

Term Information

Effective Term Spring 2023
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)?

Computer systems that interact with humans through the medium of language are called "language technology". Language technology plays an increasingly important role in mediating between humans and the environments in which they live. The lived environments this course considers are workplaces, marketplaces, and social networks. The course examines interactions between humans and these lived environments, how that interaction is mediated by language technology, and the role of language technology in ways that these interactions shape both the lived environments and people's attitudes, beliefs, and behaviors.

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 request 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 Lang Tech
Previous Value 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

COURSE CHANGE REQUEST
3803 - Status: PENDING

Last Updated: Vankeerbergen, Bernadette
Chantal
09/28/2022

Credit Available by Exam	No
Admission Condition Course	No
Off Campus	Never
Campus of Offering	Columbus, Lima, Mansfield, Marion, Newark, Wooster
<i>Previous Value</i>	<i>Columbus</i>

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 be able to discuss ethical arguments from multiple disciplines and relate them to questions that arise about language technologies.
- Students will be able to write a sustained argumentative piece on the ethics of some piece of language technology, and the way the language technology affects a lived environment (workplace, marketplace, or social network).
- Students will be able to assess the ideology and potential biases of a piece of language technology, and the ways in which these ideologies and biases lead the technology to affect human-lived environment interactions, by experimenting with it.
- Students will understand the tradeoffs and social and technical factors which make AI systems complex and give rise to unpredictable interactions between humans and their lived environments.
- Students will be able to give examples of social harms caused by language technology.
- Students will be able to relate criticisms of language technology to the philosophical tradition of ethics.
- Students will develop an understanding of the ways in which language ideology interacts with language technology systems to create or reinforce perceptions about users of different kinds of language.

Previous Value

- Students will be able to compare and critique different ideas for how to achieve ethical language technology.
- *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
- Ethics
- Statistical Learning
- Artificial Intelligence
- Speech and Language Technology
- Ethics of Speech and Language Technology

Previous Value

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

Sought Concurrence

Yes

Attachments

- ge ethics syllabus (2).pdf: UPDATED syllabus
(Syllabus. Owner: McGory, Julia Tevis)
- ge ethics syllabus (previous submission).pdf: PREVIOUS syllabus
(Syllabus. Owner: McGory, Julia Tevis)
- ge ethics justification.pdf: GE Justification
(Other Supporting Documentation. Owner: McGory, Julia Tevis)
- Screen Shot 2020-10-26 at 4.38.00 PM.png: Concurrence-Statistics/Data Analytics
(Concurrence. Owner: McGory, Julia Tevis)
- Screen Shot 2020-10-26 at 4.37.34 PM.png: Concurrence-CSE
(Concurrence. Owner: McGory, Julia Tevis)
- BA_Major_Program_Curricular_Map_10-26-2020.pdf: Curricular Map
(Other Supporting Documentation. Owner: McGory, Julia Tevis)
- Concurrence Philosophy.png: Concurrence-Philosophy
(Concurrence. Owner: McGory, Julia Tevis)
- LING 3803 ethics of lang technology -- LE syllabus 7-7-22.pdf: UPDATED syllabus 7-7-22
(Syllabus. Owner: Sims, Andrea Dorothy)
- LING 3803 ethics of lang technoogy -- LE justification 7-7-22.pdf: UPDATED GE Justification 7-7-22
(Other Supporting Documentation. Owner: Sims, Andrea Dorothy)
- LING 3803 cover letter 7-7-22.pdf: Cover Letter 7-7-22
(Cover Letter. Owner: Sims, Andrea Dorothy)

