I strongly believe that the opportunity to teach and mentor students is one of the most exciting and rewarding aspects of an academic career. I am fortunate to have had teachers and mentors who made a tremendous impact on my career path, my research, and my life. As an undergraduate student, I took a class on computational linguistics, which sparked my interest in natural language processing (NLP) and its statistical approaches. During my first two years at the University of Washington, I worked closely with senior Ph.D. students and my advisor to establish a firm foundation for my Ph.D. research career. My teaching and mentorship principles have been greatly shaped by these mentors and teachers as well as my own experiences. I eagerly anticipate 'paying it forward' by continuing to support younger generations of scholars in various stages of their career.

Teaching and Lecturing

My teaching philosophy centers around three core principles. First, **I make connections from the topic at hand to different concepts, problems, disciplines, and even daily life in my lecture.** At the UW, I assisted in teaching a graduate-level NLP course taught by Prof. Noah A. Smith. In one class, we introduced the technique of *word embeddings*, which converts words to vector representations; this technique, key to the success of deep learning models, is important for students to understand before learning more advanced material. We presented the word2vec algorithm, a standard method for creating word embeddings. The algorithm itself, not particularly complex, involves logistic regression that predicts whether words will occur in each other's neighborhood. This algorithm is motivated by theory **grounded in linguistics, distributional semantics in particular**: *a word is characterized by the company it keeps*. We highlighted this connection to linguistic theory throughout the lecture. To reinforce this lecture, we further designed a homework assignment that helps students **connect the word2vec algorithm to matrix factorization**; the algorithm can be seen as a special case of seemingly disparate matrix factorization through the lens of convex optimization. By emphasizing such connections to linguistics, mathematics, and beyond, I aim to make learning easy, memorable, and ultimately fun.

Second, I always start teaching with and repeatedly return to **big pictures** when teaching new concepts. In 2021, I gave a guest lecture on machine translation for students and scholars in the humanities and translation studies. Before explaining the details of machine translation algorithms, I discussed how current machine translation technology can impact research and applications of their interest, as well as its limitations. Rigor matters in science, but directly diving into details might drown students in a swamp of specifics. I believe that presenting a big picture helps orient learners when teaching new concepts, especially in cross-disciplinary education.

Third, beyond imparting course material, I share my own analysis and encourage students to engage in similar **critical thinking**. For example, the transformer architecture is now a backbone of recent advances in NLP, computer vision, speech, and computational biology. When I explain this architecture, I also share my thoughts on: Why did transformers become so widespread? Was it because transformers are suitable for parallel computation available on modern hardware, such as GPUs and TPUs? Were there specific empirical results that convinced many researchers of their benefits? What is coming next in the development pipeline? I also find it effective to have such discussions as interactive exercises during lecture and office hours. Raising and contemplating such questions helps students grow as independent thinkers, regardless of their future career. I am passionate about this transferable and cross-disciplinary aspect of teaching.

Invited lectures. In addition to classroom instruction, I have given many guest lectures and invited/conference talks to audiences from various fields in academia and industry. These experiences have prepared me to teach students from different fields and backgrounds as well as those who have different knowledge levels, from introductory to advanced.

Going forward. I look forward to teaching both introductory and advanced courses on NLP, machine learning, deep learning, artificial intelligence, and other related areas. I am also excited about developing new courses on special topics that have been central to my research, such as *efficient* machine learning, *multilingual* natural language processing, and *machine translation*. My prior experience makes me comfortable teaching courses of different forms, including lectures with a large audience and more intimate project- or discussion-based courses.

Specifically, I plan to develop a specialized course on **multilingual NLP**. Current NLP data creation and model development focus heavily on the English language. Many NLP courses offered in universities are also centered around methods and use cases for the English language and speakers. This limits the diversity and accessibility of AI technology and heightens the barriers to its use in international AI applications. Below are several core topics that I intend to cover in this class:

