

Designing and Implementing Responsible AI

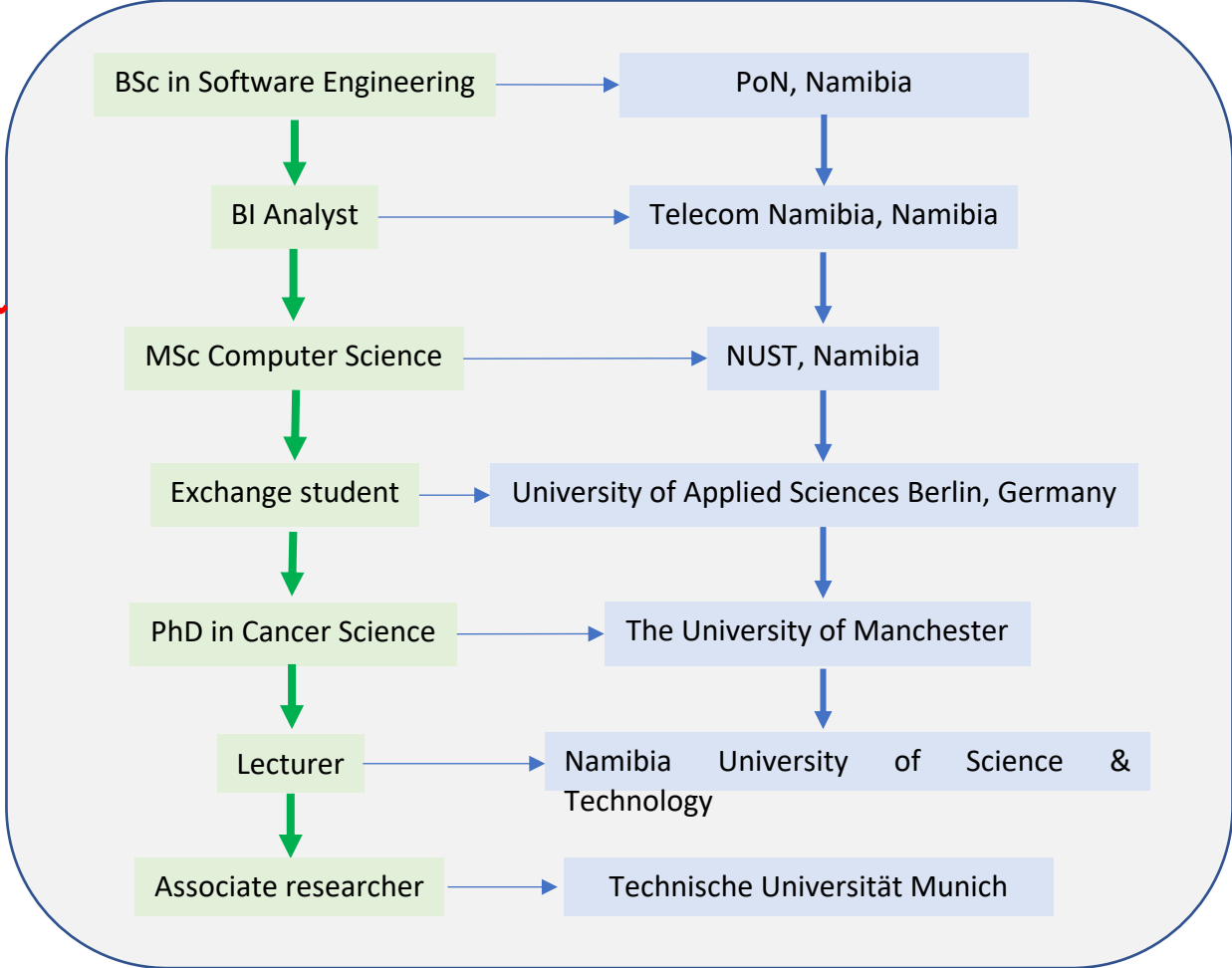
Dr. Lameck Mbangula Amugongo | Research Associate TUM Institute for Ethics in AI: Technical University of Munich

