

 [datasci-w266](#) / [2022-summer-main](#) Public[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#)[master](#) [2022-summer-main](#) / [syllabus](#) /

...

Mark H Butler Release a4 ...

19 days ago [History](#)

..

 README.md

19 days ago

 README.md

# DATASCI W266: Natural Language Processing With Deep Learning

[Course Overview](#)[Course Evaluation](#)[Final Project](#)[Course Resources](#)[Tentative Schedule and Readings](#)

## Course Overview

**Course Description** Understanding language is fundamental to human interaction. Our brains have evolved language-specific circuitry that helps us learn it very quickly; however, this also means that we have great difficulty explaining how exactly meaning arises from sounds and symbols. This course provides an introduction to natural language processing, linguistic phenomena, and our attempts to analyze them using modern deep learning approaches. We cover a wide range of concepts with a focus on practical applications such as information extraction, machine translation, sentiment analysis, and summarization. The focus of this course prioritizes practical application over theory. Each week, the course is centered on Python notebooks that include exercises and code examples. The asynchronous lectures provide a foundation to build upon in the live sessions.

## Course Prerequisites

- **MIDS 207 (Machine Learning):** We assume you know what gradient descent is. We review simple linear classifiers and softmax at a high level, but make sure you've at least heard of these! You should also be comfortable with linear algebra, which we use for vector representations and when we discuss deep learning.
- **Language:** All assignments are in Python using Jupyter notebooks, Google Colab, NumPy, TensorFlow, and Keras.
- **Time:** There are three to four substantial assignments in this course as well as a term project. Make sure you give yourself enough time to be successful! In particular, you may be in for a rough semester if you have other significant commitments at work or home, or take both this course and any of 210 (Capstone), 261, or 271.

**Course Goals and Objectives** By the completion of this course, students will be able to:

- Understand and describe multiple facets of linguistic phenomena related to natural language processing.
- Describe fundamental concepts, techniques, problems, and modern approaches in the domain of natural language processing (NLP).
- Understand the assumptions, strengths, and limitations of NLP and related deep learning techniques, and make appropriate decisions about application of techniques and solutions to NLP problems.
- Analyze textual data using a number of machine learning and deep learning-based NLP techniques.
- Demonstrate familiarity with and comprehension of existing NLP techniques for solving practical problems.
- Demonstrate an ability to stay current with a constantly evolving field: by developing a familiarity with the neural network architectures that underpin current state of the art, and an ability to seek out and understand new advances published in the field.
- Demonstrate expertise in using existing libraries and tools related to NLP work, along with an ability to familiarize oneself to new libraries and tools that are key to NLP practitioners as the domain evolves.

### Communication and Resources

- Course website: [GitHub datasci-w266/2022-summer-main](https://github.com/datasci-w266/2022-summer-main)
- **Piazza:** We'll use this for Q&A, and this will be the fastest way to reach the course staff. Note that you can post anonymously, or make posts visible only to instructors for private questions.
- Email list for course staff (expect a somewhat slower response here): [mids-nlp-instructors@googlegroups.com](mailto:mids-nlp-instructors@googlegroups.com)

## Live Sessions

- Section 1: Monday 6:30 - 8pm PST (Joachim Rahmfeld)
- Section 2: Tuesday 2 - 3:30pm PST (Natalie Ahn)
- Section 3: Tuesday 4 - 5:30pm PST (Daniel Cer)
- Section 4: Wednesday 4 - 5:30pm PST (Mark Butler)
- Section 5: Wednesday 6:30 - 8pm PST (Mike Tamir/Paul Spiegelhalter)
- Section 6: Thursday 6:30 - 8pm PST (James Kunz)

## Teaching Staff Office Hours

- **Daniel Cer:** Monday at noon PST
- **Joachim Rahmfeld:** Tuesday at 5:30pm PST
- **Mike Tamir/Paul Spiegelhalter:** Wednesday immediately after the live session
- **Natalie Ahn:** Tuesday at 4pm PST
- **James Kunz:** Thursday immediately after his live session
- **Mark Butler:** Friday at 5pm PST
- **Gurdit Chahal:** Wednesday at 2:30pm PST
- **Sandip Panesar:** Saturday at 9am PST

Office hours are for the whole class; students from any section are welcome to attend any of the times above.

