CS4333, Current Directions in Artificial Intelligence (4-0) Syllabus for Winter 2023

Course description

This course covers the practical use of AI technologies in solving problems within enterprises. Starting from a view of what problems are well solved with data-driven analysis, the course covers modern methods (including deep neural networks) and a host of practical issues that users of these technologies should be aware of. These include security and adversarial examples; privacy, data management, and data governance; testing and verification of system performance and safety; explainability, interpretability, and debugging; data issues such as measurement, preprocessing/cleaning, data bias, and causality; and legal and ethical issues with AI technologies. Prerequisites: CS3331 and CS3332.

Instructor

Prof. Neil Rowe, ncrowe@nps.edu, (831) 656-2462 (office), GE-328, faculty.nps.edu/ncrowe. No official office hours but he is usually there 900-1700 PST when not teaching. We will also have a number of guest lecturers from the AI faculty at NPS.

Classes

Classes will meet using Zoom on Tuesdays and Thursday 1700-1900 PST (2000-2200 EST). Sessions will be recorded for those who cannot attend live, and posted later on the Sakai site. Like most classes at NPS, we expect you to spend at least six hours per week outside of class, for a commitment of 10 hours per week.

Grading

Grades are based on two quizzes focused on terms and definitions, a paper presentation you give concluding with a class discussion (which can be prerecorded if you cannot make the class time), a term project including a class presentation (which can also be prerecorded), and some points on your participation in class discussions (with some alternatives for the remote students). The paper presented should come from the "Possible Readings" folder; we will ask you for your selection in advance, and negotiate as necessary. The term project will involve at least 20 hours of work and will result in a written report on some area of AI technology; it is due on Friday, March 17. The median course grade is somewhere between an A- and a B+.

Class materials

We will use a class site in Sakai that you should be able to access using your NPS account. Required readings, restricted-choice readings, and quizzes will be posted there.

Schedule

For reference on 1/3: Skim "ayoung_chinese…pptx", "baseit_survey_jnl…pdf", and rowe_decepwork18.pptx

For 1/5, Introduction: Read AI in aircraft: Read kwgF5p5p0e.htm, BWzQ2wXA.htm, and IV9r903T.htm; read also "Koslowski_preventative_maintenance …" and "JAIC Create AI …"

For 1/10, AI applications: Read "Technical Future AI ….pptx", "Understanding the Strategic Implications…"

For 1/12, AI limits with Prof. Denning; Read "intelligence_not_computable_Denning-Nov-2019.pdf" and "hofstader….pdf"

For 1/17, Prof. McClure in the first hour; in second hour, AI and surveillance: Read "acoustic_surveillance_Buswick-2021.pdf", "reading_emotions_and_personality_McQuaid_2021.pdf"

For 1/19: Deep neural networks with Prof. Orescanin, read "Multilevel_Classification_of_Heterogeneous....pdf" and "Decomposing_Satellite-Based....pdf"; read "Edge Intelligence for Rendering … .pdf

For 1/24, Reinforcement learning with Prof. C. Darken (both hours); read "boron_darken....pdf"

For 1/26, Data fusion and explainability: Read" "automatic_ai_accountability_Raji-2020.pdf" and "ai_explainability_Johnaovic-2022.pdf"

For 1/31, Four paper presentations and discussions by students (15-minute presentation, 10-minute discussion)

2/2: Two paper presentations and discussions by students (15-minute presentation, 10-minute discussion); midterm quiz in second hour

For 2/7, Prof. Kroll on trust in AI and other things; read "Selbst….Fairness-Abstraction….pdf"

For 2/9: Software engineering for AI with Prof. Michael; two paper presentations and discussions by students (15-minute presentation, 10-minute discussion)

For 2/14: Mr. Litton in the first hour; read "moving_targets_for_adversarial_ai_ Martin-2021.pdf"; two student paper presentations in the second hour

For 2/16: Prof. Davis on robotics and cybersecurity; two student paper presentations in the second hour

For 2/21: Prof. Das in the first hour; for second hour, read Regulating AI: Read "regulating_ai_Ellul-2021.pdf" and "legal_limits_to_ai_Roman-2019.pdf"

For 2/23, Adversarial AI with Prof. Barton; read "can_ethical_commitments …" and "day_zero_ethics …"

For 2/28, Prof. Kolsch in the first hour; student presentations on projects (25 minutes each) in second hour

On 3/7: Student presentations on projects in both hours (25 minutes each)

On 3/9: Student presentations on projects in both hours (25 minutes each)

On 3/14: Last class day; student presentations on projects in first hour (25 minutes each); Final quiz in second hour

On 3/17: Writeup due on class project

General criteria we will use for evaluating AI applications:
- Are the task and its metrics clearly identified?
- How difficult are issues of data acquisition?
- Is there buy-in from stakeholders?
- Have possible harms been adequately surveyed?
- How good is the process for fixing errors?
- Have adversary manipulations of the input data been adequately considered?
- How good are the procedures for reporting and assessing performance?
- How adequate are personnel resources for the application?
- Is it (or was it) a good idea, judging by what has happened so far?

- How successful is it or was it?
- How much of it depends on AI?