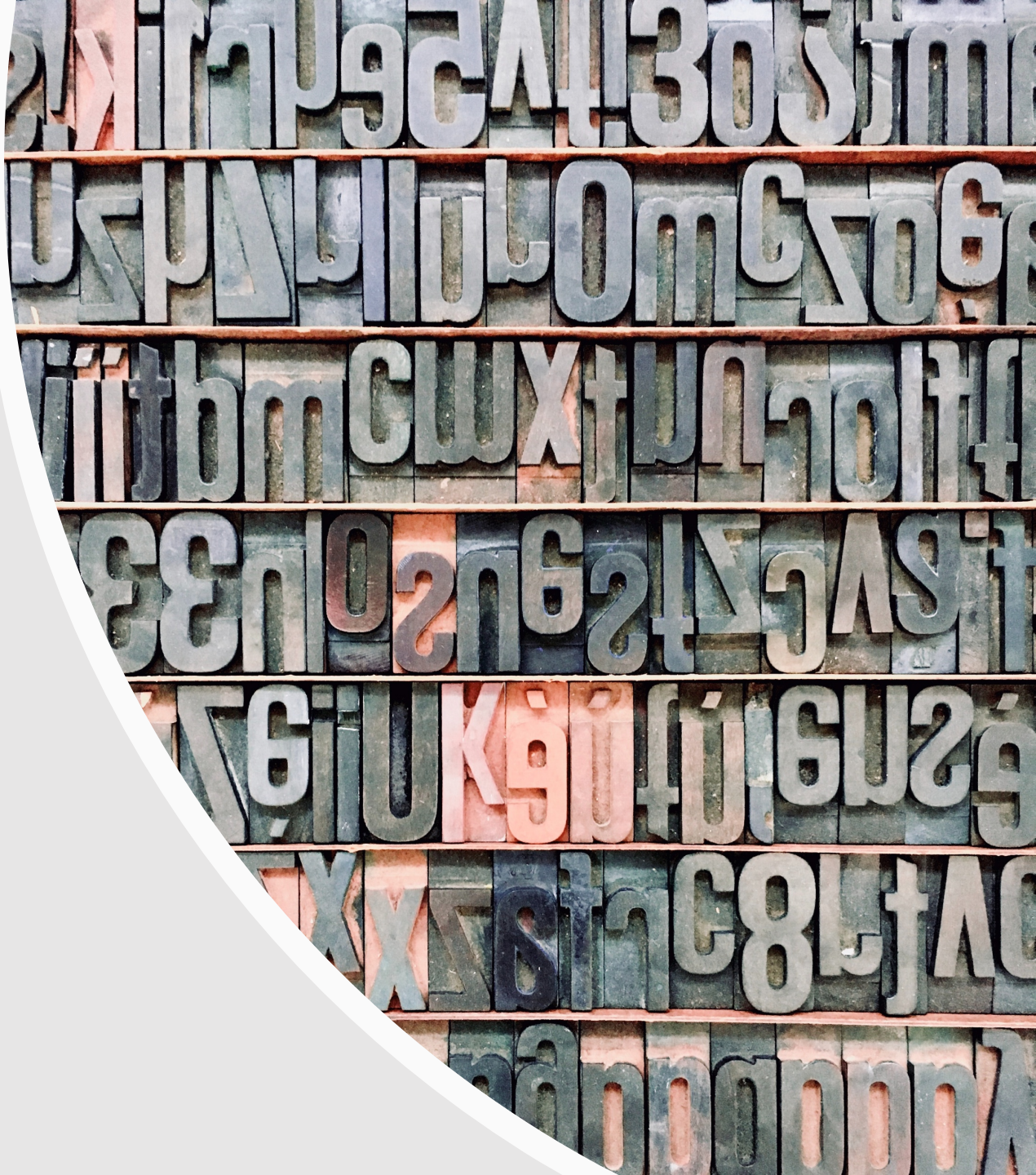
CS 421: Natural Language Processing