## Async Instructors

- James Kunz
- Joachim Rahmfeld
- Mark Butler

## Course Evaluation

---

### Breakdown

Assignment	Topic	Release	Deadline	Weight
Assignment 0	<b>Course Set Up</b> <ul style="list-style-type: none"><li>• GitHub</li><li>• Piazza</li><li>• Google Cloud</li></ul>	May 2	May 8	0%

Assignment 1	Basic Neural Nets	May 7	May 15	5%
Assignment 2	Text Classification	May 21	Jun 5	15%
Assignment 3	Question Answering	Jun 11	Jun 26	15%
Assignment 4	Multimodal NLP	Jun 25	Jul 10	15%
Final Project	<a href="#">Final Project Guidelines</a>		July 30	50%

Your assignment grade report can be found at <https://w266grades.appspot.com>.

## General Grading Philosophy

We curve the numerical grade to a letter grade. While we don't release the curve, it usually results in about a quarter of the class each receiving A, A-, B+, and B. Exceptional cases receive A+, C, or F, as appropriate.

A word of warning: Given that we (effectively) release solutions to some parts of assignments in the form of unit tests, it shouldn't be surprising that most students earn high scores. Since the variance is so low, assignment scores aren't the primary driver of the final letter grade for most students. A good assignment score is necessary but not sufficient for a strong grade in the class. A well-structured, novel project with good analysis is what makes the difference between a high B/B+ and an A-/A.

As mentioned above, this course is a lot of work. Give it the time it deserves and you'll be rewarded intellectually and on your transcript.

## Late Submission Policy

We recognize that sometimes things happen in life outside the course, especially in MIDS where we all have full-time jobs and family responsibilities to attend to. To help with these situations, we are giving you **5 "late days"** to use throughout the term as you see fit. Each late day gives you a 24-hour (or any part thereof) extension to any deliverable in the course **except** the final project presentation or report. (UC Berkeley needs grades submitted very shortly after the end of classes.)

Once you run out of late days, each 24-hour period (or any part thereof) results in a **10 percentage-point deduction** on that deliverable's grade.

You can use a **maximum of 2 late days** on any single deliverable. We will **not be accepting any submissions more than 48 hours past the original due date**, even if you have late days. (We want to be more flexible here, but your fellow students also want their graded assignments back promptly!)

We don't anticipate granting extensions beyond these policies. Plan your time accordingly!

## More Serious Issues

If you run into a more serious issue that will affect your ability to complete the course, please email the instructors mailing list and cc MIDS Student Services. A word of warning: In previous sections, we have had students ask for an Incomplete (INC) grade because their lives were otherwise busy. Mostly we have declined, opting instead for the student to complete the course to the best of their ability and have a grade assigned based on that work. (MIDS prefers to avoid giving INCs, as they have been abused in the past.) The sooner you start this process, the more options we (and the department) have to help. Don't wait until you're suffering from the consequences to tell us what's going on!

## Collaboration Policy/Academic Integrity

All students —undergraduate, graduate, professional full time, part time, law, etc.— must be familiar with and abide by the provisions of the "Student Code of Conduct" including those provisions relating to Academic Misconduct. All forms of academic misconduct, including cheating, fabrication, plagiarism or facilitating academic dishonesty will not be tolerated. The full text of the UC Berkeley Honor Code is available at:

<https://teaching.berkeley.edu/berkeley-honor-code> and the Student Code of Conduct is available at: [https://sa.berkeley.edu/student-code-of-conduct#102.01\\_Academic\\_Misconduct](https://sa.berkeley.edu/student-code-of-conduct#102.01_Academic_Misconduct)

We encourage studying in groups of two to four people. This applies to working on homework, discussing labs and projects, and studying. However, students must always adhere to the UC Berkeley Code of Conduct (<http://sa.berkeley.edu/code-of-conduct>) and the UC Berkeley Honor Code (<https://teaching.berkeley.edu/berkeley-honor-code>). In particular, all materials that are turned in for credit or evaluation must be written solely by the submitting student or group. Similarly, you may consult books, publications, or online resources to help you study. In the end, you must always credit and acknowledge all consulted sources in your submission (including other persons, books, resources, etc.)

