NAME : ALEXANDRU NICOLAE POSTU
EMAIL : A.N.POSTU@STUDENT.TUDELFT.NL

SUPERVISORS:    SARAH CARTER
                JIE YANG
                STEFAN BUIJSMAN

# FEMINIST CONSIDERATIONS ABOUT FAIRNESS IN MACHINE LEARNING
## WHAT VALUE IS THERE IN BRIDGING FEMINISM AND THE HEGEMONY WHEN IT COMES TO MACHINE LEARNING FAIRNESS?

# 1. CONTEXT

## MACHINE LEARNING (ML) BIAS

- **Over 1000 papers on ML fairness have been published per year since 2021** (Kheya et al., 2024).
- Scholars have created fairness metrics that evaluate how biased algorithms can be.
- These metrics cannot be satisfied all at once, leading people to believe that complete fairness is unachievable.

## COMMON CRITIQUE POINTS

- There are too many fairness metrics, making it difficult for developers to choose one.
- Fairness seems to be overly interpreted as a technical issue, leading scholars to neglect the social dimensions of it.
- Fairness discussions do not meaningfully analyse the impact discrimination has on the marginalised.

# 2. FAIRNESS IS SUBJECTIVE

- Fairness as a term is open to interpretation, fact proven by the different and contradicting metrics.
- Different stakeholders prefer different fairness metrics.

## HEGEMONY

- There is a dominant shared system of ideas and ethics within a period of time (Lauderdale & Amster, 2008).
- This system makes it difficult for opposing views to emerge (Bates, 1975) by labelling them as not scientific enough.
- The idea that ML fairness is a problem that can be solved technically is a hegemonic thought.
- This idea, being hegemonic, is seen as dominant, making it harder for solutions which emphasise social context to emerge.

# 2. FEMINIST PERSPECTIVE

## THE FOCUS IS ON THE MARGINALISED

- Feminism critiques hierarchical structures and the interplay with gender, race, class and heterosexism dynamics (Allen, 2022).
- A feminist perspective is focused on the well-being of marginalised groups.

## FAIRNESS FOR WHO?

- Using black feminist rhetoric, I argue that the stakeholder that should be prioritised is the marginalised groups.
- Through this line of thought, I highlight corporate lobbying practices that influence the hegemonic fairness research.

## FAIRNESS FOR WHAT?

- It is important to consider the context that an algorithm is developed in.
- If a social system is riddled with discrimination already, adding a ML tool will either keep the status quo or worsen it.

## FAIRNESS BY WHO?

- Knowledge is situated and informed by lived experience.
- Knowledge produced by people from diverse backgrounds is superior to the hegemonic knowledge, because it encompasses more aspects of society.
- If we want fair ML to include all groups, we should strive for diversity in research and companies as well.
- However, this will only be successful if we also reshape the hierarchical structures that exist in the research and corporate environments.
- In particular, developers feel unsafe advancing fair ML practices within companies (Madaio et al., 2020).

# 3. BRIDGE BUILDING

## FEMINIST ETHICS OF CARE

- The main idea is to reshape societal structures by attentively listening to all stakeholders.
- **For the marginalised,** it is important to pay more attention to already existing systemic discrimination practices and to prioritise them in ethics discussions.
- **For the ML developers,** it is important to make sure that future fairness solutions are implementable and to create an environment which allows for fairness research.

# 4. BRIDGE BUILDING USES

## DECIDING IF AN ALGORITHM IS FAIR

- Analysing who benefits the most from a fairness metric allows us to choose for a specific fairness metric.
- Analysing the context that an algorithm is developed in might help us decide that an algorithm is unfair by definition.

## ANALYSING FUTURE SOLUTIONS

### EUROPEAN UNION AI ACT

- It is important to externally pressure corporations. The EU AI Act is a step in the right direction.
- It is unclear how the risk classification was performed. There are accusations of lobbying.
- It is important to listen to both developers and the marginalised, but while avoiding unethical lobbying.
- The regulatory body should have a diverse staff in order to ensure a perspective over fairness that encompasses everyone.

## INTERDISCIPLINARY ETHICS EDUCATION

- The TU Delft bachelor's ethics course emphasises the need for interdisciplinarity, which is good.
- However, courses should incorporate more critique coming from social sciences.

# 5. CONCLUSIONS

- The hegemony focuses too much on the technical aspect, although a social context is needed.
- Feminism, which focuses on the experiences of the marginalised, provides the necessary context.
- The three feminist considerations help us decide on a metric or if an algorithm is fair or not.
- Ethics education is important. In particular, interacting more with social sciences will lead to engineers who can think more about the context of their ML solutions.

# 6. FUTURE WORK

- Anonymous surveys that explore the needs of developers are crucial towards ensuring implementable fairness guidelines and solutions.
- Mechanisms that can make the EU Committee resistant to unethical lobbying practices should be researched.
- Surveys evaluating how effective ethics education is at the moment would be a good starting point for improving current ethics courses.

# REFERENCES

- Allen, A. (2022). Feminist perspectives on power (E. N. Zalta & U. Nodelman, Eds.)
- Bates, T. R. (1975). Gramsci and the theory of hegemony.
- Kheya, T. A., Bouadjenek, M. R., & Aryal, S. (2024). The pursuit of fairness in artificial intelligence models: A survey
- Lauderdale, P., & Amster, R. (2008). Power and deviance. In Encyclopedia of violence, peace, & conflict
- Madaio, M. A., Stark, L., Wortman Vaughan, J., & Wallach, H. (2020). Co-designing checklists to understand organizational challenges and opportunities around fairness in AI.

TU Delft