My commitment to service arises from a core belief that advancing AI research requires not only technical innovation but also intentional efforts to **strengthen our academic communities**, **support emerging researchers**, and **ensure the responsible development of our field**. Throughout my Ph.D., I have contributed to service across four dimensions: (1) professional leadership within the AI and NLP community, (2) departmental and institutional engagement, (3) public outreach and education, and (4) social responsibility through research. These efforts demonstrate my commitment to building inclusive research communities, fostering scientific exchange, and aligning AI innovation with societal needs.

1 Professional Leadership and Community Service

Conference Tutorial and Workshop Organization I have been deeply committed to strengthening the AI and NLP community by leading workshops and tutorials that address critical challenges at the intersection of AI safety, ethics, and societal impact. I co-instructed the **ACL 2025 Tutorial on Guardrails and Security for LLMs**, which drew over 400 participants and helped establish a shared foundation for safe, secure, and controllable deployment of LLMs. I also co-founded the **AI Meets Moral Philosophy and Moral Psychology (MP2) Workshop at NeurIPS 2023**, which united researchers across computer science, philosophy, and psychology to explore computational approaches to moral reasoning and ethical AI development. Building on these efforts, I organized the **Socially Responsible Language Modeling Research (SoLaR) Workshop at NeurIPS 2024 and COLM 2025**, which continues to unite a global interdisciplinary community dedicated to advancing alignment, safety, and sociotechnical responsibility in AI systems.

Peer Review and Area Chair Roles I have served as a **reviewer for major conferences** including EMNLP, ACL, AAAI, NeurIPS, ICLR, ICML, and COLM, as well as for **journals** such as Language Resources and Evaluation and Applied Artificial Intelligence. I also took on expanded leadership as an **Area Chair** for the Multi-Turn Interactions in LLMs (MTI-LLM) Workshop at NeurIPS 2025, where I coordinated meta-reviews. Across these roles, I have championed rigorous evaluation standards and constructive, equitable reviewing practices that support high-quality, fair scholarship.

Grant Participation and Research Translation As a **key contributor to the DARPA In the Moment (ITM) Program**, I have helped translate foundational research into deployable, high-impact systems. My Delphi project on machine morality served as a seminal catalyst for the establishment of this influential initiative. In this capacity, I contributed to grant proposals, delivered presentations at four PI meetings (including the program kickoff), led quarterly progress reports, and coordinated with partner institutions. This experience has deepened my understanding of how academic research interfaces with funding agencies and has prepared me to effectively lead major funding efforts as a faculty member, supporting both my future lab and students.

2 Departmental and Institutional Service

Admissions and Recruitment I served as an **Area Chair and Reviewer for Ph.D. admissions at UW CSE** for three consecutive years (2021–2023). Recognizing that admissions decisions profoundly shape departmental priorities and future, I worked to promote holistic evaluation beyond traditional metrics. I managed the distribution and review of **over 200 applications annually**, provided detailed feedback to reviewers when bias emerged, and curated shortlists that deliberately surfaced strong candidates from non-traditional backgrounds. I also served as the **Liaison for Faculty Recruiting (2022–2024)**, coordinating student engagement with faculty candidates and ensuring that a broad range of student perspectives informed hiring decisions.

Diversity, Inclusion, and Community Building I contributed to DEI initiatives through multiple channels. As **Student Representative on the UW Allen School Diversity and Inclusion Committee (2022)**, I reviewed diversity statements during faculty searches. As **Co-Chair of the Prospective Student Committee (2022-2023)**, I organized visit days that showcased the inclusive culture of our department and helped admitted students envision themselves as part of our community. I also **co-organized the 2022 UW NLP Retreat**, an annual offsite event that convenes over 200 students, faculty, and external collaborators to foster cross-lab connections and a sense of belonging within the UW NLP research community.