Mbangula Lameck Amugongo



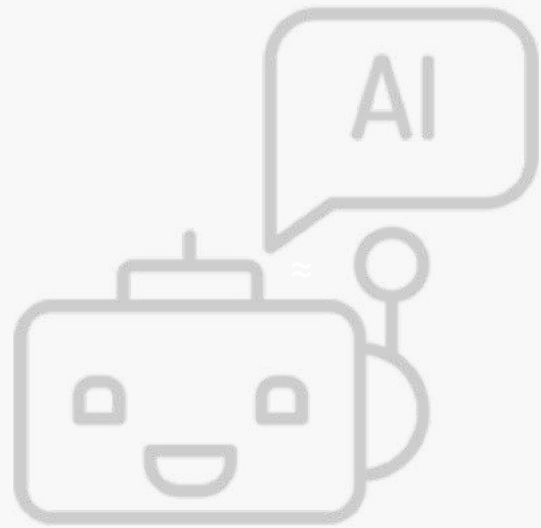
Technology Activist

I am not an ethicist



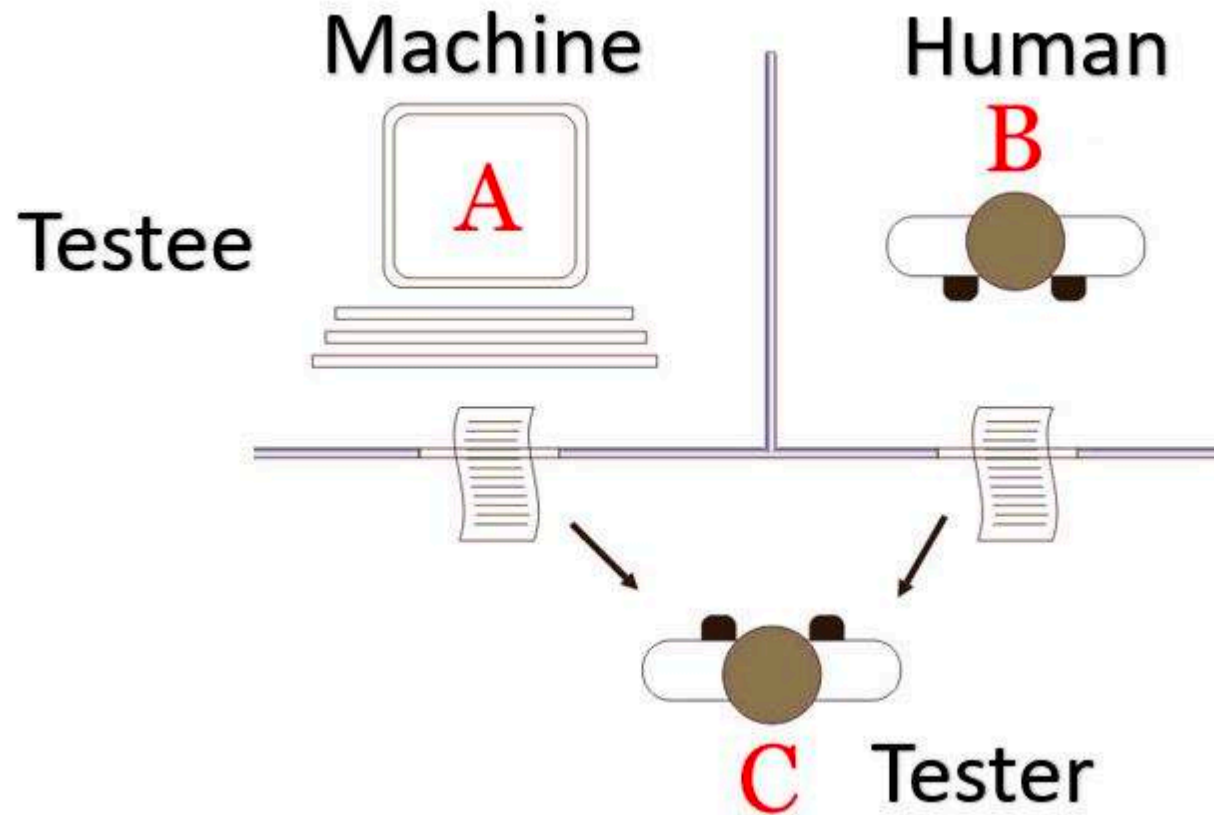
Outline

- 1** Introduction
- 2** Opportunities of AI
- 3** Challenges of AI
- 4** Ethical Principles for AI
- 5** What is Responsible AI
- 6** Implementing ethical principles
- 7** Takeaways
- 8** Conclusion



