

# Week 1: The Syllabus (CMPT 419 / 980)

Fall 2025.

Course webpage is linked from Canvas (Canvas is very minimal for now.)

Press Space for next slide

# Course Structure

# Lecture Schedule

- **Monday (1hr)**

Discuss previous week's readings

Introduce new materials

Start lecture content

- **Thursdays (2hr)**

Finish lecture (if needed)

Discussion

Class activity and/or lab time (1hr)

Occasionally: Quizzes

# Communication Policy

- Questions at start/end of each session
- Email response: 2-3 business days
- Prefer in-class questions for:
  - Faster responses
  - Benefit to all students

# Reading Components

# Reading Structure

- **Mandatory Readings:** Core course material, weekly responses required
- **Optional Readings:** Theme-organized, abstracts/intros, based on interests

Note: AI assistance discouraged for reading responses. I prefer bullet points over LLM-generated text!

# Reading Schedule

## Timeline

- Finalized Monday of Week X-1
- Complete by Monday of Week X
- Responses due before class

# Course Modules



# Course Organization

Module	Weeks	Focus
1	1–4	Admin & Human-centred / Data-centred AI Frameworks
2	5–7	Technical work in data valuation, data scaling, and algorithmic collective action
3	8–10	Online platforms, content ecosystems, and data
4	11–13	Frontiers in Data Governance

# Grading Structure

- "Reading Responses" : 10
- "Coding Assignments" : 20
- "Quizzes" : 20
- "Final Project" : 50

# Detailed Version

- 10% reading responses (12 total, drop lowest 2 using Canvas, so each of your top 10 responses effectively is worth 1%)
- 20% coding assignments (4 total; 5/5/5/5, drop lowest 1 using Canvas, so each of your top 3 assignments effectively is worth 6.6.67%)
- 20% quizzes (2 total; 10/10; may adjust scores for difficulty)
- 50% final project (5% project proposal, 45% actual project; must submit a written document and a presentation for both)

# Course FAQs

# Attendance

- Not directly graded
- Participation expected
- (Maybe) Cold-calling for reading responses
- Stay home when sick
- No need to email for absences

# Coding & Tools

FLEXIBLE.

- Python for ML/Data Science
- JavaScript for web programming
- LLM assistance allowed (with caveats)
- Free tools always available
- 4 assignments + 1 project

# Group Work Policy

- Some opportunities available. Generally **optional**
- Contribution statements required
- Code/writing review mandatory
- Guidelines per assignment

# Assessment Structure

- Likely TWO quizzes
- No midterm/final
- In-class format
- Make-up options available
- Based on lectures & readings



# AI Tool Usage

**Allowed with:** Logs, attribution, documentation

**Example citations:**

- Produced by model XYZ
- Generated with ChatGPT
- AI-assisted, heavily edited

# Example

I generated this deck by first manually writing a draft of the syllabus as `syllabus.md`, then asked Claude to style it.

## Agenda (if time)

- More details about modules
- Loose prerequisites & crash course
- Finding research papers
- Managing references & notes

# Module 1: Intro

## Goals:

- Exposure to human/data-centric AI frameworks
- Learn frameworks aimed at researchers/designers
- Key question:

When would a human- or data-centric approach change your AI product?

# Frameworks We'll See

- Human-Centered ML (Chancellor)
- DataPerf (data-centric AI)
- HCAI (Schneiderman)
- Value Sensitive Algorithm Design (Zhu)
- FairML
- "Public AI" (some of my work!)
- More to come!

# Module 1 Learning Goals

- Identify similarities/differences between approaches
- Apply concepts to scenarios (e.g., product design)

## Module 2: Technical work in data valuation, data scaling, and algorithmic collective action

- Data influence: Effect of each observation
- Scaling: Predict performance vs. dataset size
- Core idea: Reason about **data counterfactuals**. And social implications!

# Data Valuation Key Reading

Hammoudeh & Lowd (2024) Training data influence analysis and estimation: A survey



# Data Scaling Key Readings

Hestness et al. (2017) Deep learning scaling is predictable

Villalobos et al. (2024) Will we run out of data?

# Why Focus on Data?

- Some HCAI work doesn't focus on data
- Legislation, licensing, behavior change
- We'll focus more on data levers (everyone is a data creator!)

# Module 2 Learning Goals

- Explain influence calculation & applications
- Describe scaling patterns
- Interpret scaling plots
- New forms of collective action

# Module 3: Platforms & Content

- Peer production (Wikipedia, Linux)
- "User-generated content"
- Quantitative reasoning about "content ecosystems"

# Module 4: Frontiers in Data Governance

- Overlap: Governing, markets, ecosystems
- Protocols for dataset maintenance
- Opt-in/out policies
- Platform governance
- Public interest AI

# Voting & Markets

- Vote with/for data
- Market conditions change counterfactuals

If time:

- In-person and/or Google doc "interest formation" network