Ethics with NLP in a nutshell

Course Instructor: Yunze Xiao (yunzex@andrew.cmu.edu)

August 30, 2023

1 Course Description

The course offers a condensed yet impactful journey into the synergy between ethics and advanced technology. Designed with a balanced workload in mind, this course provides a streamlined exploration of the transformative potential of large language models in understanding complex social dynamics. Through concise theoretical modules and hands-on coding exercises, students will delve into the realm of natural language processing, machine learning, and their ethical implication of researches. By focusing on key concepts and practical skills, this course ensures that students can grasp the essentials without being overwhelmed. Ultimately, participants will emerge with a solid foundation in utilizing tools like ChatGPT to gain insights from textual data, paving the way for them to make meaningful contributions at the intersection of technology and societal understanding.

2 Required Text

Officially None. Ask the instructors for supplemental readings, such as:

- Introduction to Computatinal Social Science by Claudio Cioffi-Revilla
- Can Large Language Models Transform Computational Social Science? by Caleb et al.
- NLP ethics course offered by the main campus

3 Logistics

3.1 Prerequisite Knowledge

Preferably 15-112.

3.2 Key topics covered in the course

NLP, Large language Models, Ethics,

4 Attendence policy

The course follows up on the StuCo policy. Student will be warned after one unexcused absense and will result in imediate failing of the course after two or more unexcused absenece.

5 Course Outline

- Introduction to large language models
- Philosophical foundations, history:
  - medical, psychological experiments,
  - IRB
  - human subjects
- Objectivity and Bias
- Privacy, Profiling, Security
- NLP for Social Good
- Intellectual Property

6 Learning outcome

- Articulate the ethical implications of large language models and their historical context.
- Apply ethical principles to navigate challenges related to human subjects and experimentation in NLP.
- Critically assess objectivity and bias in NLP outputs, ensuring fairness and inclusivity.
- Formulate strategies to address privacy concerns, profiling, and security risks in NLP applications.
- Propose NLP-based solutions for social good while considering potential ethical dilemmas.
- Demonstrate an understanding of intellectual property considerations in the context of NLP innovations.

7 Course Grading

1. **Participation** (45%)
   (a) The course aside from the first two weeks are heavily discussion based. We would expect students to talk at least once per class.
2. **Attendence** (10%)
   (a) Attendence are mandatory. Unexcused Absense will caused a 5% deduction on the total grade.
3. **Group Presentation**(45%)
   (a) We will invite professors from different departments to listen to students' 5-10 minute of presentation on a comtemporary issue and propose their own solution using what they have learnt on class.
   (b) The professors will give a score of 1-10, the average of the score will be your presentation score.