Attention!

This is a representative syllabus.

The syllabus for the course you are enrolled in will likely be different.

Please refer to your instructor's syllabus for more information on specific requirements for a given semester.

LING 3803: Ethics of Language Technology

Instructor: Micha Elsner (elsner.14@osu.edu)
Room: 309 Campbell Hall
Time: 2:20-3:40pm, Tues/Thurs
Office hours: Mon. 2pm, Weds. 11am or by appointment, 222E Oxley Hall or on Zoom

Computer systems make up an increasingly important component which mediates between humans and the environments in which they live. The environments we will consider are workplaces, marketplaces and social networks; computer systems are now major parts of the infrastructure for navigating these environments, representing one's self within them and communicating with other humans. Many such computer systems interact with humans through the medium of language; for instance, search engines read the text of web pages when deciding how to provide information to their users, automated systems for hiring and college admissions read and assess essays, and phones use speech recognition to understand spoken commands. Systems such as these are considered "Language technology".

In this course, we will discuss interactions between humans and their environment mediated by language technology and their social implications for the lived experiences of those who interact with them. We will see that some language technology systems can have a negative impact on human lives. For instance, a college admissions system might discriminate against essays about the experiences of Black students, or a speech recognition system on a phone might misunderstand someone because they speak with an accent. The "ethics of language technology" is the social and philosophical quest to limit the negative and encourage the positive impacts of technological systems on human life and behavior. We will explore various perspectives on how these impacts come about, who is responsible for them and what can be done by decision-makers to lead to more ethical outcomes.

We will discuss philosophical and cultural attitudes, beliefs and values about how to be fair and equitable, and explore their relevance to the complex design process which creates and transforms technological systems. In a series of workshops, we will explore real language technology systems and try to understand their social and cultural impact on human life.

General education theme: Lived Environments

This course is part of the lived environments theme. The environments we will consider are workplaces, marketplaces and social networks. Human interactions with these environments are increasingly mediated and influenced by language technology: that is, by complex, data-driven statistical systems which make decisions about us based on the language we use.

As part of this theme, we will fulfill a variety of goals.

GOAL 1: Successful students will explore a range of perspectives on the interactions and impacts between humans and one or more types of environment (e.g. agricultural, built, cultural, economic, intellectual, natural) in which humans live.

1.1: Engage with the complexity and uncertainty of human-environment interactions.

1.2: Describe examples of human interaction with and impact on environmental change and transformation over time and across space.

GOAL 2: Successful students will analyze a variety of perceptions, representations and/or discourses about environments and humans within them.

2.1: Analyze how humans' interactions with their environments shape or have shaped attitudes, beliefs, values and behaviors.

2.2: Describe how humans perceive and represent the environments with which they interact.

2.3: Analyze and critique conventions, theories, and ideologies that influence discourses around environment.

Throughout the course, students will learn about the social circumstances of how complex language technologies are designed, built and deployed and how they function to shape modern life, work and social interaction. Many different stakeholders determine how these technologies function in their environments, including engineers, corporate officers, professional organizations such as the Association for Computing Machinery, data contributors and user communities. The different ideologies and values of these actors mean that human/environment interactions reflect multiple competing design imperatives, raising ethical problems about how to make systems reflect a consistent set of values and which ethical values they should uphold. Students will study the various ethical frameworks in which language technology and its social impact has been discussed.

Because most of the assignments are reflective, requiring you to discuss the readings and apply the concepts within at various levels, **you will engage with all these learning goals in each unit**, and your **reaction posts and class contributions are expected to address each of these topics** although the particular goals that are most applicable will depend on the specific readings for the given class.

Assignments and grading:

Much of your workload in this course will be spent reading. Readings for most classes will be between 10 and 40 pages. You are expected to do the reading before the day it is due.

Your **discussion points** (a few sentences to a paragraph) will be shared with the class via a Carmen discussion board, as an indication of what you'd like to focus on in class discussion.

The course is divided into five units. Each unit will begin with a **workshop** in which you and your classmates explore a piece of language technology in class. During the unit, there will be a combination of **lectures** and **discussions**. After each workshop, you will write a short (~2 page) **workshop report** on what you found, giving examples of the behavior of the system, explaining whether they represent potential ethical problems, and speculating about why they happen. You will use the data presented in class, but you will write up your opinions on your own.

