

I strongly believe that diversity and inclusion are fundamental to the responsible development of AI and to the field of computer science more generally. My commitment to these principles has been shaped by my international academic journey across Ukraine, Germany, and Switzerland, as well as my experiences working with students and communities from different backgrounds. As AI systems are getting more closely integrated into daily life, identifying their potential harms—sometimes quite subtle—and ensuring equitable impact across all communities is becoming increasingly important. Through my research in AI safety and collaborations with industry and government organizations, I have clearly realized that varied viewpoints and experiences are essential for developing AI systems that serve all of society. As a faculty member, I intend to raise awareness about these issues and work to support underrepresented communities through mentorship and outreach activities.

### Diversity Philosophy

The development of safe and reliable AI systems demands diverse perspectives and experiences to prevent systematic failures that could disproportionately impact certain demographic groups. Throughout my research career, I have witnessed both the challenges and opportunities in making computer science more inclusive. I have had a chance to work closely with excellent women researchers, both students and faculty members. However, women *still* remain underrepresented in the computer science field, which can limit the diversity of perspectives in research and potentially lead to biased and non-inclusive systems.

My experience working with displaced children from Ukraine vividly illustrated how AI systems can fail when designed from limited perspectives. For instance, I observed content filters, though intended to prevent misuse, inadvertently blocking legitimate content about these children's homeland. For example, at some point, it was impossible to generate *anything* related to Ukraine with a popular text-to-image generation model. This experience emphasized how seemingly neutral safety measures can have unintended consequences for vulnerable populations, reinforcing the importance of comprehensive evaluation across different use cases and contexts.

This commitment to diversity has shaped my mentoring philosophy. Throughout my PhD, I have supervised 13 students from varied backgrounds across multiple institutions, including EPFL, ETH Zürich, and the University of Tübingen. Rather than relegating students to secondary tasks, I emphasize empowering them to take leadership roles in their research projects. This approach has proven particularly valuable in AI safety research, where fresh perspectives often yield crucial insights. My students' achievements reflect the success of this strategy, with several publishing at top-tier conferences and earning academic recognition.

In teaching, I will prioritize creating an inclusive learning environment that accommodates diverse learning styles and backgrounds through interactive, demonstration-based methods. My experience across different cultures and educational systems has helped me develop pedagogical approaches that resonate with students regardless of their prior exposure to computer science or AI. I hope this teaching style will ensure that all students can engage meaningfully with the material while building confidence in their abilities.

### Outreach and Community Support

My outreach activities of the last few years have been mostly centered around helping people from Ukraine affected by the war. Since I am originally from Ukraine, I felt that this is where I could meaningfully contribute by teaching basic AI concepts to children, helping refugees to find temporary positions and integrate into a new academic environment.

As a national coordinator for #ScienceForUkraine in Switzerland, I have worked extensively to help displaced researchers and students continue their academic careers. This role has involved coordinating with multiple institutions, evaluating applications, and providing guidance to scholars navigating challenging transitions and cultural differences. During my experience as an admission officer for the Ukrainian Global University, I interviewed numerous talented students whose education had been disrupted by the war. Many of these students showed re-

markable resilience and creativity in continuing their studies despite extremely challenging circumstances. This experience reinforced my belief in the importance of creating pathways and opportunities in academia for those facing extraordinary challenges.

I organized AI lectures for displaced children during a summer camp in Cisnădioara, Romania organized by my friends, emphasizing interactive demonstrations to maintain their interest in STEM education during a difficult period. These lectures revealed both challenges and opportunities in making AI education accessible. The children's enthusiasm and quick grasp of technical concepts demonstrated the importance of early exposure to technology, regardless of circumstances. One particularly memorable moment was when a young participant became fascinated with prompt writing, spending a lot of time crafting increasingly sophisticated interactions with an AI system—a reminder of how natural technical aptitude can be found in unexpected places when we provide the right opportunities.

Beyond these efforts, I have contributed to broadening participation in AI research through service on program committees for 19 workshops at major conferences (ICML, NeurIPS, ICLR, CVPR). These workshops provide opportunities for emerging researchers to present work in progress and receive mentoring from more experienced colleagues. This service has given me valuable insights into the challenges faced by researchers from different backgrounds.

## Future Commitments

As a faculty member, I plan to advance diversity and inclusion through three key areas of impact:

- **Building Inclusive Research Teams:** I will actively work to recruit and retain students with different backgrounds in AI safety research, creating an environment where every member feels empowered to contribute their unique insights to addressing crucial challenges in AI development. This includes establishing mentoring networks and ensuring adequate support both in research and personal development.
- **Expanding Educational Access:** I will develop course materials and teaching methods that accommodate different learning styles, with particular focus on introductory AI courses where we can encourage students from all backgrounds to engage with technical challenges. I will continue international outreach activities and create opportunities for students from underserved communities to participate in research projects.
- **Advancing Equitable AI Development:** I will promote inclusive research practices by organizing challenges that welcome broad participation in AI safety research, fostering interdisciplinary collaborations, and ensuring our research considers impacts across different communities. This includes continuing work on developing comprehensive evaluation frameworks that help measure and mitigate potential negative impacts on different user groups.

The rapid advancement of AI technologies makes it increasingly important to ensure that the field benefits from different perspectives at all levels—from early education to advancing frontier AI models. I view fostering diversity and inclusion not just as an ethical imperative, but as a practical necessity for developing robust and reliable AI systems. By creating an inclusive environment and actively supporting underrepresented groups, I believe we can successfully advance towards this goal.

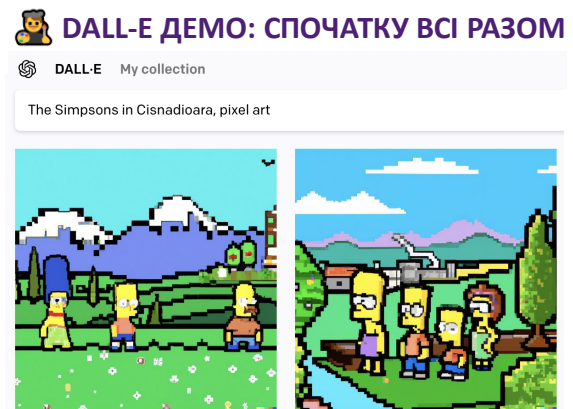


Figure 1: A slide from my AI lectures for displaced Ukrainian children. Interactive demos with generative AI tools, such as DALL-E, worked particularly well with children and kept them highly engaged.