

Data Science Capstone & Ethics (ENGI E4800), Spring 2024

Instructors

Sining Chen (sc4549) Project Instructor

Savannah Thais (st3565) Ethics Instructor

CA

Swasthi Rao (spr2139)

Behind the scenes: [Jessica Rodriguez](#) , [Eleni Drinea](#) , [Brianne Cortese](#)

Best ways to reach us for urgent requests: email both the CA/Instructor

Course overview

This course provides a unique opportunity for students in the MS in Data Science program to apply their knowledge of the foundations, theory and methods of data science to address data driven problems in industry, research, government and the non-profit sector. The course activities focus on a semester-long project sponsored by an affiliate company or a Columbia faculty member. The project synthesizes the statistical, computational, engineering and social challenges involved in solving complex real-world problems.

The course now has a well developed Ethics component supported by Dr. Savannah Thais.

Timeline and Key Dates

The project portion of the course only meets a few times throughout the semester. The first lecture and the final presentation, plus any supplementary lectures that will be announced. The first lecture and final presentation will be required to be in person.

First Class Meeting, Jan 18, 2024

Project Portion Timeline and Dates

Week 1. First lecture. Specify Preference and receive matching results

Weeks 2 First meetings with mentors. Set up logistics (git) and problem definition

Weeks 3 - 4 Data wrangling and cleaning, Exploratory Data Analysis

Weeks 5 - 7 Initial modeling and trying off-the-shelf models

Mid-Term progress report due March 14 (Thursday 11:59pm ET) Also week of Spring Break

Phase 4: (Weeks 8 - 10) In-depth modeling and algorithms
Phase 5: (Weeks 11 & 12) Model accuracy, tuning and comparison
(Very Optional 2nd progress report)

Phase 6: (Week 13 & 14) Productionizing, if applicable.
Final Presentation Session May 3rd Friday 2-5pm in person
Final Report due May 10th (Friday 11:59pm ET)

Ethics Component Timeline and Dates

Ten Mondays starting January 22, 4-6 pm [online](#): Ethics Course Meetings

Each class will include a lecture and case study around a different topic in AI Ethics and Responsible Data Science, ranging from technical considerations like quantitative fairness and explainable AI, to societal contextualization like AI governance and technological determinism. For more details please see the Ethics component at the end of the syllabus. Classes will include ample time for questions and discussion; **we will leverage Slido for course discussion and exercises so please be prepared to log in on your computer or mobile device for each class.**

Details of each assignment are included in the Ethics Component section of the syllabus.

Friday January 26, 11:59 pm: Pre-course assessment due

Monday February 19, 11:59 pm: Ethics group project topic due

Monday February 26, 4 pm: LLM Evaluation assignment due

Friday April 5, 11:59 pm: Post-course assessment due

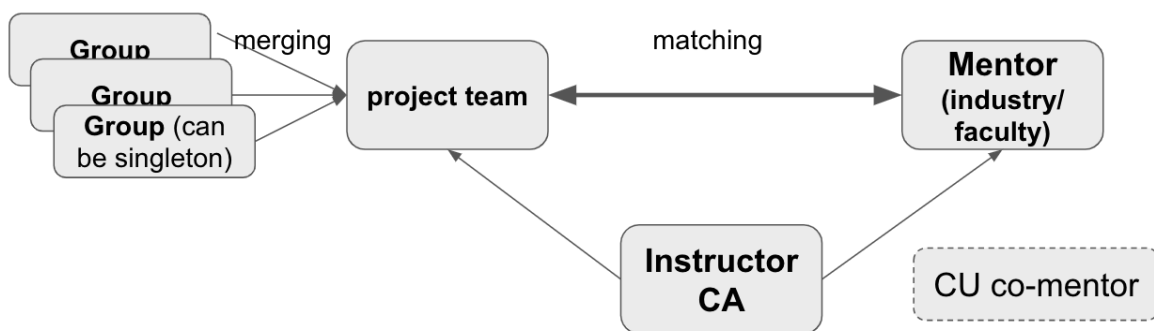
Monday April 8, 11:59 pm: Ethics group project due

Important Links ([CourseInfo](#) and [Accepted Projects](#))

Course Information folder - You can view the information about the course, i.e the syllabus, prescribed formats of the report, logs, meeting minutes, etc. [here](#).

Accepted Projects folder - You can access the different projects and proposals offered this semester [here](#).

