CS59300 Human-AI Interaction

Instructor: Tianyi Zhang, Assistant Professor of Computer Science

Email: tianyi@purdue.edu

Lecture: 12:00-1:15 TR @ Lawson B134

Office Hours: Tue and Thurs 3pm-4pm

Office: Lawson 3154H

Instructional Modality: Face-to-Face

Course Credits: 3.0

Prerequisites: Basic programming skills and basic understanding of machine learning are required.

Course Description

Have you ever wondered about these:

- Will programming jobs no longer exist because of large language models like GPT-3?[1]
- How far are we from the “black art” of natural language programming as Dijkstra called it 40 years ago?[2]
- Why does IBM suddenly seek to sell Watson Health, their AI for healthcare division, after 10 years of huge investment?[3]
- Self-driving cars are coming, but are we ready?
- How will humans and AI evolve together in the next decade?

This course will help you answer those questions. In this course, you will learn concepts and research topics about (1) how to build interactive machine learning models and systems, (2) how existing AI-based systems and applications work or clash against the strengths and weaknesses of human cognition, (3) how to design user interfaces to support human-AI partnership, and (4) how to support interpretability, transparency, trust, and fairness in AI-based systems.

This course is lecture-driven. Each lecture is accompanied with one or more research papers as optional reading materials. Throughout the semester, you should expect to

---

[1] OpenAI’s GPT-3 Can Now Generate The Code For You. *Analytics India Magazine*, July 20, 2020.

[2] E. W. Dijkstra. On the foolishness of “natural language programming”. In *Program Construction*, pages 51–53. Springer, 1979.

[3] L. Cooper and C. Lombardo. IBM Explores Sale of IBM Watson Health. *The Wall Street Journal*, Feb. 18, 2021.

select three papers and write reviews for these three papers. There will be a course project, in which you will work in groups to design and carry out research projects related to human-AI interaction. There will also be two programming assignments, one for implementing an interactive naïve bayes model and the other for implementing an interactive neural network.

This course is designed to introduce concepts and research topics in human-AI interaction. Since we have two programming assignments on interactive machine learning, students are expected to have basic programming skills and basic understanding of machine learning. Senior undergrads, master and PhD students who are interested in writing a thesis or doing a research-based project are especially encouraged to take this class. If you are not sure about your qualification for this course, feel free to research out to the instructor via email or stop by 3154H.

Learning Resources, Technology, Texts

No textbook is required for this course. Useful online learning resources include:

- Christoph Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable
- Jakob Nielson, 10 Usability Heuristics for User Interface Design
- Google PAIR, People + AI Guidebook
- Max Welling, A First Encounter with Machine Learning
- Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning
- ML + Design (a website with a collection of books, articles, tools, and tutorials on ML and design)
- Vishal Maini and Samer Sabri, Machine Learning for Humans
- Apple, Human Interface Guidelines for Machine Learning
- Microsoft, HAX Toolkit

We will use Slack to make announcements, send reminders, ask & answer questions, look for teammates, communicate to teammates, etc. You should feel free to set up a private channel for your own project in our workspace. Please join our Slack workspace via this link (to be updated).

All lecture slides, assignment instructions, and other materials will be posted on Purdue BrightSpace.

Learning Outcomes

At the end of this course, students should be able to:

- design and implement AI-based systems and applications
- know how to adapt a given machine learning model to solicit and incorporate user feedback
- evaluate an interactive AI-based system, especially through user studies in the lab or on a crowdsourcing platform like Amazon Mechanical Turks
- know the style of academic writing
- make and deliver academic presentations to the public

Grading

Two programming assignments [40%, 20% each]
Three reading assignments [15%, 5% each]
Course project [40%]
Class participation and discussion [5%]

Course Schedule

Week 1. Introduction to Human-AI Interaction	
Tue 8/23	Lecture: Introduction to Human-AI Interaction Optional paper reading: - Fails and Olsen, Interactive Machine Learning (IUI 2003) - Shneiderman and Maes, Direct Manipulation vs. Interface Agents (Interactions 1997)
Thur 8/25	Lecture: The Role of Humans in AI Optional paper reading: ● Amershi et al., Power to the People: The Role of Humans in Interactive Machine Learning (AI Magazine 2014)
Sample course project ideas released by the instructor on 8/29	
Week 2: Human Needs, Perceptions, and Experiences of Using AI	
Tue 8/30	Lecture: Everyday AI Applications Optional paper reading: ● When People and Algorithms Meet: User-reported Problems in Intelligent Everyday Applications (IUI 2019)
Thur 9/1	Lecture: Conversational Agents Optional paper reading: ● Myers et al., Patterns for How Users Overcome Obstacles in Voice User Interfaces (CHI 2018) ● Cho and Rader, The Role of Conversational Grounding in Supporting Symbiosis Between People and Digital Assistants (CHI 2020)