## Final Project

---

See the [Final Project Guidelines](#)

## Attendance and Participation

---

We believe in the importance of the social aspects of learning —between students, and between students and instructors— and we recognize that knowledge-building does not solely occur on an individual level, but is built by social activity involving people and by members engaged in the activity. Participation and communication are key aspects of this course vital to the learning experiences of you and your classmates.

Therefore, we like to remind all students of the following requirements for live class sessions:

- Students are required to join live class sessions from a study environment with video turned on and with a headset for clear audio, without background movement or background noise, and with an internet connection suitable for video streaming.
- You are expected to engage in class discussions, breakout room discussions and exercises, and to be present and attentive for your and other teams' in-class presentations.
- Keep your microphone on mute when not talking to avoid background noise. Do your best to minimize distractions in the background video, and ensure that your camera is on while you are engaged in discussions.

That said, in exceptional circumstances, if you are unable to meet in a space with no background movement, or if your connection is poor, make arrangements with your instructor (beforehand if possible) to explain your situation. Sometimes connections and circumstances make turning off video the best option. If this is a recurring issue in your study environment, you are responsible for finding a different environment that will allow you to fully participate in classes, without distraction to your classmates.

**Failure to adhere to these requirements will result in an initial warning from your instructor(s), followed by a possible reduction in grades or a failing grade in the course.**

## Diversity and Inclusion

---

Integrating a diverse set of experiences is important for a more comprehensive understanding of data science. We make an effort to read papers and hear from a diverse group of practitioners. Still, limits exist on this diversity in the field of data science. We acknowledge that it is possible that there may be both overt and covert biases in the material due to the lens through which it was created. We would like to nurture a learning environment that supports a diversity of thoughts, perspectives and experiences, and honors your identities (including race, gender, class, sexuality, religion, ability, veteran status, etc.) in the spirit of the UC Berkeley Principles of Community

<https://diversity.berkeley.edu/principles-community>



To help us accomplish this, please contact us or submit anonymous feedback through I School channels if you have any suggestions to improve the quality of the course. If something was said in class (by anyone) or you experience anything that makes you feel uncomfortable, please talk to your instructors about it. If you feel like your performance in the class is being impacted by experiences outside of class, please don't hesitate to come and talk with us. We want to be a resource for you. Also, anonymous feedback is always an option and may lead us to make a general announcement to the class, if necessary, to address your concerns. As a participant in teamwork and course discussions, you should also strive to honor the diversity of your classmates.

If you prefer to speak with someone outside of the course, the MIDS Academic Director Drew Paulin, the I School Assistant Dean of Academic Programs Catherine Cronquist Browning, and the UC Berkeley Office for Graduate Diversity are excellent resources. Also see the following: <https://www.ischool.berkeley.edu/about/community>.

## Disability Services and Accommodations

---

If you need disability-related accommodations in this class, if you have emergency medical information you wish to share with me, or if you need special arrangements in case the building must be evacuated, please inform me as soon as possible.

The I School recognizes disability in the context of diversity, and the Disabled Students' Program (DSP) equips students with appropriate accommodations and services to remove barriers to educational access. Students seeking accommodations in this class are responsible for completing the DSP application process to obtain an accommodation letter. You may reach the DSP at (510) 642-0518, or visit the website:

<https://dsp.berkeley.edu>

## Publishing Your Work

---

You are highly encouraged to use your program coursework to build an academic/professional portfolio.

