This course examines sources of and mitigation frameworks for social bias in technology (with a special focus on generative AI). We examine social bias in tech in two ways:

1. by examining structural (e.g. cultural, social and institutional) factors underlying the low-levels of diversity in decision-making roles in technology as well as technology governance
2. by examining the unequal social impact of technology in deployment

Through readings, students will gain familiarity with a wide range of previously identified structural challenges for achieving equitable representation in tech and fair outcomes when technology is integrated into social institutions.

The focus of the course will be on identifying leadership opportunities and concrete strategies for making positive changes in tech communities (both inside and outside the classroom) as well as in the way that technology is deployed, used, monitored and governed.

Fall 2024 Special Research Focus

In view of the roll-out of the [EU AI Act](#) (the world's first horizontal and standalone law governing AI) on August 1st 2024, this semester, we will take a special focus on connecting policy to technical research. Specifically, we will survey frameworks for discovering and quantifying social bias in ML/AI systems and explore ways that these technical tools can support enforcement of AI regulations. We will anchor our research to concrete goals and principles of the AI Act.

General Information

This is a seminar course, wherein students are expected to shape and direct discussions, **facilitated** by the teaching staff. This course has a service and leadership component: *Students enrolled in this course are expected to serve on the [Graduate Student Committee](#) (GAC)*, wherein they will work on implementing the strategies they developed in this course to positively impact diversity & inclusion in the broader community, as well as concretely shape conversations & actions around responsible AI usage and governance.

There will be weekly reading assignments consisting of 2-3 articles, short weekly reading/discussion responses and weekly discussion based class meetings. There will also be a semester-long research project.

Course Project

This course will include a project -- a white-paper connecting some aspects of technical research on social bias of AI systems to concrete policy recommendations for AI governance. Details to come.

Reading and Discussion Responses

For each set of weekly readings, students are asked to write a short response (approximately 3-4 paragraphs, or more if anyone so wishes) **synthesizing** and **critically engaging** with the reading materials. That is, we are **not** looking for a summary of the contents of the readings, rather we are asking for each student to engage with the ideas presented in the readings personally, intellectually and otherwise. For each class meeting students are asked to write a short reflection on their in-class discussion.

The responses are graded for effort: did the student make an effort to read the articles thoughtfully and carefully? Did the student reflect on and evaluate the ideas presented in the readings and the in-class discussions? Did the student relate the ideas presented in the readings/discussions to their own experiences or the experiences of their peers?

Each week's readings should be done and responses submitted prior to the class meeting. Discussion responses should be submitted after the in-class discussion and prior to the next class meeting.

Attendance and Participation

Students are expected to attend every class meeting and participate actively. Meaningful and constructive dialogue is crucial for this class and this requires a degree of mutual respect, willingness to listen, and tolerance of opposing points of view. Respect for individual differences and alternative viewpoints will be maintained at all times in this class. One's words and use of language should be temperate and

students are encouraged to proactively build-up the ideas of others. Simultaneously, real change (emotional, intellectual and social) does not happen without effort and discomfort, as such, students are also encouraged to challenge the ideas of their peers and their own beliefs.

Grading:

- *Reading & Discussion Responses:* 25%
- *Participation:* 25%
- *GAC Service:* 25%
- *Project: 25%*

Respect for Diversity

It is the mission of the teaching staff that students from all diverse backgrounds and perspectives be well served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. We aim to create a learning environment that is inclusive and respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions for how to better our classroom community are always encouraged and appreciated.