Comments

- This is a resubmission of a proposal to add Ling 3803 to the Lived Environments GE category. The proposal (syllabus, GE justification) has been revised in response to Panel feedback received 4/8/22. A cover letter has also been included. *(by Sims, Andrea Dorothy on 07/07/2022 04:25 PM)*
- Please see Panel feedback email sent 04/08/2022. *(by Hilty, Michael on 04/08/2022 02:25 PM)*
- I (Julia McGory) accidentally cancelled the previous request for edits, and am no longer able to modify this document. This is a re-submission of a proposal to add Ling3803 to the "Lived Environments" GE. Included attachments are (1) the 'previous syllabus' which required edits based on the committee's comments, (2) an 'updated syllable' containing new edits, (3) GE Justification also included in the previous course request. *(by McGory, Julia Tevis on 02/01/2022 09:56 AM)*

COURSE CHANGE REQUEST
3803 - Status: PENDING

Last Updated: Vankeerbergen, Bernadette
Chantal
09/28/2022

Workflow Information

Status	User(s)	Date/Time	Step
Submitted	McGory, Julia Tevis	02/01/2022 09:57 AM	Submitted for Approval
Approved	McGory, Julia Tevis	02/01/2022 09:57 AM	Unit Approval
Approved	Vankeerbergen, Bernadette Chantal	03/02/2022 02:44 PM	College Approval
Revision Requested	Hilty, Michael	04/08/2022 02:25 PM	ASCCAO Approval
Submitted	Sims, Andrea Dorothy	07/07/2022 04:27 PM	Submitted for Approval
Approved	Sims, Andrea Dorothy	07/07/2022 04:27 PM	Unit Approval
Approved	Vankeerbergen, Bernadette Chantal	09/28/2022 11:45 AM	College Approval
Pending Approval	Cody, Emily Kathryn Jenkins, Mary Ellen Bigler Hanlin, Deborah Kay Hilty, Michael Vankeerbergen, Bernadette Chantal Steele, Rachel Lea	09/28/2022 11:45 AM	ASCCAO Approval

July 7, 2022

Dear Themes Panel of ASCC,

We have revised the materials for Ling 3803 ("Ethics of language technology") in accord with your recommendations received April 8 and are resubmitting the course for the Lived Environments theme.

Comments Received from the Panel and Our Responses to the Feedback:

- Comment: The reviewing faculty see how this course could potentially fit within the GE Theme: Lived Environments but the submitted revised material did not appear to present clear human-environment interactions that will be the focus of the course (and associated readings, discussions, and assignments). They would like to see these human-environment interactions explored more clearly within the proposal, as currently it appears as if the course is exploring how algorithmic processes affect one's lived reality, and how structural inequalities can be exacerbated or mitigated through automated decision-making processes. If this is the case, they ask that how these shape the lived environments of people be brought into the forefront of the course and the course description. The first page of the syllabus should present the framework for this course as a LE course more explicitly.
 - **Response:** The course description on the first page of the syllabus now introduces the course more clearly with respect to the Lived Environments theme criteria, by explicitly defining the environments that the course will consider, explaining how humans interact with these environments using language technology, and highlighting the course's focus on human lives. This is reflected from the first sentences: "Computer systems make up an increasingly important component which mediates between humans and the environments in which they live. The lived environments we will consider are workplaces, marketplaces and social networks; computer systems are now major parts of the infrastructure for navigating these lived environments, representing one's self within them and communicating with other humans."
 - **Response:** Class topics on the syllabus's day-by-day course schedule have been edited to clarify their relevance to the Lived Environments theme -- specifically, how language technology mediating interactions between humans and workplaces/marketplaces/social networks shapes these lived environments and people's experience with them. Short descriptions (blue text) have been added to help students (and the Themes committee) understand how a particular topic connects to the Goals and ELOs of the Lived Environments theme as well as the course goals. For example: Day 2: "The social infrastructure around language technology; who creates it and contributes to its behavior? **Students learn how humans transform their lived environment by contributing to the design of language technology systems**".
 - **Response:** The GE justification statements have been expanded with additional discussion, with the goal of bringing to the forefront ways in which language technology mediating between humans and different lived environments (workplaces, marketplaces, social networks) shapes these environments and the ways in which humans interact with them. The GE justification has also added pointers (in red) to where specific topics and assignments appear on the course schedule.
- Comment: The reviewing faculty request that the Goals for the GE Theme: Lived Environments be added and that the Goals, ELOs, and a brief statement explaining how the course will meet the GE ELOs for the General Theme ELOs be added. The Goals and ELOs can be found on the Office of Academic Affairs at: <https://oaa.osu.edu/ohio-state-ge-program>
 - **Response:** The syllabus now includes (p. 2) the Goals, ELOs for the Lived Environments theme, and a paragraph-length explanation of how the course will fulfill the ELOs.

