Educating the Next Generation of Ethical AI Practitioners

Noah M. Kenney & Dr. Annie I. Antón

CISSE - 28th Colloquium
November 14, 2024

Course Origins and Purpose

- Rapid growth of AI adoption
  - 77% of companies using or exploring AI [1]
  - Industry expected to increase 13x in next 6 years [2]
  - In North America, 28% of degrees awarded to “computer-related” fields concentrated in AI or ML [3]
- Increase in AI education [4]
  - University of Southern California invested $1 billion in AI initiative; hires 90 new faculty
  - Purdue University hiring 50 new AI faculty
  - Emory University hiring 60-75 new faculty
- There is need for responsible education on the ethics & privacy of AI
- Course developed in Fall, 2023 and taught in Spring, 2024

Georgia Tech
CREATING THE NEXT

Educational Gaps and Challenges

- Started by looking at how privacy was incorporated into AI or ML courses taught at Georgia Tech
- We requested syllabi from the professors teaching these courses
- Only one course (Systems for Machine Learning) covered privacy
- Expanded search beyond Georgia Tech, but found only two courses that covered AI + privacy

University:	Course Name:
University of Pennsylvania	AI, data, and society
Western University	Artificial Intelligence and Society: Ethical and Legal Challenges

Course Design Process

- No textbook or standard curriculum for teaching AI Privacy Engineering
- Compiled a list of readings based on our research and conversations with professors
- Taught AI and privacy independently, before considering the intersection between them

Key Course Topics

- Privacy and privacy preserving technologies
- Algorithmic bias and discrimination
- Accountability
- Ethics of AI
- Risk assessments and auditing
- Data collection and handling

Learning Objectives

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. **Learning the various AI tools and technologies available, along with criteria for balancing feature benefits and [privacy, bias, economic] risks;**
4. Learning 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.**

Course Modules

Module:	Module Topics:
1: Ethical Implications	- Societal impacts of AI - Integrations with privacy regulations
2: Privacy by Design	- Data mapping - Lifecycle privacy considerations
3: Training Data and Data Integrity	- Data sourcing - Synthetic data
4: Risk Mitigation Techniques	- Privacy preserving techniques - Risk assessments
5: Privacy Frameworks	- Current frameworks - Limitations of frameworks
6: Generative AI	- Intellectual property - Data privacy in LLMs
7: Algorithmic Decision-Making	- Risks in finance, policing, etc. - Algorithmic fairness

Assignment Structure

- **In class presentations:** Students found, read, and presented three academic papers to the class related to AI, privacy, or both
- **Projects**
  - **Project 1:** Students wrote briefings of the White House Executive Order on the Safe, Secure, Trustworthy Development and use of Artificial Intelligence
  - **Project 2:** Students conducted first-hand testing of chosen AI system, and determined how the system aligned with the NIST AI RMF
  - **Project 3:** Students proposed their own project, which may be a literature review, development of an AI model, or some form of empirical testing

Grading Criteria

Grade Category:	Percentage of Grade:
Presentations (3)	24% (8% each)
Team evaluation	10%
Project 1	20%
Project 2	20%
Project 3	20%
Participation	6%

- 60% of grade came from projects
- No exams, quizzes, etc.
- Presentations were largest individual grade category

Challenges Faced

- Students had vastly different levels of prior knowledge, since we had no pre-reqs
- AI is evolving very quickly, so frameworks and legislative texts are also evolving
- We had limited computational resources available to students, which made running large AI models challenging

Planned Changes

- Emphasize more of a technical focus
- Provide computational resources to students
- Require students to consider several frameworks
- Broaden scope to cover more “privacy adjacent” topics

Broader Impact and Future Implications

The general began by sharing his assessment of cybersecurity's role in national security. His lecture then walked his audience through the national security threats he experienced throughout his career before moving to current threats like AI and cybersecurity.

"We see supply chain attacks, we see zero-day vulnerabilities, we see ransomware," he said. "Cybersecurity is national security. How we think about that is much different at our agency today."

However, confronting these security threats takes more than just technical know-how. When asked by a cybersecurity and public policy master's student about what he looks for from students like her, the general said critical thinking.

References

[1] https://www.nu.edu/blog/ai-statistics-trends/

[2] https://explodingtopics.com/blog/ai-statistics

[3] https://www.insidehighered.com/news/quick-takes/2024/05/23/ai-most-popular-speciality-computer-science-phds#:~:text=The%20Computing%20Research%20Association%27s%20annual,in%20machine%20learning%20or%20AI.

[4] https://www.insidehighered.com/news/tech-innovation/artificialintelligence/2023/05/19/colleges-race-hire-and-build-amid-ai-gold

Conclusion and Q&A

Noah M. Kenney
nkenney7@gatech.edu

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