Eve Fleisig: Diversity Statement

In teaching, mentorship, and daily life, I have met many bright, wonderful people of different backgrounds who could all be phenomenal computer scientists. It is professionally concerning, and personally frustrating, that only a few were able to pursue this path. Extensive research points to structural factors including systemic educational inequities, insufficient support, stereotyping eroding student confidence, and tacit hidden curricula that students must know to succeed [1, 2, 3, 5]. Personal experience—among students, peers, and my own Hispanic family—has further cemented my belief in the paramount importance of improving equity and belonging in computer science. To do so, I aim to amplify and prioritize underrepresented voices, and draw on evidence-based pedagogical strategies to reduce barriers in the field. In my PhD, I focused on three main areas, which I plan to continue as a professor: **supporting diverse students through mentorship and outreach, prioritizing research that builds a more equitable world, and contributing to collective initiatives for large-scale change.**

Supporting diverse students through mentorship and outreach. I strive to mentor students from diverse backgrounds in teaching and research. I mentored ten undergraduates (of whom six were women and all were students of color) during my PhD. I keep in mind that students often face outside struggles; e.g., unsupportive environments can erode confidence, making frustration harder to overcome [2], and some students face more personal obligations outside of class than others [4]. Thus, I aim to support students with guidance beyond specific projects, such as on applying to PhDs, navigating industry interviews, and seeking research-life balance. I also maintain a repository of “hidden curricula” on how to apply for research positions, find ML coursework, and learn about NLP, so that any interested student can pursue these opportunities.

I am also motivated by findings that students “are considering who they might become if they pursue computer science.... They are trying to determine what the discipline values and whether that aligns with their own values” [4]. I find that students gain confidence when their unique perspectives are valued, and they can connect their research or classwork to issues close to their heart. For example, I taught a student in CS188, Nancy Ma, who had broad interests outside of CS, particularly in literature. I gave her project guidance and encouraged her as she developed an interest in identifying copyright issues on books used for training data; she subsequently did technically impressive work on building models to identify data contamination. I enjoyed guiding another student, Vyoma Raman, to combine her perspectives on disability theory and CS in order to evaluate LLMs for performance on marginalized groups. Through Berkeley’s REU program for HBCU undergraduates, I mentored a student who spoke Amharic, Samuel Ghezae, who then became very enthusiastic about low-resource language modeling; I’ve also been guiding him on applying to graduate programs. I have been very proud to see each of these students pursue graduate studies (Nancy at Columbia, Vyoma at Stanford and Cornell Tech, and Samuel at JHU).

In addition, I prioritize outreach because not everyone who might be interested in CS is inside the CS department, often due to historical inequities [1, 4, 5]. Yet people should not need to be computer scientists to learn about and speak out on machine learning issues, especially as language models (LMs) increasingly impact all sectors of society. I’ve given several guest lectures outside computer science, such as through Berkeley’s linguistics department, business school, and BLC center on language instruction, to reach these communities more directly. I also prioritize projects that involve community outreach, such as the GRACE project on measuring LM overconfidence. We engaged in participatory design with the trivia (quizbowl) community, which I'm already immersed in; I worked with community members to create the dataset and learn from study participants. In the process, I had the pleasure of discussing and demonstrating how LMs work, why they can be confidently wrong, and the resulting societal implications at competitions across the US with a wide range of people, from teenagers in Maryland to seniors in Chicago.

Research that helps build a more equitable world. My research is fundamentally motivated by ensuring that anyone can use, understand, and benefit from language models. My work on addressing societal inequities includes developing FairPrism, a framework and dataset for diagnosing fairness-related harms from language models [Fleisig et al., 2023b]. I have also studied stereotypical demographic personas and use cases attributed to text-to-speech voices [Fleisig et al., 2025], and found that models such as ChatGPT reinforce dialect discrimination [Fleisig et al., 2024]. My work on learning from informative disagreement is also motivated by the fact that majority-vote aggregation often obscures the perspectives of already-underrepresented users [Fleisig et al., 2023a, Fleisig et al., 2024]. Finally, I consider how LMs impact groups that are already disempowered, such as crowdworkers, who often come from low-resource backgrounds. I have proposed strategies to improve data collection design by centering workers' concerns [Huang et al., 2023 - EMNLP Outstanding Paper].