Introduction

AI aims to mimic human-like intelligence



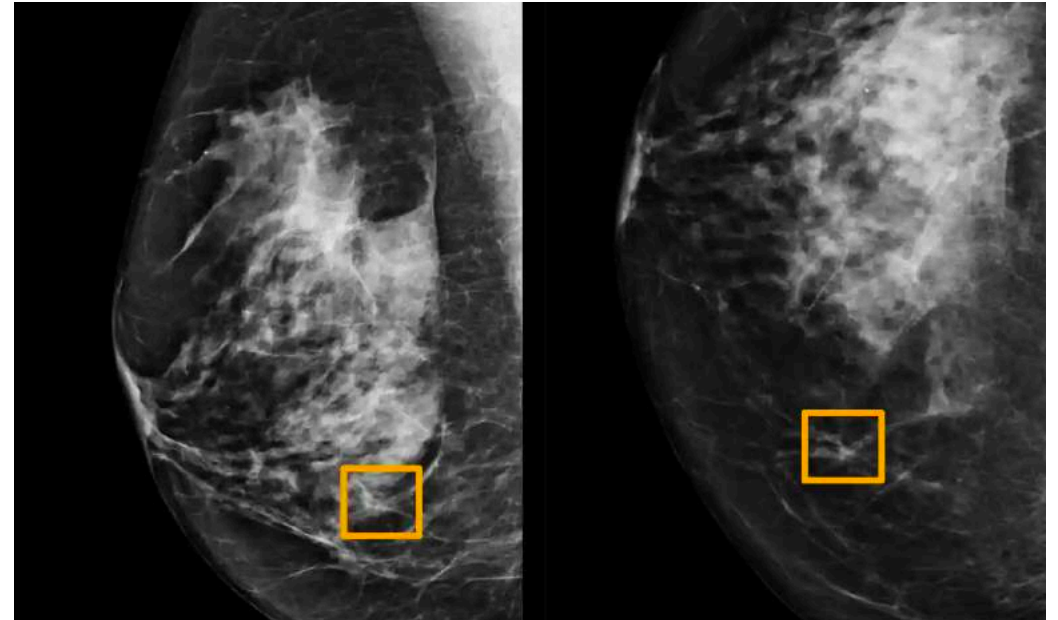
Credit: Raj

Opportunities

Numerous opportunities across various sectors ...

In healthcare: AI has the potential democratize care...

- **Intelligence augmentation:** Amplify human abilities (Jarrahi, 2018).
 - AI will not decrease jobs but to shift jobs to different tasks.
- Ethical AI for medical breakthroughs ...



Source: ACLU

Current AI algorithms

AI systems on the rise are largely pattern recognition systems.

Input: Large datasets of example (faces, text etc.)

Goal: Algorithms learn patterns.

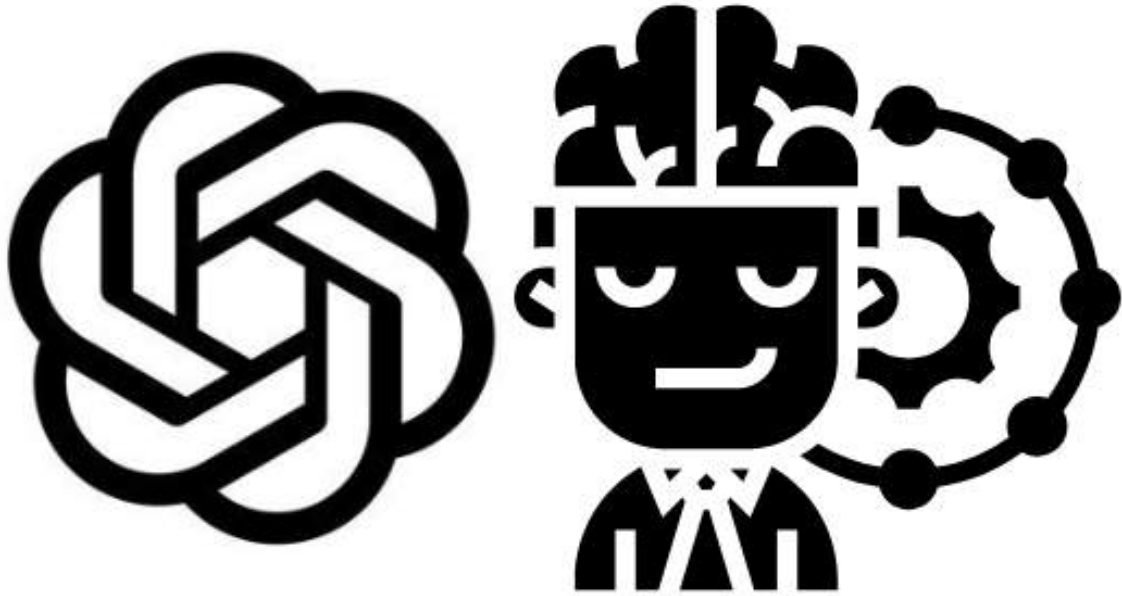
So, garbage in, garbage out.

Challenges

Given all the benefits, there are also severe challenges to AI

Challenges

I will not talk about extinction risks ...



OpenAI's ChatGPT

- Built on stolen data or unconsented data.
- Companies not sharing the training data
- Exploitation of data workers.