- Comment: The reviewing faculty request that a cover letter be provided that explains all changes made in response to the feedback provided.
 - **Response:** This is it!

Additional changes:

- Some material extraneous to the Lived Environments theme has been removed from the syllabus.

Sincerely,

Micha Elsner (Linguistics)

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 lived environments we will consider are workplaces, marketplaces and social networks; computer systems are now major parts of the infrastructure for navigating these lived 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 lived 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 lived environments we will consider are workplaces, marketplaces and social networks. Human interactions with these lived 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 lived 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/lived 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 about the relationship between humans and their lived environment 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, such that interactions between humans and their lived environments can be more equitable. 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 relating to the interaction between humans and their lived environments 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	1	15	15

discussion

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? <i>Students learn how language technology is relevant to human/lived environment interactions.</i>	-	-
13 R	The social infrastructure around language technology; who creates it and contributes to its behavior? <i>Students learn how humans transform their lived environment by contributing to the design of language technology systems</i>	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 <i>Students learn how to frame specific arguments about lived environments in terms of larger theories, conventions and ideologies</i>	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			

<p>20 R</p>	<p>Workshop 1: looking for bias in Google search results Students analyze the social impact of interacting with the internet (as a source of facts, a marketplace for products or a place to meet other users) via the medium of language technology as exemplified by Google Search, considering the values and ideologies reflected in what sites are seen as most important or valuable to a searcher</p>	<p>Angwin “Machine Bias”</p>	<p>react/disc 3</p>
<p>25 T</p>	<p>What is a model? Complexity and uncertainty in basic machine learning Students learn some basic technological terms and concepts which are necessary to understand the contributions of language technology to human/lived environment interactions later on in the course; these lessons are framed in the context of a human/lived environment interaction (deciding who to admit to college) based on a simple statistical model</p>	<p>O’Neill “Weapons of Math Destruction”, ch. 1, plus the catalog of evils in Dwork “Fairness Through Awareness”</p>	<p>react/disc 4</p>
<p>27 R</p>	<p>The problem of induction: why do humans and machines learn stereotypes? Students learn how the tendency to stereotype arises in human perception and representation of the social environment, and reflect on how these concepts also apply to representations formed with technological assistance</p>	<p>Berk et al “Fairness in Criminal Justice Risk Assessments: The State of the Art”</p>	<p>react/disc 5 Workshop 1 report</p>
<p>Feb 1 T</p>	<p>How the increasing complexity of learning technology over time is transforming how humans interact with their lived environments Students learn how human interaction with the lived environment, mediated by language technology, has changed over time as new technological systems have been developed</p>	<p>Binns “On the Apparent Conflict Between Individual and Group Fairness”</p>	<p>react/disc 6</p>
<p>3 R</p>	<p>Point / counterpoint: How/whether to design a college admissions assistant? Students discuss an interaction between humans and their lived environment (college , considered as a workplace) mediated by language technology (admissions assistant programs) and how these programs can be ethically designed to reflect human values, attitudes and beliefs</p>		

