Term Information

Effective Term	Autumn 2022
Previous Value	*Spring 2021*

Course Change Information

What change is being proposed? (If more than one, what changes are being proposed?)

We are proposing that this course be included in the upcoming new GE within the category "Lived Environments"

What is the rationale for the proposed change(s)?

Natural language processing systems are increasingly important to workplaces, marketplaces

and social networks, all of which are important environments that shape our lives.

What are the programmatic implications of the proposed change(s)?

(e.g. program requirements to be added or removed, changes to be made in available resources, effect on other programs that use the course)?

None

Is approval of the requrest contingent upon the approval of other course or curricular program request? No

Is this a request to withdraw the course? No

General Information

Course Bulletin Listing/Subject Area	Linguistics
Fiscal Unit/Academic Org	Linguistics - D0566
College/Academic Group	Arts and Sciences
Level/Career	Undergraduate
Course Number/Catalog	3803
Course Title	Ethics of Language Technology
Transcript Abbreviation	Ethics Language
Course Description	Students will learn about how language processing systems are created, and at what steps in the process bias and unfairness might creep in. They will learn about efforts to define, detect and quantify bias, and how different ethical principles can lead to different results. Finally, students will discuss different ways to remedy the ethical problems of language technology.
Semester Credit Hours/Units	Fixed: 3

Offering Information

Length Of Course	14 Week, 12 Week, 8 Week, 7 Week, 6 Week, 4 Week
Flexibly Scheduled Course	Never
Does any section of this course have a distance education component?	No
Grading Basis	Letter Grade
Repeatable	No
Course Components	Lecture
Grade Roster Component	Lecture
Credit Available by Exam	No
Admission Condition Course	No
Off Campus	Never
Campus of Offering	Columbus

Prerequisites and Exclusions

Prerequisites/Corequisites	
Exclusions	
Electronically Enforced	No

Cross-Listings

Cross-Listings

Subject/CIP Code

Subject/CIP Code	16.0102
Subsidy Level	Baccalaureate Course
Intended Rank	Freshman, Sophomore, Junior, Senior

Requirement/Elective Designation

Lived Environments
The course is an elective (for this or other units) or is a service course for other units

Previous Value

The course is an elective (for this or other units) or is a service course for other units

Course Details

Course goals or learning objectives/outcomes

- Students will recognize and be able to describe the potential harms which can be caused by AI and language technology.
- Students will be able to discuss language as a key component of social systems and point out effects of language ideology on the collection and annotation of language datasets.
- Students will have a high-level understanding of the technical / statistical framework used for modern speech and language technology, and how aspects of this framework can lead to harmful consequences.
- Students will understand the ethical frameworks in which language technology has been discussed, be familiar with their analyses of existing ethical dilemmas, and be able to apply them to practical case studies.
- Students will be aware of current proposals for "ethical NLP" (on both technical and societal levels) and arguments for and against them.

Content Topic List

- Natural Language Processing
- Statistical Learning
- Artificial Intelligence
- Speech and Language Technology
- Ethics of Speech and Language Technology

Sought Concurrence No

Previous Value **Yes**

Attachments

- ge ethics syllabus.pdf: syllabus

  (Syllabus. Owner: McGory,Julia Tevis)

- ge ethics justification.pdf: justification

  (Other Supporting Documentation. Owner: McGory,Julia Tevis)

Comments

- We sought concurrence when the course was initially developed. I was prompted to include one here, but it seems unnecessary. Please let me know if you have any questions. McGory.1@osu.edu *(by McGory,Julia Tevis on 08/06/2021 11:13 AM)*

Workflow Information

Status	User(s)	Date/Time	Step
Submitted	McGory,Julia Tevis	08/06/2021 11:15 AM	Submitted for Approval
Approved	McGory,Julia Tevis	08/06/2021 11:15 AM	Unit Approval
Approved	Vankeerbergen,Bernadette Chantal	08/17/2021 04:23 PM	College Approval
Pending Approval	Cody,Emily Kathryn Jenkins,Mary Ellen Bigler Hanlin,Deborah Kay Hilty,Michael Vankeerbergen,Bernadette Chantal Steele,Rachel Lea	08/17/2021 04:23 PM	ASCCAO Approval

LING 3803: Ethics of Language Technology

Rapid increases in the capabilities of Natural Language Processing (NLP) systems and other language technologies are leading us toward a world in which computers make many of the decisions which affect our everyday lives. NLP systems are already involved in hiring workers, filtering our words online and deciding how political campaigns choose to approach us. These systems have immense power--- but all too often, they make unfair decisions that reflect or even amplify the biases of the society that created them.

