University of South Carolina
College of Engineering and Computing

CSCE 590-01: Trusted AI
Fall 2021

Section 001: **Traditional**

Additionally, Asynchronous Online (via Blackboard Recording after scheduled class),

Biplav Srivastava, Ph.D. Professor, AI Institute, Computer Science & Engg., 541 Main St, Horizon 1, 4th Floor, Univ. South Carolina, Columbia, SC 29208 (518) 496-0128 **biplav.s@sc.edu**	Course Website: **https://blackboard.sc.edu** Supplementary Website: **https://sites.google.com/site/biplavsrivastava/teaching/csce-590-trusted-ai**
	Teaching Assistant: N/A
	Office Hours: WF 11:30 am – 12:30 pm or by appointment
	Class Timings: TTh 2:50 pm – 4:05 pm

Course Syllabus

Catalog Description

This will be an advanced Artificial Intelligence (AI) course focused on understanding reasons why Artificial Intelligence (AI) systems can be problematic and what can be done to make them trust-worthy. We will briefly cover AI as a decision support technology involving data processing, generating analysis and communicating insights to users. Then we will study how output of AI can be sensitive to issues in data, algorithmic steps (pre-, during and post-) and interaction with people of diverse background. We will conclude with techniques to remove or mitigate issues. The students will be exposed to latest tools and will do a project using AI technique of their choosing. The course will cover AI sub-fields of learning, reasoning, representation, preferences and uncertainty.

Course Outcomes

As a result of successful participation in this course, undergraduate students will be able to:

1. L1: Explain, execute and create AI-based analytical methods to process data: (a) unstructured data, (b) semi-structured data, (c) structured data
2. L2: Explain AI methods in data analysis: (a) Learning methods, (b) Reasoning, (c) Representation and standardization – knowledge graphs/ ontology, (d) Preferences, (e) Handling Uncertainty
3. L3: Identify trust issues in AI methods: (a) fairness and bias, (b) harmful language, (c) data privacy
4. L4: Methods and tools to promote trust: (a) Data sampling and synthetic data, (b) Testing and rating for communication, (c) Algorithmic innovations like differential privacy and explanations

As a result of successful participation in this course, graduate students will be able to do all of the above, and:

1. L5: Evaluate gaps in Trusted AI tools and create new datasets to handle them
2. L6: Explain emerging standards, frameworks and laws.
3. L7: Explain research findings in open areas and critique their contributions

Graduate students will be assigned additional responsibilities and evaluated correspondingly.

Prerequisites

Experience with a first course in data structures (CSCE 350), programming (CSCE 330) and introductory AI course (CSCE 580, 590 or equivalent) is needed.

Textbooks and Reading Materials

1. AI Fairness
   Trisha Mahoney, Kush R. Varshney, and Michael Hind, Available at:
   https://krvarshney.github.io/pubs/MahoneyVH2020.pdf
2. In AI We Trust: Ethics, Artificial Intelligence, and Reliability, Mark Ryan. Available at:
   https://link.springer.com/article/10.1007/s11948-020-00228-y
3. Python for Data Analysis
   a. Latest: Python for Data Analysis Book, by Wes McKinney, 2nd Edition. On Amazon at:
      https://www.amazon.com/gp/product/1491957662/, ISBN-13: 978-1491957660, ISBN-10:
      1491957662
   b. Book Data and Code Notebooks: https://github.com/wesm/pydata-book
   c. 1st edition (free download): https://bedford-computing.co.uk/learning/wp-content/uploads/2015/10/Python-for-Data-Analysis.pdf
4. Artificial Intelligence: A Modern Approach (Fourth edition, 2020)
   Stuart Russell and Peter Norvig
   http://aima.cs.berkeley.edu/
   ISBN-13: 978-0134610993
5. Tutorial on Trusting AI by Testing and Rating Third Party Offerings, in conjunction with 29th International Joint Conference on Artificial Intelligence (IJCAI 2020), Biplav Srivastava, Francesca Rossi, Yokohoma, Japan, Jan 2021, https://sites.google.com/view/ijcai2020tut-aitrust/home
6. Open Datasets
   a. US: https://www.data.gov/ or any US state
   b. Text of legislations - LegiScan, https://legiscan.com/
   c. Kaggle datasets: https://www.kaggle.com/datasets
   d. Google datasets search: https://datasetsearch.research.google.com/

All readings/materials comply with copyright and fair use policies.