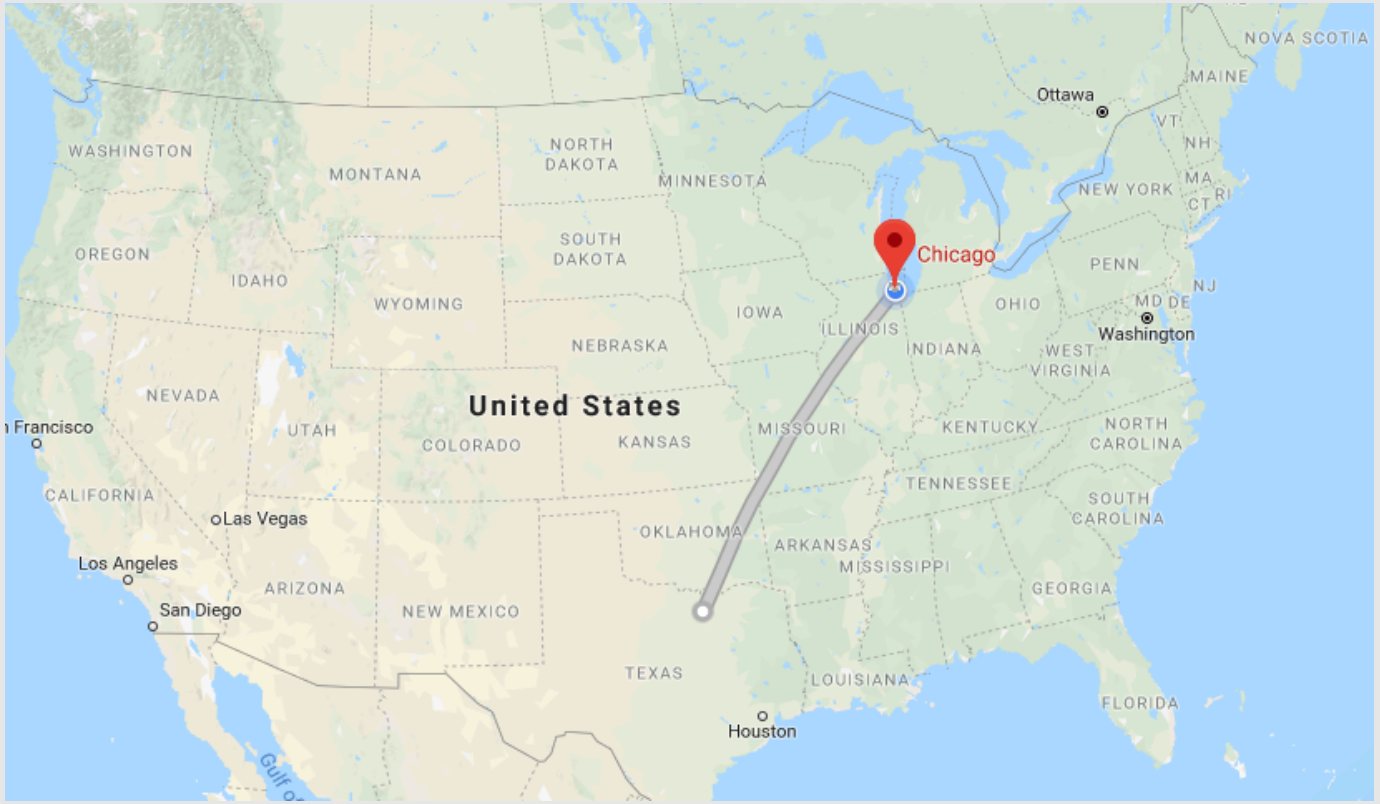
Natalie Parde, Ph.D.

