Diversity Statement

Daniel Khashabi

November 2021

Advancing diversity and inclusion is essential to our collective prosperity. For academic institutions to fulfill their mission of empowering future societies, they must reflect the communities they serve. We must strive to create environments that welcome and celebrate people of different races, genders, sexual orientations, religions, nationalities and cultures.

As an immigrant who has been a victim of inhumane immigration policies, I want to be the change I wish to see in the world around me. I plan to tackle these ideas in three different fronts of my life: as a leader, as an individual, and as an academic researcher:

- **(A)** as a **leader**, I promote diversity and inclusion in the management of my projects.
- **(B)** as a **citizen**, I endeavor to give back by promoting access to science and technology.
- **(C)** as a **researcher**, I use research tools to contribute to the discussion of bias & fairness.

(A) Leading a Diverse and Inclusive Team

As a manager, I am committed to ensuring that our educational environments provide an **inclusive** and safe space for everyone, especially minorities and under-represented groups. I have always tried to build an environment where the unique qualities of my mentees and teammates are accepted and appreciated. Inclusion is what attracts future diverse talents and encourages their engagement.

During my research **advising experience**, I have mentored over 10 undergraduate or junior graduate students with a variety of backgrounds (over 30% women; overwhelmingly underrepresented or international students). Most of these students, with my help and their perseverance, have successfully published based on their research experience in prestigious conferences such as ACL and EMNLP.

I have also organized events that aimed to **increase access to science**. This year, I am the volunteer chair of NAACL'22 where diversity (in terms of skill set, background, nationality, gender, etc.) is one of our key criteria for distributing awards. In 2019, I co-chaired an ACL student workshop which provided an excellent forum for student researchers interested in computational linguistics to present their work and receive valuable feedback from the international research community — especially those who did not have easy access to top-notch mentors, typically from top schools.

In the future, I hope to help improve the representation problem in computer science and engineering. Many students, particularly students from underrepresented groups (including women, African-Americans, and Hispanics) face a variety of *structural* and *social* barriers to get exposure to computers and receive an education in CS.[1] To bridge the diversity gap, we need to make CS more accessible to everyone regardless of their background or financial status. I am committed to ensuring that we create research environments that welcome all backgrounds, especially those that have historically been denied.

---

[1] A 2016 Gallup report indicates that Black students are less likely than White students to have classes dedicated to CS at the school they attend (47% vs. 58%, respectively). The article reports that males are more likely than female students to be told by a parent that they would be good at CS (46% vs. 27%). Another study predicts that, with the current pace, it may take more than a century before we get equal participation from males and females in CS (Wang et al., 2020).

(B) Giving Back to the Community

As an individual living in this society and using its benefits to grow, I feel a responsibility to give back.

I have **volunteered to teach** basic computer science in various communities. For a number of semesters in 2018, I worked with Tech-It-Out to teach the basics of web design to middle school students in Philadelphia, a city that suffers from enormous generational social and economic inequality.[2] Similarly, during the pandemic, I gave online lectures about AI to elementary school students through the SkypeAScientist program. After one of these lectures, I received an email from the teacher. An excerpt read *"... was great! I haven't seen [...] that excited about anything for a long time - so that was worth it right there!"* Feedback like this indicates that no matter how small these contributions might seem, they have the power to inspire the students to pursue something bigger for a prosperous future for themselves and their communities.

Looking ahead, I hope to continue to give back to our communities. Seeing the excitement of youngsters learning about computers gives me immense fulfillment and gratitude. I am also interested in working with organizations[3] that focus on refugee resettlement and assimilation efforts. I believe helping refugees learn about the basics of computer and data science can be an enormous benefit for their job prospects.

(C) Evidence-driven Research About Bias & Fairness

I am encouraged by our growing collective awareness of bias & fairness in AI because of the research dedicated to this topic. This is particularly important now that we are witnessing a massive adoption of AI and language technologies by various industries.

In the past few years, I have researched formalizing approaches to **quantifying social stereotypes in language models** (Li et al., 2020). The findings of this project[4] indicated troubling trends about biases of our cutting-edge technologies, especially regarding larger models which are known to lead to better performance. Furthermore, we have studied the **alignment of model values** with human values (Christian, 2021; Zhao et al., 2021), which has proven to be difficult and demands more research before we can make them a reality.

Last but not least, I have developed algorithms for discovering **diverse perspectives** related to controversial topics (Chen et al., 2019). In a world that suffers from polarization and information silos, we need AI-driven systems that help us see the perspectives of those who might be thinking differently than us.

In the coming years, I would like to lead research into the **unintended consequence of AI algorithms** in perpetuating bias. There are many instances of racial disparities in the quality of language technologies (Kirkpatrick, 2020; Demszky et al., 2021). There are many others that we probably don't know, as new technologies emerge and people adapt their usage patterns.

Beyond biases of our models, I see opportunities in **human-AI partnership** for improving societal fairness. For example, algorithms that observe our behavior (say, in interactions with others) can discover blind spots in our unconscious biases (Voigt et al., 2017). I want to play a role in introducing such tools to help improve our collective behavior.

Last but not least, the quest for more diversity and inclusion is not a sprint, but a marathon. Maintaining these values requires significant effort and diligence, as they are often endangered by many factors. We must actively rethink our advocacy, and reevaluate our goals as we make further progress.

---

[2] The poverty rate in Philadelphia is roughly 26%, while the national average is around 13% (Fontenot et al., 2017).
[3] Such as IRC (https://www.rescue.org) and Upwardly Global (https://www.upwardlyglobal.org).
[4] https://unqover.apps.allenai.org

References

S. Chen, **D. Khashabi**, W. Yin, C. Callison-Burch, and D. Roth. Seeing things from a different angle: Discovering diverse perspectives about claims. In *Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL), 2019.

B. Christian. *The Alignment Problem: How Can Machines Learn Human Values?* Atlantic Books, 2021.

D. Demszky, D. Sharma, J. H. Clark, V. Prabhakaran, and J. Eisenstein. Learning to recognize dialect features. In *Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL), pages 2315–2338, 2021.

K. Fontenot, J. Semega, and M. Kollar. Income and poverty in the united states. 2017.

K. Kirkpatrick. Natural language misunderstanding. *Communications of the ACM*, 63(11):17–18, 2020.

T. Li, **D. Khashabi**, T. Khot, A. Sabharwal, and V. Srikumar. Unqovering stereotypical biases via underspecified questions. In *Conference on Empirical Methods in Natural Language Processing* (EMNLP) - *Findings*, pages 3475–3489, 2020.

R. Voigt, N. P. Camp, V. Prabhakaran, W. L. Hamilton, R. C. Hetey, C. M. Griffiths, D. Jurgens, D. Jurafsky, and J. L. Eberhardt. Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences*, 114(25):6521–6526, 2017.

L. L. Wang, G. Stanovsky, L. Weihs, and O. Etzioni. Gender trends in computer science authorship. *Communications of the ACM (CACM)*, 2020.

J. Zhao, **D. Khashabi**, T. Khot, A. Sabharwal, and K.-W. Chang. Ethical-advice taker: Do language models understand natural language interventions? In *Annual Meeting of the Association for Computational Linguistics* (ACL) - *Findings*, 2021.