Students from underrepresented groups often encounter substantial barriers in academic careers, sometimes without fully recognizing their impact. At the University of Tokyo, where women comprised less than 10% of engineering students, I was frequently one of the only women in lecture halls of over 100 peers. Transitioning to the Ph.D. program at the University of Washington, with its supportive advisors, mentors, female role models, and a student body that is over 33% female, was a transformative experience. For the first time, I felt fully empowered to voice my thoughts and engage with peers. This experience highlighted the unconscious suppression I had felt and the critical need for diversity and inclusion. I am deeply committed to fostering environments where all students, particularly those from marginalized backgrounds, can reach their full potential.

Throughout my Ph.D., I have actively contributed to DEI initiatives at UW (University of Washington) CSE and in the broader research community. As a professor, I am dedicated to furthering these efforts by building inclusive communities within my research group, department, and university. Specifically, I plan to enhance diversity and inclusion in computer science through three key avenues:

- **Outreach and Admissions:** Expanding Access and Encouraging Applications from Underrepresented Groups
- **Mentoring and Environment:** Fostering an Inclusive and Supportive Academic Community
- **Research:** Advancing Equity by Empowering Underrepresented Communities through Research

Outreach and Admissions

At UW, I played a key role in outreach and admissions efforts aimed at encouraging students from underrepresented backgrounds to participate in research and pursue Ph.D. programs.

Admission: I was heavily involved in Ph.D. admissions at UW, **serving as a student area chair for two years**. I view admissions as a critical step for advancing diversity and inclusion. As competition grows, I have noticed a trend among applicant reviewers to prioritize metrics like publication count, GPA, and school names [1]. To address this, I: (1) Supervised over 20 NLP student reviewers, providing feedback on biased or incomplete reviews. (2) Created shortlists from 500+ applications each year, ensuring no high-potential students facing external challenges were overlooked. (3) Carefully reviewed "borderline" cases, providing detailed meta-reviews to help faculty consider diverse factors. (4) Consistently reminded reviewers to be mindful of unconcious biases, such as affinity bias, and shared resources for reviewing.

Outreach: Since 2021, I have hosted **weekly virtual office hours** open globally, supporting around 90 participants from countries like Japan, China, India, Turkey, Iran, and Brazil. These sessions offered guidance on research, Ph.D. admissions, and navigating Ph.D. journeys. Five international attendees were later accepted into top Ph.D. programs, including UW and CMU, attributing these sessions to helping refine their research goals and applications. As a mentor for UW CSE's Pre-Ph.D. Application Mentorship Program (PAMS), I also helped applicants polish their materials, with one mentee now a Ph.D. student at UW. Additionally, I actively promote these mentorship programs through minority-focused networks.

Moving Forward: I am committed to advancing diversity and inclusion through outreach and admissions. Specifically, I aim to actively participate in the admissions process and lead outreach efforts to encourage applications from underrepresented groups. I also plan to design systematic improvements to ensure fair evaluation of candidates from all backgrounds.

Mentoring and Environment

Mentoring: Hands-on mentorship is crucial for student success [2], yet students from underrepresented backgrounds often lack access to such guidance. During my Ph.D., I closely mentored seven students across multiple institutions, including four women, six people of color, and five first-generation immigrants. Beyond recruiting students from diverse backgrounds, I aim to provide holistic support that encompasses not only their research but also their long-term career goals and personal challenges outside academia. Many students face unique hurdles, such as hesitancy in expressing opinions—especially in a non-native language—or stress from being far from family in a new country [3]. Having experienced similar challenges early in my Ph.D., I am sensitive to these issues and prioritize offering encouragement and practical support to help them overcome obstacles. I consistently remind my mentees that their voices are valued and that they possess exceptional skills worthy of confidence.

Inclusive Environments: I have actively advocated for student perspectives to enhance inclusiveness and diversity within my department, contributing to a more inclusive environment in three key ways: (1) representing graduate students as a DEI representative at UW CSE and redesigning the student-led DEI review process for faculty hiring, (2) promoting inclusive teaching practices across two courses, and (3) ensuring speaker diversity in the UW NLP annual speaker series and workshops.

As a DEI student representative in 2022-2023, I played a central role in redesigning the DEI review process for CSE faculty candidates, establishing guidelines for evaluating DEI statements, and facilitating discussions on DEI during student meetings. In my role as head TA for two courses, I introduced systematic changes to support students facing challenges. For an undergraduate AI course, I introduced digital lecture notes for easier class review and collaborated closely with Professor Hannaneh Hajishirzi to accommodate students with disabilities. Additionally, in my nine invited lectures at various universities, I used anonymous Q&A platforms such as Sli.do to encourage questions from students who may feel hesitant to speak publicly. As the organizer of the UW NLP seminar series for two consecutive years, I prioritized inviting speakers from diverse backgrounds in terms of gender, ethnicity, and career stage. Notably, of the 12 speakers invited between 2022 and 2024, five were women, supporting a broader representation within the series.

Moving Forward: I am committed to fostering an inclusive, supportive environment where all students can thrive. Moving forward, I plan to stay active in DEI efforts by joining DEI committees, promoting open communication, and creating spaces for students to comfortably share feedback, allowing me to effectively address their evolving needs. I will ensure that our students reflect diverse genders, races, countries, and backgrounds, and I will fully support all students in their academic and professional growth.