Current AI Harms

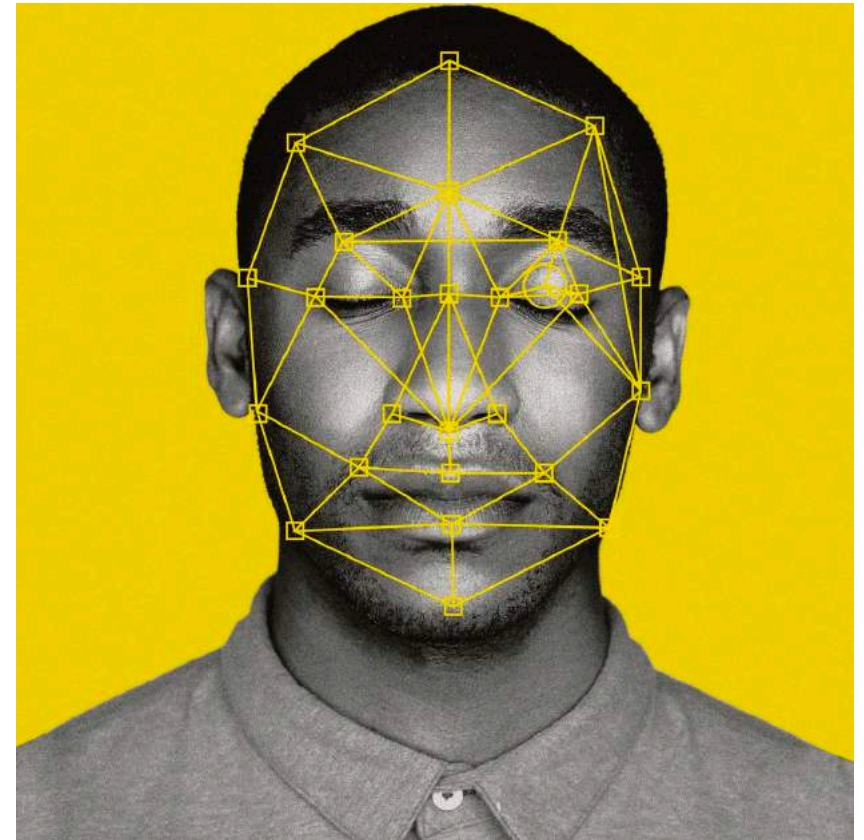
AI is built on the foundation of oppression

Example: Racism

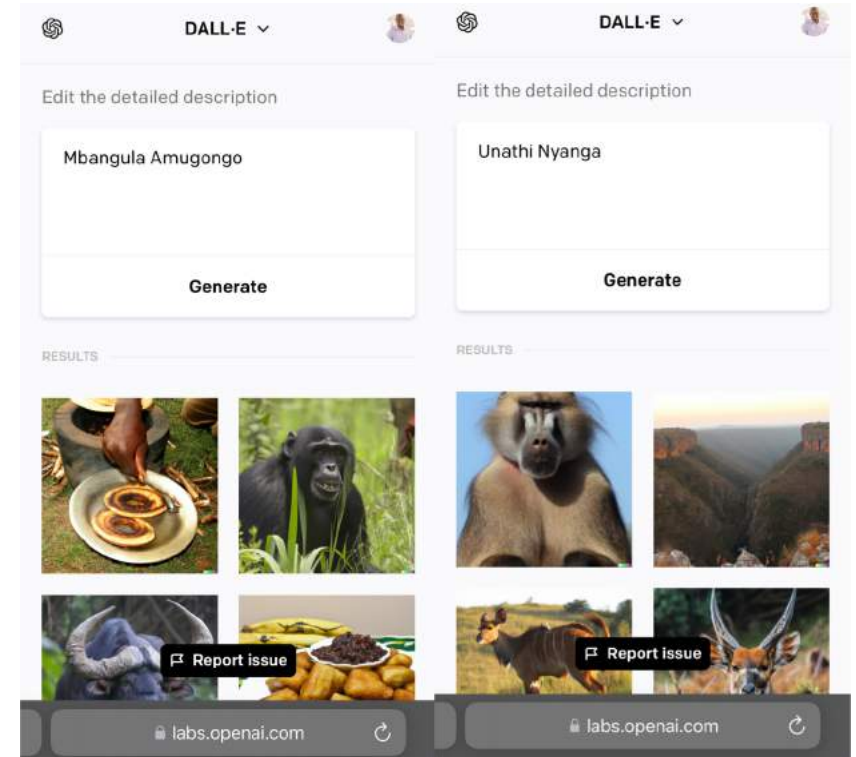
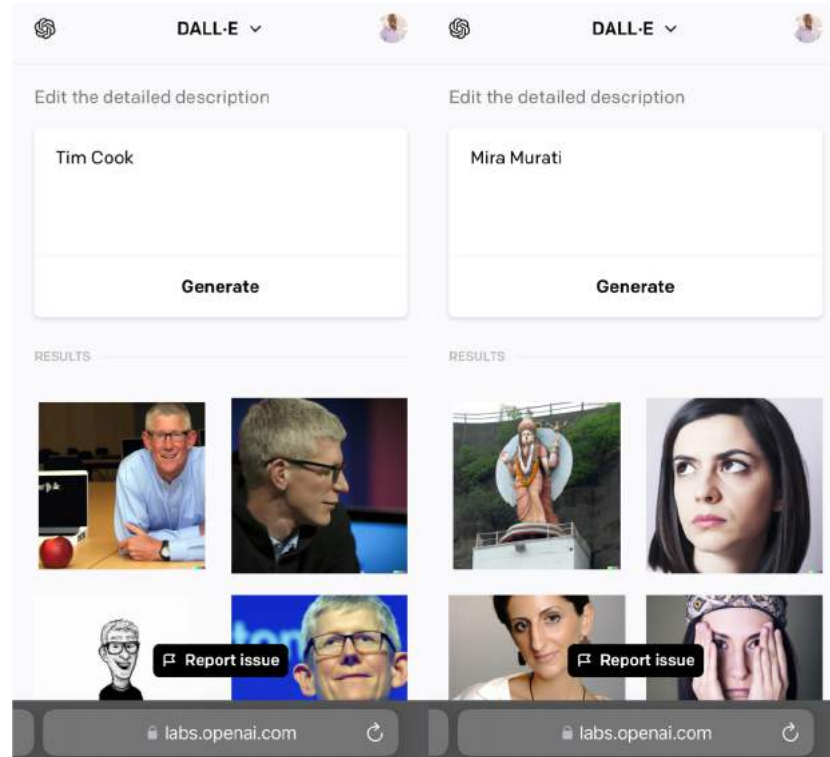
AI-powered biometrics wrongfully arrested a black man, Robert Williams



