Educating the Next Generation of Ethical AI Practitioners

Noah M. Kenney
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia USA
nkenney7@gatech.edu
0009-0000-5972-7552

Annie I. Antón
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia USA
aianton@gatech.edu
0000-0002-4397-9613

Abstract—Artificial intelligence (AI) technologies are rapidly advancing, increasing concerns about data privacy harms in AI models. We discuss how ethical AI can be incorporated into computer science curricula. This paper describes the design process for the first 'AI Privacy Engineering' course, to the best of our knowledge, taught in the United States. The course is designed for both undergraduate and graduate students at Georgia Tech. Throughout this course, students examine ethical implications of AI system design, development, deployment, and utilization, using the ACM's General Ethical Principles as an ethical framework. Recognizing that data privacy represents only one possible form of harm, the course blends theoretical and conceptual lectures with hands-on projects that require students to address ethical issues, including bias, fairness, and accountability in AI systems. Herein, we discuss the course design process, including selecting the appropriate body of knowledge, establishing learning objectives, creating assignments, and considering pedagogical methodologies we employed. We explain the empirical methods used to inform our design, including a systematic review of courses teaching AI development at over 40 universities. Our structured curriculum can be used to effectively teach ethical and safe AI, and we propose how these topics may be incorporated more broadly into computer science courses. Finally, we discuss early successes and the challenges faced while teaching the course, particularly in maintaining relevance despite fast-paced changes in the field of AI, an evolving legislative landscape, accessing computational systems to run AI models, and varying levels of student preparedness.

Keywords—*AI, Privacy, Ethics*

I. INTRODUCTION

Artificial Intelligence (AI) technologies are becoming increasingly ingrained in daily life. Generative AI has led to over 100 million users of emerging generative AI systems as of early 2023 [1]. Similarly, the use of AI technologies in the workplace has also increased significantly, with over 60% of employees using AI at work [2], and LinkedIn reporting a 74% annual increase in demand for AI specialists [3].

In response to the growing demand for AI specialists, students are increasingly seeking AI education curricula. In fact, approximately 28% of students who earned a doctoral degree in North America in computing or a related field specialized in AI or machine learning, making it the most popular specialty in the 2023 Taulbee survey produced by the Computing Research Association [4] [5]. Universities are expanding AI education offerings and hiring new AI faculty. For example, the University of Southern California invested $1 billion in an AI initiative and is hiring 90 new AI faculty, the University of Albany is hiring 27 new AI faculty, Purdue University is hiring 50 new AI faculty, and Emory University is hiring 60-75 new faculty [6].

At the Georgia Institute of Technology (Georgia Tech), the number of students seeking to enroll in AI and ML courses has made it challenging to keep up with the demand. In response, Georgia Tech announced a new minor in AI and machine learning in 2024, and unveiled an AI makerspace that provides students with access to "one of the most powerful computational accelerators capable of enabling and supporting advanced AI and machine learning efforts" [7] [8].

As often happens with the early excitement that accompanies evolving, powerful new technologies the course work initially lacks a much-sed focus on ethics and safety. Herein, we discuss the crucial need for curricula to prepare students for professions that demand they understand how to design and engineer ethical and safe AI. To this end, our curricula specifically focuses on data privacy and AI.

Without privacy as a foundation principle in an AI model, the potentially resulting harms can be quite serious. First, AI opens the door to data aggregation and inference, increasing the risk of an AI model inferring information beyond the purpose for which the data was provided by the user or originally collected by an AI model. This raises questions about whether the aforementioned data was collected and used without explicit consent [9][10]. Second, certain AI tools (such as generative AI) may reveal sensitive information from raw inputs to other users, or even the general public, against the will of its users [11]. Third, AI models can generate and spread false or misleading information, leading to reputational harms [12]. Beyond these harms, additional risks include phrenology and physiognomy (inferring personality and social attributes and possibly lead to discriminatory practices), biometric data risks (such as AI used for surveillance), and opaque decision-making, possibly leading to unfair or biased decisions.

This paper is premised on the need for developers to be well equipped to consider and address privacy risks at every stage of the development lifecycle, working to mitigate those risks. Not surprisingly, regulation and industry standards for AI model privacy have struggled to keep pace with AI model advancements, putting additional responsibility on the developer to address these risks and harms in a responsible manner. In this paper, we discuss the extent to which privacy appears to be taught in traditional computer science curricula, describe the design process leading to our AI Privacy Engineering course at Georgia Tech, and propose a curriculum that can be adopted by other institutions.