Unit 2: Censorship: Free speech, hate speech and speech communities			
10 T	<p>Workshop 2: the language ideology of the Perspective comment toxicity system</p> <p>Students analyze interactions between humans and social media platforms as mediated by systems for labeling certain kinds of language as toxic or problematic, and consider the perceptions and ideologies about language which motivate these labels</p>	<p>“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”</p>	react/disc 7
15 R	<p>Human perception of others via the language they use</p> <p>Students reflect on how humans perceive and represent other humans or computer systems based on language usage, and learn some linguistic theories of how ideologies and conventions labeling language as “good” or “bad” can develop</p>	<p>Mill “On Liberty”, ch. 2</p>	react/disc 8
17 T	<p>Speech on the internet and its impact on society</p> <p>Students learn how interactions between humans and workplaces/social spaces are mediated by the technological infrastructure of the internet, how humans have shaped the internet over time, and how it has affected human beliefs and values</p>	<p>Syed “Real talk”</p>	<p>react/disc 9</p> <p>Workshop 2 report</p>
22 R	<p>Theoretical and ideological perspectives on free speech: liberalism and post-liberalism</p> <p>Students learn how liberal and post-liberal ideas about free speech have shaped the technological architecture of the internet and thus the human/lived environment interactions which it mediates; students learn that differing perspectives on liberalism have led to complex and hard-to-settle debates on how the internet should be regulated</p>	<p>Sap et al “Annotators with Attitudes: How Annotator Beliefs And Identities Bias Toxic Language Detection”</p>	react/disc 10
24 T	<p>Point / counterpoint:</p> <p>How/whether to design an ethical comment filter?</p> <p>Students discuss an interaction between humans and their lived environment (social media) mediated by systems for filtering problematic or toxic language; students</p>		

	argue for different perspectives on how such systems ought to work, informed by differing theories and ideologies of free speech		
Unit 3: Representational harms: Does Google think “Mexican” is an insult?			
March 1 T	Workshop 3: assessing the worldview of word embeddings Students analyze interactions between humans and various lived environments mediated by technologically created representations of those lived environments, by assessing what true or misleading information those representations contain and how these may impact the interactions	Speer “How to make a racist AI”	react/disc 11
3 R	Word embeddings as representations of lived environments Students learn about how technologically mediated representations of language are created, and discuss how the complex process of forming these representations can inadvertently incorporate undesired stereotypes which can affect human beliefs and values about social environments	Crawford “The trouble with bias”	react/disc 12
8 T	Intersectionality: the complexity of representing social identities Students discuss the complex nature of social identities which tend to incorporate elements of many different social categories; students reflect on how this makes it difficult to represent a person’s identity in a technological system, and the consequences for human/lived environment interactions mediated by these technologies	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? Students discuss some proposals for humans to transform interactions with the lived environment as mediated by word representation technology by altering this technology in accord with particular ideologically defined goals; students consider the degree to which the complexity of human identities makes achieving these ideological goals difficult	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	<p>Point / counterpoint: How/whether to debias word embeddings?</p> <p>Students discuss how word representation systems can or should be altered to affect human/lived environment interactions</p>		
Unit 4: Privacy: Big Brother is reading your twitter			
24 R	<p>Workshop 4: how much does your phone know about you?</p> <p>Students explore the ways in which predictive text, a kind of language technology, forms a representation of their personalities and lives, and discuss the ways in which this representation, if made available to social or marketplace contacts, can affect their interactions with their lived environment</p>	Schneier “Data and Goliath” ch. 3, 8	react/disc 15
29 T	<p>The panopticon: a theory of the importance of privacy</p> <p>Students learn about an influential theory which has affected discourses around how lived environments such as marketplaces ought to be designed</p>	Foucault “Discipline and Punish” ch. 3	react/disc 16
31 R	<p>Differential privacy: a technological proposal, the values that shaped it and the consequences for human-environment interactions</p> <p>Students learn about a technical proposal for protecting privacy, understand how it reflects particular values related to the importance of privacy, and consider how its use affects real and potential human/lived environment interactions such as the US census</p>	Wood et al “Differential privacy: a primer for a non-technical audience”	react/disc 17 Workshop 4 report
April 5 T	<p>Privacy as a legal right: a social proposal, the values that shaped it and the consequences for human-environment interactions</p> <p>Students learn about some legal proposals for protecting privacy, understand how they reflect particular values related to the importance of privacy, and consider how their use affects real and potential human/lived environment interactions such as the data which can be collected by companies doing business in Europe</p>	Blanchette et al “Data retention and the panoptic society: The social benefits of forgetfulness”	react/disc 18

