CSCI 4800/5800 Fall 2024
Explainable Artificial Intelligence (XAI)
Syllabus

Instructor:	Dr. Doug Williams
Class Hours:	MW 9:30 AM – 10:45 AM
Room:	North 3205
Office Hours:	MW 8:00 AM – 9:15 AM
Office:	Lawrence Street Center, 3rd Floor LW-310D
Phone:	303-710-2488 (cell)
Email:	milton.williams@ucdenver.edu

Catalog Data:

As artificial intelligence/machine learning (AI/ML) models are increasingly being employed to aid critical decision making in high-stakes domains such as government, healthcare, finance, and law, it becomes important to ensure that relevant stakeholders can understand the behavior of these models and the decisions they recommend. Such an understanding helps determine if, when, and how much to rely on the outputs generated by these models. This special topic course familiarizes students with recent advances in the emerging field of eXplainable Artificial Intelligence (XAI). In this course, we will review seminal research papers in the field, understand the concept of explainability from the perspective of different end users, discuss in different classes of interpretable models and post hoc explanations (e.g., rule-based and prototype-based models, feature attributions, counterfactual explanations, mechanistic interpretability).

Prerequisites:

CSCI 4800: CSCI 3412 – Algorithms (required), CSCI 4202 – Introduction to Artificial Intelligence (recommended)

CSCI 5800: Graduate standing, an undergraduate course in AI or machine learning is recommended.

Note: Students must have satisfied all prerequisite requirements with a grade of C- or better. If not, they will not receive credit for the course. Each student must sign the Prerequisites Agreement form to receive any credit for any assignment or exam. If this form is not signed by the end of the 1st week of class, the student will be administratively dropped from the course.

Expected Knowledge at the Start of the Course:

Students are expected to be fluent in basic linear algebra, probability, and algorithms. Knowledge of artificial intelligence (AI) and machine learning (ML) are recommended.

Students are also expected to have programming and software engineering skills with data sets using Python with packages such as numpy and sklearn.

Expected Knowledge Gained at the End of the Course:

By the end of the course, students will be able to:

1. Understand what Explainable AI us, its scope, and impact on various domains.
2. Understand global vs local explanations and their applications.
3. Identify and evaluate the most used XAI techniques and algorithms.
4. Use Python to apply Explainer algorithms and models and interpret the results.
5. Critically evaluate and contextualize the performance and reliability of explanations and identify their limitations and biases.

Textbook:

There is no textbook for this course. Topic readings will be poses on Canvas.

Topics:

- Inherently interpretable models
- Developing XAI-based applications
- Taxonomy of XAI techniques
- Black box vs white box model techniques
- Model-specific vs model-agnostic techniques
- Global vs local interpretations
- Model explainability
- Feature-based techniques
- Design considerations for implementing XAI

Course Outline: [TBD]

#	Date	Topic	Reading	Assignments
1	8/19			
2	8/21			
3	8/26			
4	8/28			
	9/2	Labor Day		
5	9/4			
6	9/9			
7	9/11			
8	9/16			
9	9/18			
10	9/23			
11	9/25			
12	9/30			
13	10/2			
14	10/7			
15	10/9			
16	10/14			
17	10/16			
18	10/21			
19	10/23			

20	10/28			
21	10/30			
22	11/4			
23	11/6			
24	11/11			
25	11/13			
26	11/18			
26	11/20			
	11/25			
	11/27			
27	12/2			
29	12/4			
30	12/11	Final Exam		

Grading Policy:

Assessment	Percent
Programs (5)	30%
Exams (2)	40%
Projects (1)	30%

94% - 100%	A
90% - 93.9%	A-
87% - 89.9%	B+
84% - 86.9%	B
80% - 83.9%	B-
77% - 79.9%	C+
74% - 76.9%	C
70% - 73.9%	C-
67% - 69.9%	D+
64% - 66.9%	D
60% - 63.9%	D-
00% - 59.9%	F

Examinations:

Examinations are intended to measure your individual mastery of the material. Exams concentrate on your understanding of the important concepts, rather than your ability to memorize details. For this class we will have two in-class exams: a midterm and a final. The exams will test your knowledge of assignment material, so you are responsible for mastering all lab, homework, and programming material submitted with other partners, as if you did all the work by yourself. The nature of the course material is such that the final exam must be cumulative.

Note that the examinations will include additional questions specifically for graduate students enrolled in CSCI 5800.

Term Project:

[Undergraduate] Undergraduate students enrolled in CSCI 4800 will have a term project that implements some XAI technique(s). Undergraduate students may work together in teams of up to four (4) students. Team compositions must be determined by the end of the second week of class.

[Graduate] Graduate students enrolled in CSCI 5800 will have individual term projects where they select an XAI technique, research the literature on the techniques, implement the techniques and replicate the original authors results, and documents the research in a research paper.

Programming Assignments:

There will be five (5) programming assignments during the term to implement various XAI techniques. An example of what is expected for a programming assignment will be provided.

E-Mail:

All email communication by students must use ucdenver.edu as the email domain, emails from Gmail, Hotmail, Yahoo, etc. are not considered valid methods of communication.

Class Conduct:

No phone calls, no text messaging, and no operation of computers for web reading, chatting, emailing, or gaming will be tolerated. Using computers to visit web sites distracts other students. Use of computers to read lecture notes and power point slides is appropriate. No loud talking between students is allowed. The goal is to create an excellent environment for learning and teaching.

Extensions/Make-ups:

In general, late work will not be accepted. Turn in all work by the established deadline. In case you have difficulties finishing an assignment contact the instructor before the deadline. Late work can be accepted only under circumstances beyond student's control and after arrangement with the Instructor, prior to the deadline. Note: work turned-in on time is eligible for partial credit. It will always be better to turn work in by the deadline, as trying to "perfect" it and turn it in late will give you no points at all. You have to follow the submission and media policies and guidelines published on the web. Plagiarism is the passing of someone else's work as one's own, without giving the original author due credit. Scholastic dishonesty will be treated very strictly as per University of Colorado Denver rules.

Lectures:

Class attendance is required, per department policy. Lecture material will be made available on the web prior to class. Lecture will also consist of chalk drawings, overhead drawings, and content not explicitly present in slides and notes.