COSI-230B: Natural Language Annotation for Machine Learning

Course Introduction & Overview

Jin Zhao

Brandeis University
Computational Linguistics Program

Spring 2026

Today's Agenda

1. Introduction
2. Course Overview
3. Learning Objectives
4. Lecture Schedule
5. Assignments & Grading
6. Tools & Resources
7. Course Policies
8. What's Next

Welcome to COSI-230B!

Lecturer: Jin Zhao ("Jin"; she/her)
jinzhao@brandeis.edu
Volen 109
Wed 1–3pm ET

Teaching Assistant: Richard Brutti ("Ricky"; he/him)
brutti@brandeis.edu
Abelson Lower Level
Thu after lab or by appointment

Course Platform: MOODLE

Essential Logistics

Location & Time

Room: Volen National Center for Complex Systems, Room 106
Time: Mon, Wed, & Thu from 9:05 AM - 9:55 AM ET

Mon & Wed:
Lectures led by Jin

Thursday:
Lab sessions led by Ricky

Prerequisites

COSI 115b, or COSI 114a and COSI 115b (concurrent)

Let's Get to Know Each Other

Quick introductions:

- Your name and preferred pronouns
- Your current program and previous academic background
- One NLP task you find interesting
- Have you used ChatGPT/Claude/other LLMs for anything?

What is This Course About?

Course Description

This course covers the **theory and practice** of creating annotated datasets for natural language processing and machine learning.

Key themes:

- Annotation methodology and best practices
- Quality measurement (inter-annotator agreement)
- Annotation tools and workflows
- **[NEW]** LLM-assisted annotation approaches
- **[NEW]** Human-AI collaborative annotation
- **[NEW]** Preference data for RLHF

Why Does Annotation Matter?

"Data is the new oil" — but it needs **refining**

Machine Learning Pipeline:

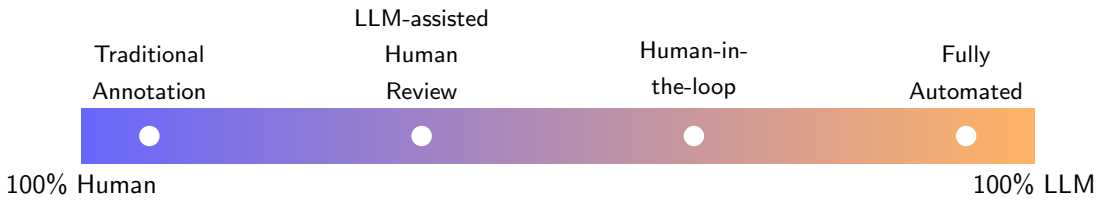
1. Collect raw data
2. **Annotate data** ← *This course!*
3. Train model
4. Evaluate model
5. Deploy & iterate

Annotation is everywhere:

- Sentiment analysis
- Named entity recognition
- Machine translation
- Question answering
- ChatGPT/Claude training (RLHF!)

The Annotation Landscape in 2025

The Human-AI Annotation Spectrum



Key question we'll explore:

When should we use humans, LLMs, or a combination?

What's New in 2025?

Traditional Topics (Still Essential!)

- Annotation fundamentals
- Task design & guidelines
- Annotation tools
- Inter-annotator agreement
- Quality control
- Gold standard creation

[NEW] New Topics for 2025

- LLM-based annotation
- Prompt engineering for annotation
- Human-AI collaboration
- RLHF & preference data
- LLM-as-judge evaluation
- Safety & red teaming

The fundamentals haven't changed—but the **tools** and **workflows** have evolved dramatically!

Course Learning Objectives

By the end of this course, you will be able to:

1. **Design** annotation schemas for various NLP tasks
2. **Write** clear guidelines for humans *and* prompts for LLMs
3. **Calculate** and interpret inter-annotator agreement metrics
4. **Use** modern annotation tools effectively
5. **Evaluate** trade-offs between human and LLM annotation
6. **Create** high-quality annotated datasets
7. **Understand** preference data collection for RLHF

Lecture Schedule: Weeks 1–5

Wk	Dates	Lecture 1	Lecture 2
1	Jan 12, 14	Course Introduction	Annotation Fundamentals
2	Jan 21 *(MLK Day)*	When to Annotate — Tools & Formats *(No class Jan 19)*	
3	Jan 26, 28	MATTER/MAMA Cycle	Corpus Selection
4	Feb 2, 4	Tasks I: Classification	Tasks II: Sequence Labeling
5	Feb 9, 11	Task Formalization	Tasks III: Relations

Lecture Schedule: Weeks 6–10

Wk	Dates	Lecture 1	Lecture 2
6	Feb 16, 18	*February Break — No Classes*	
7	Feb 23, 25	Tasks IV: LLM Tasks	Writing Guidelines
8	Mar 2, 4	Group Presentations	LLM-Based Annotation
9	Mar 9, 11	Annotation Tools I	Annotation Tools II
10	Mar 16, 18	Tools Advanced	Human-AI Collaboration

Lecture Schedule: Weeks 11–15

Wk	Dates	Lecture 1	Lecture 2
11	Mar 23, 25	IAA I: Cohen's Kappa	IAA II: Fleiss, Krippendorff
12	Mar 30	IAA — Modeling Intro *(No class Apr 1)*	
—	Apr 6, 8	*Passover Break — No Classes*	
13	Apr 13, 15	Modeling I	Modeling II
14	Apr 20, 22	Preference Data & RLHF	Safety & Red Teaming
15	Apr 27, 29	Low-Resource Languages	Best Practices

Weeks 16–17 (Apr 30 – May 12): Final presentations & reports due

Homework Assignments

Assignment	Topic
HW0	Dataset Exploration
HW1	Annotation Tools Exploration (brat, Label Studio)
HW2	Data Wrangling with Pandas
HW3	Inter-Annotator Agreement
HW4	Sentiment Analysis Fine-tuning

Each homework assignment is of **equal weight**.