Week 3: Human Needs, Perceptions, and Experiences of Using AI	
Tue 9/6	Lecture: AI-based Programming Tools Optional paper reading: • Weisz et al., Perfection Not Required? Human-AI Partnerships in Code Translation (IUI 2021)
Thur 9/8	Lecture: AI-based Healthcare Applications Optional paper reading: • Liberati et al., What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation (Implementation Science 2017)
Reading assignment 1: Select a paper from Week 1 to Week 5 and submit a paper review (1-2 pages) by 9/11 midnight	
Week 4. Design Principles and Guidelines for Human-AI interaction	
Tue 9/13	Lecture: Guidelines for Human-AI Interaction Optional paper reading: • Amershi et al., Guidelines for Human-AI Interaction (CHI 2019) • Horvitz, Principles of Mixed-Initiative User Interfaces (CHI 1999)
Thur 9/15	Lecture: User Interface Design for Interactive Machine Learning Optional paper reading: • Dudley and Kristenssion, A Review of User Interface Design for Interactive Machine Learning (TIIS 2018)
Week 5. Communicating Model Confidence and Uncertainty	
Tue 9/20	Lecture: Methods to Measure Model Confidence and Uncertainty Optional paper reading: • Bhatt et al., Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty (AEIS 2021)
Thur 9/22	Lecture: Visualizations for Communicating Model Uncertainty Optional paper reading: • (Student Presentation) Kay et al., When (ish) is My Bus? User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems (CHI 2016)
Project proposal due on 9/25 midnight	
Week 6. Supporting Model Customization, Refinement, and Correction	
Tue 9/27	Lecture: Supporting Model Customization via Active Learning

	Optional paper reading: • Fu et al., A Survey on Instance Selection for Active Learning (Knowledge and Information Systems 2013)
Thur 9/29	Lecture: Supporting Model Customization via Direct Manipulation Optional paper reading: • Kulesza et al., Principles of Explanatory Debugging to Personalize Interactive Machine Learning (IUI 2015) • Koh et al., Concept Bottleneck Models (ICML 2020)
Week 7. Explaining Model Behavior (Part I)	
Tue 10/4	Lecture: An Overview of Explainable AI Optional paper reading: • Miller, Explanation in artificial intelligence: Insights from the social sciences (Artificial Intelligence 2019)
Thur 10/6	Lecture: Model-Agnostic Methods for Explainable AI Optional paper reading: • Ribeiro et al., "Why Should I Trust You?": Explaining the Predictions of Any Classifier (KDD 2016) • Ribeiro et al., Anchors: High-Precision Model-Agnostic Explanations (AAAI 2018) • Wexler et al., The What-If Tool: Interactive Probing of Machine Learning Models (VAST 2019)
Programming assignment 1 due on 10/9 midnight	
Week 8. Explaining Model Behavior (Part II)	
Tue 10/11	**No Class (October Break)**
Thur 10/13	Lecture: Model-Aware Methods for Explainable AI Optional paper reading: • Bau et al., Network Dissection: Quantifying Interpretability of Deep Visual Representations (CVPR 2017) • Strobelt et al., LSTMVis: A Tool for Visual Analysis of Hidden State Dynamics in Recurrent Neural Networks (TVCG 2017)
Reading assignment 2: Select a paper from Week 6 to Week 10 and submit a paper review (1-2 pages) by 10/16 midnight	
Week 9. Interactive Visual Analytics for Deep Learning	