Department of Computer Science

University of Illinois at Chicago

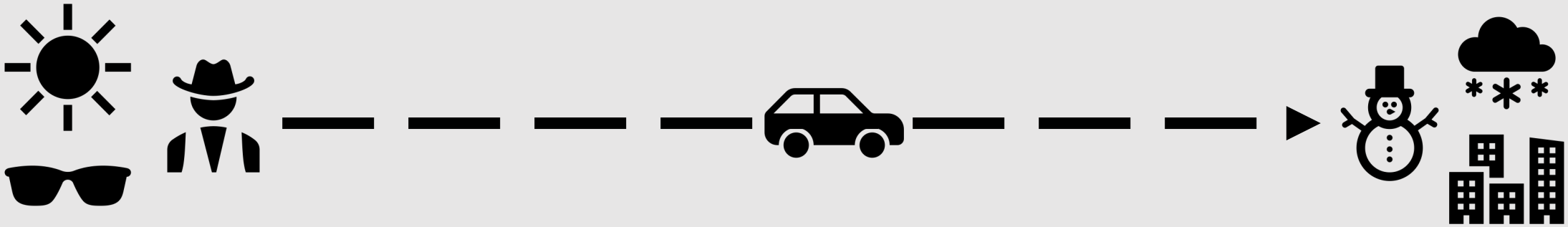
Fall 2019





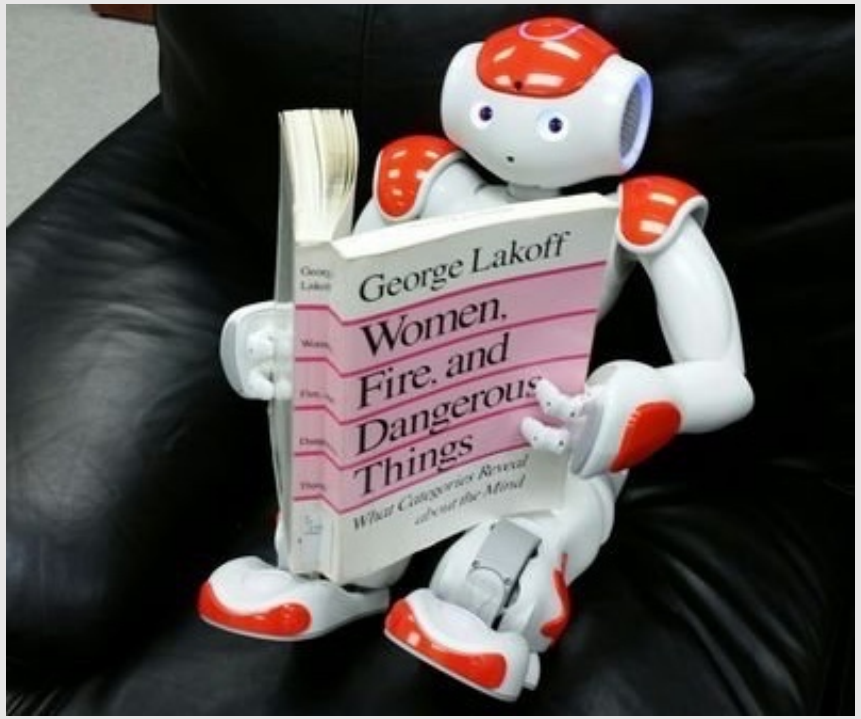
Who is teaching this class?

- **Assistant Professor**
- **Co-Director**, Natural Language Processing Laboratory: nlp.lab.uic.edu
- New **Chicagoan**



