NETHOPE

AI Ethics for Nonprofits Workshop

As we get started

Please use the Chat window to say hello, tell us where you're joining from, and what organization you work for.

About the workshop

The Artificial Intelligence (AI) Ethics Workshop for Nonprofits was developed by NetHope, USAID, MIT D-Lab, and Plan International.

The goal of the workshop is to build capacity in the nonprofit sector to design and use AI responsibly and ethically.

NETHOPE

USAID
FROM THE AMERICAN PEOPLE

PLAN INTERNATIONAL

MIT D-Lab
designing for a more equitable world

Purpose of this workshop

Learn how to **practically apply ethical considerations related to the principle of Fairness** in the context of humanitarian and international development use cases.

Workshop Agenda

- Brief overview of the key AI Ethics concepts
- Breakouts: Practical application of ethical considerations
- Report out from breakouts and discussion
- Resources and next steps

Tools for today's workshop

- **Zoom for communication**
  - Video, audio, chat
- **Miro for collaboration**
  - Virtual whiteboarding, with sticky notes

NETHOPE

Ethics Primer

AI in the Nonprofit Sector

- AI can help us do our work better:
  - Reach more people with services and information they need.
  - Make decisions and act faster in emergencies.
  - Predict emergencies before they spread.
  - Amplify human effort and free up limited human resources to focus on high-priority work.
  - Improve outcomes through real-time feedback on the effectiveness of programs and recommendations for improvements.
- Challenge: Responsibly design and use AI technology - maximizing its benefits while minimizing risks and protecting human rights

AI Ethics

"A set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies."

The Alan Turing Institute

What is an ethical AI system?

An AI system that supports individual and collective well-being and enhances our ability to tackle global challenges.

Responsible Innovation

Responsible Innovation is a transparent, interactive, sustainable process by which organizations proactively evaluate how they can design and use technology in ways that are aligned with their values and missions.

What are some of the ethical issues surrounding technology use today?

- Intentional harms such as hate speech, misinformation, weaponization of technologies like AI.
- Infringement on rights and values such as surveillance.
- Unfair outcomes like discrimination and prejudice stemming from bias.

Bias & Fairness

Bias - Systematically favoring one group relative to another. Bias is always defined in terms of specific categories or attributes (eg gender, race, education level).

Fairness - Just and equitable treatment across individuals and/or groups.

Three general questions to ask:

- How might ML model design and implementation cause disproportionate harm?
- How well do we understand how ML models are working? Would we recognize bias or inequities when (or before) they occur?
- What happens when things go wrong?

Key considerations relevant to Fairness and AI

How might data and ML model implementation cause disproportionate harm?

- **Equity** - Does the model work better, or do model failures have significantly worse consequences for one group than another?
- **Representativeness** - To what extent is the training data representative of the population that will be affected by the use of the ML/AI model? To what extent are the people developing the ML/AI model?
- **Bias** - What biases may be embedded in the data? (*consider real-world power dynamics likely to shape what data is available and about whom)*

Where to look for potential disproportionate harm…

Systematic Differences in Failure Rates Between Groups of Interest

	Men	Women
% Accurate predictions		
% Inaccurate predictions		

**Your organization should work collaboratively to identify the groups across which you are concerned about fairness - it could be across more than two groups. The examples above are illustrative only.*

Where to look for potential disproportionate harm…

Systematic Differences in Failure Rates Between Groups of Interest

	Men	Women
% Accurate predictions		
% Inaccurate predictions		

Systematic Differences in Error Types Between Groups of Interest

	Men	Women
FALSE POSITIVE: Given loan, but won't repay		
FALSE NEGATIVE: Denied loan, but would have repaid		

**Your organization should work collaboratively to identify the groups across which you are concerned about fairness - it could be across more than two groups. The examples above are illustrative only.*

Where to look for potential disproportionate harm…

Systematic Differences in Failure Rates Between Groups of Interest

	Men	Women
% Accurate predictions		
% Inaccurate predictions		

Systematic Differences in Error Types Between Groups of Interest

	Men	Women
FALSE POSITIVE: Given loan, but won't repay		
FALSE NEGATIVE: Denied loan, but would have repaid		

Codified Social Bias

**Your organization should work collaboratively to identify the groups across which you are concerned about fairness - it could be across more than two groups. The examples above are illustrative only.*

