TELL US ABOUT YOUR RESEARCH INTERESTS AND THE SORTS OF TOPICS YOU WOULD LIKE TO EXPLORE IF YOU WERE TO JOIN DEEP MIND

I'm interested in fairness and bias in natural language processing (NLP) applications. I am particularly interested in collaborating with Sebastian Ruder, Shakir Mohamed, Marie-Therese Png, and William Isaac to help to extend their work on text representation, transfer learning and ethical AI towards the evaluation of bias and harms in NLP applications. For example, I am interested in the following projects:

1. **Bias in word embeddings:** Using counterfactual datasets to understand the influences of stereotypical social bias in word embeddings (fixed and contextualized) on the downstream tasks. This will lead to developing the most effective ways to mitigate the bias instead of superficially covering it up.
2. **Intersectionality of Bias:** Investigating how to include the intersectionality of sexism and racism in developing fairer methods to measure gender bias in ML and NLP models. This research project will lead to understanding how NLP models exhibit gender bias towards women from different backgrounds, while the current measures of gender bias are tailored only towards white women.
3. **The bias introduced by crowd workers:** Studying the bias of crowd workers who label datasets when it comes to hate speech versus freedom of speech. In particular, studying the bias of crowd workers labelling certain textual content as islamophobic, anti-sematic, or anti-feminist while it criticizes extreme religious interpretations, government policies, or the history of a movement's decisions and racial inclusion. This project should help lead to developing hate speech detection models that block actual hateful people instead of blocking people who express their freedom of speech rights to criticize particular views or decisions respectfully.

These projects fit with the research goal of the safety research theme at DeepMind that aims to understand the behaviour of AI systems, including unintended behaviours and side effects.

PLEASE LIST YOUR TOP 5 RESEARCH INTERESTS SEPARATED BY COMMAS

Bias and fairness in AI, Intersectionality of Bias, Ethics and social exclusion in AI, Hate speech, Transfer learning.

IS THERE ANYTHING ELSE YOU'D LIKE TO ADD IN SUPPORT OF YOUR APPLICATION?

I'm passionate about supporting my communities, especially under-represented people. This is why I started the women_in_nlp talk series in February 2021. This is a monthly event where I invite women who successfully carved their career path in NLP, either in academia or in the industry, to share their experiences and advise early-stage colleagues. It is open to everyone, not only women. The idea is to show the success stories of women in NLP, which hopefully inspire younger women to pursue a career in NLP. Since February, I have hosted 7 talks of inspiring women: Rachael Tatman (Rasa), Khyathi Chandu (Carnegie Mellon University), Vered

Schwartz (Allen AI), Jasmijn Bastings (Google), and Sabine Weber (University of Edinburgh), Maria Antoniak (Cornell University), and Alexandra Olteanu (Microsoft Research).

I also volunteer at the Scottish Informatics and Computer Science Alliance (SICSA) as a PhD peer support member, where PhD students in computing science schools across Scotland can talk and express themselves in a safe, friendly, and nonjudgmental environment.

I recently joined the volunteers' team for the ACL year-round mentorship scheme, which aims to support junior researchers in NLP and promote equal access to advice on NLP career choices, and initiating research in NLP.