# Ethics in AI, human rights and equality

**Rose Nakasi**
**Makerere University, Uganda**

22nd/6/2021

# Ethics is Important

Ethical AI is artificial intelligence that adheres to well-defined ethical guidelines regarding fundamental values, including such things as individual rights, privacy, non-discrimination, and non-manipulation.

The application of AI can help organizations;

- Operate more efficiently
- Produce cleaner products
- Reduce harmful environmental impacts
- Increase public safety, and improve human health.

However, misuse of AI/ML can cause problems or unintended consequences

# Examples of Bias

World Health Organization

Amazon hiring

https://www.businessinsider.com/amazon-built-ai-to-hire-people-discriminated-against-women-2018-10?IR=T

Recidivism

https://www.wired.com/story/crime-predicting-algorithms-may-not-outperform-untrained-humans/

# Reasons for ethical challenges

- Representativeness in data
- Explainability
- Inherent biases in the training data

# Representativeness in Data

Training data is from a different context

- Obvious case: do not use European training data for medical diagnostics in East Africa.
- Less obvious case: Medical diagnostics data could use training data from East Africa for an algorithm in West Africa – may under-diagnose malaria

Training data is imbalanced

- Data that is not balanced in terms of gender, race, age, etc may bias certain algorithms to favor certain groups

# Explainability

Ethical fading

- Organizations may not think of ethics when using data/outcomes from algorithms, resulting organizational values not being captured in outcomes.

Relevancy / Value

- Should organizations be using ML and does ML provide enough value that offsets potential complications?

Transparency

- How transparent are the algorithms/outputs? Different burden for different situations (prison sentencing vs providing prisoners support services)

# Societal Biases

Does your data have biases reflected in it? If so, your algorithms may amplify these biases.

- Past prison sentencing may be inherently racist or unfair to certain minorities
- Hiring practices may favor men over women, or be unfair to minority groups

# What can we do?

- Auditable algorithms
- Accountability and responsibility
- Implement fairness

# Auditable algorithms

Bias is introduced at the algorithm level

- Developers are making decisions to choose between equity, accuracy, speed

Ensure that algorithms are auditable

- Can they be monitored by external actors? How can you test the algorithms?

# Accountability and responsibility

What mechanisms can we add to make sure that we are holding ourselves and algorithms accountable?

- We work in complex situations and downstream effects are hard to predict
- Build in monitoring to make sure that we can adapt to and fix inequalities that result
- Above all, an AI Assessment Framework to guide AI Solutions is needed.

# Fairness

Data will never be perfect. How will we deal with imperfect data?

- Getting better data
- Fairness through unawareness is rarely enough
  - ML algorithms are typically sophisticated enough to recreate omitted data such as gender, age
- Implementing fairness/equity at the algorithm level
  - Techniques such as demographic parity, equalized odds, equalized opportunities may help. Organizations need to decide what is important to them and work with algorithm developers to implement it.

# Ethical dilemmas

Data will never be perfect. How will we deal with imperfect data?

- Getting better data
- Fairness through unawareness is rarely enough
  - ML algorithms are typically sophisticated enough to recreate omitted data such as gender, age
- Implementing fairness/equity at the algorithm level
  - Techniques such as demographic parity, equalized odds, equalized opportunities may help. Organizations need to decide what is important to them and work with algorithm developers to implement it.

# Ethical dilemmas

9 core scenarios-----Who will you spare?

- Sparing humans (versus pets),
- Staying on course (versus swerving),
- Sparing passengers (versus pedestrians),
- Sparing more lives (versus fewer lives),
- Sparing men (versus women),
- Sparing the young (versus the elderly),
- Sparing pedestrians who cross legally (versus jaywalking),
- Sparing the fit (versus the less fit) and
- Sparing those with higher social status (versus lower social status)

# Conclusion

- Current AI ethics are quite undefined
- Everyday new controversial discussions are held around AI in the future
- AI wants to create something we do not really know: intelligence
- What is intelligence could be discovered by AI researching
- We cannot think about AI without ethics.

# For more info, contact:

Rose Nakasi
g.nakasirose@gmail.com