Source: ACLU

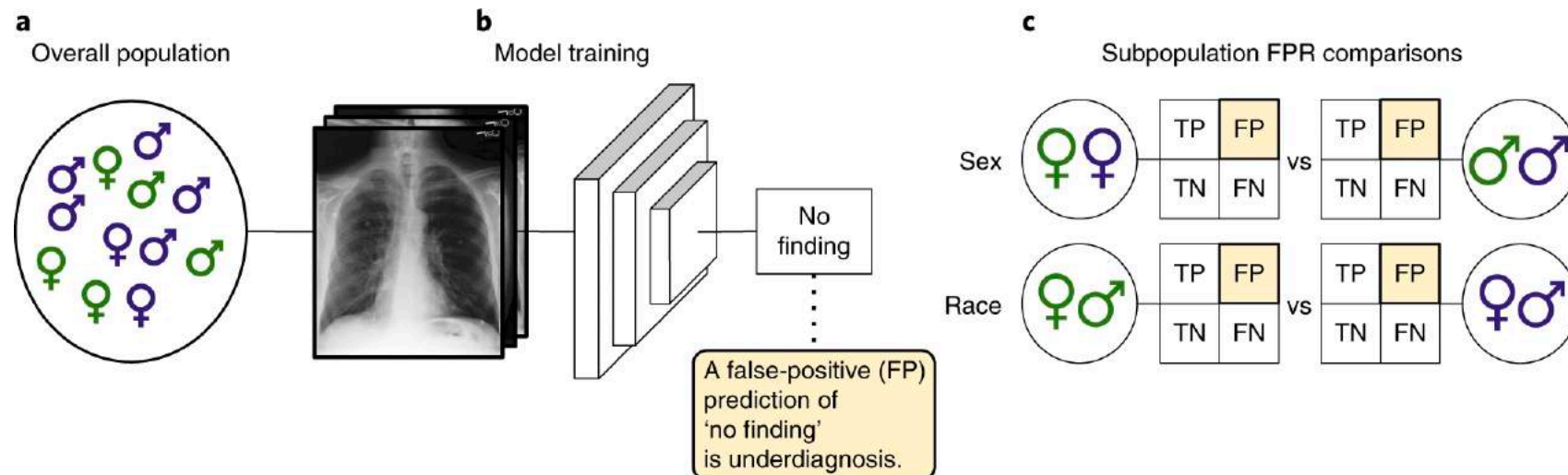


Example: Racism



AI models should not perform any action when they do not understand the request.

Example: Fairness



(Seyyed-Kalantari et al., 2020)



“Black patients had to be a lot sicker than white patients before they received the extra care.”

– an AI algorithm predicted future healthcare costs based on historic data

(Obermeyer et al., 2019).

Growing call for Ethical principles to guide and ground AI ...

Design of Responsible Hybrid Intelligence



Ethical Principles for AI

AI4People recommendations reference for AI ethics in the West ...

- adapted from bioethical principles.

- 1 Beneficence**
Promoting well-being, preserving dignity and sustaining the planet
- 2 Non-maleficence**
Ensuring privacy, security and “capability caution” (upper limit of future AI capabilities)
- 3 Autonomy**
Striking a balance between the decision-making power we retain for ourselves and which we delegate to AI.

- 4 Justice**
Creating benefits that are (or could be) shared, preserving solidarity
- 5 Explicability**
Enabling the other principles through intelligibility and accountability

Source: Floridi et al. (2018)

Global AI Ethics

A study identified 84 documents on ethical principles/guidelines for AI (Jobin, Ienca and Vayena, 2019).

- Existing ethical principles are based on western values
 - lack emphasis on communal values.
- Communal values are essential to ensuring
 - Beneficial AI for all.

