Responsible AI Engineering (CS 594)

Monday/Wednesday from 3:30 pm to 4:45 pm (Behavioral Sciences Building 281)

Spring 2025

Dr. Saeid Tizpaz-Niari

saeid@uic.edu

COURSE DESCRIPTION

The objective of this course is to familiarize students with the state-of-the-art artificial intelligence engineering techniques that incorporate deep neural network (DNN) and Large Language Models (LLMs) as well as their safety, security, and accountability implications. The course first overviews the fundamentals of engineering AI systems that includes requirements, architecture design, validation, and operation of AI-enabled software. Then, the course covers responsible AI-Software development that include topics such as security, privacy, fairness, interpretability, explainability, transparency, and trust, etc.

COURSE OBJECTIVES

Upon completion of this course, students can answer the following questions:

- How to reliably deploy and update AI models in production? How can we test the entire machine learning pipeline? How can MLOps tools help to automate and scale the deployment process?
- How do we scale production ML systems? How do we design a system to process huge amounts of training data, telemetry data, and user requests? Should we use stream processing, batch processing, lambda architecture, or data lakes?
- How to test, debug, and repair production ML systems? How can we evaluate the quality of a model's predictions in production? How can we test the entire AI-enabled system, not just the model? What lessons can we learn from software testing, automated test case generation, simulation, and continuous integration for testing for production machine learning?
- Which qualities matter beyond a model's prediction accuracy? How can we identify and measure important quality requirements, including learning and inference latency, operating cost, scalability, explainablity, fairness, privacy, robustness, and safety?
- What does it take to build responsible products? How to think about fairness of a production system at the model and system level? How to mitigate safety and

security concerns? How can we communicate the reasons of an automated decision or explain uncertainty to users?

PREREQUISITES

Some machine-learning experience required:

- Basic understanding of data science process, incl. data cleaning, feature engineering, using ML libraries.
- High-level understanding of machine-learning approaches.
- supervised learning, regression, decision trees, neural networks accuracy, recall, precision, ROC curve
- Ideally, some experience with notebooks, sklearn or other frameworks.

Basic programming and command-line skills will be needed, but no further soware-engineering knowledge required.

COURSE Topics

The topics of course include:

- *From Models to AI-Enabled Systems (weeks 1-2)*
- *Model Quality and Unit Testing (weeks 3-4)*
- *Architectural Design for AI-enabled Systems (week 5)*
- *Data and Infrastructure Quality (weeks 6-7)*
- *Safety and Security Concerns (weeks 8-10)*
- *Ethics and Fairness (weeks 11-12)*
- *Emerging Topics (weeks 13-14)*
- *Project demo and presentations (weeks 15-16)*

REQUIRED MATERIALS

The primary book for this course is Building Intelligent Systems by Geoff Hulten https://www.buildingintelligentsystems.com/.The second half of course will also take

topics from Trustworthy Machine Learning by Kush R. Varshney: http://www.trustworthymachinelearning.com/. The material for class lectures is covered through "Machine Learning in Producon: From Models to Products".

COURSE ASSIGNMENTS AND GRADING

This is a research-oriented and discussion-based course, which also includes hands-on exercises. The students are required to write a review for assigned chapters and papers prior to the class so that they can participate in class discussions. Students will also work on a major project in groups of 3 students and deliver in phases.

Category	Percentage
Homework Assignments	20%
Reading Assignments	25%
Participations	10%
Lab Assignments	10%
Final Project (Write-up, code, and presentation)	35%

Homework Assignments (20% of the grade)

There will be 4 individual assignments throughout the semester. The assignments will cover ML-enabled software products, Requirements, MLOps tools, and Explainability. The lowest grade of assignments will be dropped.

Reading Assignments (25% of the grade)

There will be reading assignments for each class. Students are required to write the paper summary and submit it before the class. The lowest grade of paper assignments will be dropped.

Class Discussion Participations (10 % of the grade)

Since the course is discussion-based, participation in the class discussion (or online forum) is required.

Lab Assignments (10 % of the grade)

Labs typically introduce tools and have a task with one or more clear deliverables. Lab assignments are designed to take about one hour of work and can be completed by the end of each week. The labs are passed/failed, and up to 2 failed lab assignments will be dropped.