You are expected to **participate** in the class, by attending class regularly and punctually and speaking up during discussions. I expect to assign you full marks for participation, but if you plan to be absent for a large number of class periods, you must contact me ASAP, and by the end of the term, I should remember you making useful contributions during class at least a few times!

Each unit will end with a **point/counterpoint discussion** in which a group of students lead a discussion on how to design a more ethical version of the system discussed in the unit. The group is responsible for applying the ideas of the scholars discussed in the unit to the problem at hand, explaining what different answers they would give, and leading a discussion on which one is better.

Finally, you will write up a **brief** (~8 pages) arguing for a specific solution to the design question raised in one of the units. You will respond to the various arguments raised by the readings and in the class discussions. You may choose which unit to do the brief on, but it may not be the same one in which your group lead the point/counterpoint discussion. The brief is due at the end of class (during finals week).

Assignment values:

Assignment type:	**How many:**	**Each one worth:**	**Total worth:**
Discussion posts	22	1.5	33
Class participation	1	7	7
Workshop reports	5	6	30
Lead point/counterpoint discussion	1	15	15

Brief	1	15	15
			100

Grading scale: The course uses the standard OSU grading scale.

Course format: The course meets in-person, twice a week.

Required materials: There is no textbook for this course. Readings will be made available via Carmen.

Expected conduct: This class deals with sensitive topics, including racism and sexism. Some readings will come with content warnings; if the content of a reading is likely to be problematic for you, contact the instructor. You are expected to write and speak about these topics in a mature and responsible manner. In particular, we will not insult or denigrate each other, or the scholars whose work we read. A more detailed code of conduct will be provided to you on the first day of class.

Date	Class topic	Read before class	Due today
Unit 0: Whose language? Whose ethics? Whose technology?			
Jan 11 T	Course intro: Language technology mediates interactions with lived environment; what is ethics?	-	-
13 R	The social infrastructure around language technology; who creates it and contributes to its behavior?	Noble “Algorithms of oppression”, ch 2	react/disc 1 **Code of conduct**
18 T	Meta-ethics: How conventions, theories and ideologies of ethics shape arguments	White “Getting good results vs doing the right thing”; reading TBA	react/disc 2 **Point / counterpoint group preferences**
Unit 1: Allocative harms: He goes to Harvard, she goes to prison			
20 R	**Workshop 1:** looking for bias in Google search results	Angwin “Machine Bias”	react/disc 3
25 T	What is a model? Complexity and uncertainty in basic machine learning	O’Neill “Weapons of Math Destruction”, ch. 1, plus the catalog of evils in Dwork “Fairness	react/disc 4

		Through Awareness”	
27 R	The problem of induction: why do humans and machines learn stereotypes?	Berk et al “Fairness in Criminal Justice Risk Assessments: The State of the Art”	react/disc 5 **Workshop 1 report**
Feb 1 T	How the increasing complexity of learning technology over time is transforming how humans interact with their environments	Binns “On the Apparent Conflict Between Individual and Group Fairness”	react/disc 6
3 R	**Point / counterpoint:** How/whether to design a college admissions assistant?		
Unit 2: Censorship: Free speech, hate speech and speech communities			
10 T	**Workshop 2:** the language ideology of the Perspective comment toxicity system	“One of Europe’s Largest Gaming Platforms is Tackling Toxicity with Machine Learning”, Blue “Google’s comment-ranking system will be a hit with the alt-right”	react/disc 7
15 R	Human perception of others via the language they use	Mill “On Liberty”, ch. 2	react/disc 8
17 T	Speech on the internet and its impact on society	Syed “Real talk”	react/disc 9 **Workshop 2 report**
22 R	Theoretical and ideological perspectives on free speech: liberalism and post-liberalism	Sap et al “Annotators with Attitudes: How Annotator Beliefs And Identities Bias Toxic Language Detection"	react/disc 10
24 T	**Point / counterpoint:** How/whether to design an ethical comment filter?		
Unit 3: Representational harms: Does Google think “Mexican” is an insult?			