In this course, we'll learn about how language processing systems are created, and at what steps in the process bias and unfairness might creep in. We'll learn about efforts to define, detect and quantify bias, and how different ethical principles can lead to different results. Finally, we will discuss different ways to remedy the ethical problems of language technology, to what extent they can be 'fixed', and whether there are problems for which it is too dangerous to use NLP at all.

This course is intended for upper-level students from multiple disciplines, and does not require any specific background in linguistics, mathematics, programming or philosophy. This course is for you if:

- You are a linguist who wants to learn how language ideologies can embed themselves within language technology
- You want to work on language technologies yourself (within linguistics or another data analysis subfield) and want to make sure you do so responsibly
- You come from a cultural studies or philosophy background and want to equip yourself to critique and report on language technology
- You plan to live in our society, and want to understand how language technology will affect you!

Discussions about ethics in the language processing community also tend to draw on parallel issues in other data science areas. While our main focus will be on language, we will also draw on research and writing from these areas (including criminal justice, machine vision and statistical analysis) where it is most relevant.

The conversation on NLP ethics draws from a variety of communities and perspectives, and I believe it is important to represent all of them in the course. We will read some foundational works in ethics and philosophy, some discussions of language technology and its effects by critics and observers outside the field, and some proposals by NLP practitioners themselves. This does mean that some of our readings will be technical papers involving mathematics--- but the focus will be on high-level understanding of what is going on, not on the details or implementation. Each technical reading will come with a study guide intended to make it accessible to students from all backgrounds.

General education theme: Lived Environments:

Natural language processing systems are increasingly important to workplaces, marketplaces and social networks, all of which are important environments that shape our lives.

As part of this theme, we will:
1.1: Engage with the complexity and uncertainty of human-environment interactions, focusing on the inherent difficulties in understanding the behavior of complex, data-driven statistical systems.
1.2: Describe examples of human interaction with and impact on environmental change and transformation over time and across space. We will learn about potential and actual harms and benefits caused by language technology.
2.1: Analyze how humans' interactions with their environments shape or have shaped attitudes, beliefs, values and behaviors. We will study the various ethical frameworks in which language technology has been discussed, become familiar with their analyses of existing ethical dilemmas, and apply them to practical case studies.
2.2: Describe how humans perceive and represent the environments with which they interact. We will consider languages, and attitudes towards them, as a component of social systems, and discuss the effects of language ideology on the collection and annotation of language datasets
2.3: Analyze and critique conventions, theories, and ideologies that influence discourses around environment. We will read current proposals for "ethical NLP" (on both technical and societal levels) and arguments for and against them; you will form and present your own arguments on their merits.

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 type of assignment**, 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. After each reading, you will complete two small assignments.

Your **reaction** (three sentences) will be sent to the instructor, and will indicate:

- How hard you felt the reading was
- How much you felt you learned from it
- Whether you liked it

This is intended to calibrate the syllabus for future revisions of the class.

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:**
Reaction posts	22	1/2	11
Discussion posts	22	1	22
Class participation	1	7	7
Workshop reports	5	6	30
Lead point/counterpoint discussion	1	15	15
Brief	1	15	15
			100

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

Required materials: “Weapons of Math Destruction”, by Cathy O’Neill, which should cost about $14 for a new copy. (Try to replace?)