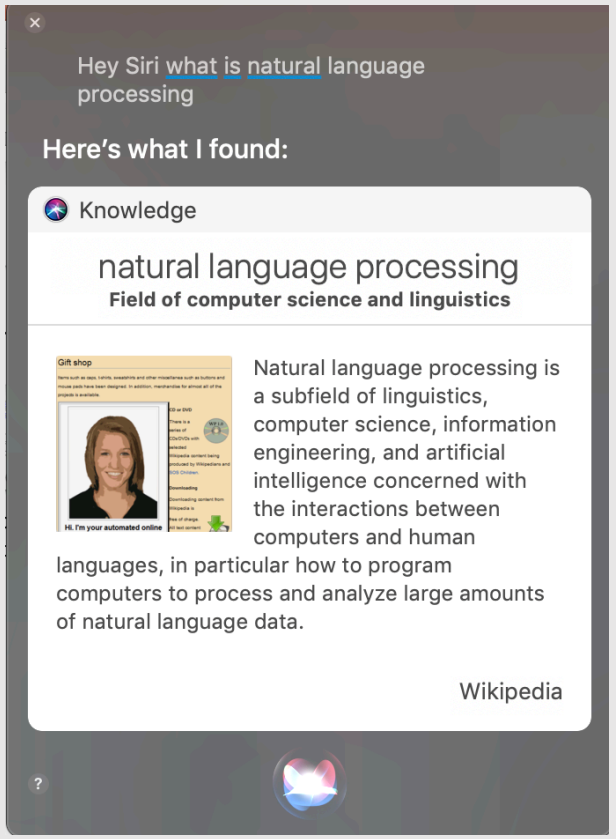
Core Research Areas

- **Semantics**
  - Metaphor, sarcasm
- **Multimodal NLP**
  - Grounded language learning, visual storytelling
- **Robotics**
  - Human-robot dialogue, interactive language learning, social robotics
- **Healthcare**
  - Cognitive health promotion, dementia detection



Tell me about yourself!

- Based on the roster:
  - 39% Grad
  - 61% Undergrad
- What made you interested in this class?
- What are you hoping to get out of it?



What is natural language processing?

- Subfield of **artificial intelligence**
- Focus: Automatically **interpreting** and **generating** natural **language**
- Most popular examples: Alexa, Siri, Google Assistant, Cortana….

Topics We'll Cover This Semester

- Text preprocessing
- Word similarity
- Part-of-speech tagging
- Parsing
- Language modeling
- Word embeddings
- Text classification
- Sequence processing
- Dialogue systems
- Question answering
- Machine translation

Syllabus Time

Contact Info

- Professor: Natalie Parde
- Email: parde@uic.edu
- Office: SEO 1132
- Office Hours: Tuesday 1:30 - 2:30 p.m. / Thursday 3:00 - 4:00 p.m.

- TA: Usman Shahid
- Email: hshahi6@uic.edu
- Office: SEL 4029
- Office Hours: Wednesday 12:00 - 2:00 p.m.

Course Communications

- **Piazza:** https://piazza.com/uic/fall2019/cs421
  - Also linked directly from Blackboard
- **Blackboard:** CS 421 Natural Language Processing (Undergrad: 43946; Graduate: 43947) 2019 Fall
- **Gradescope:** https://www.gradescope.com/
- **Course Website:** http://www.natalieparde.com/teaching/cs421_fall2019.html

- Try to keep email to a minimum! Instead, post questions on **Piazza**.
- All assignments should be submitted on **Gradescope**.

Prerequisite

- **CS 301** (Languages and Automata) or **MCS 441** (Theory of Computation I)
- If you have not taken one of those courses, you may want to email CS Student Affairs (ugrad@cs.uic.edu or grad@cs.uic.edu) to make sure that you meet the enrollment requirements for CS 421!

What will we be reading?

- Daniel Jurafsky and James H. Martin. Speech and Language Processing (2nd Edition). Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2009: https://uic.verbacompare.com/comparison?id=201042
- Daniel Jurafsky and James H. Martin. Speech and Language Processing (3rd Edition). Draft, 2019: https://web.stanford.edu/~jurafsky/slp3/
  - Still being written! The draft can be freely accessed at the link above.

What will be graded?

Assignments

- Five total
- Mix of theoretical and coding questions

Project

- Required for grad students, optional for undergrads

Exams

- Two (non-cumulative) midterms
- One (cumulative) final exam

How much will each of these items contribute to the final grade?

Undergrads

Exams: 50%

- 15% for each midterm
- 20% for the final exam

Assignments: 50%

- 10% for each assignment

Grads

Exams: 40%

- 12% for each midterm
- 16% for the final exam

Assignments: 40%

- 8% for each assignment

Project: 20%

- 6% for the implementation
- 6% for the presentation
- 8% for the report

Project Details

Select from the following topics:

- Build your own chatbot
- Build your own essay grader
- Custom project (requires pre-approval)

Detailed assignment overview coming soon

Implementation:

Should be documented well enough for the Professor/TA to run it on their own machines!