II. COURSE DESIGN PROCESS

In August of 2023, we proposed a new AI Privacy Engineering course at Georgia Tech for Spring 2024 semester. We first evaluated the extent to which privacy was covered in current Georgia Tech AI courses, or in related disciplines such as machine learning and deep learning. Table I shows the course codes and course titles for these courses covering AI or AI-related topics in Spring of 2023 or Fall of 2024.

TABLE I. Courses taught at Georgia Tech that incorporate AI or AI-related concepts.

Course Code:	Course Title:
CS 3600	Introduction to Artificial Intelligence
CS 3630	Introduction to Robotics and Perception
CS 4476	Intro Computer Vision
CS 4635/7637	Knowledge-based Artificial Intelligence
CS 4644/7643	Deep Learning
CS 4731	Game AI
CS 8803	Explainable AI
CS 8803	Advanced Natural Language Processing
CS 8803	Conversational AI
CS 8803	Systems for Machine Learning
PHIL 4803	AI Ethics and Policy
PUBP 8833	Public Policy for a Digital Age

We used the course descriptions to eliminate courses that only touched on AI in a brief way. Next, we contacted the instructor for each remaining course and shared our plan to create the new AI and privacy course. We requested a copy of each syllabus to determine whether and how privacy is taught throughout Georgia Tech's computer science curriculum. Additionally, we asked these professors to suggest any specific privacy topics, problems, or technology that would be beneficial for us to cover, to align with the AI curricula in their own courses. We contacted 11 professors in total, who taught the courses with an asterisk (*) after the course code in Table I. We received replies from all 11 professors.

Only one course had any real focus on privacy: CS 8803 (Systems for Machine Learning). However, it did so in a single lecture that covered federated learning and model extraction attacks and defense. Several professors we had contacted explained privacy-related concerns they had in their areas of domain expertise, which informed our decisions about lecture topics to include in our own course. Our goal was first to create a much-needed course, and second to ensure that the course would complement and become a valuable course for GT AI and ML students to take as part of their curriculum.

We further expanded our search to identify course syllabi from other universities, again seeking courses that teach privacy in the context of either AI or machine learning. Despite an extensive search, we found only two courses, summarized in Table II.

TABLE II. Two AI and ethics courses taught at other universities as of Fall of 2023.

University:	Term Taught:	Course Name:	Textbook:
University of Pennsylvania	Spring, 2020	AI, data, and society	No textbook
Western University	Spring, 2022	Artificial Intelligence and Society: Ethical and Legal Challenges	No textbook

The University of Pennsylvania course was primarily designed for those without a computer science background, and focused on broad implications of AI instead of strictly focusing on privacy. Based on the course syllabus, it did briefly discuss ethics, bias, and constraints of algorithmic decision making [13]. The course taught at Western University focused heavily on ethical theory, and included discussion of ethical and legal frameworks.

With few courses available at the university level, we expanded our search to industry and found two relevant courses. The first was 'AI Privacy and Convenience', offered by LearnQuest via Coursera, and the second was 'Accomplishing AI Privacy and Compliance with IBM Privacy Toolkits', offered by IBM. Our course differs from both these courses in that we focused more on how to implement privacy in the context of specific AI models, while they focused on theoretical applications of privacy in AI. None of these courses required or suggested a textbook, and upon further research, we could find no available textbook covering AI privacy engineering at an in-depth level, suitable for a university course. Instead, we compiled a list of readings that were relevant to the topics for our course and supportive of the privacy-related topics that professors, with whom we spoke, suggested would be helpful to cover. These readings and suggested topics from professors, informed the key topics we would cover and formed the basis for a 16-week course aligned with the Georgia Tech university calendar.

We began with a broad overview of AI and privacy, taught as independent concepts before merging the topics to discuss specific applications.

The course calendar and content is outlined below in Table III. Our class met twice a week, so the first bullet point for each week represents the topic covered in the first class of the week, and the second bullet point represents the topic covered in the second class of the week.

TABLE III. Weekly schedule of topics covered in Georgia Tech's new AI Privacy Engineering Course.