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? Big questions: to whom are we responsible, and for what? Technical concepts: social architecture of an NLP project Ethical concepts: deontological vs utilitarian ethics Linguistic concepts: disciplinary standards for research ethics in linguistics Case study: search results			
A 24	Course intro, practical ethics	-	-
26	NLP in social context // class discussion of reading	Noble “Algorithms of oppression”, ch 2 (44 pgs)	react/disc 1 **Code of conduct**
31	Applying philosophy to real life // class discussion of reading	White “Getting good results vs doing the right thing”; Leidner et al “Ethical by Design: Ethics Best Practices for Natural Language Processing”	react/disc 2 **Point / counterpoint group preferences**
Unit 1: Allocative harms: He goes to Harvard, she goes to prison Big questions: what is “fair” decision-making and how do we know if we’re doing it? Technical concepts: supervised learning, models, objectives, true and false positives Ethical concepts: rights of groups vs individuals Linguistic concepts: none in this unit Case studies: sentencing guidelines, academic assessment			
S 2	**Workshop 1:** Google search	Aguera et al “Physiognomy’s New	react/disc 3

		Clothes", Angwin "Machine Bias"	
7	Basics of supervised learning	O'Neill "Weapons of Math Destruction", ch. 1 (17 pgs), plus the catalog of evils in Dwork "Fairness Through Awareness" (1 pg)	react/disc 4
9	Base rates, sources of error // class discussion of reading	Berk et al "Fairness in Criminal Justice Risk Assessments: The State of the Art" (42 pgs)	react/disc 5 **Workshop 1 report**
14	Different approaches to fairness // class discussion of reading	Binns "On the Apparent Conflict Between Individual and Group Fairness" (11 pgs)	react/disc 6
16	**Point / counterpoint:** How/whether to design an ethical sentencing assistant?		
Unit 2: Censorship: Free speech, hate speech and speech communities Big questions: should social media be censored? Can we trust NLP as the censor? Technical concepts: bias and variance, annotator versus dataset bias Ethical concepts: ethics of free speech Linguistic concepts: language varieties, language ideology, slurs Case studies: abusive language, pornography detection			
21	**Workshop 2: abusive language detection**	Matsakis "Tumblr's Porn-Detecting AI Has One Job—and It's Bad at It"	react/disc 7
23	Abusive language and language ideology // class discussion of reading	Mill "On Liberty", ch. 2	react/disc 8
28	Technical background for statistical language learning	Syed "Real talk" (21 pgs)	react/disc 9 **Workshop 2 report**

30	Liberalism // class discussion of reading	Sap et al “The risk of racial bias in hate speech detection” (9 pgs)	react/disc 10
O 5	**Point / counterpoint:** How/whether to design an ethical comment filter?		
Unit 3: Representational harms: Does Google think “Mexican” is an insult? Big questions: what is “representational harm” and who suffers from it? Technical concepts: unsupervised learning, word embeddings Ethical concepts: intersectionality Linguistic concepts: distributional semantics Case studies: word embedding spaces			
7	**Workshop 3: word embeddings**	Larson et al “Breaking the black box”, Speer “How to make a racist AI”	react/disc 11
12	Word embeddings: how and why	Crawford “The trouble with bias” (40 mins)	react/disc 12
14	**Fall break**		
19	Debiasing	Bolukbasi et al “Man is to Computer Programmer as Woman is to Homemaker?”, Gonen et al “Lipstick on a pig”	react/disc 13 **Workshop 3 report**
21	Intersectionality // class discussion of reading	Crenshaw “Mapping the margins”	react/disc 14
26	**Point / counterpoint:** How/whether to debias word embeddings?		
Unit 4: Privacy: Big Brother is reading your twitter Big questions: is privacy important? If so, how should we protect ourselves? Technical concepts: data mining, differential privacy Ethical concepts: the panopticon, the right to be forgotten			