March 1 T	**Workshop 3:** assessing the worldview of word embeddings	Speer “How to make a racist AI”	react/disc 11
3 R	Word embeddings as representations of lived environments	Crawford “The trouble with bias”	react/disc 12
8 T	Intersectionality: the complexity of representing social identities	Crenshaw “Mapping the margins”	react/disc 13 **Workshop 3 report**
10 R	**Spring break**		
15 T	**Spring break**		
17 R	Proposals for debiasing word embeddings: what do they do and how well do they work?	Bolukbasi et al “Man is to Computer Programmer as Woman is to Homemaker?”, Gonen et al “Lipstick on a pig”	react/disc 14
22 T	**Point / counterpoint:** How/whether to debias word embeddings?		
Unit 4: Privacy: Big Brother is reading your twitter			
24 R	**Workshop 4:** how much does your phone know about you?	Schneier “Data and Goliath” ch. 3, 8	react/disc 15
29 T	The panopticon: a theory of the importance of privacy	Foucalt “Discipline and Punish” ch. 3	react/disc 16
31 R	Differential privacy: a technological proposal, the values that shaped it and the consequences for human-environment interactions	Wood et al “Differential privacy: a primer for a non-technical audience”	react/disc 17 **Workshop 4 report**
April 5 T	Privacy as a legal right: a social proposal, the values that shaped it and the consequences for human-environment interactions	Blanchette et al “Data retention and the panoptic society: The social benefits of forgetfulness”	react/disc 18
7 R	**Point / counterpoint:**		

	How/whether to protect ourselves from surveillance?		
Unit 5: Dual-use technologies: Are we enabling "fake news" and should we stop?			
12 T	**Workshop 5:** how convincing is machine-generated fake news and propaganda?	Vincent "AI researchers debate the ethics of sharing potentially harmful programs"	react/disc 19
14 R	Pretrained language models: an uncertain future of threats and opportunities	Ehni "Dual use and the ethical responsibility of scientists"	react/disc 20
19 T	Release statements for language models: values and behaviors in the modern community of language technologists	McGuffie et al "The Radicalization Risks of GPT-3 and Advanced Neural Language Models"	react/disc 21 **Workshop 5 report**
21 R	End-of-semester wrap-up	Leins et al "Give me convenience and give her death"; ACL ethics checklist	react/disc 22
	Point / counterpoint: How/whether to work on dual-use technologies?		
	End of class		
			Brief

See Carmen for a full list of sources, including additional resources for further reading.

Remaining required material:

Academic misconduct: It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct http://studentlife.osu.edu/csc/.

Disability services: The University strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on your disability (including mental health, chronic or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: slds@osu.edu; 614-292-3307; slds.osu.edu; 098 Baker Hall, 113 W. 12th Avenue.

Mental health: As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a student's ability to participate in daily activities. The Ohio State University offers services to assist you with addressing these and other concerns you may be experiencing. If you or someone you know are suffering from any of the aforementioned conditions, you can learn more about the broad range of confidential mental health services available on campus via the Office of Student Life's Counseling and Consultation Service (CCS) by visiting ccs.osu.edu or calling 614-292-5766. CCS is located on the 4th Floor of the Younkin Success Center and 10th Floor of Lincoln Tower. You can reach an on call counselor when CCS is closed at 614-292-5766 and 24 hour emergency help is also available through the 24/7 National Suicide Prevention Hotline at 1-800-273-TALK or at suicidepreventionlifeline.org.

Sexual harassment: Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories (e.g., race). If you or someone you know has been sexually harassed or assaulted, you may find the appropriate resources at http://titleix.osu.edu or by contacting the Ohio State Title IX Coordinator at titleix@osu.edu

Diversity: The Ohio State University affirms the importance and value of diversity in the student body. Our programs and curricula reflect our multicultural society and global economy and seek to provide opportunities for students to learn more about persons who are different from them. We are committed to maintaining a community that recognizes and values the inherent worth and dignity of every person; fosters sensitivity, understanding, and mutual respect among each member of our community; and encourages each individual to strive to reach his or her own

potential. Discrimination against any individual based upon protected status, which is defined as age, color, disability, gender identity or expression, national origin, race, religion, sex, sexual orientation, or veteran status, is prohibited.