Tue 10/18	Lecture: Interactive Visual Analytics for Deep Learning Optional paper reading: • Hohman et al., Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers (TVCG 2019) • Kahng et al., ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models (VAST 2017)
Thur 10/20	Lecture: Visualization Methods for Neural Networks Optional paper reading: • Yosinski et al., Understanding Neural Networks Through Deep Visualization (ICML Workshop on Deep Learning 2015) • Ohal et al., Feature Visualization (Distill 2017)
Week 10. AI Ethics, Fairness, and Equity	
Tue 10/25	Lectures: AI Bias and Fairness: Definitions, Metrics, and Root Causes Optional paper reading: • Angwin et al., Machine Bias (ProPublica 2016) • Buolamwini and Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification (FACCT 2018)
Thur 10/27	Lectures: Methods for Detecting and Fixing AI Bias Optional paper reading: • Galhotra et al., Fairness Testing: Testing Software for Discrimination (ESEC/FSE 2017) • Cabrera et al., FAIRVIS: Visual Analytics for Discovering Intersectional Bias in Machine Learning (VIS 2019) • Zhao et al., Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints (EMNLP 2017)
Mid-point project summary due on 10/30 midnight	
Week 11. Human-AI Partnership	
Tue 11/1	Lectures: An Overview of Human-AI Partnership Optional paper reading: • Bansal et al., Is the Most Accurate AI the Best Teammate? Optimizing AI for Teamwork (AAAI 2021) • Kamar, Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence (2016)
Thur 11/3	Lectures: Example Systems of Human-AI Partnership Optional paper reading:

	• Clark et al., Creative writing with a machine in the loop: Case studies on slogans and stories (IUI 2018) • Nguyen et al., Believe it or not: Designing a Human-AI Partnership for Mixed-Initiative Fact-Checking (UIST 2018) • Felix et al., The Exploratory Labeling Assistant: Mixed-Initiative Label Curation with Large Document Collections (UIST 2018) • Louie et al., Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models (CHI 2020)
Week 12. Calibrating User Trust and Reliance on AI	
Tue 11/8	Lecture: Trust in Automation Optional paper reading: • Merritt et al., I Trust It, but I Don't Know Why: Effects of Implicit Attitudes Toward Automation on Trust in an Automated System (Human Factors 2012)
Thur 11/10	Lecture: Methods for Adjusting User Trust in AI Optional paper reading: • Häuslschmid et al., Supporting Trust in Autonomous Driving (IUI 2017) • Kocielnik et al. Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-User Expectations of AI Systems (CHI 2019)
Reading assignment 3: Select a paper from Week 11 to Week 15 and submit a paper review (1-2 pages) by 11/13 midnight	
Week 13. Human Decision-Making Heuristics and Biases	
Tue 11/15	Lecture: The Dual Process Theory Optional paper reading: • Evans, Heuristic and Analytic Processes in Reasoning (British Journal of Psychology 1984)
Thur 11/17	Lecture: Human Biases Optional paper reading: • Kahneman and Tversky, Judgment under Uncertainty: Heuristics and Biases (Science 1974)
Programming assignment 2 due on 11/20 midnight	
Week 14. No Class (Enjoy Thanksgiving!)	
Week 15. Mental Models of AI Agents	

Tue 11/29	Lecture: Mental Models: What it is and Why it matters? Optional paper reading: ● Chapter 2 – Mental Models in Human-Computer Interaction from *Handbook of Human-Computer Interaction*
Thur 12/1	Lecture: The Role of Mental Models in HAI Optional paper reading: ● Beyond Accuracy: On the Role of Mental Models in Human-AI Teams (HCOMP 2019)
Week 16. Final Project Presentation	
Final project report due on 12/12 midnight	

Paper Reading Assignment

Each lecture is accompanied with one or more research papers to help you get a deep understanding of the topic covered by the corresponding lecture. While you are not required to read all these papers, you should expect to select and read three papers you are interested in. For each selected paper, you need to submit a paper review (one or two pages). As you read a paper or write your review, focus on the following perspectives.

- **Motivation of the work.** If the paper presents a new tool, who are the target users? Do they really need such a tool? What pain points does this tool address for those users? If the paper presents an empirical study, what are the research questions this study aims to answer? How important are these studies? Who will care about the findings and why should they care?
- **Novelty and significance of the work.** What is new here? What are the main contributions of the paper? What did you find most interesting?
- **Limitations, flaws, and blind spots.** Are there any unrealistic or false assumptions about the target users or the approach? Are there flaws or mistakes in the tool design, technical approach, or the study design?
- **Future work.** How would you improve on this work? Does this paper inspire any new ideas in your own research?

The grading of your paper review will depend on the overall quantity and quality of your questions and comments.