Week Number:	Topic(s) Covered:
Week 1	• Intro to AI and ethics • Intro to privacy and ethics
Week 2	• Overview of general data privacy • Overview of AI
Week 3	• Privacy impact assessments and data flow diagrams • Key privacy regulations
Week 4	• Training data • Synthetic data generation and supervised ML
Week 5	• Data usage and data profiling • Risk mitigation in model training
Week 6	• Differential privacy and data de-identification • Differential privacy and data de-identification (continued)
Week 7	• AI privacy frameworks and guides • AI privacy case study
Week 8	• Cybersecurity side of data privacy • Cybersecurity side of data privacy (continued)
Week 9	• Data privacy in generative AI • Privacy in generative AI case study
Week 10	• AI in healthcare and corresponding risks • Algorithmic decision making and corresponding risks
Week 11	• Spring break • Spring break
Week 12	• AI in financial lending and corresponding risks • General methods of risk mitigation
Week 13	• Robotics and sensors and corresponding risks • General methods of risk mitigation
Week 14	• User privacy controls and pay for privacy solutions • Ethics of privacy in AI
Week 15	• Panel discussion (guest speakers) with AI experts • Panel discussion (guest speakers) with privacy engineers
Week 16	• Final presentations

Based on the content we planned to cover and the gaps we had identified in the existing courses offered, we developed eight course objectives, with the goal of enabling students to understand how privacy is defined, protected, and managed, particularly within the context of AI. These objectives are outlined below, exactly as they were worded in our syllabus:

1. Examining the state-of-the-art research and practice in information privacy, including methods, tools, notations and processes used in information systems;
2. Gaining a grounding for future technical research in AI and privacy via the examination of current research issues and problems;
3. Learn the various AI tools and technologies available, along with criteria for balancing feature benefits and [privacy, bias, economic] risks;
4. Learn how personally identifiable data is utilized in the training of AI models and the associated risks;
5. Gaining experience in reading, analyzing, and presenting various forms of academic papers within AI and privacy;
6. Identify how to evaluate and implement current and future AI privacy frameworks;

7. Gaining experience in handling real-world privacy challenges through practical case studies and examples; and
8. Learning tools and methodologies for approaching privacy concerns, such as data collection, data storage, data usage in model training, and differential privacy.

To align these learning objectives with the grading criteria for the course, we emphasized presentations and projects when choosing learning assessments for the course. An important feature in grading for the ethics aspects of the course is that the General Ethical Principles codified in ACM Code of Ethics [REF] grounded our discussions throughout the semester and served as an important consideration in evaluating students analyses of AI and privacy technologies. For example, students were encouraged to consider how technologies supported or could undermine the ACM general ethical principles. This guided and shaped our grading rubrics and criteria.

When we reviewed the syllabi of other AI courses sent to us by Georgia Tech professors, we found a significant emphasis on exams and programming-based assignments. We noted that assessments of communication skills and assignments requiring students to evaluate current research, demonstrating critical thinking, were seldom utilized. Given the pace of innovation in AI, we found it particularly important for students to be able to read and understand current research and technological advancements. Additionally, we wanted to ensure students were capable of understanding key frameworks and legislative text as they relate to AI. Perhaps most important, we sought to ensure that students would be able to convey technical information to a wide variety of stakeholders, including technologists, business managers, and policy-makers through strong written and verbal communication.

These objectives were accomplished, in part, through student presentations, which we considered an important aspect of our pedagogical approach. Students individually choose an academic paper centered on AI, privacy, or both AI and privacy. Students were required to present three papers over the course of the semester, and received feedback on their presentations based on their ability to accurately and effectively present the key information from the paper, respond to questions afterwards, and connect the paper to AI and privacy.

In addition to presentations, students completed three projects over the course of the semester. Students were required to work in a group for at least one project, emphasizing the importance of communication skills, collaboration, and teamwork. In the first project, the primary goal is to give students an opportunity to apply and hone their critical thinking skills. To this end, they were required to evaluate a proposed White House executive order on the safe and ethical use of AI. The project required students to take the role of an advisor to the President by writing two briefings to President Biden. The first briefing supporting the executive order and the other opposing the executive order with the intent to persuade or dissuade the president from signing it. This project required students to objectively consider how technology and policy intersect, how technology can be implemented to serve public interests, how to evaluate a topic from multiple perspectives, and make arguments in favor of each.

In the second project, students worked in groups to evaluate an open-source AI model for data privacy harms. We allowed students to choose which model they wanted to evaluate and found that the majority chose a Large Language Model (LLM). This project provided them with hands-on experience evaluating a model, identifying possible harms, and communicating those harms in the form of a written report.

