

## Term Information

Effective Term Spring 2021

## 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 None  
Exclusions None  
Electronically Enforced No

## Cross-Listings

Cross-Listings None

## Subject/CIP Code

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

## Requirement/Elective Designation

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

Yes

## Attachments

- Screen Shot 2020-10-26 at 4.38.00 PM.png: statistics/data analytics  
*(Concurrence. Owner: McGory,Julia Tevis)*
- Screen Shot 2020-10-26 at 4.37.34 PM.png: CSE  
*(Concurrence. Owner: McGory,Julia Tevis)*
- ethics syllabus (1).pdf: syllabus  
*(Syllabus. Owner: McGory,Julia Tevis)*

## Comments

## Workflow Information

Status	User(s)	Date/Time	Step
Submitted	McGory,Julia Tevis	10/26/2020 04:46 PM	Submitted for Approval
Approved	McGory,Julia Tevis	10/26/2020 04:47 PM	Unit Approval
Approved	Haddad,Deborah Moore	10/26/2020 05:17 PM	College Approval
Pending Approval	Jenkins,Mary Ellen Bigler Hanlin,Deborah Kay Oldroyd,Shelby Quinn Vankeerbergen,Bernadette Chantal	10/26/2020 05:17 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.

**Course goals:** You will...

- Recognize and be able to describe the potential harms which can be caused by AI and language technology.
- 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.
- 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.
- 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.
- Be aware of current proposals for "ethical NLP" (on both technical and societal levels) and arguments for and against them.

### **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:**

<b>Assignment type:</b>	<b>How many:</b>	<b>Each one worth:</b>	<b>Total worth:</b>
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 <b>Code of conduct</b>
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 <b>Point / counterpoint group preferences</b>

### 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	<b>Workshop 1:</b> automated essay scoring	Aguera et al "Physiognomy's New Clothes", Angwin "Machine Bias"	react/disc 3
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:	react/disc 5 <b>Workshop 1 report</b>

		The State of the Art” (42 pgs)	
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	<b>Point / counterpoint:</b> How/whether to design an ethical sentencing assistant?		
<b>Unit 2: Censorship: Free speech, hate speech and speech communities</b>			
<p>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</p>			
21	<b>Workshop 2: abusive language detection</b>	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 <b>Workshop 2 report</b>
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	<b>Point / counterpoint:</b> How/whether to design an ethical comment filter?		
<b>Unit 3: Representational harms: Does Google think “Mexican” is an insult?</b>			
<p>Big questions: what is “representational harm” and who suffers from it?          Technical concepts: unsupervised learning, word embeddings          Ethical concepts: intersectionality          Linguistic concepts: distributional semantics</p>			



Case studies: word embedding spaces			
7	<b>Workshop 3: word embeddings</b>	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	<b>Fall break</b>		
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 <b>Workshop 3 report</b>
21	Intersectionality // class discussion of reading	Crenshaw "Mapping the margins"	react/disc 14
26	<b>Point / counterpoint:</b> How/whether to debias word embeddings?		
<b>Unit 4: Privacy: Big Brother is reading your twitter</b>			
<p>Big questions: is privacy important? If so, how should we protect ourselves?</p> <p>Technical concepts: data mining, differential privacy</p> <p>Ethical concepts: the panopticon, the right to be forgotten</p> <p>Linguistic concepts: collection and annotation of language corpora, language and identity</p> <p>Case studies: targeted advertising</p>			
28	<b>Workshop 4: targeted advertisements</b>	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 Foucault "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 <b>Workshop 4 report</b>
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	<b>Veterans day</b>		
16	<b>Point / counterpoint:</b> How/whether to protect ourselves from surveillance?		
<p><b>Unit 5: Dual-use technologies: Are we enabling “fake news” and should we stop?</b></p> <p>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</p>			
18	<b>Workshop 5: GPT2</b>	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	<b>Thanksgiving</b>		
30	Dual-use technology // class discussion of reading	Leins et al “Give me convenience and give her death”	react/disc 21 <b>Workshop 5 report</b>
2	Ethical proposals // class discussion of reading	Ehni “Dual use and the ethical responsibility of scientists”	react/disc 22

7	<b>Point / counterpoint:</b> How/whether to work on dual-use technologies?		
	<b>End of class</b>		
			<b>Brief</b>

**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](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 therapeuticae 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 L&J 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](mailto:slds@osu.edu); 614-292-3307; [slds.osu.edu](http://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](http://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](http://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](mailto: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.

Teodorescu, Radu

Fri 10/16/2020 2:43 PM

To: McGory, Julia

Cc: Soundarajan, Neelam; Fosler-Lussier, Eric; Arora, Anish



Dear Julia,

The CSE Curriculum Committee discussed this request and agreed to granting **concurrence** to the Department of Linguistics offering **Ethics of Language Technology LING3803**.

Have a good weekend!

--Radu

—  
**Radu Teodorescu**

Associate Professor

Computer Science and Engineering

The Ohio State University

(614) 292-7027

[arch.cse.ohio-state.edu](http://arch.cse.ohio-state.edu)

**From:** Craigmile, Peter <pfc@stat.osu.edu>  
**Sent:** Tuesday, September 29, 2020 1:28 PM  
**To:** McGory, Julia <mcgory.1@osu.edu>  
**Cc:** Parthasarathy, Srinivasan <srini@cse.ohio-state.edu>; Hans, Christopher <hans@stat.osu.edu>  
**Subject:** Re: Linguistics seeking concurrence from Data Analytics

Dear Julia,

The Department of Statistics concurs with the course Ethics of Language Technology, LING3803. We think that this course is timely in highlighting the real-life ethical dilemmas underlying the use of AI and Machine/Statistical learning that our world currently faces. All the best with the course.

Regards,  
Peter

---

Peter Craigmile, Ph.D.,  
Professor, Department of Statistics, The Ohio State University.