Required Software

In order to access course materials and complete the course assignments you must have access to:

- A personal computer (PC) or laptop with the **Microsoft Office Suite** (Word, Excel, PowerPoint).
- The Internet (for using email, browsing the web, accessing the course website, and submitting assignments)
- Programming environment setup using open source. Python using Jupyter Notebook or Java using Eclipse will be the supported languages.
- For any proprietary tool, please consult the instructor. As a general rule, the instructor will ensure that no student may get an undue advantage due to usage of a particular technology that is unavailable to others.

Course Format

This course will be delivered in class. The primary medium of course delivery will be in class and their recordings available afterwards for **asynchronous viewing**.

- Student-to-Instructor (S2I) Interaction: Students must attend class unless they have taken permission from instructor to skip a lecture due to any reason and review the lecture via Blackboard later. The professor will post regular announcements, provide individual feedback to students, and hold office hours. One office hour every week will be in person and one will be online via Blackboard Collaborate Ultra.

- Students-to-Student (S2S) Interaction: Students will engage in discussions in class. Further, they can engage through email, the discussion board, and Blackboard Collaborate Ultra.

- Student-to-Content (S2C) Interaction: Students will engage with course content by completing assignments and participating in class activities. They can also use discussion boards and video conference meetings.

The instructor will reply to all feedback in a reasonable amount of time; the same is expected of the students. Specifically,

- Communication: Responses to email communication and questions will be provided within 48 hours. In subject, please prefix with "**CSEC590**: "If you have a question about any deliverable (e.g., project, assignment, presentation) with a deadline, you are advised to email me at least 5 days before it is due so that there is sufficient time between response and deadline.

- Test Grading: Grades for assignments will be returned within 1 week of due date.

Attendance Policy

You are expected to attend class lectures and participate in class discussions. If you expect to miss class for any reason you should contact the instructor by email as soon as possible. You are responsible for all material covered in lectures whether you are present or not. Lectures will not only be used to illustrate and expand on the material in the textbook, but will also include material available only during lecture that will appear on the assignments, quizzes, and exams.

Lecture presentations assume that you have read the assigned material **before** coming to class and are prepared to ask questions during class. If you don't ask questions, then I will assume that you understand the material. If there is a topic you do not understand, **it is your responsibility** to seek clarification from me during lectures or during office hours, or from other students. If you miss a lecture, **it is your responsibility** to view recorded class notes from Blackboard, get the notes and announcements from a classmate.

Time Commitment and Planning

Any university senior undergraduate or graduate course requires a large amount of work outside of lecture. I assume that when you register for this course you will allocate an average of at least 8-10 hours per week, in addition to lectures, to study the textbook material, complete the homework assignments, and prepare for course project. It is your responsibility to manage your workload. If you procrastinate starting your assignments, you may find that you do not have enough time to complete the project or assignments, or that a technology problem may prevent you from completing your assignment. Note that not being able to access a computer or network will not be considered an acceptable excuse for submitting your assignment late.

Time Allocation Plan

- Week 1: Introduction
- Week 2: Background: AI - Common Methods
- Week 3: The Trust Problem
- Week 4: Machine Learning (Structured data) - Classification
- Week 5: Machine Learning (Structured data) - Classification – Trust Issues
- Week 6: Machine Learning (Structured data) – Classification – Mitigation Methods
- Week 7: Machine Learning (Structured data) – Classification – Explanation Methods
- Week 8: Machine Learning (Text data) - Classification
- Week 9: Machine Learning (Text data) - Classification – Trust Issues
- Week 10: Machine Learning (Text data) – Classification – Mitigation Methods
- Week 11: Machine Learning (Text data) – Classification – Explanation Methods

- Week 12: Emerging Standards and Laws
- Week 13: Project presentations
- Week 14: Project presentations, Conclusion

Assessments

For undergraduate students

- Project: 50% + 10%: project report (50%) and code, for elevator presentation to class (10%)
  - Data analysis project
    - Dataset must be from given catalog
    - Use analytical methods to present new insights
- Quiz: 20%
  - 4 based on preceding lectures
- Exam: 20%
  - For undergraduate, final examination. Total 20%

Tests	**1000 points**
• Course Project – report, in-class presentation	600 points
• Quiz – best of 3 from 4	200 points
• Final Exam	200 points
Total	**1000 points**

Your overall final course letter grade will be determined by your points on the above assessments.

