Cover Letter: Jaydeep Borkar
jaijborkar@gmail.com
Machine Learning Internship program, Cohere http://jaydeepborkar.github.io/

I'm a Computer Science Ph.D. student at Northeastern University in Boston and one of the organizers of the Trustworthy ML Initiative. My research interests fall broadly in the field of trustworthy machine learning focusing on specific areas like privacy, security, and interpretability. Lately, I've been learning more about the capabilities of large language models and the accompanying issues for their safe deployment. In order to contribute to this specific space, I'm interested in tackling problems on both the engineering and research side to make LLMs safe and more efficient.

Why Cohere?

Vision Large language models are computationally very expensive to train from scratch which makes it difficult for different entities to afford it. I like Cohere's vision of offering the capabilities of large language models to clients for different use cases to adapt the models for a variety of tasks. Needing no machine learning expertise to leverage the capabilities of LLMs at a massive scale makes it even more impactful as it pushes the boundary and doesn't limit the usage to only people with machine learning expertise. Giving access to LLMs through simple and easy-to-use APIs gives different organizations an opportunity to do meaningful work in society by leveraging the power of LLMs. I share Cohere's vision of democratizing the resources and capabilities of LLMs.

Values I have been admiring Cohere for AI's efforts to form a community across the globe to create more "entry points" for machine learning research. Democratizing the opportunities to conduct research and re-defining by whom the research is done is incredibly important in my opinion. On a technical side, I'm intrigued by Cohere's efforts on making large language model training more efficient and greener. At Cohere, I'd like to contribute to the efforts to reduce the carbon footprint by training more efficient and smaller models. I would like to be a part of the environment where I get to take multiple roles such as in engineering, research, and community efforts. I believe Cohere would be a perfect fit to grow in all of these areas.

Why am I a good fit for Cohere?

Shared values and vision Taking lessons from my own journey, I really believe in creating more opportunities and lowering the entry barriers for getting into research, especially for people from low-resource environments. I had to overcome a lot of challenges to get into research as I had no resources to do hands-on research during my undergrad. I cold emailed at least 50 people for a year before getting my first research collaboration. This makes me appreciate all the community driven efforts even more. These experiences helped me develop qualities such as perseverance and pushing myself amidst uncertainties. My strengths that I always back myself on are kindness and empathy. I believe some of these values would make me a caring and thoughtful co-worker at Cohere.

Making LLMs safer and more efficient I'm interested in solving problems related to efficient training and safe deployment of large language models such as model compression, reducing carbon emissions, and privacy/security issues such as training data extraction attacks. I have followed along and enjoyed reading some of Cohere's recent works on compression (Ogueji et al., 2022) and efficient training (Yoo, Perlin, Kamalakara, & Araújo, 2022). There has been a very important work on extracting training data from large language models (Carlini et al., 2020). But not a lot of work has focused on studying training data extraction attacks on fine-tuned models and if we can extract the pre-trained data after fine-tuning phase. Lately, I've been trying to frame and organize my efforts to work on this problem. Through projects and self-learning, I have developed skills that will be useful to work with LLMs. I have gotten familiar with Transformers and libraries such as PyTorch and JAX. Working on my PhD so far has given me skills like organizing projects, and working independently as well as in a team as a joyful teammate.

Taking initiatives I like taking initiatives whenever I feel the need of one in order to solve or improve something. Two years back, I figured that there was no dedicated channel on Twitter to disseminate latest research and news on trustworthy machine learning. There were handles for HCI and NLP and other areas, but there was nothing on trustworthy ML. This led me to creating a Twitter handle (@trustworthy_ml) and disseminate all the research and resources through the channel. After some months I met a couple more like-minded folks and we ended up launching the Trustworthy ML Initiative to lower the entry barriers for

getting into trustworthy machine learning. With the field growing so quickly, I realized the need to highlight up and coming researchers and so I also initiated a Twitter highlight series to highlight students and post-docs (specifically from lesser known institutions).

I admire Cohere for its work on making LLMs efficient, community efforts, and its core values in order to create a diverse team. I'd be overjoyed to get an opportunity to learn and grow at Cohere with these shared values.

Sincerely,
Jaydeep

*

References

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... Raffel, C. (2020). *Extracting training data from large language models.* arXiv. Retrieved from https://arxiv.org/abs/2012.07805 doi: doi:10.48550/ARXIV.2012.07805

Ogueji, K., Ahia, O., Onilude, G., Gehrmann, S., Hooker, S., & Kreutzer, J. (2022). Intriguing properties of compression on multilingual models. Retrieved from https://arxiv.org/abs/2211.02738 doi: doi:10.48550/ARXIV.2211.02738

Yoo, J., Perlin, K., Kamalakara, S. R., & Araújo, J. G. M. (2022). *Scalable training of language models using jax pjit and tpuv4.* arXiv. Retrieved from https://arxiv.org/abs/2204.06514 doi: doi:10.48550/ARXIV.2204.06514