions, the Discussion Leader format was initially chosen by the lead author as the course format both because of students' familiarity with completing readings before coming to class and to foster open conversations among individuals from varying perspectives on a topic so new and rapidly evolving. Furthermore, the lead author wanted to steer clear of being the 'expert' in the room in terms of determining aspects of class conversations to be correct or incorrect, and simply participate in conversations. There is also evidence [30], albeit in a different field, which suggests that students can learn more through such interactive engagement as opposed to listening to instructor lectures, further strengthening the choice of Discussion Leader format. While the modality was a resounding success based on student comments above and a strong contributor towards their learning, it is still important to note that sections of interviewees were dissatisfied with what they considered a low level of instructor involvement. In particular, interviewees such as Camus (GAIC-B) and Adorno (GAIC-M) mentioned their disappointment at receiving more of an 'expert' perspective from an instructor they chose to take the class with based on their research in the field. Future iterations of this course, especially if taught by an instructor whose research centers around AI ethics, should consider dedicating more portions of class time to instructor-led lecturing or discussions, if even to provide their own perspectives on the topics at hand and refraining from entering conversations about correctness or incorrectness about ethics-based topics. Instructors may also consider teaching their own research on AI ethics, if relevant, something which the lead author decided not to do because of their discomfort with that idea and to not put students in an awkward position of feeling unable to critique a reading because the author is the lead instructor.

Reflecting as instructors, we believe that the course was a great success even from our points of view, as we observed students meeting the learning goals we set out at the beginning of each quarter. This is especially exemplified by students finding internships and full-time opportunities in AI-related industry roles based on their time in this course, over and above the fact that in-class conversations across the board were rich and engaging. This course can have a bright future, and we believe that it should become a part of the departmental curriculum beyond a simple topics course. We are currently advocating for such a change, especially given how GAI tools are becoming more and more prevalent in our daily lives and proficiency in a wide range of such tools is being sought for industry positions. Such a course should be a part of engineering curricula across universities, and we welcome the opportunity to collaborate with educators who wish to design such a course for their contexts.

Limitations

One of the limitations we acknowledge in our work is the possibility that some participants might have been hesitant to discuss feelings of displeasure with their course experience with their interviewer, especially since

the interviewer (and lead author) was also their instructor, over and above the fact that this type of interview-style research likely only features students who had positive experiences. While we did uncover some elements of student dissatisfaction with the course, as mentioned above, it is entirely possible that still more opinions could have existed. Additionally, in relying on this voluntary interview method, we might be altogether missing out opinions of students who did not like the course overall and therefore did not bother to sign up for interviews. Future iterations of such work could consider interviewers being different from course instructors if possible, as well as including formal course evaluations as a source of student opinions, although those too can suffer from the same problem of dissatisfied students choosing to not fill out evaluations.

Conclusion

In this paper, we present a case study of an AI ethics course taught at both the Bachelor's and Master's level, which we designed from the ground up. We lay out the course syllabus and learning objectives, provide information about mechanics, and a full readings list. We also document patterns of success within the course based on 12 interviewees across two cohorts of students who graduated from the class, as we uncover how they found the course useful in teaching them crucial AI ethics and trained them to advocate for safer AI systems. We also reflect upon the ways in which the course can be improved in future iterations.

We hope this paper inspires others across Engineering departments and institutions to develop AI ethics courses. As we write this, the dangers of AI being used ubiquitously are growing by the day, and calls for slowing down and taking stock of potential impacts are falling upon deaf ears. The responsibility, as always, falls upon us as educators and practitioners to prepare our future generations to be more responsible in their AI work, and that work begins in the classroom.

Acknowledgments

Anonymized for this submission.

References

[1] S. Ghosh et al., ‘Do Generative AI Models Output Harm while Representing Non-Western Cultures: Evidence from A Community-Centered Approach’, *Proc. AAAIACM Conf. AI Ethics Soc.*, vol. 7, pp. 476–489, Oct. 2024 [Online]. Available: 10.1609/aies.v7i1.31651.

[2] S. Ghosh, ‘Interpretations, Representations, and Stereotypes of Caste within Text-to-Image Generators’, *Proc. AAAIACM Conf. AI Ethics Soc.*, vol. 7, pp. 490–502, Oct. 2024 [Online]. Available: 10.1609/aies.v7i1.31652.

[3] K. A. Mack et al., ‘“They only care to show us the wheelchair”: Disability Representation in Text-to-Image AI Models’, presented at the CHI Conference on Human Factors in Computing Systems (CHI ’24), Honolulu, HI, USA, 2024, p. 23 [Online]. Available: 10.1145/3613904.3642166.

[4] R. Qadri et al., ‘AI’s Regimes of Representation: A Community-centered Study of Text-to-Image Models in South Asia’, in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, New York, NY, USA, 2023, pp. 506–517[Online]. Availablehttps://dl.acm.org/doi/10.1145/3593013.3594016[Accessed: 24September2024].

[5] J. Coleman, ‘AI’s Climate Impact Goes beyond Its Emissions’, *Scientific American*, 2023[Online]. Availablehttps://www.scientificamerican.com/article/ais-climate-impact-goes-beyond-its-emissions/[Accessed: 11January2025].

[6] S. McLean, ‘The Environmental Impact of ChatGPT’, *Earth.Org*, Apr. 28, 2023[Online]. Availablehttps://earth.org/environmental-impact-chatgpt/[Accessed: 11January2025].

[7] P. Verma and S. Tan, ‘A bottle of water per email: the hidden environmental costs of using AI chatbots’, *Washington Post*, Sep. 18, 2024[Online]. Availablehttps://www.washingtonpost.com/technology/2024/09/18/energy-ai-use-electricity-water-data-centers/[Accessed: 11January2025].