Prioritise rationality

*Individualistic: focus
on individual rights*

*Lack cultural and
historic diversity*

AI Ethics: From Rationality to Relationality

Ubuntu ethics promotes “abantu/omuntu”

Personhood is inextricably linked to other people

Umuntu ngumuntu ngabantu (Nguni proverb) - A person is a person because of other people

AI Ethics: From Rationality to Relationality

Western ethics begins with individual (I think therefore I am)

Ubuntu begins with the group (I am because we are).

Opposes the capitalist ethos of AI development

Shift ethics from rationality to relationality

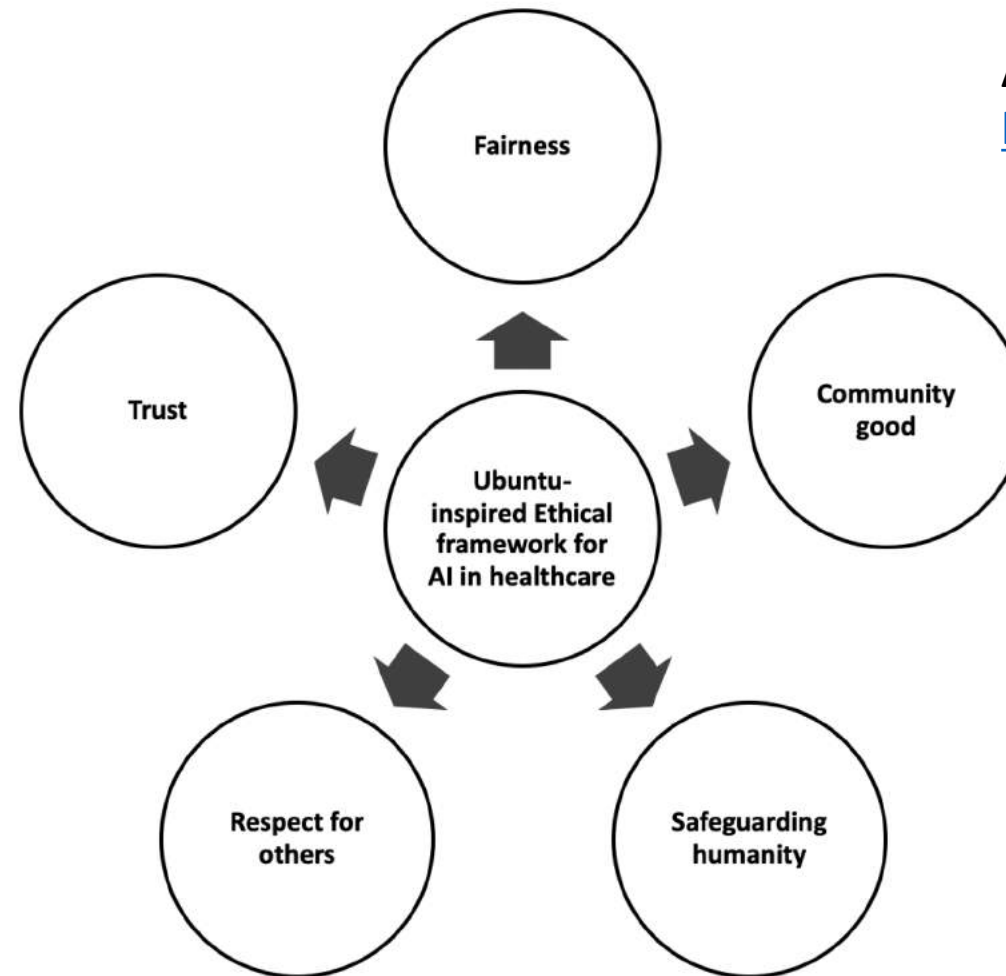
Aligns best with UN Declaration of Human Rights

Ubuntu is an African perspective's contribution to global AI ethics ...

- Decolonise the influence of western values in AI ...

Ethical framework based on Ubuntu

Amugongo et al. FAccT '23:
<https://doi.org/10.1145/3593013.3594024>



Scan code to read the paper

*Ethical principles
do not guarantee
ethical AI*
- Mittelstadt (2019)



Two schools of thought

Self regulation

Big tech will not self-regulate
as they profit-driven.

Enforceable regulation

Regulations such as EU AI Act
are useful.

However, regulations will not solve all issues.

What is Responsible AI (RAI)?

An approach to designing, deploying and using AI in a safe, trustworthy, and ethical way.