In teaching, I will continue to develop accessible practices, such as anonymous Q&A tools and flexible participation options, to engage students from diverse backgrounds. I also plan to share these methods with colleagues to build a more inclusive teaching culture within the department. Beyond the classroom, I will prioritize diversity in events, from seminars to workshops, to ensure a range of perspectives is represented. As a professor, I aim to establish mentorship opportunities that support underrepresented students in their academic and professional journeys.

Research

The core aim of my research is to make essential information accessible to everyone using state-of-the-art technologies. My work advances language technologies for linguistic minorities, addresses global information inequality, and develops responsible, reliable systems for safety-critical domains.

Developing NLP for Linguistic Minorities: Linguistic minorities face major barriers to essential information, creating disparities in education, healthcare, and social participation [4]. Most scientific and technical content is available only in widely spoken languages like English, limiting access for speakers of underrepresented languages. Additionally, NLP technologies are often developed exclusively for English [5], resulting in suboptimal performance in many world languages [6]. To address information scarcity in many world languages, I pioneered multilingual retrieval-augmented generation, enabling systems to retrieve and generate responses across languages. I developed the first cross-lingual Retrieval-Augmented LM, CORA [7], and the XOR QA dataset [8], which is a large-scale open-retrieval Question Answering (QA) dataset covering seven diverse languages. This work expanded into the first shared task [9] at NAACL 2022, which evaluated multilingual Retrieval-Augmented LMs in 16 languages, including low-resource languages like Khmer and Tagalog. I also co-created AfriQA [10], the first open-retrieval QA dataset for African languages, and conducted a meta-survey on world language datasets [5], suggesting the need for high-quality data across more languages. Most recently, I contributed to Pangea [11], a cutting-edge multilingual, multimodal LM that excels in low-resource languages.

Reliable and Responsible LLMs that Address Real-world Problems: I have analyzed the limitations of parametric LMs and advocated for Retrieval-Augmented LMs to make LM-based systems more reliable and responsible. I showed that LLMs hallucinate more in long-tail cases (e.g., more expert domains, non-North American cultures), alerting the NLP community to their disproportionate harms on certain demographics and use cases. Moreover, my work sheds light on limitations of standard Retrieval-Augmented LMs (e.g., their answers are not always supported by citations), facilitating research on building more advanced techniques to further improve factual precisions of such LM-based systems and applications to expert-domain [12], [13].

Moving Forward: I am dedicated to advancing fair and responsible language technology to ensure equitable access to essential information for people from all backgrounds. Building on my previous work, I will expand efforts in multilingual NLP, particularly for extremely low-resourced languages, by developing accessible, high-quality resources and tools that can serve linguistic minorities. My goal is to further reduce barriers for underrepresented languages by creating more robust models, datasets, and benchmarks that address the unique needs of these communities. I will conduct rigorous research on the critical risks posed by large language models, such as factual inaccuracies, data fairness, and biased outputs. I plan to integrate advanced retrieval-augmented and self-reflective methodologies to enhance transparency, citation accuracy, and accountability in NLP systems, especially in safety-critical domains.

References

[1] Posselt, "Toward inclusive excellence in graduate education: Constructing merit and diversity in phd admissions," *American Journal of Education*, 2014.

[2] Summers and Hrabowski III, "Preparing minority scientists and engineers," *Science*, 2006.

[3] Sawir, Marginson, Deumert, Nyland, and Ramia, "Loneliness and international students: An australian study," *Journal of Studies in International Education*, 2008.

[4] Al Shamsi, Almutairi, Al Mashrafi, and Al Kalbani, "Implications of language barriers for healthcare: A systematic review," *Oman medical journal*, 2020.

[5] Yu*, **Asai***, Chatterjee, Hu, and Choi, "Beyond counting datasets: A survey of multilingual dataset construction and necessary resources," in *Findings of EMNLP*, 2022.

[6] **Asai**, Kudugunta, Yu, Blevins, Gonen, *et al.*, "BUFFET: Benchmarking large language models for few-shot cross-lingual transfer," in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2023.

[7] **Asai**, Yu, Kasai, and Hajishirzi, "One question answering model for many languages with cross-lingual dense passage retrieval," in *Neural Information Processing Systems (NeurIPS)*, 2021.

[8] **Asai**, Kasai, Clark, Lee, Choi, *et al.*, "XOR QA: Cross-lingual open-retrieval question answering," in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2021.

[9] **Asai**, Choi, Clark, Hu, Lee, *et al.*, "The 1st workshop on multilingual information access (mia)," *NAACL 2022 Workshop*, 2022.

[10] Ogundepo, Gwadabe, Rivera, Clark, Ruder, *et al.*, "AfriQA: Cross-lingual open-retrieval question answering for african languages," in *Findings of EMNLP*, 2023.

[11] Yue, Song, **Asai**, Kim, Nyandwi, *et al.*, "Pangea: A fully open multilingual multimodal llm for 39 languages," *arXiv*, 2024.

[12] **Asai**, He, Shao, Shi, Singh, *et al.*, "OpenScholar: Synthesizing scientific literature with retrieval-augmented lms," *Arxiv*, 2024.

[13] Wang, **Asai**, Yu, Xu, Xie, *et al.*, "CodeRAG-Bench: Can retrieval augment code generation?" *ArXiv*, 2024.