Programming Assignment

We will have two programming assignments in this course. Each assignment takes 20% of your final grade.

- Assignment 1: Building an interactive naïve bayes model.
    - Write code to implement a naïve bayes model to classify emails
    - Write a simple front-end user interface that visualizes the importance of each word based on the weights learned by the naïve bayes model
    - Modify your naïve bayes model to allow users to directly adjust the weight of a word in the naïve bayes model
- Assignment 2: Building an interactive neural network.
    - You will be given a dataset of bird images labeled with bird species and some concepts about a bird, such as wing color and beak length.
    - First, build and train two CNN models---one model that takes a bird image as input and outputs concepts and another model that takes the predicted concepts of the first model as input and outputs bird species
    - Second, build a simple front-end user interface that allows users to change the incorrect concept predictions of the first model so that users can influence the prediction of the second model

For each assignment, detailed instructions and code skeletons will be provided via Brightspace.

Course Project Instructions

You are expected to work on a course project either alone or in groups (no more than 3 students in a group). You can pick any topics related to human-AI interaction, e.g., building an interactive intelligent application, building a new user interface for an existing ML technique, evaluating existing AI-base techniques with user studies, etc. At the end of Week 1, I will release a list of sample project ideas with paper references, but feel free to work on your own ideas. Between Week 2 and Week 5, please stop by during office hours to discuss your project ideas with the instructor to get early feedback on the relevance, novelty, feasibility, and significance of your ideas.

A short project proposal is due on Sept 25 midnight (Week 5). This proposal should describe the project idea, the motivation of this idea, and (optional) a usage scenario if you propose to build a new tool. The proposal could be any length but no longer than 4 pages. It will be evaluated based on the quality of the idea and writing, not the length of your writing.

A mid-point project summary is due on Oct 30 midnight (Week 9). This summary should describe the envisioned approach/methodology/design as well as which parts have been done so far. The summary could be any length but no longer than 4 pages.

In Week 16 (Dec 6-Dec 10), each team will deliver a presentation of their project. The presentation will be about 20 minutes. You will get another 5 minutes for Q&A.

A final project report is due on Dec 15 midnight (max 10 pages plus references). Your final project report should be built upon your proposal and project summary. Feel free to reuse sections from those two reports in your final report. You may include an appendix beyond 10 pages, but your paper must be understandable without it. Submissions should be in the ACM format.

Your final report should be structured like a conference paper. It should contain:

- Abstract
- A well-motivated introduction
- Related work with proper citations
- Description of your methodology
- Evaluation results
- Discussion of your approach, threats to validity, and additional experiments
- Conclusions and future work

If you are doing a project that involves implementation, please include a link to your Github repository in your final report. Please also add a README file in your repository to describe how to run and test your code.

Important dates:

- Project proposal due on 9/25 midnight
- Mid-point project summary due on 10/30 midnight
- Final project presentation on Week 16 (12/5-12/9)
- Final project report due on 12/12 midnight

Course Policies and Expectations

Late submissions
No late submission is allowed.

Attendance

This course follows Purdue's academic regulations regarding attendance, which states that students are expected to be present for every meeting of the classes in which they are enrolled. Your final grade will depend on your participation in the class. Please come to the class continuously, read the assigned papers, and participate in discussions.

If you feel sick, have any symptoms associated with COVID-19, or suspect you have been exposed to the virus, you should stay home and contact the Protect Purdue Health Center. Please also notify the instructor so that the instructor can arrange remote participation for you. If you miss classes because of COVID-related reasons, your final grade will not be affected by your absence of classes. For more guidance on class attendance related to COVID-19 are outlined in the Protect Purdue Pledge for Fall 2021 on the Protect Purdue website.

For other conflicts or absences, such as for many University-sponsored activities and religious observations, the student should inform the instructor of the situation as far in advance as possible. When the student is unable to make direct contact with the instructor and is unable to leave word with the instructor's department because of circumstances beyond the student's control, and in cases falling under excused absence regulations, the student or the student's representative should contact or go to the Office of the Dean of Students website to complete appropriate forms for instructor notification. Under academic regulations, excused absences may be granted for cases of grief/bereavement, military service, jury duty, and parenting leave. For details, see the Academic Regulations & Student Conduct section of the University Catalog website.