To achieve RAI, we need explicit decision on:

Values

Ethics by design

Design

Design is political. So we must question our own design

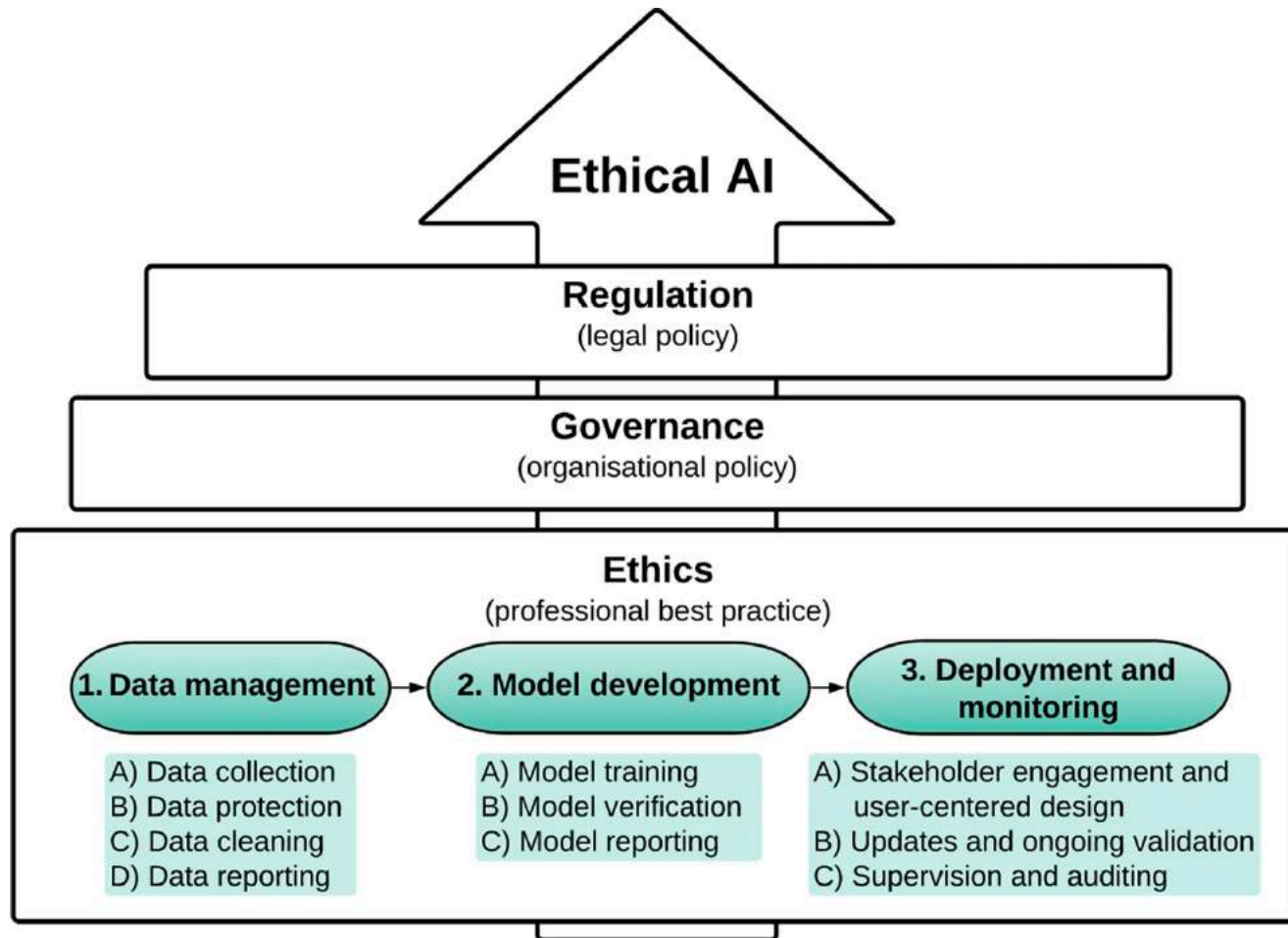
Governance

Regulations
Accountability mechanisms

Operationalizing AI ethics

How do we implement responsible AI in practice ...

Operationalising AI Ethics in AI pipeline



Source: Solanki et al. (2022)

AI development is closely related to software engineering.

Software engineering has well established methods.

We should not reinvent the wheel -> integrate AI ethics in SDLC

- Periodical evaluation of system to ensure that it functions in an ethical manner.
- Operationalize accountability mechanisms.
- Continuous evaluation of ethical principles.
- Provide equitable care.
- System logs, including failure.
- Cause no harm (**Safe guard humanity**).

Deployment



Requirements Elicitation

- Identify and priorities ethical principles.
- Data gathering (**Fairness**)
 - Involve user in data curation.
- Transparent reporting of data used.
- Establish data governance framework.
- Align system requirements with identified ethical principles (**ethics by design approach**).



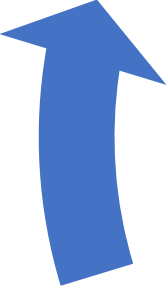
Design and development

- Design with the user (**trust**).
- User autonomy (**Human-in-the loop**).
- Balance system performance with ethical principles such as privacy.
- Robust and reproducible.
- Develop human-centric explanations.






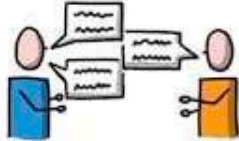



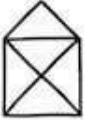
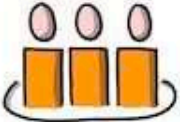



Testing

- Evaluate the system with users.
- Evaluate the accuracy of the system (**Precision**).
- Evaluate system explanations.
- Evaluate whether system transparently report failed tests.
- Determine whether there mechanism for accountability.
- Evaluate privacy and consent.
- Evaluate the robustness and reproducibility.