Linguistic concepts: collection and annotation of language corpora, language and identity Case studies: targeted advertising			
28	**Workshop 4: targeted advertisements**	Angwin et al “Facebook Enabled Advertisers to Reach ‘Jew Haters’	react/disc 15
N 2	Language and identity // class discussion of reading	O’Neil “Weapons of Math Destruction” ch 10 (19 pgs); possibly a short selection from Foucalt “Discipline and Punish”	react/disc 16
4	Privacy and research practice	Wood et al “Differential privacy: a primer for a non-technical audience”	react/disc 17 **Workshop 4 report**
9	Rights-based approaches to privacy // class discussion of reading	Blanchette et al “Data retention and the panoptic society: The social benefits of forgetfulness”	react/disc 18
11	**Veterans day**		
16	**Point / counterpoint:** How/whether to protect ourselves from surveillance?		
Unit 5: Dual-use technologies: Are we enabling “fake news” and should we stop? Big question: is it ethical to work on dual-use technology? How can it be controlled? Technical concepts: “deep fakes”, targeted propaganda Ethical concepts: dual-use technology Linguistic concepts: language modeling Case studies: GPT2, face recognition			
18	**Workshop 5: GPT2**	Crawford “Halt the use of facial recognition”; Vincent “AI researchers debate the ethics of sharing potentially harmful programs”	react/disc 19

23	Pretrained language models: theory and hype	Zellers et al “Defending against neural fake news”	react/disc 20
25	**Thanksgiving**		
30	Dual-use technology // class discussion of reading	Leins et al “Give me convenience and give her death”	react/disc 21 **Workshop 5 report**
2	Ethical proposals // class discussion of reading	Ehni “Dual use and the ethical responsibility of scientists”	react/disc 22
7	**Point / counterpoint:** How/whether to work on dual-use technologies?		
	End of class		
			Brief

Sources:

Aguera y Arcas, Blaise et al. “Physiognomy’s New Clothes” https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a 2017.
Angwin, Julia et al. “Facebook Enabled Advertisers to Reach ‘Jew Haters’” https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters 2017.
Berk, Richard, et al. "Fairness in criminal justice risk assessments: The state of the art." Sociological Methods & Research (2018): 0049124118782533.
Binns, Reuben. "On the apparent conflict between individual and group fairness." Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. 2020.
Blanchette, Jean-François, and Deborah G. Johnson. "Data retention and the panoptic society: The social benefits of forgetfulness." The Information Society 18.1 (2002): 33-45.
Bolukbasi, Tolga, et al. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." Advances in neural information processing systems. 2016.
Crawford, Kate. “The trouble with bias.” https://www.youtube.com/watch?v=fMym_BKWQzk NIPS 2017 keynote.
Crawford, Kate. “Halt the use of facial-recognition technology until it is regulated”, Nature, Aug 27. https://www.nature.com/articles/d41586-019-02514-7
Crenshaw, Kimberle. "Mapping the margins: Intersectionality, identity politics, and violence against women of color." Stan. L. Rev. 43 (1990): 1241.
Dwork, Cynthia, et al. "Fairness through awareness." Proceedings of the 3rd innovations in theoretical computer science conference. 2012.

Ehni, Hans-Jörg. "Dual use and the ethical responsibility of scientists." Archivum immunologiae et therapiae experimentalis 56.3 (2008): 147.
Gonen, Hila, and Yoav Goldberg. "Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them." arXiv preprint arXiv:1903.03862 (2019).
Larson, Jeff et al. “How Machines Learn to Be Racist” https://www.propublica.org/article/breaking-the-black-box-how-machines-learn-to-be-racist?word=Trump 2016.
Leidner, Jochen L., and Vassilis Plachouras. "Ethical by design: Ethics best practices for natural language processing." Proceedings of the First ACL Workshop on Ethics in Natural Language Processing. 2017.
Leins, Kobi, Lau, Jey Han, and Timothy Baldwin. "Give Me Convenience and Give Her Death: Who Should Decide What Uses of NLP are Appropriate, and on What Basis?." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.
Matsakis, Louise. “Tumblr's Porn-Detecting AI Has One Job—and It's Bad at It.” https://www.wired.com/story/tumblr-porn-ai-adult-content/ 2018.
Mill, John Stuart. "On liberty." A selection of his works. Palgrave, London, 1966. 1-147.
Noble, Safiya Umoja. “Algorithms of oppression.” New York University, 2018.
O’Neill, Cathy. “Weapons of math destruction.” Broadway Books, 2016.
Sap, Maarten, et al. "The risk of racial bias in hate speech detection." Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019.
Speer, Robyn. “How to make a racist AI without really trying” http://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/ 2017.
Syed, Nabiha. "Real talk about fake news: towards a better theory for platform governance." Yale LJF 127 (2017): 337.
White, Mark. “Getting good results vs doing the right thing.” https://www.learnliberty.org/blog/getting-good-results-vs-doing-the-right-thing/ 2016.
Wood, Alexandra, et al. "Differential privacy: A primer for a non-technical audience." Vand. J. Ent. & Tech. L. 21 (2018): 209.
Vincent, James. “AI researchers debate the ethics of sharing potentially harmful programs.” https://www.theverge.com/2019/2/21/18234500/ai-ethics-debate-researchers-harmful-programs-openai 2019.
Zellers, Rowan, et al. "Defending against neural fake news." Advances in Neural Information Processing Systems. 2019.