7 R	<p>Point / counterpoint: How/whether to protect ourselves from surveillance?</p> <p>Students discuss how technological surveillance alters their interactions with the social environment and make proposals for technological and social changes which would lead to more equitable outcomes, appealing to the theoretical and ideological perspectives on privacy from earlier in the unit</p>		
Unit 5: Dual-use technologies: Are we enabling “fake news” and should we stop?			
12 T	<p>Workshop 5: how convincing is machine-generated fake news and propaganda?</p> <p>Students explore some computer programs which can generate fluent text and consider the likelihood that these programs could be used to generate misinformation or extremist content. Students reflect on how the ability to generate high-quality fake news without effort could affect trust in human/lived environment interactions involving information transfer.</p>	Vincent “AI researchers debate the ethics of sharing potentially harmful programs”	react/disc 19
14 R	<p>Pretrained language models: an uncertain future of threats and opportunities</p> <p>Students learn how technologies for generating fluent text have changed in recent years, the impact this has had on human/lived environment interactions, and some potential changes which are expected in the near future. Students discuss the difficulty in predicting the impact of new or emerging technologies on human/lived environment interactions.</p>	Ehni “Dual use and the ethical responsibility of scientists”	react/disc 20
19 T	<p>Release statements for language models: values and behaviors in the modern community of language technologists</p> <p>Students learn how the emergence of new language technologies has affected the computer science community, motivating it to develop new ethical guidelines and restrictions; students discuss the values and ideologies which motivate these guidelines and how they impact the continued deployment of new technologies with the potential for changing the way humans interact with their lived environment</p>	McGuffie et al "The Radicalization Risks of GPT-3 and Advanced Neural Language Models"	react/disc 21 Workshop 5 report
21 R	Point / counterpoint:	Leins et al “Give me	react/disc 22

	How/whether to work on dual-use technologies? Students discuss how new technologies ought to be developed and released in order to minimize the likelihood of detrimental effects on human/lived environment interactions, arguing for their positions in terms of the theories and conventions covered earlier in the unit	convenience and give her death”; ACL ethics checklist	
	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.

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)

Throughout the course, students will learn how language technology mediates between humans and environments (workplaces, markets and social media networks) in which they live their lives. For instance, when searching for information online, most people use a search engine, which is a kind of language technology--- search engines can find valuable information, but they might also direct users to biased or extremist content, or fail to find minority viewpoints. By reading articles on the social impact of these kinds of technological systems, and then reacting, discussing and debating their ideas, students will think critically about how they impact human-environment interactions.

The ethical issues surrounding these human-environment interactions have come to the forefront of multiple disciplines in recent years. The course engages with the full complexity of debates over how to design and deploy language technology systems in an ethical way by asking students to perform close reading of primary sources from different communities and different points of view. Papers such as Sap et al “Annotators with attitudes: How annotator beliefs and identities bias toxic language detection” (2021; [Unit 2, day 4](#)) and Gonen et al “Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them” (2019; [Unit 3, day 4](#)) are drawn from the recent research literature in computational linguistics. Selections such as Foucault’s “Discipline and Punish” ([Unit 4, day 2](#)) and Mill’s “On Liberty” ([Unit 2, day 2](#)) are foundational works of ethics and philosophy. Crenshaw’s “Mapping the Margins” ([Unit 3, day 3](#)), written in the legal academic community, introduces the concept of intersectionality.