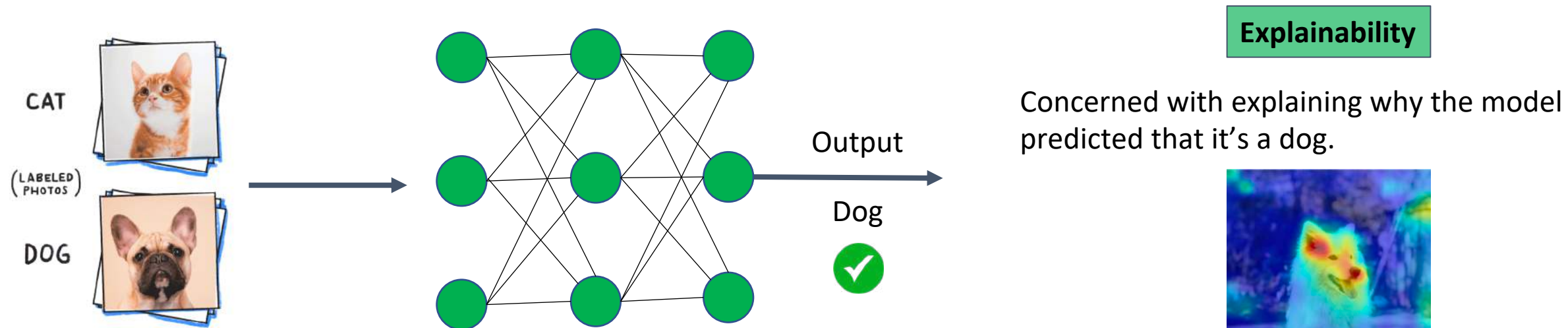
Agile principles

12 Agile Principles @OlgaHeismann			
<p>Satisfy the customer through early and continuous delivery of valuable software.</p> 	<p>Welcome changing requirements, even late in development.</p> 	<p>Deliver working software frequently.</p> 	<p>Business people and developers must work together.</p> 
<p>Build projects around motivated individuals. Give them the support they need. Trust them.</p> 	<p>The most efficient and effective method of conveying information is face-to-face conversation.</p> 	<p>Working software is the primary measure of progress.</p> 	<p>The sponsors, developers, and users should be able to maintain a constant pace indefinitely.</p> 
<p>Continuous attention to technical excellence and good design.</p> 	<p>Simplicity—the art of maximizing the amount of work not done—is essential.</p> 	<p>The best architectures, requirements, and designs emerge from self-organizing teams.</p> 	<p>The team reflects on how to become more effective and adjusts its behavior accordingly.</p> 

Source: [OlgaHeismann](#)

Example: Transparency and Explainability

- Transparency and explainability are often used interchangeably
 - However, they are not the same.



Transparency is not technical

Concerned with how AI system is developed, trained, operates, and deployed in the relevant application domain.

- Dataset
- Foster general awareness

Explainability and transparency does not imply trustworthy.

Arrieta et al. (2019)

Trust

- Trust is complex
- Ethical principles view trust from the Anglo-American jurisprudence
- Concerned about assessing the trustworthiness of AI ...
- Ubuntu: Trust is rooted in the interconnectedness and interdependence of individuals within a community.
 - Trust dependent on long term relationships with the community.

BUILD LONG-TERM TRUST
*Long-term collaboration with
communities*

Way forward

We need to rethink how we design and implement AI

Human-Centered AI: Towards long-term trust

Account for diverse cultural values and viewpoints.

Engage stakeholders through the AI system development cycle.

Address power dynamics by empowering marginalised.

Human centred evaluations and community vetting.

Next steps



Inclusive **regulatory framework**.

- International cooperation on AI regulation to address ethical concerns.
- Establish cooperation on accountability.
- Set up global risk based approaches for developing and deploying AI.



- Incorporate AI ethics into the data science curriculum
- Create tools to test, evaluate and monitor the application of the AI Principles.

Key Take-aways

- We cannot expect people to trust AI with all concerns surrounding AI ...
- We need AI ethics to shift from rationality to relationality.
 - Foster the collaborative spirit for effective AI ethics
- We need an interdisciplinary approach to RAI
- Human rights is intrinsically linked to capacity building
 - Thus, we need to create spaces for interdisciplinary capacity building to address AI concerns

The Responsible AI Forum

The IEAI hosts the Responsible AI Forum in Munich on 13-15th September 2023.

The Responsible AI Forum

 Munich

 13-15 September 2023

Join us for a 3-day conference that brings together research, policy and practice on topics related to responsible use of AI

<https://responsibleaiforum.com>





Acknowledgement

- IEAI Team

“It takes a village to raise a child”

Thank you

Any questions