Report:

2200-2800 words

Presentation:

Week before finals week

Assignments and Exams

- Required programming language: **Python**
- Each assignment includes material covered up to one week before the deadline
- Assignment 1 (due 9/6) available very soon!
- Exams will contain a mix of **true/false**, **multiple choice**, and **free-response** questions

Missed Class

Missed Presentation:

- If you discover that you'll miss your scheduled presentation, contact me as soon as possible. I'll either:
  - Reschedule it
  - Assign a video presentation due the same day
- If you don't contact me ahead of time, I'll deduct:
  - 25% as soon as class begins
  - An additional 50% at that time the following day
  - The remaining 25% at that time the day after that

Missed Exam:

- If you need to miss class on a day when an exam is scheduled, contact me as soon as possible. I'll:
  - Reschedule it to take place under my or the TA's supervision at a time earlier than the original exam date/time
- If you do not contact me early enough to reasonably reschedule the exam to an earlier time, no make-up will be given

Late Cards

- A late card allows you to turn in your assignment up to 24 hours after the deadline with no grade penalty
- 4 per semester
- It's fine to use multiple late cards for the same assignment

- If you have no late cards remaining (or choose not to use them for a given assignment):
  - 25% deducted for each day the assignment is late, starting a minute after it is due

Honor Code

You are expected to do your own work on all assignments!

If I or the TA discover that you cheated:

You will receive a 0 on the assignment

If a repeat offense, we will also report you to CS Student Affairs

Examples of cheating:

Copying and pasting code from the internet

Copying and pasting answers from solutions that you obtained from prior CS 421 students

Saving solutions with the intention of providing them to future CS 421 students

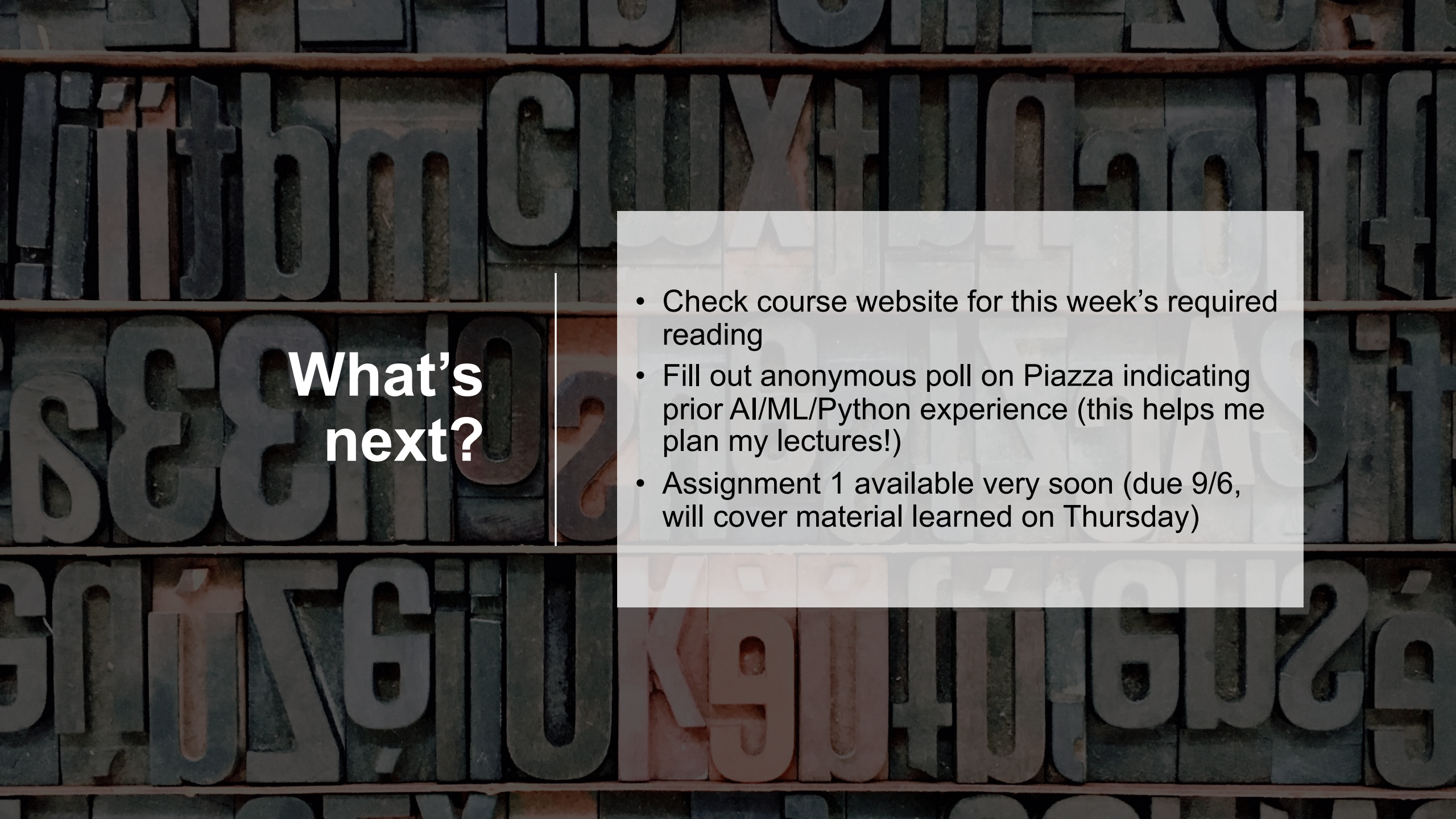
Copying and pasting text that you did not write without surrounding the copied text in quotation marks and following it with a citation