Many of the case studies are drawn from technological advances which have appeared in recent years. For instance, beginning in 2019, companies which were designing “language models”--- computer systems which can generate fluent text on a variety of topics--- were warned that such models could be used to generate fake news and propaganda articles with more variety and convincing detail than human propagandists, leading to risks if they were made available to the public. Because such language models are also a foundational technology for cutting-edge search engines and translation systems, writers were divided on how widely accessible they ought to be (see [Unit 5, day 1 and 3](#)). New recommendations for safe AI publications are still being adopted by professional organizations; for instance, students will read the new Association for Computational Linguistics publication checklist ([Unit 5, day 4](#)).

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 about human interactions with their environment as mediated by language technology through posting reactions to the readings and engaging in discussion on this topic. Class discussion is scheduled for each class in the 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, which in each case is focused on a particular human-environment interaction and the technology which affects it (for instance, **Unit 1, day 5**, on how college admissions are affected by technological recommendations for which students to admit) 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. The brief is required to outline a social issue related to a particular human-environment interaction, explain the relevance of language technology to the issue, and point out an ethical problem, then argue for a particular solution. Students are expected to supplement the class readings with additional sources in framing their description of the interaction and its social implications. 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 in human-environment interactions. These activities should teach the students new ways to consider other technological systems they have or will encounter in daily life.

For instance, in the first workshop, students will examine potential biases in Google search responses. By synthesizing criticism of search results from the readings (especially “Algorithms of Oppression” by Safiya Noble, [Unit 0 day 1](#)) with their own experience as internet users, students will design an individualized set of questions to investigate, for instance, how Google search results characterize poverty, or how Google responds to queries about radical leftist versus rightist political groups. Such questions analyze the social impact of interacting with the internet (as a source of facts, a marketplace for products or a place to meet other users) via the medium of language technology, which inevitably colors such interactions with its own point of view. In the second workshop, students will analyze what sorts of language are determined to be “toxic” by the Perspective comment moderation API (see [Unit 2, day 1 and 4](#)), and synthesize their findings with the literature on social attitudes towards the language typically used by African-Americans and other minoritized groups ([Unit 2, lecture 2](#)). Students will submit a writeup for each workshop which describes the potential social and cultural impact of the technology being investigated.

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)

Throughout the course, students will learn about the social circumstances of how complex language technologies are designed, built and deployed, and how the different stakeholder groups involved consider their potential impact on human-environment interactions. For instance, students will study the decision-making process through which the OpenAI Foundation released a new language generation program capable of generating extremely convincing fake news articles and propaganda for extremist viewpoints (Unit 5). Some groups warned that this technology was too dangerous to release because of the threat that it might overwhelm social media networks with hate speech or misleading information, spreading mistrust and apprehension among internet users. Others argued that public benefit of these programs, which can also be used to make human-computer interaction more flexible and accessible by allowing computers to produce more naturalistic language, outweighed the risk.

Students will learn about the many different stakeholders that interact to make such decisions, including engineers, corporate officers, professional organizations such as the Association for Computing Machinery, data contributors and user communities. They will also learn about the statistical and technological complexity that makes it challenging to understand the potential behaviors of the system itself and how it will mediate interactions between humans and their lived environment. This will include some insight into modeling technology and its history (Unit 1) and the trend towards increasing complexity and decreasing transparency in language generation systems (Unit 5).

Students will understand how social power dynamics can lead to particular stakeholders being ignored or under-served, and how the complexities of system design can obscure biases in systems or datasets (Units 2 and 3). This will cross over with readings which highlight the complexities of human identities (especially the concept of intersectionality in Unit 3, but also language ideology and various approaches to privacy in Units 2 and 4).

For instance, the paper “Scaling fair learning to hundreds of intersectional groups” (Zhao, Huang, Liu, Yu, Liu, Russakovsky, Anandkumar 2022, covered in Unit 3, lecture 3) explains how search technology can under-represent the intersectional complexity of a community. Searches for “programmers” on the internet mostly represent white, male programmers; the computer can be instructed to rebalance these results to increase gender diversity, but when doing so, the number of dark-skinned programmers represented decreases. Similarly, balancing to represent the racial diversity of programmers leads to under-representation of women. When such technology mediates between a community of users and their environment (potentially including job searches, educational sites for computer scientists, and other information on how to become a programmer), the result is an interaction which is biased against particular intersectionally defined groups, such as dark-skinned females. Students will learn to analyze the impact of such biases, how they arise and how or whether they can be controlled.