Academic Guidance in Event of Quarantine or Isolation

If you must miss class at any point in time during the semester, please reach out to me via email so that we can communicate about how you can maintain your academic progress. If you find yourself too sick to progress in the course, notify your adviser and notify me via email or Brightspace. We will make arrangements based on your particular situation. Please note that, according to Details for Students on Normal Operations for Fall 2021 announced on the Protect Purdue website, "individuals who test positive for COVID-19 are not guaranteed remote access to all course activities, materials, and assignments."

Feedback to the instructor

During this course, I will be asking you to give me feedback on your learning in both informal and formal ways. Occasionally, at the end of a lecture, I will hand out index cards to collect anonymous comments and questions about this class and your learning experience. In the middle of the semester, I will send out an anonymous midpoint

survey about how my teaching strategies are helping or hindering your learning. It is very important for me to know your reaction to what we are doing in the class, so I encourage you to respond to these surveys, ensuring that we can create an environment effective for teaching and learning.

Classroom Guidance Regarding Protect Purdue

Any student who has substantial reason to believe that another person is threatening the safety of others by not complying with Protect Purdue protocols is encouraged to report the behavior to and discuss the next steps with their instructor. Students also have the option of reporting the behavior to the Office of the Student Rights and Responsibilities. See also Purdue University Bill of Student Rights and the Violent Behavior Policy under University Resources in Brightspace.

Academic Integrity

Please read and follow the academic integrity policy from the CS department (link) and from Purdue University (link). For paper reading assignments and the course project, you are encouraged to discuss with your classmates and course instructor. However, you should ensure that any written work you submit for evaluation is the result of your own research and writing and that it reflects your own approach to the topic. You must also adhere to standard citation practices in this discipline and properly cite any books, articles, websites, lectures, etc. that have helped you with your work.

Nondiscrimination Statement

Purdue University is committed to maintaining a community which recognizes and values the inherent worth and dignity of every person; fosters tolerance, sensitivity, understanding, and mutual respect among its members; and encourages each individual to strive to reach his or her potential. In pursuit of its goal of academic excellence, the University seeks to develop and nurture diversity. The University believes that diversity among its many members strengthens the institution, stimulates creativity, promotes the exchange of ideas, and enriches campus life. A hyperlink to Purdue's full Nondiscrimination Policy Statement can be found here.

Accessibility

Purdue University is committed to making learning experiences accessible. If you anticipate or experience physical or academic barriers based on disability, you are

welcome to let me know so that we can discuss options. You are also encouraged to contact the Disability Resource Center at: drc@purdue.edu or by phone: 765-494-1247.

Mental Health/Wellness Statement

If you find yourself beginning to feel some stress, anxiety and/or feeling slightly overwhelmed, try WellTrack. Sign in and find information and tools at your fingertips, available to you at any time.

If you need support and information about options and resources, please contact or see the Office of the Dean of Students. Call 765-494-1747. Hours of operation are M-F, 8 am- 5 pm.

If you find yourself struggling to find a healthy balance between academics, social life, stress, etc. sign up for free one-on-one virtual or in-person sessions with a Purdue Wellness Coach at RecWell. Student coaches can help you navigate through barriers and challenges toward your goals throughout the semester. Sign up is completely free and can be done on BoilerConnect. If you have any questions, please contact Purdue Wellness at evans240@purdue.edu.

If you're struggling and need mental health services: Purdue University is committed to advancing the mental health and well-being of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of mental health support, services are available. For help, such individuals should contact Counseling and Psychological Services (CAPS) at 765-494-6995 during and after hours, on weekends and holidays, or by going to the CAPS office on the second floor of the Purdue University Student Health Center (PUSH) during business hours.

Basic Needs Security

Any student who faces challenges securing their food or housing and believes this may affect their performance in the course is urged to contact the Dean of Students for support. There is no appointment needed and Student Support Services is available to serve students 8 a.m.-5 p.m. Monday through Friday. Considering the significant disruptions caused by the current global crisis as it is related to COVID-19, students may submit requests for emergency assistance from the Critical Needs Fund.

Emergency Preparation

In the event of a major campus emergency, course requirements, deadlines and grading percentages are subject to changes that may be necessitated by a revised semester calendar or other circumstances beyond the instructor's control. Relevant changes to this course will be posted onto the course website or can be obtained by contacting the instructors or TAs via email or phone. You are expected to read your @purdue.edu email on a frequent basis.