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

##### 

| **Instructors** | **CAs** |
| --- | --- |
| Sining Chen (sc4549) (Project main) | Swasthi Rao (spr2139, assisting Savannah and Sining) |
| Savannah Thais (AI Ethics) | Tina Cao (tc3334 assisting Adam and Sining) |
| Adam Kelleher (Project) |  |
|  |  |

## 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 and ethics 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.

**Course Format**

The Capstone has two components, a **project component (75%)** and an **ethics component (25%).** Both components are team-based.

Please note that in-person attendance is required for classroom meetings. Please contact your instructors/CAs if you need to be absent.

### **Project Component Timeline and Dates**

**Sept 9. First lecture on project logistics. Combined with Ethics First lecture.**

Week 1-2. Receive matching results. First meetings with mentors. Set up logistics (git) and problem definition.

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

**Collaboration Workflow Workshop (Sept 30 exact date needs to be confirmed)**

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

**Mid-Term progress report due October 26 (Saturday 11:59pm ET)**

Weeks 8 - 10 In-depth modeling and algorithms

Weeks 11 & 12 Model accuracy, tuning and comparison

**(Optional 2nd progress report)**

Week 13 & 14 Productionizing, if applicable.

**Poster file due: Dec 5, 2024(**or 4 business days before the poster session)

**Poster Session Week of December 9 (exact date TBA pending venue availability), 2-5pm in person**

**Final Report due December 19th (Thursday 11:59pm ET)**

### **Ethics Component Timeline and Dates**

**Mondays September 9 - November 18, 4-6 pm in NWC 501**

**(Detailed syllabus see Ethics Section Later in This Document)**

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; **due to the large class size, discussion will be facilitated through an online platform, please bring a laptop or mobile device.**

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

**Friday September 13, 11:59 pm:** Pre-course assessment due

**Monday October 7, 8:00 pm:** Deadline for final group project topic approval

**Monday October 21, 4 pm**: LLM Evaluation assignment due

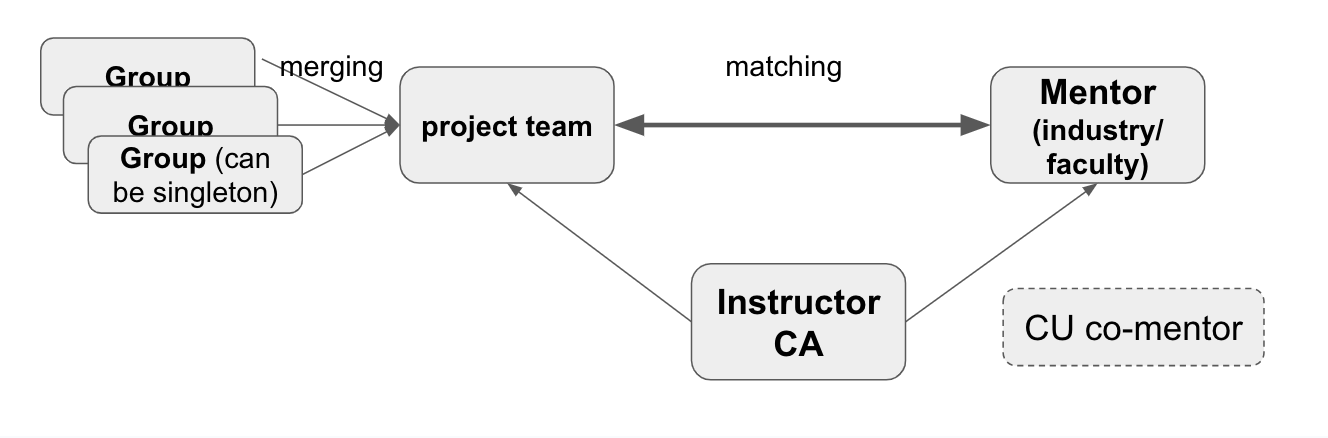
**Wednesday November 20, 11:59 pm**: Post-course assessment due

**Wednesday November 27, 11:59 pm**: Ethics group project due

## 

## 

## Projects [(CourseInfo shared folder)](https://drive.google.com/drive/folders/1sBv1EySZc7H2Qb0teD-PA7ekEjRAuxBi?usp=sharing)



### 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**

4-6 students will work together as a team on each project. Students will be asked to form *groups* of sizes 1 to 6 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 4-6 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.

Note that you **need to select X or more Loves and fewer than Y 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 a team will get a project that they love.** The instructors will do **everything** to not assign students to their DoNotLikes. Please keep in mind that the success and amount learned of a project does not always correlate with the attractiveness of the proposal. Many factors play into the success of a project.

**Team Work**

Students are expected to meet at least weekly 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 (rotation is recommended)

--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.

--Slack (mentor, supervisors’ participation are optional)

– [Weekly meeting minute](https://docs.google.com/document/d/1kPUfk1TkVJZ2mZlnZgOvLS4Li3pd-gZd/edit?usp=sharing&ouid=111187559732535721269&rtpof=true&sd=true) and [weekly progress log](https://docs.google.com/document/d/1SPqgOoIzLp3TrAJtV5VAYxTbZGvISWiG/edit?usp=sharing&ouid=111187559732535721269&rtpof=true&sd=true) in a shared folder (rotate loggers).

## 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.

1. **Reports (40%)**

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

* The first *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.

1. **Final Poster Presentation (15%)**

Students will summarize and present their work in a final poster presentation. **Attendance is mandatory**. The poster presentation will account for **15%** of the final grade. Industry affiliates and faculty members will all be invited to attend the presentation session.

1. **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.

