

1 Teaching and Lecturing

Philosophy Education is an act of empowerment. I view teaching as both a privilege and a civic responsibility to bridge barriers in access to knowledge and foster equitable participation in science. My goal as an educator is to lower the threshold for understanding complex ideas, making cutting-edge research in AI accessible, inspiring, and transformative. I strive to cultivate an environment where students feel challenged yet supported, guided yet independent, and motivated to connect their technical learning with societal impact.

Experience During my Ph.D., I have served **twice as lead or head TA at UW CSE**, delivered over **20 guest lectures and invited talks**, and co-instructed the **ACL 2025 Tutorial on Guardrails and Security for LLMs**. My teaching emphasizes rigor, clarity, and timeliness, reflecting the rapid advances of AI.

Teaching Assistant (TA): At UW, I served as the **head TA** for CSE 447/517: *Natural Language Processing—An LLM Version* with over 230 undergraduate and graduate students, and as one of two **lead TAs** for CSE 599 D1: *Exploration on Language, Knowledge, and Reasoning* with 30+ graduate students. As head TA for CSE 447/517, I **led a team of 12 TAs** and took a central role in **revamping the entire NLP curriculum into a modern, LLM-centered course**. I helped redesign the full course structure, including 27 lectures, 5 assignments, and a final project with 5 milestones, while collaborating closely with the instructor and TAs to develop and review lecture materials. I also authored slides for key topics such as **NLP Overview, Transformers, and Pre-Training**. Beyond content development, I managed all course logistics, including the website, timeline, grading systems, and communication across 15+ Slack channels, ensuring that TAs actively supported students through Ed forum and office hours. The result was a comprehensive, modernized NLP curriculum now publicly available for future offerings. **My TA leadership received outstanding feedback** from both the instructor and students:

- **Prof. Yejin Choi (instructor):** “Thank you a ton Liwei for being the most amazing TA I’ve ever worked with! / Your performance as the head TA was above and beyond exceptional (huge thank you again!!!)”
- **Student post on Ed Discussion:** “The Bob Bandes Nomination (*Note: UW’s best TA award*) portal has opened! Let’s vote for goat TAs [redacted] and Liwei for the Bob Bandes Award.”

Guest Lectures and Talks: I have delivered over **20 guest lectures and invited talks** to diverse audiences, ranging from undergraduate and graduate students at leading computer science institutions (e.g., CMU, UCLA, UIUC, KAIST) to high school students in Seattle as part of outreach efforts. My lectures span topics from **my research on LLM morality, safety, and value alignment to core concepts such as reasoning and prompting**, typically lasting between 50 and 90 minutes. I strive to make each lecture both content-rich and up-to-date, continually revising my materials to reflect the latest research and best practices, and always dedicating time for audience questions. Here are a few examples of positive feedback from course instructors.

- **Prof. Maarten Sap (instructor of course 11-430/830: Ethics, Safety, and Social Impact in NLP and LLMs at CMU):** “Chatted with students and they LOVED your guest lecture! They loved how you taught them cutting edge stuff, they were super super happy.”
- **Hudson Harper (instructor of the Ethics and Citizenship course at The Downtown School, Seattle):** “It was incredibly informative and provided great food for thought that our students continued to talk about.”

Conference Tutorial: I served as a main co-instructor for the *ACL 2025 Tutorial on Guardrails and Security for LLMs*, which attracted around 400 in-person and 50 online participants. I designed and delivered three core units on **LLM Security, Alignment and Misalignment, and Agent Safety**, covering 60 minutes of the 3-hour session. I also organized the overall structure, presentation templates, and coordination among co-instructors to ensure coherence and up-to-date content. The tutorial received highly positive feedback, with participants praising its engaging and dynamic format that sustained attention throughout. It was also highlighted in the *ACL 2025 Keynote* by Verena Rieser, who described it as “one of my highlights so far.”

Future Looking ahead, I am passionate about **teaching a broad range of core and advanced courses** in Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML). I am particularly eager to **design and teach advanced special-topic courses in both lecture and seminar formats** on areas such as AI Safety and LLM Security, Value Alignment, Human-Centered NLP, Reinforcement Learning and (Multi-)Agent LLMs, and Socially Beneficial and Human-Empowering AI. I will **keep course content current with the latest research**, integrating modern insights and real-world case studies even into classical topics. I also plan to **make all teaching materials, including slides, assignments, and notes, publicly available** and share recorded lectures online for broader accessibility. My teaching will **build on strong computational foundations while emphasizing interdisciplinary connections**, using computational perspectives to bridge the humanities and sciences and to foster human-centered progress in AI.

2 Mentorship and Advising

Philosophy Students pursue graduate study not just to acquire skills, but to spend their most formative years tackling meaningful problems, defining enduring goals, and cultivating a lifelong commitment to growth and fulfillment. As an advisor, I view students' success as inseparable from my own and am dedicated to inspiring their curiosity, nurturing their ambition, and helping them turn challenges into lasting confidence and purpose. I strive to align my academic pursuits with students' unique paths to foster mutual success.

Experience Mentorship has been central to my Ph.D., and guiding my mentees' growth has been deeply rewarding. I have mentored 26 junior researchers across 10 academic institutions and co-authored over 15 papers at top academic venues, including 9 Ph.D. students, 10 master's students, 6 undergraduates, and 1 high school student. I devote significant time to individualized mentorship, offering hands-on research training and broader career guidance. I strive to be encouraging and empowering. Having myself entered the NLP/AI field only in my second year of Ph.D., I deeply value mentoring students who may not start with the strongest background but show dedication and enthusiasm to learn and grow. Under my guidance, several have published their first lead-author papers in AI, and many mentorships have evolved into long-term collaborations resulting in multiple top-tier conference publications. Representative highlights:

Note: Individual mentee stories have been redacted in this public version to protect privacy. Please feel free to reach out for the full version.

Future My advising will focus on cultivating the core skills of academic excellence, e.g., identifying meaningful research problems, designing and executing rigorous studies, thinking creatively beyond conventions, communicating with clarity and impact, and engaging diverse audiences through collaboration and leadership. I also aim to nurture enduring qualities that transcend academia, e.g., perseverance and confidence in pursuit, a collaborative spirit in teamwork, and a civic commitment to harness science for positive change. In my future lab, I will foster a culture grounded in integrity, respect, and fairness, where each student's unique strengths and aspirations are valued and mentorship is guided by care, well-being, and mutual respect.