Projects



Industry vs Faculty-Sponsored, Mentors, Co-Mentors

There are two types of projects. One type is sponsored by **industry** (or government/non-profit). Most of these projects are mentored by a mentor from the sponsor, sometimes co-mentored by a Columbia faculty member.

The other type is offered by Columbia **faculty**. These projects are mentored by the professor who proposed the project.

The mentor is a person or team from the sponsor side of the project, typically an employee of the sponsoring company or a faculty member and/or research fellow who proposed the project. A mentor's responsibilities include: set goals and milestones for the team, meet with the students weekly, provide knowledge and guidance, monitor the progress, and provide the instructors with detailed assessment of the team and team members' work as a basis for grading. Every project has a mentor.

A co-mentor of an industry project is a Columbia faculty member who helps with the mentoring. Sometimes the co-mentor also happens to be one of the instructors. A co-mentor's

responsibilities include: facilitate all aspects of the project, meet with and provide additional guidance to the students as needed, resolve issues. Not every project has a co-mentor.

Project Logistics

Instructors/CAs: provide course level information, coordination and supervision. Every project has an instructor/CA pair.

Self-Grouping

4-5 students will work together as a team on each project. Students will be asked to form *groups* of sizes 1 to 5 before the start of the semester. Each person will need to register themselves to a group via a google form (including singleton groups). Project-based preference (as opposed to topic-based preference) needs to be submitted as a group. Note that groups of size less than 5 will likely be merged by the instructors to form a *team* of size 4-5 at the time of project assignment.

Project Matching

This semester, student preferences will be solicited on **topic and project basis**. You can specify your preferences by choosing Loves, OKs, and DoNotLikes. Forms will be provided after the first class meeting.

Note that you **need to select 2 or more Loves and fewer than 3 DoNotLikes**. **Failure to follow this rule will result in all projects being OKs. Empty entries (NAs) are treated as OKs.**

Groups (if smaller) will be merged into teams and will be assigned to projects by the instructors based on preference, background, experience and learning opportunities.

Note that there is NO guarantee that everyone on a team will get a project that s/he loves. The instructors will do everything in order to not assign students to their true dislikes. Please keep in mind that the success of a project does not always correlate with the attractiveness of the proposal. Many factors play into the success of a project, for example mentor commitment.

Team Work

Students are expected to meet (online or offline) often as a team in order for projects and deliverables to take shape. It is up to the team to manage their tasks and time effectively.

The modes of collaboration is recommended:

--Selection of a team captain (with or without help from mentor/instructor/supervisor)

- A mechanism to select weekly presenter at the mentor meetings (rotated)
- A Github repository to organize and communicate about their work. Each team should add the instructor, and possibly industry mentor and/or faculty supervisor, to their Github repository.
- Regularly update Github Project Board updating regarding task status in project
- Regularly update [Weekly meeting minute](#) and [weekly progress log](#) in a shared folder (rotate loggers).
- Slack (mentor, supervisors' participation are optional)

Grades

At a high-level, the student's letter grade will be determined by the quality of their professionalism, effort, the written reports, the final poster presentation, adjusted by the self/peer-assessment results. It is likely that members of the same team do not receive the same grades.

1. Meetings with Industry Mentors and professionalism (20%)

Participation will account for 20% of the final grade. Evaluation will be based on clear and concise thinking towards achieving the research goals of the project, quality of discussions, participation and professionalism. In particular, students should attend and actively participate in **all** weekly team meetings in person or online, unless there is an emergency. Students should always be on time for meetings: students will be assigned letter grades individually, not as a team, and individual grades may be lowered if a student misses meetings or comes late or leaves early without prior approval.

Further, team members (i.e., students) are responsible for (a) meeting all deadlines, including those set by the instructor and/or team mentor even if they do not appear in the syllabus, and (b) maintaining adequate progress throughout the semester. Teams should keep in mind that their team mentors will discuss with the instructors the quality of their work. Teams are encouraged to email/post summaries of weekly or bi-weekly discussions, to keep all parties informed.

Other components of participation include: activity on github repos (for example, organizing the repo, actual code commits, resolving issues & answering comments, etc); attendance of team hacking sessions, and more. Github Classroom organization will be used to keep track of this activity.

2. Reports (45%)

Students are expected to write one original *progress* report, one ethics report and one *final* report.

- The midterm *progress* (6-12 pages) report aims at synthesizing phases 1-3 of the project **(15%)**.
- The *final* report (12-20 pages) is a culmination of how the project synthesizes the statistical, computational, social and ethical challenges involved in solving complex real-world problems. **(30%)**

Expected length of 1st report: 6-12 pages, not including appendices. Final report: 12-20 pages, not including appendices. Late reports will not be accepted.