Mentorship Infrastructure Beyond individual mentoring as detailed in my Teaching & Mentoring Statement, I have helped build scalable mentorship infrastructure. As a **Mentor for the Pre-Application Mentorship Service (PAMS, 2021)**, I guided applicants, many from historically underrepresented backgrounds, in preparing research statements and Ph.D. applications, while expanding the program's reach to international students through social and professional networks.

3 Public Engagement and Outreach

Educational Outreach to K-12 Students I am deeply committed to inspiring the next generation of AI researchers and making AI research accessible to young learners. In 2022 and 2023, I delivered guest lectures on "Can We Teach Machines Human Ethics and Values?" to **high school students at The Downtown School in Seattle**, presenting complex AI concepts in an engaging and age-appropriate way that sparked thoughtful discussions on the ethical and societal implications of AI. Beyond classroom outreach, I have **met individually with high school students** who reached out with questions about AI research and **mentored a high school student**, Abhay Gupta (John Jay Senior High School), guiding his research exploration about the field.

Public Communication of Research I have **actively engaged with media to communicate AI research to broad public audiences**, recognizing that clear communication is essential for informed discourse on AI policy and governance. My work has appeared in leading outlets including *The New York Times*, *The New Yorker*, *Vox*, *Nature Outlook*, *Wired*, *IEEE Spectrum*, and *The Guardian*. Through interviews, I offer accessible explanations of technical concepts that help journalists convey AI's capabilities and limitations and support more informed public dialogue on AI safety and ethics.

Open Science and Resource Sharing I am committed to **advancing open and equitable access to AI research and education**. Throughout my Ph.D., I have open-sourced datasets, models, and interactive demos such as Delphi, WildGuard, and WildTeaming that have supported over 4 million queries and 230K downloads. I have also shared extensive teaching materials from my roles as Head TA for UW's CSE 447/517 and CSE 599 D1, along with slides from more than 20 invited lectures at CMU, UCLA, UIUC, KAIST, and the ACL 2025 Tutorial on Guardrails and Security for LLMs. Through these resources, I aim to democratize AI safety and NLP research, broaden participation, and reduce barriers to entry.

4 Social Responsibility through Research

My research advances responsible AI by **developing systems that are safe, inclusive, and morally and culturally grounded**. I design models and benchmarks that operationalize human morality, pluralistic values, and safety robustness in LLMs. Through open-source AI safety toolkits, I promote equitable access to research resources worldwide. Collectively, these efforts bridge technical innovation with ethical responsibility, ensuring that progress in AI benefits diverse societies rather than a privileged few.

Future Service Commitments

Institutional Leadership I plan to contribute actively to departmental committees on admissions, curriculum, and DEI. In particular, I hope to serve on graduate admissions and faculty hiring committees to promote holistic evaluation and help recruit and support scholars whose excellence may not be reflected in traditional metrics. I am also committed to fostering a fair, student-empowering culture by strengthening mentoring structures, advancing equitable evaluation, and helping shape curricula that prepare students for responsible engagement with evolving AI technologies.

Professional Society Engagement I plan to continue organizing workshops and tutorials at major conferences while taking on senior service roles such as program chair, general chair, or member of conference organizing committees. I also aim to contribute to professional societies by serving on boards or steering committees that shape research priorities, promote inclusivity, and guide the ethical advancement of AI.

Policy and Interdisciplinary Bridge-Building I will collaborate with policymakers, industry leaders, and civil society organizations to ensure AI research addresses real-world needs while upholding human values. This includes participating in expert panels, offering technical consultation on AI policy proposals, and fostering dialogue between technologists, social scientists, and other stakeholders.

Mentorship at Scale I will create mentorship programs that extend beyond my research group, including mentorship opportunities for motivated students from underrepresented backgrounds. I will also share publicly accessible online materials to broaden participation and ensure equitable access to mentorship.

Through these service activities, I aim to strengthen the foundations of our research community, broaden participation in AI scholarship, and ensure that our field evolves in ways that genuinely benefit humanity. **Service is not ancillary to my academic mission; it is central to building the collaborative, inclusive, and socially responsible research ecosystem that the future of AI demands.**