For the first two projects, students had a prescribed scope of work. However, for the third project, we allowed students to propose their own project topic and scope of work. This gave students practice pitching a project, defining a project scope and deliverables, and gaining approval to proceed with their project. We used over half of the class meeting time for several weeks to provide students an opportunity to collaborate, ask us questions, and get early feedback on their project. As another benefit, students were able to choose a project that related to the work they planned to do in the future. Some students focused on projects applicable to a particular industry or to industry at large, while others focused on research-oriented projects to prepare themselves for more advanced research in future education. As the semester progressed, we saw significant growth of our students in the learning objectives we previously outlined.

Throughout various class discussions and grading of student projects, it also became clear how certain topics we covered would be particularly applicable to certain industries. For example, algorithmic bias and fairness was of particular concern in finance and lending, given that algorithmic bias could lead to marginalized communities not having equal access to lending. Thus, students taking courses covering the development of financial models, algorithm development, or data science may benefit from a discussion of algorithmic bias in AI models. As a result, we propose seven modules that collectively encapsulate the topics we suggest including in a course on AI privacy, ethics, and safety. Table IV shows how each module could be incorporated as a single (or multiple) lecture(s) in specific courses, both inside and outside computer science disciplines.

TABLE IV. Summary of proposed course modules and how they map to relevant existing curricula.

Module:	Summary and Recommended Mapping:
Foundations of AI and Data Privacy	**Summary:** Module focuses on the ethical implications of AI system design and use, including an overview of the societal impact of privacy regulations as they relate to AI **Recommended Mapping:** Courses in computer science ethics, and public policy courses emphasizing technology
Privacy by Design in Data Science	**Summary:** Module focuses on privacy impact assessments and mapping of data, teaching how to integrate privacy considerations in the development lifecycle **Recommended Mapping:** Courses in data science, artificial intelligence, and machine learning
Training Data and Data Integrity	**Summary:** Module focuses on the collection of data, generation of synthetic data, and mitigation of risks during model training, teaching how to balance tradeoffs between accuracy and data integrity **Recommended Mapping:** Courses in data science, artificial intelligence, and machine learning
Privacy Mitigation Techniques	**Summary:** Module focuses on applied methods of mitigating data privacy risks in AI models, including those utilized by autonomous systems, robotics, and Internet of Things (IoT) devices **Recommended Mapping:** Courses in robotics, IoT, and data privacy
AI Privacy Frameworks	**Summary:** Module focuses on AI privacy frameworks, implications of frameworks on industry, and practical case studies, with specific emphasis on applying frameworks to high risk scenarios (healthcare, finance, education, etc.) **Recommended Mapping:** Courses in data science, healthcare, and analytics
Generative AI	**Summary:** Module focuses on ethical concerns in generative AI systems, including concerns of data privacy, intellectual property, and copyright ownership **Recommended Mapping:** Courses in public policy
Algorithmic Decision-Making	**Summary:** Module focuses on risks of algorithmic decision-making, including algorithmic bias and unfair outcomes, specifically within lending, finance, and policing **Recommended Mapping:** Courses in data science, algorithms, economics, and finance

We note that the modules we proposed above differ slightly in scope and content from our initial instruction of the course. These changes were made in response to student feedback. Most significantly, we saw a need to broaden the curriculum beyond a strict focus on privacy in AI, to also include discussion of other forms of harm in the context of AI models. Examples of such forms of harm include algorithmic bias, fairness, and accountability. Many of the papers students presented during their class presentations included a discussion of these privacy adjacent topics, and we determined they were essential to incorporate in the course curriculum. The modules we outlined above represent the structure we intend to use in future instruction of the course.

III. SUMMARY AND LESSONS LEARNED

This course is intended to help future AI practitioners understand and mitigate the ethical risks associated with AI systems thereby advancing responsible AI usage. While teaching this course, we faced several challenges. First, we acknowledge the difficulty of keeping course content relevant in a fast-paced field. We found that incorporating discussion of current news articles related to AI and ethics was helpful in addressing this challenge. Second, the legislative landscape in AI is evolving, which makes it challenging to teach. As a result, we focused on teaching students the skills of ethical decision making and privacy by design, reading, understanding, and determining how future legislation may impact compliance standards, instead of focusing only on current legislation or proposed legislation. Third, we faced some challenges in providing adequate computational resources to students in the course. In future semesters, we plan to request supercomputer access for the class, enabling students to run large models if their personal hardware lacks the computational power necessary to do so. Finally, we note the challenge of managing diverse levels of student preparedness. We offered the course with no prerequisites and did not restrict it to computer science students. As a result, we did not require students to program for any assessments, though some chose to do so in their third project. While it was a challenge to manage this diverse level of student preparedness, we ultimately found that it contributed to more broad and enlightening class discussions, and do not intend on adding any additional restrictions on course registration in the future.