At the end of each report please include each member's contributions to the work.

3. Final Presentation (10%)

Students will summarize and present their work in a final presentation. **Attendance is mandatory.** Industry affiliates and faculty members will all be invited to attend the presentation session.

4. Ethics component (25%)

Students' grade for the ethics component will be based on successful completion to the pre- and post-course assessments and their grades on the LLM evaluation project and final group project. Additional credit will be given for active participation in in-class discussions.

Self/Peer/Mentor Assessments

To promote fairness and collaboration among team members, we ask the students to fill out a [self/peer-assessment form](#) for the team at the end of the semester, or **at any time** during the semester as supporting evidence of any teamwork problem that should be brought to the attention of the instructor.

Mentors will be asked to fill out a similar form for **each member** at the end of the semester.

The instructors will summarize the results and factor the assessments into the final grade. Please note that the students should not feel that they should score high on all items to receive a good grade. The best scenario is when each member is engaged and brings his/her unique strength to the team.

Ethics Component

Course Outline

This section outlines the topics that will be covered during the ethics course. Each class session will include a lecture, a case study (a code-based demo or guided discussion/exploration), and time for questions and open discussion. Each week will include suggested readings that, while not required to be read before class, will deepen your understanding of the material we will cover in lectures and can provide a basis for your group project.

Week 1: Introduction to AI Ethics and Responsible Data Science (01/22/24)

- Taxonomy of AI Ethics
- Frameworks for responsible data science

Case Study: PredPol (Predictive Policing)

Suggested Readings:

- [Do Artifacts Have Politics?](#) (Langdon Winner)
- [The Fallacy of AI Functionality](#) (Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz, Andrew D. Selbst)

Week 2: Algorithmic Bias (01/29/24)

- Types of data and model bias
- Measuring bias in data sets and models
- Quantitative fairness definitions and limitations
- Methods for computationally addressing bias:
 - Adversarial de-biasing
 - Re-balancing data sets
 - Synthetic data

Case Study: COMPAS Recidivism Prediction

Suggested Readings:

- [Machine Bias](#) (Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner)
- [Fairness Definitions Explained](#) (Sahil Verma and Julia Rubin)
- [On the \(im\)possibility of fairness](#) (Sorelle A. Friedler, Carlos Scheidegger, Suresh Venkatasubramanian)

Week 3: Building Trustworthy Models (02/05/24)

- Data set construction considerations

- Inferential statistics
 - Correlation vs causation
 - Confidence and uncertainty
- Data leakage
- Model evaluation

Suggested Readings:

- [Leakage and the Reproducibility Crisis in ML-based Science](#) (Sayash Kapoor and Arvind Narayanan)
- [Statistical and Causal Models](#), from Causal Inference in Statistics, A Primer (Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell)

Case Study: [ML to Predict Suicide Attempts](#)

Week 4: Explainability and Transparency (02/12/24)

- Why does explainability matter
- Inherently interpretable models
- Global explainability methods
- Local explainability methods
- Neural network explainability methods
- Limits of explainability/using explainability for inference
- Model documentation and auditing

Suggested Readings

- [REFORMS: Reporting Standards for Machine Learning Based Science](#) (Sayash Kapoor et al)
- [Interpretable Machine Learning](#) (Christoph Molnar)
- [Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead](#) (Cynthia Rudin)

Case study: [GPT4 Model Card](#)

Week 5: Privacy (02/19/24)

- Differential privacy
- Federated learning
- Model monitoring

Case study:

Week 6: Current AI Models: Large Language Models and Generative AI (02/26/24)

- Technical foundations
- Emergent capabilities
- Hallucinations
- Biases
- Evaluation and red-teaming
- Future technical capacity of AI

Suggested Readings:

- [Evaluating LLMs is a minefield](#) (Arvind Narayanan and Sayash Kapoor) [Annotated slides]
- [How should AI systems behave, and who should decide?](#) (OpenAI)

Case study: guided discussion of LLM evaluation exercise, [AI Creativity Evaluation](#)

Week 7: Current Applications of AI (03/04/24)

- Training data and generative AI
- Social media (misinformation, silo-ing, mental health)
- AI and politics (military, information manipulation)
- AI for Social Good (humanitarian aid, AI and medicine, materials and energy science, etc)