Students will 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. Some results on the difficulty of assessing fairness are given in Berk et al "Fairness in criminal justice" (Unit 1, day 3). For instance, it is mathematically impossible to ensure that a system is fair in the sense of being equally accurate for members of two different groups, and also fair in the sense that when it makes errors, it is equally likely to assign members of both groups to undesirable outcomes. The choice between these two criteria, therefore, requires complex reasoning about what sorts of outcomes are genuinely fair (in a social sense) which cannot be resolved mathematically, a point made in Binns's article "The apparent conflict between individual and group fairness" (Unit 1, day 4).

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 problems caused by allowing language technology to mediate between humans and their environment, 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 became big news stories. There are also academic readings which focus on particular philosophical perspectives on how these interactions are shaped by different technological frameworks (such as Noble's "Algorithms of Oppression"; Unit 0, day 2, Foucault's concept of the Panopticon; Unit 4, day 2, and McGuffie's article on radicalization; Unit 5, day 3). When possible, these are paired with workshops which test the same or a similar system in a hands-on way (Workshops 1 and 4).

For instance, Foucault argues that the ability to constantly monitor the behavior of a prisoner, worker or student imposes constraints on their behavior, even when they cannot know for certain whether they are being observed at any particular moment. This argument, originally made with respect to visual surveillance, has been extended to the long-term retention and searchability via language technology of posts on social media networks like Facebook and Twitter. Because such posts can potentially be used to discredit a person years after they are made, writers such as Alice Marwick argue that many social media users impose extremely harsh standards of self-censorship (research covered in Unit 2, day 3 and Unit 4, day 2).

On the other hand, students will see that humans shape their ability to interact with the environment they live in by designing and contributing to language technologies. For instance, students will see that Google's Perspective comment filtering tool reflects the ideology that mainstream American English sounds more polite than African-American English (via Sap et al. "Annotators with Attitudes"; Unit 2, day 4), and this in turn determines which users are permitted to speak and be heard on a variety of social media platforms, including video gaming platforms and comment threads for Washington Post articles. A simple system for evaluating restaurant reviews on Yelp learns a socially motivated bias against Mexicans (Speer "How to make a racist AI"; Unit 3, day 1), which can have a substantial impact on the market for restaurant food and

which dining establishments thrive or fail. Students will analyze this feedback loop between human culture and technology; in the workshops, students will evaluate how critiques have impacted the current design of language technology systems and how language technology systems affect our own decision-making and lifestyle.

Several parts of the course take a historical perspective on the evolution of the lived environment as different technological pieces were added or removed. For instance, in **Unit 2**, students will learn some history of speech and speech regulation on the internet, beginning with early communities like Usenet which were created by free-speech absolutists and ending with the modern, corporatized infrastructure of social media. Students will discuss how the different values and incentives of the stakeholders created an environment in which different kinds of users were privileged. In **Unit 4**, the course will take a similar perspective on surveillance, looking at who is surveilled, by whom and for what purposes, drawing on readings by Schneier (**day 1**) and Blanchette et al (**day 4**). **Unit 4** also draws a geographic contrast between American and European legal ideas on privacy, showing that American law tends to prioritize a liberal framework emphasizing freedom of speech and contract while European law places more value on individual dignity and privacy.

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)

The field of AI/language technology ethics represents a reaction to the increasing power of technological systems in our lives. Students 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. For instance, in **Unit 2, day 2**, students will read John Stuart Mill's defense of free speech in *On Liberty* and see how it has been applied to the debate over how to regulate abusive language on social media. In **Unit 4, day 2**, students will learn about Michel Foucault's theory of the panopticon and how it has become a central pillar of modern critiques of technological surveillance.