Final Project (35% of the grade)

The final project is the most important component for the course. Students need to form a group of 3 students and deliver materials in phases. Deliveries include write-ups, code, and presentations.

Note: Your grade is independent of anyone else's grade in this class; that is, we do not grade on a curve. Everyone can get an A in this class. The instructor reserves the right to adjust these criteria, e.g., so that 88% or higher represents an A, based on overall class performance.

ATTENDANCE POLICY

This is a discussion-based graduate level class. Participation in class is absolutely required, but up to 4 absences will be waived. Occasionally, we will meet online via Zoom. In the event of an online class, zoom link will be provided in advance.

STANDARDS of CONDUCT

You are expected to conduct yourself in a professional and courteous manner, as prescribed by Student Disciplinary Policy. All graded work (except the final project with your classmate) is to be completed independently and should be unmistakably your own work. You may not represent as your own work material that is transcribed or copied from another source, including persons, books, or

Web pages. Plagiarism is a serious violation of university policy and will not be tolerated. You are welcome and encouraged to work together in learning the material. However, whatever you submit must be your own. In other words, cutting and pasting or copying verbatim from another source be it a classmate, an online source or even something that the TA/instructor showed you is strictly forbidden. All cases of suspected plagiarism will be reported to the Dean of Students for further review.

The use of generative AI tools such as ChatGPT is permitted. However, you should not use AI tools to ask for the exact questions. Students must cite any borrowed content sources to comply with all applicable citation guidelines, copyright law, and avoid plagiarism. Instances that violate these guidelines will be referred to the Office of Student Conduct and Conflict Resolution.

Cite Your Sources: If you worked with someone on an assignment, or if your submission includes quotes from a book, a paper, or a web site, you should clearly acknowledge the source. Bottom line: feel free to use resources that are available to you as long as the use is reasonable, and you cite them in your submission. However, copying answers directly or indirectly from solution manuals, web pages, or your peers is certainly forbidden.

Inspiration is free: you may discuss homework assignments with anyone. You are especially encouraged to discuss in online forum with your instructor and your classmates.

Plagiarism is forbidden: the assignments and code that you turn in should be written entirely on your own. You should not need to consult sources beyond your textbook, class notes, posted lecture slides and notebooks, programming language documentation, and online sources for basic techniques. Copying/soliciting a solution to a problem from the internet or another classmate constitutes a violation of the course's collaboration policy and the honor code and will result in an F in the course and a trip to the honor council.

Do not search for a solution online: You may not actively search for a solution to the problem from the internet. This includes posting to sources like StackExchange, Reddit, Chegg, etc. Searching for basic techniques in Python/Pandas/Numpy is totally fine. If you want to post and ask "How do I group by two columns, then do

something, then group by a third column" that's fine. What you cannot do is post "Here's the problem my professor gave me. I need to convert Age in Earth years to Martian years and then predict the person's favorite color. Give me code!" That's cheating.

When in doubt, ask: We have tried to lay down some rules and the spirit of the collaboration policy above. However, we cannot be comprehensive. If you have doubts about this policy or would like to discuss specific cases, please ask the instructor. If it has not been described above, you should discuss it with us first

Please also pay attention to the following netiquettes:

Always consider audience: Remember that members of the class and the instructor will be reading any postings. Respect and courtesy must be provided to classmates and to instructor at all times. No harassment or inappropriate postings will be tolerated.

Blackboard is not a public internet venue; all postings to it should be considered private and confidential. Whatever is posted on in these online spaces is intended for classmates and professor only. Please do not copy documents and paste them to a publicly accessible website, blog, or other space. If students wish to do so, they have the ethical obligation to first request the permission of the writer(s).

ACCOMMODATIONS POLICY

The University of Illinois at Chicago ("the University" or "UIC") is committed to full inclusion and participation of people with disabilities in all aspects of university life. The University seeks to provide an academic, social, and physical environment that makes persons with disabilities integral to the diversity of perspectives that is vital to an academic community. The course will follow the policy as described in https://drc.uic.edu/uic-student-accommodation-policy/