Late Homework Policy:

- **3 grace days** total for the semester
- **Max 1 grace day** per assignment without penalty
- Additional extensions require approval from Jin and Ricky
- Late submissions: **20% penalty per day**

Semester Project (50% of Grade)

Groups of 3–4 students working collaboratively on a semester-long annotation project following the MATTER/MAMA cycle.

What You'll Do:

- Design an annotation specification and guidelines for an NLP task
- Annotate a dataset as a group
- Evaluate inter-annotator agreement
- Refine guidelines and create gold standard
- Train and evaluate baseline NLP models

Evaluation:

- Evaluated as groups *and* based on individual contributions
- Peer evaluation component

Project Deliverables

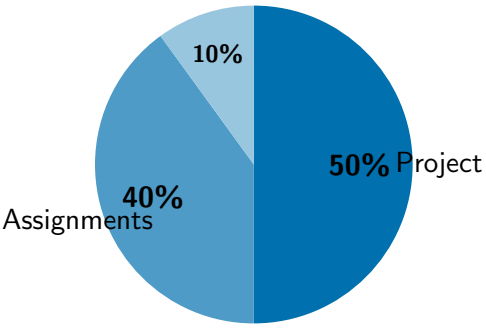
Final Deliverables:

- **4-page paper** using LaTeX ACL/LREC format
- **15-minute presentation** in class
- **Dataset submission** with documentation
- **Peer evaluation** of group members

Milestones Throughout Semester:

- Form groups, choose dataset (Week 3)
- Conceptualize annotation task (Week 4)
- Present chosen tasks (Week 7)
- Draft guidelines + pilot annotations (Week 9)
- Inter-annotator agreement evaluation (Week 11)
- Adjudicated gold standard dataset (Week 13)
- Final report & presentation (Weeks 16–17)

Grading Breakdown



Component	%
Participation	10%
Assignments (HW0–4)	40%
Semester Project	50%

Important:
Stay on top of the semester project—it's essential to perform well in this course!

Expectation:
∼9 hours/week study time
(4-credit course)

Tools We'll (Likely) See in this Course

Annotation Tools:

- Label Studio (recommended)
- Argilla (for RLHF)
- brat (traditional)
- Prodigy (if licensed)

Programming:

- Python 3.9+
- pandas, scikit-learn
- Hugging Face transformers

LLM Access:

- OpenAI API (GPT-4)
- Anthropic API (Claude)
- Open-source options: Llama 3, Mistral

Course Platform:

- MOODLE (announcements, submissions)
- GitHub (code, materials)

Recommended Readings

Textbook:

- Pustejovsky & Stubbs (2012). *Natural Language Annotation for Machine Learning*. O'Reilly.

Key Papers (we'll read throughout):

- Gilardi et al. (2023). "ChatGPT Outperforms Crowd-Workers for Text-Annotation Tasks"
- Ouyang et al. (2022). "Training language models to follow instructions with human feedback" (InstructGPT)
- Zheng et al. (2023). "Judging LLM-as-a-Judge with MT-Bench"
- Bai et al. (2022). "Constitutional AI" (Anthropic)

Full reading list available on MOODLE.

Attendance & Participation

Attendance:

- Attending lectures and lab sessions is **mandatory**
- Impacts your ability to contribute to discussions, activities, and group project
- Results in lower participation grade if you miss class
- Reasonable accommodations for excused absences

Missed Classes: Contact us ASAP if you have an emergency!

Generative AI Policy

LLMs may be used *as objects of analysis* or *as limited experimental tools* when explicitly permitted.

Permitted uses:

- Pilot annotation to stress-test guidelines
- Error analysis and comparison with human annotations
- Exploratory analysis of ambiguity or disagreement

Disallowed uses:

- Submitting model-generated annotations as human-produced
- Generating assignment write-ups or analyses for submission
- Using models to replace required human annotation work

Any use of generative models must be clearly disclosed.
Undisclosed use = violation of academic integrity policy.

Academic Honesty & Support

Academic Honesty:

- Follow Brandeis University's policies on academic integrity
- Violations forwarded to Student Rights and Community Standards
- Sanctions can include failing grades and/or suspension

Accommodations:

- Work with Student Accessibility Support (SAS)
- Contact: 781-736-3470 or access@brandeis.edu
- Provide accommodation letter as soon as possible

Student Support:

- Care Team, Academic Services, Graduate Student Affairs
- University Ombuds, Office of Equal Opportunity
- Visit: `brandeis.edu/support`

Communication

How to Reach Us

- **MOODLE:** Course materials, announcements, submissions
- **Email:** Best way to communicate outside of class
- **Office Hours:** Face-to-face (or Zoom) clarification

Response Time: We aim to respond within **one business day** (by 5pm ET the next weekday)

Tips for Success

- Start homework early
- Form project groups by Week 3
- Attend office hours—they're there to help you!
- Check MOODLE regularly for announcements

This Week's Plan

Today (Lecture 1):

- Course overview ✓
- Introductions ✓
- Activity: What makes a dataset trustworthy?

Next Class (Lecture 2):

- Annotation fundamentals
- Types of NLP annotation tasks
- The MATTER cycle

Action Items:

1. Access MOODLE and check the course page
2. Install Python 3.9+ if you haven't
3. Start thinking about project interests!

Questions?

jinzhao@brandeis.edu

Office Hours: Wed 1–3pm (Volen 109)

MOODLE for announcements