- Blog about your coursework (and other ideas) and share on the I School Medium channel
  - Instructions are here on the intranet for students:  
<https://www.ischool.berkeley.edu/intranet/connect>
  - Instructions here are public for alumni:  
<https://www.ischool.berkeley.edu/alumni/stay-connected>

- Publish projects to your I School project portfolio gallery (for more than just the capstone).
- Publish your work on LinkedIn and tag @UC Berkeley School of Information. Do **NOT** publish your homework assignments!
- Publish in academic journals: Contact your professors for assistance. (Note that multiple review iterations are usually required; this can be a time-intensive endeavor.)
  - For help writing professional academic papers students are encouraged to contact Sabrina Soracco, the Director of the Graduate Writing Center, in the Graduate Division -- see <https://grad.berkeley.edu/staff/sabrina-soracco/> and <https://grad.berkeley.edu/professional-development/graduate-writing-center/> -- the latter has links to resource guides, appointments with consultants, workshops, etc.
- Publish your news (e.g., conference talks, awards, scholarships) to the I School internal newsletter.

## Course Resources

---

We are not using any particular textbook for this course. We'll list some relevant readings each week. Here are some general resources:

- [Speech and Language Processing \(3rd edition draft\)](#) by Daniel Jurafsky and James H. Martin - *free online!*
- [Speech and Language Processing \(2nd edition\)](#) by Daniel Jurafsky and James H. Martin - *This edition does not cover neural net approaches*
- [NLTK Book - Natural Language Processing with Python](#) - accompanies NLTK (Natural Language Tool Kit) and includes useful, practical descriptions (with python code) of basic concepts.
- [Deep Learning](#) (Goodfellow, Bengio, and Courville)

We'll be posting materials to the course [GitHub repo](#).

*Note:* This syllabus may be subject to change. We'll be sure to announce anything major on Piazza.

## Code References

The course will be taught in Python, and we'll be making heavy use of NumPy, TensorFlow, Keras, and Jupyter (IPython) notebooks. We'll also be using Git for distributing and submitting materials. If you want to brush up on any of these, we recommend:

- **Git tutorials:** [Introduction / Cheat Sheet](#), or [interactive tutorial](#)



- **Python / NumPy:** Stanford's CS231n has [an excellent tutorial](#).
- **TensorFlow:** We'll go over the basics of TensorFlow and Keras in [Assignment 1](#). [Effective TensorFlow 2](#) is a great reference, ranging from the absolute basics through advanced topics like multi-GPU training, `tf.learn`, and debugging. You can also check out the [tutorials](#) on the TensorFlow website, but these can be somewhat confusing if you're not familiar with the underlying models. Also, look at the [TensorFlow Keras Guide](#) as we will be using Keras in this class.
- **Keras:** The [Keras Site](#) has a variety of useful guides as well as code examples.

## Miscellaneous Deep Learning and NLP References

Here are a few useful resources and papers that don't fit under a particular week -- all optional, but interesting!

- (Optional) [Chris Olah's blog](#) and [Distill](#)
- (Optional) [GloVe: Global Vectors for Word Representation \(Pennington, Socher, and Manning, 2014\)](#)
- (Optional) [Jay Alammar's blog](#)

## Tentative Schedule and Readings

We'll update the table below with assignments as they become available, as well as additional materials throughout the semester. Keep an eye on GitHub for updates!

*Dates are tentative:* Assignments in particular may change topics and dates. (Updated slides for each week will be posted during the live session week.)

**Live Session Slides** [\[available with @berkeley.edu address\]](#)

## Deliverables

Note: we will update this table as we release assignments. Each assignment will be released around the last live session of the week and due approximately 1 week later (for simple assignments) or 2 to 3 weeks later (for complex assignments).

