Measuring Bias in Language Models **Against Overlooked Marginalized Groups in Regional Contexts UNDP** Artificial Intelligence Trust & Safety Re-Imagination Programme

Fatma Elsafoury 5.05.2025



Introduction

Fatma Elsafoury

Post-Doctoral Researcher at Weizenbaum Institute. I work on ulletbias and unfairness in AI and Language models. In particular I study bias against marginalised communities in the Arab world and Arabic Language technologies.

David Hartman

A PhD Researcher at Weizenbaum Institute.







Context Setting

Importance



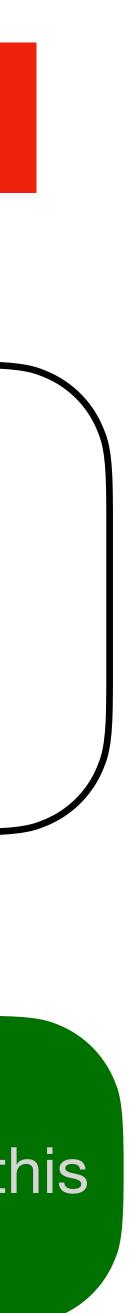
- systems or studying its negative impact.
- It serves marginalised groups in the Arab world.
- Governments, NGOs and civil societies can use my research.

Experience

research.

The under-representation of Arabic as a language and as an identity in developing AI

 I'm an Arab woman and a postdoctoral researcher in bias and fairness in AI which gives me the academic, technical and personal experience to work on this





Key Research Findings

- 2. The importance of **including marginalised groups** in Arab world in **the** or further stereotype them in a negative way.
- 3. The impact of **power dynamics within the Arab region** on the Arabic Al dialects and political and economical interests.

1. The importance of building AI tools using authentic Arabic data and not translated data to avoid reproducing negative stereotypes about Arabs.

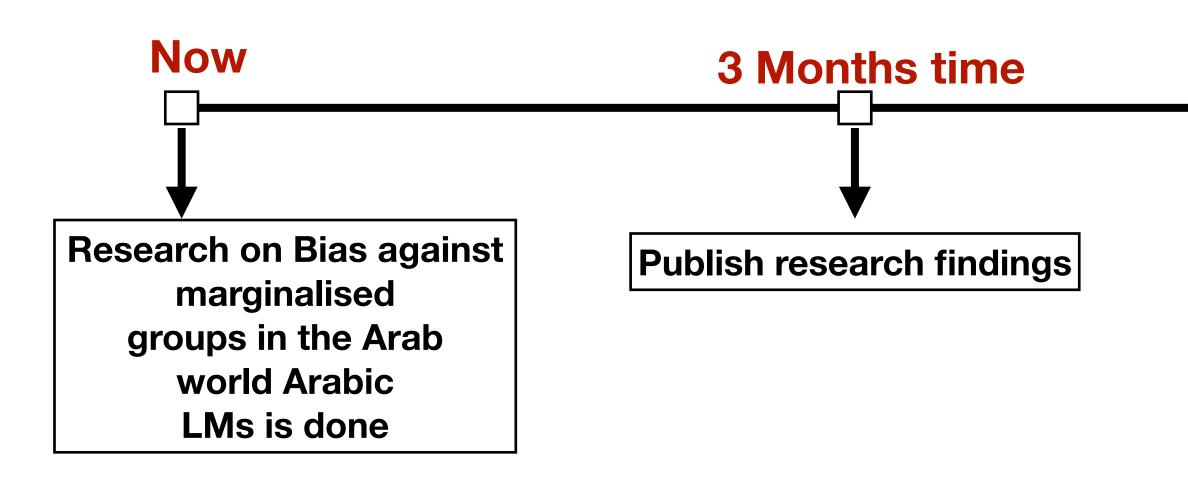
discussion around AI systems and how they might negatively impact their lives

systems and how their might be used as tool to represent specific nations, cultures, dialects or political agendas and exclude other Arab nations, cultures,





Research Plan



12 Months time

Do research on AI governance in the context of the Arab world with a potential outcome to provide a framework for AI governance specific for the region that works for the benefits of the Arab world and its rich diverse communities.



