Algorithmic Fairness in Clinical Natural Language Processing: Challenges and Opportunities

Daniel Anadria, Anastasia Giachanou, Jacqueline Kernahan,
Roel Dobbe and Daniel Oberski

Utrecht University

UMC Utrecht

TU Delft

We are

Daniel Anadria

Anastasia Giachanou

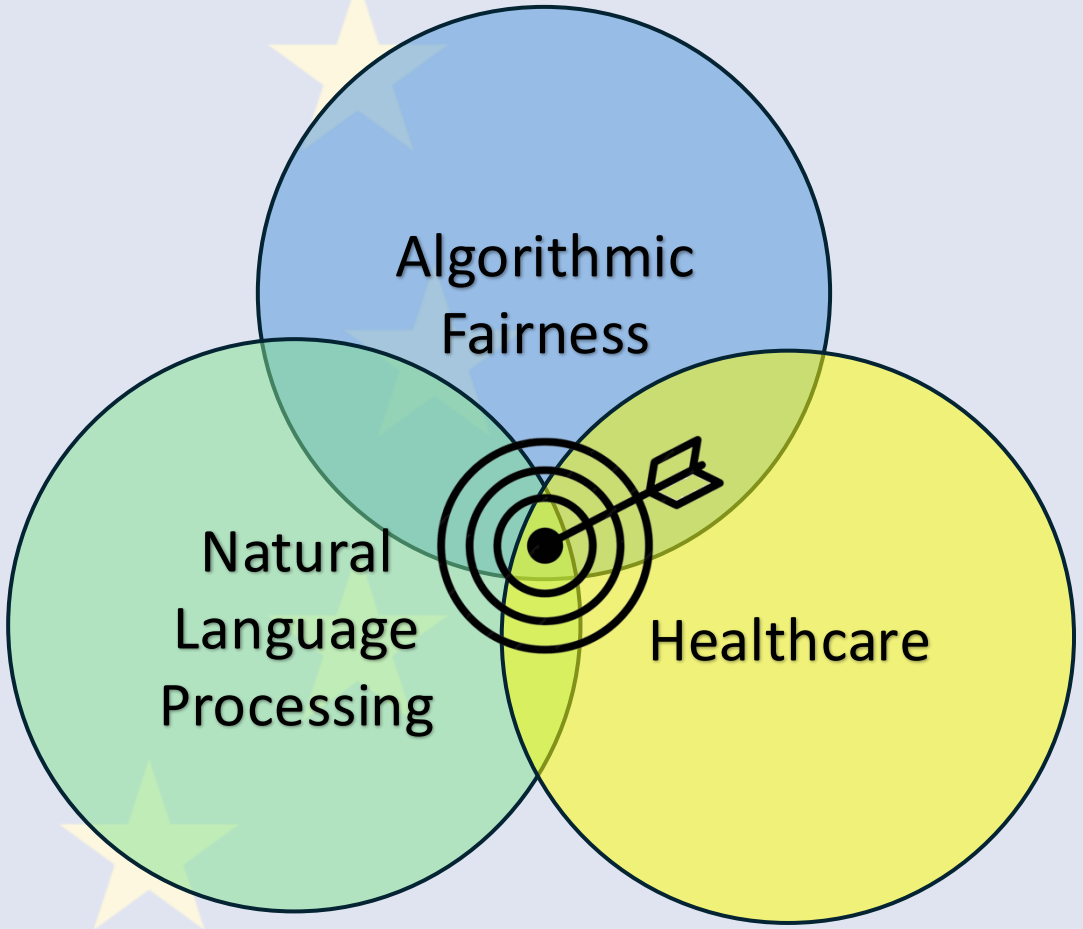
Jacqueline Kernahan

Roel Dobbe

Daniel Oberski

What and Why?

- **Identify gaps in knowledge:**
- Conducted a scoping review of research at the intersection of algorithmic fairness and natural language processing in the clinical domain
- **Discover opportunities:**
- Identified four key areas that require further input from the research and regulatory community.
- **Set the research agenda** aimed at closing the identified gaps



Gap 1: Protected Groups

- Examined groups are narrow in scope (sex, ethnicity/race, and age). Studies focus on the American landscape. Vulnerable groups such as individuals with various forms of disability, mental health diagnoses, or traditionally overlooked groups such as individuals admitted during the weekend as opposed to on a weekday remain understudied.
- Limited attention is given to the differences in geographical and cultural context on which local groups ought to be protected.
- Studies rarely report how the protected attribute was constructed. Some EU healthcare systems encode very limited demographic information.

Gap 2: Method Selection

- Researchers rarely report their motivations for the selection of particular operationalization of fairness and bias mitigation methodology.
- Not every computationally feasible approach has clinical legitimacy. Some of the proposed bias mitigation approaches inadvertently erase medical signal together with the attribute information.
- The presence of bias should be corroborated with an understanding of its source as this can inform the appropriate mitigation approach.

Gap 3: Data Sharing and Privacy

- Acquisition of clinical datasets is a major challenge, especially for fairness auditing as protected attributes tend to be anonymized.
- Construction of accurate outcome labels for supervised learning tasks requires medical expert input which is expensive and time-consuming.
- Publicly available real-world datasets are very limited but necessary for the development of fairness tools and methodologies. Patient privacy concerns need to be addressed, perhaps with synthetic data approaches. Transfer learning and weak supervision could help alleviate the problem of the missing gold standard.

Gap 4: Generalizability

- MIMIC and MIMIC-derived datasets represent the vast majority of publicly available free text clinical data. We have identified only three publicly available datasets not based on MIMIC-notes.
- While some of the studies had access to non-public data, in all studies the hospitals were based in the US. Supplementary search of Physiobank for publicly available medical databases has revealed that the only languages with representation other than English were Spanish and Portuguese, each with a single database.
- There's a major gap in languages other than English, and countries other than the US. It is unknown how well the existing bias detection and mitigation approaches generalize across languages and countries.

Let’s chat in the break!