**Please note, class attendance is expected. Attending fewer than 6 classes in person will result in you receiving a 0 for the ethics component, regardless of your performance on the assignments.**

**Self/Peer/Mentor Assessments**

To promote fairness and collaboration among team members, we ask the students to fill out a [self/peer-assessment form](https://docs.google.com/spreadsheets/d/13BXOWGuY5ToqvGtL6HRUyvQS_lhOI803/edit?usp=sharing&ouid=111187559732535721269&rtpof=true&sd=true) 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.

Note: this portion of the syllabus will be updated throughout the semester as readings and case studies are finalized.

#### Week 1: Introduction to AI Ethics and Responsible Data Science (09/09/24)

* Taxonomy of AI Ethics
* Frameworks for responsible data science

**Case Study**: PredPol (Predictive Policing)

**Suggested Readings**:

* [Do Artifacts Have Politics?](https://www.jstor.org/stable/20024652) (Langdon Winner)
* [The Fallacy of AI Functionality](https://arxiv.org/abs/2206.09511) (Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz, Andrew D. Selbst)

#### Week 2: Algorithmic Bias (09/16/24)

* 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](https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing) (Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner)
* [Fairness Definitions Explained](https://fairware.cs.umass.edu/papers/Verma.pdf) (Sahil Verma and Julia Rubin)
* [On the (im)possibility of fairness](https://arxiv.org/abs/1609.07236) (Sorelle A. Friedler, Carlos Scheidegger, Suresh Venkatasubramanian)

#### Week 3: Building Trustworthy Models (09/23/24)

* 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](https://www.cell.com/patterns/pdfExtended/S2666-3899(23)00159-9) (Sayash Kapoor and Arvind Narayanan)
* [Statistical and Causal Models](http://bayes.cs.ucla.edu/jp_home.html), from Causal Inference in Statistics, A Primer (Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell)

**Case Study**: [ML to Predict Suicide Attempts](https://qz.com/1001968/artificial-intelligence-can-now-predict-suicide-with-remarkable-accuracy?ref=jeremyjordan.me)

#### Week 4: Explainability and Transparency (9/30/24)

* 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](https://reforms.cs.princeton.edu/draft-paper.pdf) (Sayash Kapoor et al)
* [Interpretable Machine Learning](https://christophmolnar.com/books/interpretable-machine-learning/) (Christoph Molnar)
* [Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead](https://arxiv.org/abs/1811.10154) (Cynthia Rudin)

**Case study**: [GPT4 Model Card](https://cdn.openai.com/papers/gpt-4-system-card.pdf)

#### Week 5: Privacy (10/07/24)

* Differential privacy
* Federated learning
* Model monitoring

Case study:

#### Bonus Week: Guest Lecture on GitHub Best Practices from Daniel Woulfin (10/14/24)

#### Week 6: Current AI Models: Large Language Models and Generative AI (10/21/24)

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

**Suggested Readings**:

* [Evaluating LLMs is a minefield](https://www.cs.princeton.edu/~arvindn/talks/evaluating_llms_minefield/) (Arvind Narayanan and Sayash Kapoor) [Annotated slides]
* [How should AI systems behave, and who should decide?](https://openai.com/blog/how-should-ai-systems-behave) (OpenAI)

**Case study**: guided discussion of LLM evaluation exercise, [AI Creativity Evaluation](https://www.nature.com/articles/s41598-023-40858-3)

#### Week 7: Current Applications of AI (10/28/24)

* 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](https://www.nature.com/articles/s41467-020-15871-z) (Nenad Tomašev et al)
* [How to Design AI for Social Good: Seven Essential Factors](https://pubmed.ncbi.nlm.nih.gov/32246245/) (Floridi et al)

**Case study:** Exploration of NYC Algorithmic Hiring Policy

#### Week 8: Governance and Policy (11/04/24)

* 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](https://sciencepolicyreview.org/2023/08/mitspr-191618004007/) (Thais et al)
* [AI Index Report Governance and Policy Chapter](https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report-2023_CHAPTER_6-1.pdf)

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

#### Week 9: Real and Imagined Futures (11/11/24)

* 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)](https://www.noemamag.com/artificial-general-intelligence-is-already-here/)
* [Talking about a ‘schism’ is ahistorical](https://medium.com/@emilymenonbender/talking-about-a-schism-is-ahistorical-3c454a77220f) (Emily Bender)

**Case study**: Yoshua Bengio’s [Catastrophic Risk Estimate](https://yoshuabengio.org/2023/06/24/faq-on-catastrophic-ai-risks/?utm_source=ADSA&utm_campaign=5995eb0f57-EMAIL_CAMPAIGN_2022_10_04_06_04_COPY_01&utm_medium=email&utm_term=0_5401c7226a-5995eb0f57-453703484)

#### Week 10: The Role of Technology in Shaping Society (11/18/24)

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

**Suggested Readings:**

* [AI Colonialism](https://www.technologyreview.com/supertopic/ai-colonialism-supertopic/) (Karen Hao et al)
* [The Siren Song of AI Safety](https://techpolicy.press/the-siren-song-of-ai-safety/) (Brian Chen)
* [Situating Search](https://dl.acm.org/doi/10.1145/3498366.3505816) (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.

#### Class Participation (Individual)

Every student is expected to attend and participate in class. We will use Slido to facilitate class polls and other assessments, in addition to open ended discussion questions. Each week, a small number of project teams will be ‘on call’ and may be called on to answer questions or offer their thoughts on the course material. We will share the on call schedule once the project teams have been finalized.

#### Pre + Post Course Assessments (Individual)

**Pre-course assessment due Friday September 13, 11:59 pm**

**Post-course assessment due Wednesday November 20, 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 October 21, 4 pm**

This assignment will involve generating and evaluating content using ChatGPT or a similar large language model and writing a short summary of your explorations. Further details will be provided in Week 4.

#### Group Project (Group)

**Due Wednesday November 27, 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 October 7.**

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)