[8] S. Ghosh and S. M. Coppola, ‘Making a Case for HyFlex Learning in Design Engineering Classes’, presented at the 2023 ASEE Annual Conference & Exposition, 2023[Online]. Availablehttps://peer.asee.org/making-a-case-for-hyflex-learning-in-design-engineering-classes[Accessed: 21October2023].

[9] The University of Chicago Harris School of Public Policy, *The Ethics and Governance of Artificial Intelligence*. [Online]. Available: https://harris.uchicago.edu/academics/programs-degrees/courses/spring-2025/38850/1. [Accessed: 20 Jan. 2025].

[10] University of Illinois at Urbana-Champaign, *Ethics of AI Certificate*. [Online]. Available: https://philosophy.illinois.edu/academics/ethics-ai-certificate. [Accessed: 20 Jan. 2025].

[11] | University of Colorado Boulder, *CSCA 5224: Ethical Issues in AI and Professional Ethics*. [Online]. Available: https://www.colorado.edu/cs/academics/online-programs/mscs-coursera/csca5224. [Accessed: 20 Jan. 2025].

[12] Brown University, *AI Governance and Ethics*. [Online]. Available: https://programs.professional.brown.edu/ai-governance-ethics. [Accessed: 20 Jan. 2025].

[13] S. Blum, *Ungrading: Why Rating Students Undermines Learning (and What to Do Instead)*. West Virginia University Press, 2020[Online]. Availablehttp://wvupressonline.com/ungrading[Accessed: 25October2023].

[14] S. M. Coppola and J. Turns, ‘Developing a Grounded Framework for Implementing Ungrading in a Disciplinary Context’, *ASEE Annu. Conf. Expo.*, 2023.

[15] S. Ferns et al., ‘Ungrading, Supporting Our Students through a Pedagogy of Care’, vol. 12, Sep. 2021.

[16] S. Ghosh, ‘ChatGPT as a Tool for Equitable Education in Engineering Classes’, in *2024 ASEE Annual Conference & Exposition Proceedings*, Portland, Oregon, 2024, p. 48458[Online]. Availablehttp://peer.asee.org/48458[Accessed: 12January2025].

[17] E. Brockman and M. Taylor, ‘Four College-Level Writing Assignments: Text Complexity, Close Reading, and the Five-Paragraph Essay’, 2016.

[18] S. Wang et al., ‘Naming Research Participants in Qualitative Language Learning Research: Numbers, Pseudonyms, or Real Names?’, *J. Lang. Identity Educ.*, pp. 1–14 [Online]. Available: 10.1080/15348458.2023.2298737.

[19] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, 2018.
[20] G. E. M. Anscombe, ‘Modern Moral Philosophy’, *Philosophy*, vol. 33, no. 124, pp. 1–19, Jan. 1958 [Online]. Available: 10.1017/S0031819100037943.
[21] T. Gebru and R. Denton, ‘Beyond Fairness in Computer Vision: A Holistic Approach to Mitigating Harms and Fostering Community-Rooted Computer Vision Research’, *Found. Trends® Comput. Graph. Vis.*, vol. 16, no. 3, pp. 215–321, Sep. 2024 [Online]. Available: 10.1561/0600000102.
[22] A. Birhane, ‘Algorithmic injustice: a relational ethics approach’, *Patterns*, vol. 2, no. 2, p. 100205, Feb. 2021 [Online]. Available: 10.1016/j.patter.2021.100205.
[23] K. Crawford, ‘Critiquing Big Data: Politics, Ethics, Epistemology | Special Section Introduction’, 2014.
[24] D. G. Johnson, *Engineering ethics*. Yale University Press, 2020[Online]. Availablehttps://books.google.com/books?hl=en&lr=&id=SKXeDwAAQBAJ&oi=fnd&pg=PP1&dq=Deborah+Johnson%27s+Engineering+Ethics+(2020&ots=7_9ED7s0NM&sig=Q9aCj2ezHxdMUMA-f2voZnrH1gQ[Accessed: 19April2025].
[25] J. Jung et al., ‘Part-time master’s students’ attitudes towards study and work’, *Stud. Contin. Educ.*, pp. 1–18, 2023 [Online]. Available: 10.1080/0158037X.2023.2254244.
[26] M. A. O. Cohen and S. Greenberg, ‘The Struggle to Succeed: Factors Associated with the Persistence of Part-Time Adult Students Seeking a Master’s Degree’, *Contin. High. Educ. Rev.*, vol. 75, pp. 101–112, 2011.
[27] F. Heidari, ‘Master of Science Degree in Indsutrial Management Designed for Technical College Instructors in Engineering and Technology’, presented at the 2014 ASEE Annual Conference & Exposition, 2014, p. 24.886.1-24.886.10[Online]. Availablehttps://peer.asee.org/master-of-science-degree-in-indsutrial-management-designed-for-technical-college-instructors-in-engineering-and-technology[Accessed: 20January2025].
[28] K. Jared et al., *Promoting Active Learning through the Flipped Classroom Model*. IGI Global, 2014.
[29] C. Van Klaveren, ‘Lecturing style teaching and student performance’, *Econ. Educ. Rev.*, vol. 30, no. 4, pp. 729–739, Aug. 2011 [Online]. Available: 10.1016/j.econedurev.2010.08.007.
[30] J. K. Knight and W. B. Wood, ‘Teaching More by Lecturing Less’, *Cell Biol. Educ.*, vol. 4, no. 4, pp. 298–310, Dec. 2005 [Online]. Available: 10.1187/05-06-0082.