IV. ACKNOWLEDGEMENT

This work was sponsored by NSF Award #2153481 (SafeInsights: A National Research Infrastructure for Large-Scale Learning Science and Engineering).

REFERENCES

[1] U. S. G. A. Office, “Artificial Intelligence’s Use and Rapid Growth Highlight Its Possibilities and Perils | U.S. GAO.” Accessed: Sep. 03, 2024. [Online]. Available: https://www.gao.gov/blog/artificial-intelligences-use-and-rapid-growth-highlight-its-possibilities-and-perils

[2] “AI study: Over 60 per cent use Artificial Intelligence at work – almost half of all employees are worried about losing their jobs,” Deloitte Switzerland. Accessed: Sep. 03, 2024. [Online]. Available: https://www2.deloitte.com/ch/en/pages/press-releases/articles/ai-study-almost-half-of-all-employees-are-worried-about-losing-their-jobs.html

[3] "The Future Of Work: Embracing AI's Job Creation Potential." Accessed: Sep. 03, 2024. [Online]. Available: https://www.forbes.com/councils/forbestechcouncil/2024/03/12/the-future-of-work-embracing-ais-job-creation-potential/

[4] "CRA Taulbee Survey," CRA. Accessed: Sep. 03, 2024. [Online]. Available: https://cra.org/resources/taulbee-survey/

[5] L. Coffey, "AI Most Popular Speciality for Computer Science Ph.D.s," Inside Higher Ed. Accessed: Sep. 03, 2024. [Online]. Available: https://www.insidehighered.com/news/quick-takes/2024/05/23/ai-most-popular-speciality-computer-science-phds

[6] S. D'Agostino, "Colleges Race to Hire and Build Amid AI 'Gold Rush,'" Inside Higher Ed. Accessed: Sep. 03, 2024. [Online]. Available: https://www.insidehighered.com/news/tech-innovation/artificial-intelligence/2023/05/19/colleges-race-hire-and-build-amid-ai-gold

[7] "Atlanta," Georgia Tech Catalog. Accessed: Sep. 03, 2024. [Online]. Available: https://catalog.gatech.edu/programs/minor-artificial-intelligence-machine-learning/

[8] "Georgia Tech Unveils New AI Makerspace in Collaboration with NVIDIA." Accessed: Sep. 03, 2024. [Online]. Available: https://coe.gatech.edu/news/2024/04/georgia-tech-unveils-new-ai-makerspace-collaboration-nvidia

[9] F. Wu, L. Cui, S. Yao, and S. Yu, "Inference Attacks: A Taxonomy, Survey, and Promising Directions," Jun. 27, 2024, *arXiv*: arXiv:2406.02027. Accessed: Sep. 09, 2024. [Online]. Available: https://arxiv.org/abs/2406.02027

[10] "Privacy attacks on AI systems: A current concern for organizations." Accessed: Sep. 09, 2024. [Online]. Available: https://iapp.org/news/a/privacy-attacks-on-ai-systems-a-current-concern-for-organizations

[11] S. Ikeda, "Security Researchers: ChatGPT Vulnerability Allows Training Data to be Accessed by Telling Chatbot to Endlessly Repeat a Word," CPO Magazine. Accessed: Sep. 09, 2024. [Online]. Available: https://www.cpomagazine.com/cyber-security/security-researchers-chatgpt-vulnerability-allows-training-data-to-be-accessed-by-telling-chatbot-to-endlessly-repeat-a-word/

[12] J. Zhou, Y. Zhang, Q. Luo, A. G. Parker, and M. De Choudhury, "Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg Germany: ACM, Apr. 2023, pp. 1–20. doi: 10.1145/3544548.3581318.

[13] P. Prasanna, "Please note: This syllabus is from Spring 2020. Spring 2021 will be online/synchronous and there will be modifications to the syllabus."