	Topic	Release	Deadline
Assignment 0	<b>Course Set-up</b> <ul style="list-style-type: none"> <li>• GitHub</li> <li>• Piazza</li> </ul>	May 2	May 8

	<ul style="list-style-type: none"> <li>Google Cloud</li> </ul>		
<a href="#">Assignment 1</a>	<b>Assignment 1</b> <ul style="list-style-type: none"> <li>Neural Networks</li> </ul>	May 7	May 15
<a href="#">Assignment 2</a>	<b>Assignment 2</b> <ul style="list-style-type: none"> <li>Text Classification</li> </ul>	May 21	Jun 5
<a href="#">Project Proposal</a>	<a href="#">Final Project Guidelines</a>		June 11
<a href="#">Assignment 3</a>	<b>Assignment 3</b> <ul style="list-style-type: none"> <li>Multiclass Classification</li> <li>Summarization</li> <li>Question Answering</li> </ul>	June 11	June 26
<a href="#">Assignment 4</a>	<b>Assignment 4</b> <ul style="list-style-type: none"> <li>Image Captioning</li> </ul>		July 17
<a href="#">Project Reports</a>			Due July 30 (hard deadline)
<a href="#">Project Presentations</a>			In-class Aug 1-5

## Course Schedule

	Async Material to Watch	Topics	Materials
<b>Week 1</b> (May 2)	Introduction	<ul style="list-style-type: none"> <li>Overview of NLP applications</li> <li>NLP tasks, model structures and neural architectures</li> <li>Ambiguity and grounding in language</li> </ul>	<ul style="list-style-type: none"> <li>Read: <a href="#">AI's Language Problem (Technology Review)</a></li> <li>Read: <a href="#">The Rise and Fall of the English Sentence</a></li> <li>Skim: <a href="#">NLTK book chapter 1 (python and basics)</a></li> <li>Skim: <a href="#">NLTK book chapter 2 (data resources)</a></li> </ul>

		<ul style="list-style-type: none"> <li>• Introduction to word embeddings</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Optional:</i> <a href="#">The Interpreter (New Yorker)</a></li> <li>• <i>Optional:</i> <a href="#">Introduction to Linguistic Typology</a></li> <li>• <i>Optional / fun:</i> <a href="#">Tensorflow Playground</a></li> </ul>
<b>Week 2</b> (May 9)	Text Classification	<ul style="list-style-type: none"> <li>• Text classification approaches with neural networks</li> <li>• Text classification with CNNs</li> </ul>	<ul style="list-style-type: none"> <li>• Read: <a href="#">Understanding Convolutional Neural Networks for NLP (via Wayback Machine link)</a></li> <li>• Read: <a href="#">Convolutional Neural Networks for Sentence Classification (Kim, 2014)</a></li> <li>• <i>Skim:</i> <a href="#">The Illustrated Word2vec</a></li> <li>• <i>Optional / fun:</i> <a href="#">Tensorflow Playground</a></li> <li>• <i>Optional:</i> <a href="#">Natural Language Processing (almost) from Scratch (Collobert et al., 2011)</a></li> </ul>
<b>Week 3</b> (May 16)	Language and Context	<ul style="list-style-type: none"> <li>• Recurrent Neural Nets (RNNs) and language modeling</li> <li>• Attention</li> <li>• Context awareness and embeddings</li> </ul>	<ul style="list-style-type: none"> <li>• Read: <a href="#">Deep contextualized word representations (Peters et. al., 2018)</a></li> <li>• <i>Skim:</i> <a href="#">The Unreasonable Effectiveness of Recurrent Neural Networks</a></li> <li>• <i>Optional:</i> <a href="#">Attention and Augmented Recurrent Neural Networks</a> (section on "Attentional Interfaces" has an awesome visualization of an MT example, showing alignments)</li> </ul>