- **Machine translation.** How did machine translation research start? How does neural machine translation work, and how did it revolutionize the field? What aspects of the field work well, and what are its limitations? What are the promising applications of machine translation in the future?
- **Typology and linguistic diversity.** What is the space of language? What linguistic properties and challenges would be overlooked if NLP researchers fixate on the English language? How should NLP models handle complex morphology and diverse scripts that are not present in English? What are the challenges towards massively multilingual NLP for the 7000+ global languages beyond English?
- **Low-resource language processing and crosslingual transfer.** Most non-English languages have only small amounts of data for many natural language tasks. Is it possible to perform crosslingual transfer from English or other high-resource languages? How can multilingual NLP benefit from the recent advances in large language models?
- **Multilingual information access and data creation.** How do we build models and resources to provide information access tools for speakers of various languages (e.g., question answering, information retrieval, fact verification, and information extraction)? The field of NLP has yet to sufficiently consider varying use cases from diverse cultures and languages; what recent efforts and future research directions can mitigate this community-wide problem? How can academics effectively collaborate to create resources across different languages and cultures? Can we simulate such collaborations in our classroom?

Mentoring

During my Ph.D. at the UW, I have been privileged to mentor more than 10 undergraduate, Master's, and junior Ph.D. students over a wide range of topics in NLP and machine learning. Many of these projects turned into papers published at top-tier venues [1, 2, 3, 4, 5, 6, 7, 8], one of which won the best paper award at NAACL 2022 [1]. I played various mentoring roles, depending on the students' interests and backgrounds; in some cases, I advised students who led their own projects, while other students occasionally contributed to my projects, such as by helping me with manual annotations and analysis.

As a mentor, I begin by helping students translate abstract concepts and research questions into practical experiments. I guide them in breaking down long-term objectives into manageable milestones, which we discuss on a regular basis. When it is helpful, I also advise students regarding mathematical and experimental details. As the research progresses, I assist them with technical writing and presentations. I particularly enjoy supporting my mentees as they develop into independent researchers. Indeed, after graduation, six former mentees pursued a Ph.D. in computer science or related fields (the University of Michigan and the University of Hong Kong) or joined an industrial research lab (Meta AI, Apple, and the Allen Institute for AI). I also enjoy exploring new research areas and learning mentoring skills through advising students. I look forward to building longer-term mentoring relationships with my students as a faculty member.

References

[1] Ximing Lu, Sean Welleck, Peter West, Liwei Jiang, **Jungo Kasai**, Daniel Khashabi, Ronan Le Bras, Lianhui Qin, Youngjae Yu, Rowan Zellers, Noah A. Smith, and Yejin Choi. Neurologic A*esque decoding: Constrained text generation with lookahead heuristics. In *Proc. of NAACL*, 2022. **Best Paper Award**.

[2] **Jungo Kasai**, Keisuke Sakaguchi, Ronan Le Bras, Lavinia Dunagan, Jacob Morrison, Alexander R. Fabbri, Yejin Choi, and Noah A. Smith. Bidimensional leaderboards: Generate and evaluate language hand in hand. In *Proc. of NAACL*, 2022.

[3] **Jungo Kasai**, Keisuke Sakaguchi, Lavinia Dunagan, Jacob Morrison, Ronan Le Bras, Yejin Choi, and Noah A. Smith. Transparent human evaluation for image captioning. In *Proc. of NAACL*, 2022.

[4] Zeyu Liu, Yizhong Wang, **Jungo Kasai**, Hannaneh Hajishirzi, and Noah A. Smith. Probing across time: What does RoBERTa know and when? In *Findings of ACL: EMNLP*, 2021.

[5] Hongjin Su, **Jungo Kasai**, Chen Henry Wu, Weijia Shi, Tianlu Wang, Jiayi Xin, Rui Zhang, Mari Ostendorf, Luke Zettlemoyer, Noah A. Smith, and Tao Yu. Selective annotation makes language models better few-shot learners. Under review.

[6] Akari Asai, Xinyan Yu, **Jungo Kasai**, and Hannaneh Hajishirzi. One question answering model for many languages with cross-lingual dense passage retrieval. In *Proc. of NeurIPS*, 2021.

[7] **Jungo Kasai**, Keisuke Sakaguchi, Yoichi Takahashi, Ronan Le Bras, Akari Asai, Xinyan Yu, Dragomir Radev, Noah A. Smith, Yejin Choi, and Kentaro Inui. RealTime QA: What's the answer right now?, 2022. Under review.

[8] Haoxin Li, Phillip Keung, Daniel Cheng, **Jungo Kasai**, and Noah A. Smith. NarrowBERT: Accelerating masked language model pretraining and inference. Under review.