Contributing to collective initiatives for large-scale change. I help facilitate university- and community-wide initiatives that undo structural barriers in research and education. I co-chaired ACL 2024's Student Research Workshop, which provides research mentorship, a dedicated workshop for publication, and travel funds to attend ACL for young researchers worldwide who might not be able to otherwise. This experience underscored to me that addressing inequities in CS involves many moving parts, both individual and collective. In addition to co-organizing the workshop program, I organized the review of 200+ workshop submissions and meta-reviewed 50+ papers. In parallel, reviewing PhD applications for the Berkeley NLP Group made me acutely conscious of the difficulties in encouraging others to consider students from nontraditional backgrounds, because those students often have less conventional paths to the PhD and less support in preparing applications. As a professor, I will be especially conscious to review applications from nontraditional students in depth, and will look to networks such as WiNLP, Latinx in AI, Queer in AI, and Black in AI for recruitment. I have also seen how much support groups can help students adjusting to the PhD; I was fortunate to participate in Berkeley's LAGSES group, which brings together Latine graduate engineers. I'd be excited to get involved with similar clubs, such as by connecting them with speakers and funding, organizing events, or serving as a faculty sponsor.

Empowering diverse voices in computer science is central to my mentorship style, my research direction, and my personal experience. I strongly believe that as computer scientists, we must all work to undo structural barriers and support underrepresented students and communities, thus contributing to collective, positive change.

References

Eve Fleisig, Rediet Abebe, and Dan Klein. 2023. When the Majority is Wrong: Modeling Annotator Disagreement for Subjective Tasks. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6715–6726, Singapore. Association for Computational Linguistics.

Eve Fleisig, Aubrie Amstutz, Chad Atalla, Su Lin Blodgett, Hal Daumé III, Alexandra Olteanu, Emily Sheng, Dan Vann, and Hanna Wallach. 2023. FairPrism: Evaluating Fairness-Related Harms in Text Generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6231–6251, Toronto, Canada. Association for Computational Linguistics.

Eve Fleisig, Su Lin Blodgett, Dan Klein, and Zeerak Talat. 2024. The Perspectivist Paradigm Shift: Assumptions and Challenges of Capturing Human Labels. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Volume 1: Long Papers), pages 2279–2292, Mexico City, Mexico. Association for Computational Linguistics.

Eve Fleisig, Genevieve Smith, Madeline Bossi, Ishita Rustagi, Xavier Yin, and Dan Klein. 2024. Linguistic Bias in ChatGPT: Language Models Reinforce Dialect Discrimination. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 13541–13564, Miami, Florida, USA. Association for Computational Linguistics.

Eve Fleisig, Julian Vargo, Nikolai Schwarz, Veronica Grajeda, Abigail Roberts, Rhosean Asmah, Nicole Holliday. 2025. Perceptions of OpenAI's Whisper Text-to-Speech Technology. *Linguistic Society of America Annual Meeting*.

Olivia Huang, Eve Fleisig, and Dan Klein. 2023. Incorporating Worker Perspectives into MTurk Annotation Practices for NLP. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1010–1028, Singapore. Association for Computational Linguistics.

[1] Paulo Freire. 2017. *Pedagogy of the oppressed.* Penguin Classics.

[2] Victoria Beth Sellers and Idalis Villanueva. 2021. What Strategies do Diverse Women in Engineering Use to Cope with Situational Hidden Curriculum? In *2021 ASEE Virtual Annual Conference Content Access*, online. doi:10.18260/1-2--38051.

[3] Victoria Beth Sellers and Idalis Villanueva. 2023. From Message to Strategy: A Pathways Approach to Characterize the Hidden Curriculum in Engineering Education. *Studies in Engineering Education*, 4(2), pages 176–200. doi:10.21061/see.113.

[4] Parth Sarin. 2023. Levers for Improving Diversity in Computer Science. Center for Security and Emerging Technology.

[5] Shannon Wongvibulsin. 2019. Educational strategies to foster diversity and inclusion in machine intelligence. *Nature machine intelligence*, 1(2), pages 70-71. doi:10.1038/s42256-019-0021-8.