Schedule

Week	Topic	Deliverables
8/26-8/30	Introduction, Text Preprocessing, and Edit Distance	—
9/2-9/6	Automata, Transducers, and Hidden Markov Models	Assignment 1: 9/6 by 12 p.m.
9/9-9/13	Part-of-Speech Tagging and Formal Grammars	—
9/16-9/20	Syntactic and Dependency Parsing	Assignment 2: 9/20 by 12 p.m.
9/23-9/27	First-Order Logic and Review/Catch-Up	—
9/30-10/4	**Exam 1 (10/1)** and N-Gram Language Modeling	—
10/7-10/11	Word Embeddings	Assignment 3: 10/11 by 12 p.m.
10/14-10/18	Naïve Bayes, Text Classification, and Evaluation Metrics	—
10/21-10/25	Neural Networks and Neural Language Models	Assignment 4: 10/25 by 12 p.m.
10/28-11/1	Sequence Processing with Recurrent Networks and Review/Catch-Up	—
11/4-11/8	**Exam 2 (11/5)** and Information Extraction	—
11/11-11/15	Dialogue Systems and Chatbots	—
11/18-11/22	Question Answering and Summarization	Assignment 5: 11/22 by 12 p.m.
11/25-11/29	Machine Translation	—
12/2-12/6	Project Presentations and Review	—
12/9-12/13	**Exam 3 (Finals Schedule TBD)**	—

Classroom Environment

- Treat everyone with respect!
  - Silence all devices
  - Don’t chat with one another during lectures unless explicitly asked to do so
  - Don’t bring disruptive food
- Inform me and UIC’s Disability Resource Center (http://drc.uic.edu/) about any disabilities for which you would like to request accommodation.
- Let me know if you’d like me to update any of the information (e.g., name or pronouns) I’ve received about you from the class roster.
- Feel free to reach out to me with feedback throughout the semester!



What's next?

- Check course website for this week's required reading
- Fill out anonymous poll on Piazza indicating prior AI/ML/Python experience (this helps me plan my lectures!)
- Assignment 1 available very soon (due 9/6, will cover material learned on Thursday)