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.

GE rationale: LING 3803 (Ethics of Language Technology)
Theme: Lived Environments

https://oaa.osu.edu/sites/default/files/uploads/general-education-review/new-ge/submission-lived-environments.pdf

GOAL 1: Successful students will analyze an important topic or idea at a more advanced and in-depth level than the foundations. Please briefly identify the ways in which this course represents an advanced study of the focal theme. In this context, “advanced” refers to courses that are e.g., synthetic, rely on research or cutting-edge findings, or deeply engage with the subject matter, among other possibilities. (50-500 words)

This course relies on close reading of primary sources in multiple disciplines. Papers such as Sap et al “The risk of racial bias in hate speech detection” (2019) and Gonen et al “ "Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them” (2019) are drawn from the recent research literature in computational linguistics. Selections such as Foucault’s “Discipline and Punish” and Mill’s “On Liberty” are foundational works of ethics and philosophy. Crenshaw’s “Mapping the Margins” introduces the concept of intersectionality.

Students are expected not only to read these sources closely, but to synthesize across the different ways of thinking and writing that they represent, enabling them to apply philosophical concepts to computational dilemmas.

ELO 1.1 Engage in critical and logical thinking about the topic or idea of the theme. Please link this ELO to the course goals and topics and indicate specific activities/assignments through which it will be met. (50-700 words)

Students will engage in critical thinking through posting reactions to the readings and engaging in discussion. Class discussion is scheduled for approximately two classes per unit. Students will also be responsible for leading a point/counterpoint discussion in which they summarize various ethical approaches to the main questions of the unit, and for writing an 8-page brief arguing for a specific ethical proposal related to one of the units. Completing these assignments will require students to consider the points of view they have encountered in the reading and evaluate their logical and ethical argumentation, then construct arguments of their own in dialogue with those of their sources.

ELO 1.2 Engage in an advanced, in-depth, scholarly exploration of the topic or idea of the theme. Please link this ELO to the course goals and topics and indicate specific activities/assignments through which it will be met. (50-700 words)

The written brief is expected to represent an in-depth, scholarly discussion of a particular issue. Students will supplement the class readings with additional sources. They are expected to respond to issues raised in the readings and class discussions with original, well-written

argumentation. In arguing for a specific solution to a design question, they will have to apply general philosophical ideas which they have learned about, but do so in an original way.

GOAL 2: Successful students will integrate approaches to the theme by making connections to out-of-classroom experiences with academic knowledge or across disciplines and/or to work they have done in previous classes and that they anticipate doing in future.

ELO 2.1 Identify, describe, and synthesize approaches or experiences as they apply to the theme. Please link this ELO to the course goals and topics and indicate specific activities/assignments through which it will be met.
(50-700 words)

The workshop assignments which begin each unit ask students to work hands-on with a piece of real-world language technology, such as a search engine or text generation system, evaluate it from an ethical standpoint and consider how the underlying technology was designed to create or defuse potential problems. These activities should teach the students new ways to consider other technological systems they have or will encounter in daily life.

