March 30, 2025

Professor Sina Fazelpour
Assistant Professor of Philosophy and Computer Science
Northeastern University
Ethics Institute
Boston, MA 02115

Dear Professor Fazelpour,

Thank you so much for talking with me Friday. I'm really looking forward to being in Boston and sincerely hope we can work together. In this letter, you'll find a brief bio and description of my research. Following that, I've listed three possible joint projects I'd like to work on if I were to join your lab. Of course, these ideas are not meant to be exhaustive or restrictive. The hope is only that they give you an idea of the directions I see my work moving in.

Bio: As you know, I'm a Philosophy PhD student at Vanderbilt University and am also involved with the ANU MINT Lab. My research sits at the intersection of philosophical analysis and applied AI ethics. My work comes to AI ethics by way of long-standing engagement with Islamic ethics and a fairly thorough classical training in Greek philosophy. These foundations explain some of my interest in value pluralism, discussed below. Most recently, I've been working in contemporary ethics and political philosophy, focusing on non-ideal theory, as you'll see in my writing sample. I'm also just spinning up a project now with a friend and colleague in the sociology department to use topic modeling and other computational methods to study the relationship between funding sources and the ideological content of AI governance proposals.

My technical proficiency in AI is strong, and I'm still learning! I began by developing NLP assistants for my work in ethics, which gave me firsthand experience with the practical challenges of aligning AI systems with human values. I've built several text analysis and content-scraping tools using Python (leveraging text-embeddings, semantic similarity, and multi-shot API prompting), which has provided me valuable insights into how AI-generated outputs can influence human understanding. I've also developed some simple tools for scraping and translating Arabic texts at scale, tools for conducting cross-linguistic source-analysis, and I'm currently working on an agentic tool to allow MINT Lab members to dialogue with a database of texts drawn from the #news channel. Last weekend I also built a simple ABM to model the effects of bill-splitting behavior on friend-group stinginess, just for fun and as a way to explore ABM a little more. Some of the tools I've built can be seen here: https://campattison.github.io/bio/ai-tools.html.

Conceptual Analysis of Value Pluralism: In recent months there has been an increase in interest in the degree to which foundation models can approach value pluralism. Researchers find the responses of LLMs to be well aligned with high SES liberal values characteristic of those who design and deploy foundation models, and seek to design models that instantiate a more pluralistic approach to ethical values. Some attempts have been made along these lines, but there remains a conceptual gap between the practical implementation of value pluralism and the philosophical analysis of value pluralism. I'd like to bridge this gap by developing a more precise and comprehensive conceptual analysis of value pluralism, along the lines of what you did with Maria De-Arteaga in defining diversity. Conferences that would be interested in this kind of work include the Paris Conference on AI & Digital Ethics (PC-AIDE), the TeXne Conference (TeXne), the International Conference on Large-Scale AI Risks (ICSAIR), and the AAAI Conference on AI, Ethics, and Society (AIES).

Fairness in Language Models: While the fairness literature has ballooned in ML in recent years, there has not been the same level of attention paid to fairness in the literature concerning LMs in particular. I'd like to change this by developing a more precise and comprehensive conceptual analysis of fairness in LMs. Some attention is paid to this issue in recent system cards etc. but they tend to focus on contributions dating to 2021 and earlier. The almost exclusive focus in these documents on data blinding and similar issues leaves a gap in the literature that could be filled by a carefully written guide to sociotechnical fairness evaluation. I have not seen any work that attempts to give guidance on how to evaluate LM deployments in context (per the kind of non-ideal, sociotechnical evaluations you've discussed in the past). Deliverables from this kind of project could thus include a guide to fairness evaluation both pre- and post-deployment, a set of tools for conducting such fairness evaluations, and a series of case studies. It might also be interesting to develop a detailed prompt for use in no-code platforms like Cursor and Windsurf. Amateur use of these tools that allow users to generate complex code with natural language prompts has already been demonstrated and this opens up new areas of risk for fairness (and everything else!) in the lifecycle of apps developed using these tools. Venues that are interested in this kind of work include the ACM Conference on Fairness, Accountability, and Transparency (FAccT), the Workshop on Advancing Fairness in Machine Learning (WAFML), the European Workshop on Algorithmic Fairness (EWAF), the Workshop on Sociotechnical AI Governance (STAIG@CHI), and NeurIPS.

Harm-reduction in Goverment Use of AI: Given the recent layoffs in government, there may be an opportunity to focus on a few government agencies and conduct an exposure analysis of risks associated with individual use of AI. The underlying assumption here is that major reductions in workforce size will lead to a greater reliance on AI tools, and that this will be done without a thorough risk assessment. I'd be interested in working with you to develop a typology of harms that might be associated with the use of AI in government, and in developing a framework for assessing and mitigating these harms. We'd focus narrowly on certain use cases, develop a full understanding of the types of tasks associated with those bureaus, and develop a sociotechnical analysis of the risks associated with the use of AI in those specific contexts. The topical nature of this project making choosing a venue a little more difficult, but I'm sure we could find a good fit if we were to pursue this. FAccT would likely be interested.

I expect that these projects may be a little too large for a single summer, but I hope they inform you as to what I find interesting and exciting here. As also noted above, these suggestions are not listed as the only possible projects! I look forward to learning from you and your team and would be happy to play any catchup I need to prior to arriving in Boston. If either of the above proposals are of interest or if they prompt adjacent inquiries, please let me know. I'm also happy to fill out these suggestions with lit reviews etc. so that we can hit the ground running come summer.

As we discussed, I'm committed to being in Boston for the entire duration of the internship. If there's anything else you'd like to talk about, please don't hesitate to reach out. Of course, if things don't align well this year, I completely understand and would be very happy to stay in touch for future opportunities, or for anything I might be able to do while I'm in Boston for AIDE.

Thanks so much for considering my application and I look forward to hearing from you soon.

Very best,

Cameron Pattison
PhD Student, Vanderbilt University
cameron.pattison@vanderbilt.edu — (802) 461-5799