Suggested Readings:

- [AI for social good: unlocking the opportunity for positive impact](#) (Nenad Tomašev et al)
- [How to Design AI for Social Good: Seven Essential Factors](#) (Floridi et al)

Case study: Exploration of NYC Algorithmic Hiring Policy

Week 8: Governance and Policy (03/18/24)

- Auditing and transparency
- Approaches to AI regulation
- Current state of US policy
- Regulatory capture

Suggested Readings:

- [Algorithmic Bias: Looking Beyond Data Bias to Ensure Algorithmic Accountability and Equity](#) (Thais et al)
- [AI Index Report Governance and Policy Chapter](#)

Case study: Generative AI (opting out, copyright lawsuits, and water marking)

Week 9: Real and Imagined Futures (03/25/24)

- AGI and AI Safety/Alignment
- AI hype
- Environmental impacts of AI

Suggested Readings:

- [Artificial General Intelligence Is Already Here \(BLAISE AGÜERA Y ARCAS AND PETER NORVIG\)](#)
- [Talking about a 'schism' is ahistorical](#) (Emily Bender)

Case study: Yoshua Bengio's [Catastrophic Risk Estimate](#)

Week 10: The Role of Technology in Shaping Society (04/01/24)

- Technological determinism and technosolutionism
- Digital colonialism
- TESCREAL
- Participatory design
- Alternative technological frameworks (data feminism, indigenous AI, non-western philosophies)

Suggested Readings:

- [AI Colonialism](#) (Karen Hao et al)
- [The Siren Song of AI Safety](#) (Brian Chen)
- [Situating Search](#) (Shah and Bender)

Assignments

Below you will find descriptions of each assignment for the ethics course. The pre and post-course assessments should be completed and submitted individually by each student; the LLM exploration and final projects will be completed with your capstone team and should be submitted by the team captain.

Pre + Post Course Assessments (Individual)

Pre-course assessment due Friday January 26, 11:59 pm

Post-course assessment due Friday April 5, 11:59 pm

After both the first class and the last class, submit a 1-2 paragraph response to one of the following questions (use the same question for both posts). Please include the question you choose to address at the beginning of your submission.

- Does AI function?
- What are the biggest limitations of AI right now? Are they addressable?

- What are the biggest risks of AI right now? Are they addressable?
- Does AI/data science have politics?
- How do/can you practice responsible/ethical data science?
- What role does technology play in society?

LLM Exploration/Evaluation (Group)

Due Monday February 26, 4 pm

Ask ChatGPT to write an essay or blog post on the topic of your choice, including references and supporting information. Evaluate the response with your group: is the information all accurate? How would you rate this essay or post? Does it provide realistic and informative information about the topic? Then try at least one of the following

- Feed the generated content back to ChatGPT and ask if it was AI generated. Repeat multiple times with the content and other non-AI-generated content.
- Ask ChatGPT to create a rubric for grading an essay or evaluating a blog post on the topic you selected. Then ask it to evaluate the previously generated content.
- Ask ChatGPT to identify weaknesses in the generated content and explain how it would correct them.
- Repeat the content generation process with a slightly varied prompt and compare the output. How can you create maximal variation?
- Repeat the content generation process with an additional stylistic or worldview prompt and compare the output. Does the content change?
- Another red-teaming experiment of your choice

Submit a 1-2 paragraph post about your exploration. The full assignment should not take more than 2 hours to complete.

Group Project (Group)

Due Monday April 8, 11:59 pm

A ~3000 word report or 10-12 minute presentation (recorded and submitted digitally) on one of the following topics. Other project formats such as a public-focused blog post, explainer tool, or others may be allowed with permission. We expect this project to take approximately 10 hours over the course of the semester. You may begin working on it at any point after your capstone team is formed. **Project topic must be submitted and approved by Dr. Thais by February 19.**

Possible topics:

- A specific application area of AI (criminal justice, housing, content generation, etc) including the technical approaches, data collection, evaluation, and societal contextualization
- What is AGI, how will we know if it is achieved, and what are its implications?
- An evaluation or case study of a specific AI or data science model or system
- What is the role of technology in general, and AI specifically, in a just society?

- A description and evaluation of a specific technical subject with the AI ethics wheelhouse (e.g. an explainable AI method, a model auditing framework, a red teaming approach, etc)
- A summary of an 'ethical' AI system
- A summary of ethical considerations about your capstone project, engaging with other relevant materials from the course
- Another relevant topic of your choice (with approval)