ELO 2.2 Demonstrate a developing sense of self as a learner through reflection, self-assessment, and creative work, building on prior experiences to respond to new and challenging contexts. Please link this ELO to the course goals and topics and indicate specific activities/assignments through which it will be met. (50-700 words)

Several assignments offer the opportunity for reflection and self-assessment. The structure of posting a reading comment, then discussing the reading in class, then applying lessons from the reading in the point/counterpoint discussion is intended to allow students to formulate more sophisticated understanding of the material by learning from their classmates. Similarly, the structure of in-class workshop followed by written workshop report will allow students to take time to reconsider their first impressions and reach more nuanced conclusions. The brief, which focuses more deeply on a unit of the student's choice, also offers a chance to recapitulate and improve upon earlier ideas.

Specific Expectations of Courses in Lived Environments

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.

ELO 1.1 Engage with the complexity and uncertainty of human-environment interactions. Please link this ELO to the course goals and topics and indicate specific activities/assignments through which it will be met.
(50-700 words)

The course will focus on the inherent difficulties in understanding the behavior of complex, data-driven statistical systems. Readings and lectures will spell out how machine learning works (basics of supervised learning, base rate effects, bias and variance, etc.), how they make errors, and why it is so difficult to understand how language processing systems will behave on any particular example. This will cross over with readings which highlight the complexities of human identities (intersectionality, language ideology and various approaches to privacy). Students should learn that even determining whether a system is fair, or understanding the perspectives of different stakeholder communities on a system's fairness, is a challenging task.

ELO 1.2 Describe examples of human interaction with and impact on environmental change and transformation over time and across space. Please link this ELO to the course goals and topics and indicate specific activities/assignments through which it will be met. (50-700 words)

Students will learn about actual and potential harms caused by language technology, as well as some cases in which ethical critiques led to successful and unsuccessful system redesigns. The popular-press readings beginning each unit focus on particular cases where ethical issues with language technology became big news stories. There are also academic readings which focus on particular systems (e.g. Noble's critique of Google search and Angwin's of Facebook's advertisement service). When possible, these are paired with workshops which test the same or a similar system in a hands-on way (Workshops 1 and 4). We will evaluate whether these critiques seem to have had an impact on the current design of these systems.

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

ELO 2.1 Analyze how humans' interactions with their environments shape or have shaped attitudes, beliefs, values and behaviors. Please link this ELO to the course goals and topics and indicate specific activities/assignments through which it will be met. (50-700 words)

Thefield of AI/language technology ethics represents a reaction to the increasing power of technological systems in our lives. We will see how this field has drawn on pre-existing philosophical frameworks to create systematic proposals for "ethical technology". Such proposals often conflict, due to the different values or approaches taken by their proponents. In the point/counterpoint discussions and the brief, we will apply these abstract ideas to concrete case studies.

ELO 2.2 Describe how humans perceive and represent the environments with which they interact. Please link this ELO to the course goals and topics and indicate specific activities/assignments through which it will be
met. (50-700 words)

Many of the readings discuss particular cases of human interaction with technology, for instance the articles by Julia Angwin about the COMPAS sentencing system and Facebook's advertisements, Crawford's article about face recognition, etc. In addition, there are lectures and

readings focusing on language as an aspect of human identity, which are intended to show how the kinds of language used by people and technological systems can position them in different cultural ways. Sap's article on racial bias in hate speech, for instance, shows that some hate speech detection systems are more inclined to mark messages written in African-American English as abusive, regardless of their content. This kind of bias links human perception of their environment (language ideology) with system design (hate speech detection), creating feedback from the environment that reinforces the original ideology.

ELO 2.3 Analyze and critique conventions, theories, and ideologies that influence discourses around environments. Please link this ELO to the course goals and topics and indicate specific activities/assignments through which it will be met. (50-700 words)

The course will discuss different approaches to the problem of "ethical AI/language technology" in each unit. For instance, in the privacy unit, students will consider a technological solution (differential privacy) and a legal solution (the right to be forgotten), contrasting their different notions of what privacy is and how to accomplish it. These solutions, rooted in different communities and different ideological priors, assume very different things about how privacy protects people and whose job it should be to ensure it. Other units also juxtapose different points of view in the same way.

The point/counterpoint discussion at the end of each unit is intended to give students a forum to discuss the conflicting assumptions and consequences of these various approaches. The group leading the discussion is asked to explain the different answers each of their sources might offer to the topic under discussion and guide the class in structuring arguments for and against each one.