For graduate students

- Project: 50% + 10%: project report (50%) and code, for elevator presentation to class (10%)
  - Data analysis project OR
    - Dataset must be from given catalog
    - Use analytical methods to present new insights
  - Create or explore new methods (preferred for graduate students) project
    - Problem to be discussed with instructor
    - Example: Analyze sound signals to estimate crowd
- Quiz: 20%
  - 4 based on preceding lectures
- Exam:
  - Research paper reading (10%) and presentation to class (10)% - Total 20%
    - Read a paper accepted at a top Data / AI conference: AAAI 2019-2021, IJCAI 2019-2021, NeurIPS 2019-2021, KDD 2019-2021, SIGMOD 2019-2021. Make a 1-slide summary based on given presentation template.
    - Present a 1-slide summary to class (10%)

Tests	**1000 points**
• Course Project – report, in-class presentation	600 points
• Quiz – best of 3 from 4	200 points
• Final Exam – Paper summary, in-class presentation	200 points

Total	**1000 points**

Your final grade is based on the total points you have earned over the semester. Letter grades will be assigned as follows:

A = [900-1000]
B+ = [870-899]
B = [800-869]
C+ = [770-799]
C = [700-769]
D+ = [670-699]
D = [600-669]
F = [0-599]

If everyone performs very well, I do not have a problem with assigning everyone A's. However, poor performance (particularly failure to do project on time) will result in a low grade.

Important Note Regarding Grade Appeals
My teaching assistant or I will grade all assignments. If you have a question about a grade you have received, or you believe that you were graded incorrectly, please see me during **office hours** or set up an online appointment. If you wish to appeal an assignment grade you must do so within one (1) week of my posting the grade to Blackboard. If you want to make a case for re-grading your work based on another student's grade on the same task, I will review and then re-grade your work as well as the other student's work entirely from scratch.

Missing a Quiz: There will be 4 quizzes on announced dates. A student's lowest score from the 4 will be removed and the rest will be considered for assessment. If a person misses 1 quiz, the score of that quiz will be ignored. If a person misses more quizzes, this will impact their assessment.

Missing slot on project or paper presentation: Unless the instructor is informed 1 week in advance, missing the presentation slot will lead to a zero on presentation component.

Missing the project report: A delay of 7 days will be allowed with a penalty of 20% (200 points). No point will be awarded for any delay beyond a week for project report.

Request for Accommodations
The University of South Carolina is committed to providing access to programs and services for qualified students with disabilities. If you are a student with a disability and require accommodation to participate and complete requirements for this class, notify me immediately and contact the Student Disability Resource Center (http://www.sa.sc.edu/sds, 1523 Greene Street, LeConte College Room 112A, 803-777-6142, sasds@mailbox.sc.edu) for verification of eligibility and determination of specific accommodations. In addition, please provide me the required accommodation letter from the Student Disability Resource Center. *All course materials are available in alternative format upon request*.

Academic Integrity
University policies and procedures regarding academic integrity are defined in policy STAF 6.25, Academic Responsibility - The Honor Code (see http://www.sc.edu/policies/ppm/staf625.pdf). Prohibited behaviors include
plagiarism, cheating, falsification, and complicity. All potential Honor Code violations will be reported to the Office of Student Conduct and Academic Integrity, which has the authority to implement non-academic penalties as described in STAF 6.25. Academic penalties for Honor Code violations in this course range from a zero on the assignment to failure of the course.

In reference to this course, students are expected to do their own work when assignments require individual work. For example, students may not copy the work of others, either manually or electronically, under these conditions. Further, students who allow their work to be copied by others risk violation of the

University Honor Code. If situations arise in which the application of the University Honor Code is unclear, students should seek the interpretation of the instructor.
The faculty takes violations of the University Honor Code (http://www.sc.edu/policies/ppm/staf625.pdf) seriously. Students are encouraged to review the Honor Code and to understand the consequences of any action that is proven to be a violation of the code.

Remember that the first tenet of the Carolinian Creed is, "I will practice personal and academic integrity."