			<ul style="list-style-type: none"> <li>• <i>Optional:</i> <a href="#">Natural Language Corpus Data</a> (Peter Norvig)</li> </ul>
<b>Week 4</b> (May 23)	Pretrained Transformers	<ul style="list-style-type: none"> <li>• Transformer Architecture</li> <li>• Pretrained Transformers and Language Models</li> <li>• Pretrained Transformers and Context-Based Embeddings</li> </ul>	<ul style="list-style-type: none"> <li>• Read: <a href="#">Attention is All You Need</a> (Vaswani et. al., 2017)</li> <li>• <i>Skim:</i> <a href="#">BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding</a>", (Devlin et al, 2018)</li> <li>• <i>Skim:</i> <a href="#">The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning)</a>", (Alammar, 2018)</li> </ul>
<b>Week 5</b> (May 30)	Text Generation Models	<ul style="list-style-type: none"> <li>• Sequence to Sequence Architectures</li> <li>• Pretrained Encoder-Decoder Transformer Architectures</li> </ul>	<ul style="list-style-type: none"> <li>• Read: <a href="#">Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer</a> (Raffel et. al. 2019)</li> <li>• <i>Skim:</i> <a href="#">Language Models are Few Shot Learners</a> (Brown et. al. 2020)</li> </ul>
<b>Interlude (Extra Material)</b>	Units of Meaning: Words, Morphology, Sentences	<ul style="list-style-type: none"> <li>• Edit distance for strings</li> <li>• Tokenization</li> <li>• Sentence splitting</li> </ul>	<ul style="list-style-type: none"> <li>• Read: <a href="#">Sentence Boundary Detection and the Problem with the U.S.</a></li> <li>• <i>Skim:</i> <a href="#">NLTK book chapter 3</a> (processing raw text)</li> <li>• <i>Skim:</i> <a href="#">Natural Language Corpus Data</a> (Peter Norvig) (if you didn't read in Week 3)</li> </ul>
<b>Week 6</b> (Jun 6)	Machine Translation	<ul style="list-style-type: none"> <li>• Challenges of Language</li> </ul>	<ul style="list-style-type: none"> <li>• Read: <a href="#">Google's Neural Machine Translation</a></li> </ul>

		<ul style="list-style-type: none"> <li>Encoder Decoder architectures for Neural Machine Translation</li> <li>Evaluation</li> <li>Subword Models</li> </ul>	<p><a href="#">System</a></p> <ul style="list-style-type: none"> <li><i>Skim:</i> <a href="#">Sequence to Sequence Learning with Neural Networks</a></li> <li><i>Skim:</i> <a href="#">Neural Machine Translation by Jointly Learning to Align and Translate</a></li> <li><i>Skim:</i> <a href="#">Beyond English-Centric Multilingual Machine Translation</a></li> </ul>
<b>Week 7</b> (Jun 13)	Question Answering and Summarization	<ul style="list-style-type: none"> <li>Question Answering Architectures</li> <li>Extractive and Abstractive Summarization</li> <li>Evaluation</li> </ul>	<ul style="list-style-type: none"> <li>Read: <a href="#">Get To The Point: Summarization with Pointer-Generator Networks</a> (See et al. 2017)</li> <li><i>Skim:</i> <a href="#">Know What You Don't Know: Unanswerable Questions for SQuAD</a> (Rajpurkar et al. 2018)</li> <li><i>Skim:</i> <a href="#">DROP: A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs</a> (Dua et al. 2019)</li> <li><i>Skim:</i> <a href="#">Hurdles to Progress in Long-form Question Answering</a> (Krishna et al. 2021)</li> <li><i>Skim:</i> <a href="#">A Survey on Automatic Text Summarization</a> (Das and Martins, 2007)</li> </ul>
<b>Week 8</b> (Jun 20)	Linguistic Representation	<ul style="list-style-type: none"> <li>Elements of Language</li> <li>Grammars</li> </ul>	<ul style="list-style-type: none"> <li>Read: <a href="#">SyntaxNet (Parsey McParseface)</a></li> <li>Read: <a href="#">A Fast and Accurate Dependency Parser using</a></li> </ul>