In each case, however, beliefs and values have also changed to respond to technological innovations. Unrestricted speech on the internet has been blamed for the spread of extremist ideas and fraudulent propaganda. Nabiha Syed's article "Real talk" (**Unit 2, day 3**) suggests that the free speech idealism of the First Amendment might need to be modified in its application to social media. We also discuss the concept of "sousveillance" (**Unit 4, lecture 3**), in which data analysis is used by the less powerful to assess and describe the behavior of the powerful; the ability to do this has emerged as a consequence of the increasing availability of surveillance technology for all levels of society, and complicates Foucault's picture of the panoptic society.

In the point/counterpoint discussions and the brief, students 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)

Language is a major way in which humans perceive and represent environments like workplaces, markets and social networks. For instance, humans use language to assess people socially; the language a person uses can indicate their place of origin, their gender and their race, as well as indicating how friendly, formal or confrontational they are being. Articles covered in lectures in **Units 1 and 4** will discuss how a person's language can be used to predict various features about them. When technological systems form part of the infrastructure of these environments, they may have their own voices (as digital assistants like Siri and Alexa do) or express themselves via automatically generated text such as news articles. They may also decide which human language is allowed to propagate through the environment, as Facebook's algorithmic filters do. This means that human perception of their environment is mediated by these technological systems.

The concepts of language varieties (dialects) and language ideologies (stigma or prejudice towards a dialect) form a major part of **Unit 2**. In this unit, students will discuss how language-based comment filtering systems shape the environment of social media networks like Facebook. In **Unit 3**, students will discuss how perception of language can affect attitudes towards racial groups like Black Americans based on how they speak, and how this can lead to mis-perception and exclusion of Black people from certain environments. 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.

When technological systems generate their own language, this also affects how humans perceive the environments in which they live. For instance, computer systems can be used to flood a chat group with extremist propaganda, creating the false impression that extremist views are common and accepted there. On the other hand, computer systems such as Siri and Alexa adopt particular language styles (such as feminine voices) to represent themselves as helpful and unassuming; this use of language affects how humans perceive them, but also helps to reinforce the socially determined link between femininity and meekness.

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, White’s article “Getting good results vs doing the right thing” (**Unit 0, day 3**) introduces the concept of deontological fairness (treating people in ways that seem inherently equitable) versus consequential fairness (trying to ensure equality of outcomes among different social groups), which students are invited to discuss in their analysis of bias in Google search results (**Unit 1 workshop**) and subsequently throughout the course. In **Unit 2’s workshop**, students will contrast different approaches to the problem of free speech and its relevance to comment filtering technology on the internet. In **Unit 3** students will discuss the theory of intersectionality, which states that social categories like “Black female” are more than the sum of their parts (Crenshaw “Mapping the Margins; **day 3**), and explain how this complicates the task of removing gender bias from systems that try to represent the meanings of words such as “nurse” or “astronomer” (**day 4 and workshop**). In **Unit 4**, students will learn about philosopher Michel Foucault’s concept of the Panopticon, a system in which social conventions are enforced by the absence of privacy (**day 2**), and evaluate various technical and legal proposals for protecting privacy to see if they provide the protections he was advocating for (**days 3 and 4**). In the **workshop**, they will analyze how much their phone seems to know about them personally and whether this surveillance affects their use of texting and other phone-based applications to communicate. In **Unit 5**, students will discuss technologies with the potential to be used for harm, comparing the precautions taken in language technology to those from medicine and physics.

In many cases, solutions rooted in different communities and different ideological priors assume very different things about what an ethical solution is and whose job it should be to ensure it, an example of how differing ideologies influence the discourses around environments. For instance, there is a clear contrast between deontological and consequential ethics, between free speech absolutism and the post-liberal program of limiting hateful or extremist speech via filters, and between technological privacy protections like Differential Privacy and legal approaches like the Right to Be Forgotten (**Unit 5, days 3 and 4**). These contrasts will force students to confront the complexity of ethical arguments in realistic case studies.

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.