		<ul style="list-style-type: none"> <li>• Dependency Parsing</li> <li>• Phrase based Parsing</li> <li>• Evaluation</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Neural Networks</a> (Chen &amp; Manning 2014)</li> <li>• Read: <a href="#">A Minimal Span-Based Neural Constituency Parser</a> (Stern et al. 2017)</li> <li>• <i>Skim</i>: <a href="#">NLTK book chapter 8 (analyzing sentence structure)</a></li> <li>• <i>Optional</i>: <a href="#">NLTK book chapter 5: Categorizing and Tagging Words</a></li> <li>• <i>Optional</i>: <a href="#">A Universal Part-of-Speech Tagset</a></li> <li>• <i>Optional</i>: <a href="#">Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?</a></li> <li>• <i>Optional</i>: <a href="#">Accurate Unlexicalized Parsing</a> (Klein &amp; Manning 2003)</li> <li>• <i>Optional / reference</i>: <a href="#">Penn Treebank Constituent Tags</a></li> <li>• Play: <a href="#">Stanford parser</a> (online demo)</li> </ul>
<b>Week 9</b> (Jun 27)	Entities and Linking	<ul style="list-style-type: none"> <li>• Named Entity Recognition</li> <li>• Coreference Resolution</li> <li>• Entity Linking</li> <li>• Relation Extraction</li> </ul>	<ul style="list-style-type: none"> <li>• Read: <a href="#">NLTK Book Chapter 7 (Extracting Information from Text)</a></li> <li>• Read: <a href="#">SpanBERT: Improving Pre-training by Representing and Predicting Spans</a> (Joshi et al. 2019)</li> <li>• <i>Optional</i>: <a href="#">Simple Coreference Resolution with Rich Syntactic and Semantic Features</a></li> </ul>



			<p>(Haghighi and Klein 2009, rule-based coreference)</p> <ul style="list-style-type: none"> <li>• <i>Optional:</i> <a href="#">Improving Coreference Resolution by Learning Entity-Level Distributed Representations</a> (Clark and Manning 2016, neural coreference)</li> </ul>
<b>Week 10</b> (Jul 4)	Embedding-based Retrieval	<ul style="list-style-type: none"> <li>• Single- vs. multi-document retrieval</li> <li>• Two tower models</li> </ul>	<ul style="list-style-type: none"> <li>• Read: <a href="#">Web Search for a Planet</a> (Google)</li> <li>• Read: <a href="#">The Anatomy of a Large-Scale Hypertextual Web Search Engine</a></li> <li>• <i>Skim:</i> <a href="#">"An Introduction to Information Retrieval"</a>, sections 6.2 and 6.3</li> <li>• <i>Optional:</i> <a href="#">PageRank</a> (Page, et al. 1999)</li> </ul>
<b>Week 11</b> (Jul 11)	Multimodality in NLP	<ul style="list-style-type: none"> <li>• Multimodal Applications</li> <li>• Image Captioning</li> <li>• Visual Question Answering</li> </ul>	<ul style="list-style-type: none"> <li>• TBD</li> </ul>
<b>Week 12</b> (Jul 18)	ML Fairness and Privacy	<ul style="list-style-type: none"> <li>• Fairness</li> <li>• Privacy</li> </ul>	<ul style="list-style-type: none"> <li>• Read: <a href="#">How to make a racist AI without really trying</a></li> <li>• Read: <a href="#">Data Decisions and Theoretical Implications when Adversarially Learning Fair Representations</a> (Beutel et. al. 2017)</li> </ul>

			<ul style="list-style-type: none"> <li>• Read: <a href="#">Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings</a> (Bolukbasi et. al. 2016)</li> <li>• Read: <a href="#">Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them</a> (Gonen &amp; Goldberg 2019)</li> </ul>
<b>Week 13</b> (Jul 24)	NLP in the Real World	<ul style="list-style-type: none"> <li>• Productization at Scale</li> <li>• Infrastructure</li> <li>• Metrics</li> <li>• Failure Modes</li> </ul>	<ul style="list-style-type: none"> <li>• TBD</li> </ul>
<b>Week 14</b> (Aug 1)		In-class project presentations	

Thanks for a great semester!