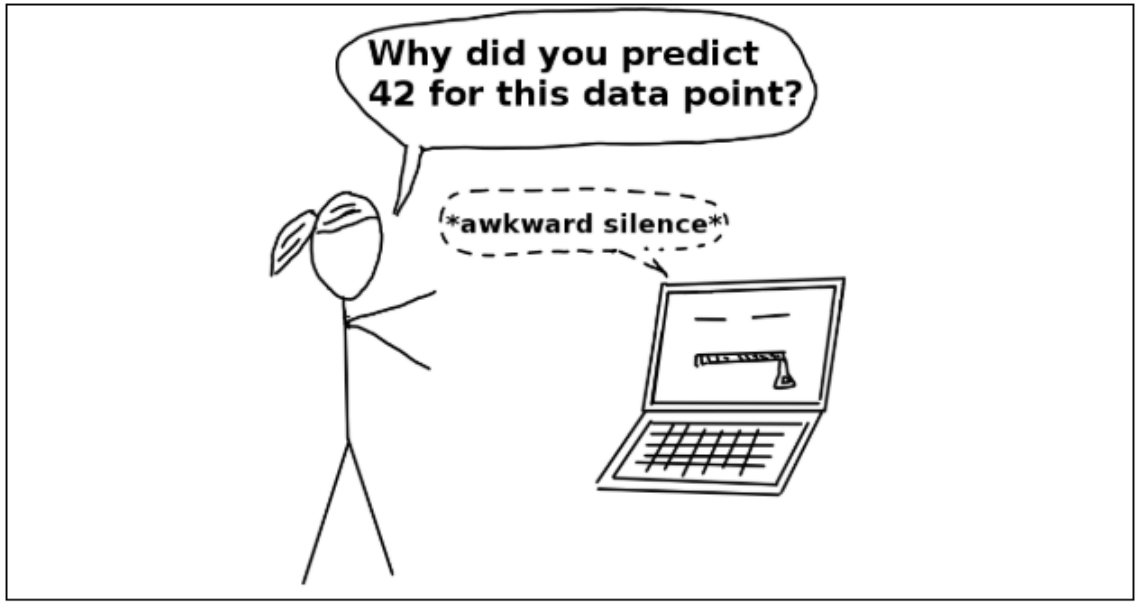
Key considerations relevant to Fairness and AI

How well do we understand how ML models are working? Would we recognize bias or inequities when (or before) they occur?

- **Explainability** - to what extent can the predictions made by ML model be understood in non-technical terms? Can we interpret the relationships underlying the model's predictions?
- **Auditability** - to what extent can outside actors query AI/ML models (eg, to check for bias)?

How confident can we be that model results are not based on underlying biases in the data?

To what extent could we figure out what would need to change to get a different result?



Source: Interpretable Machine Learning, a book by Christopher Molnar

Key considerations relevant to Fairness and AI

What happens when things go wrong?

- **Accountability**
  - What mechanisms are in place to identify when mistakes are made?
  - To what extent will feedback be sought from those affected by the predictions the model makes?
  - What can be done to redress possible harms that result from mistakes?

What are some things we can we do to mitigate concerns?

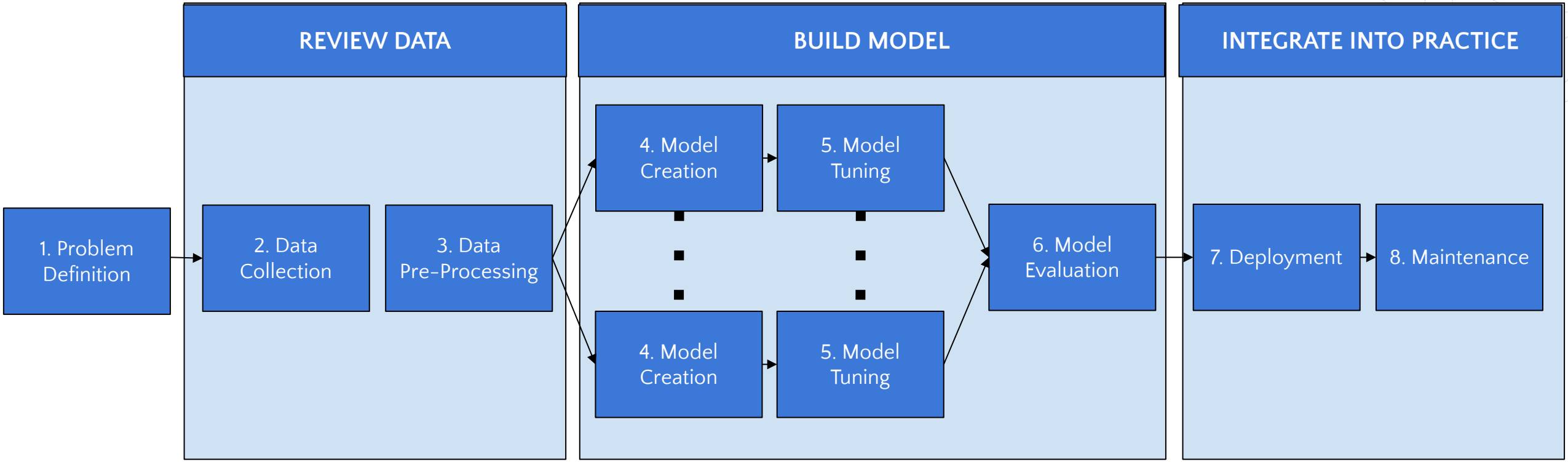
Project Level

- Ask the right questions
- Define which attributes you don't want to bias model predictions
- Identifying sources of bias (historical biases, individual biases, biases in data)
- Exploring technical approaches to testing for bias and implementing fairness

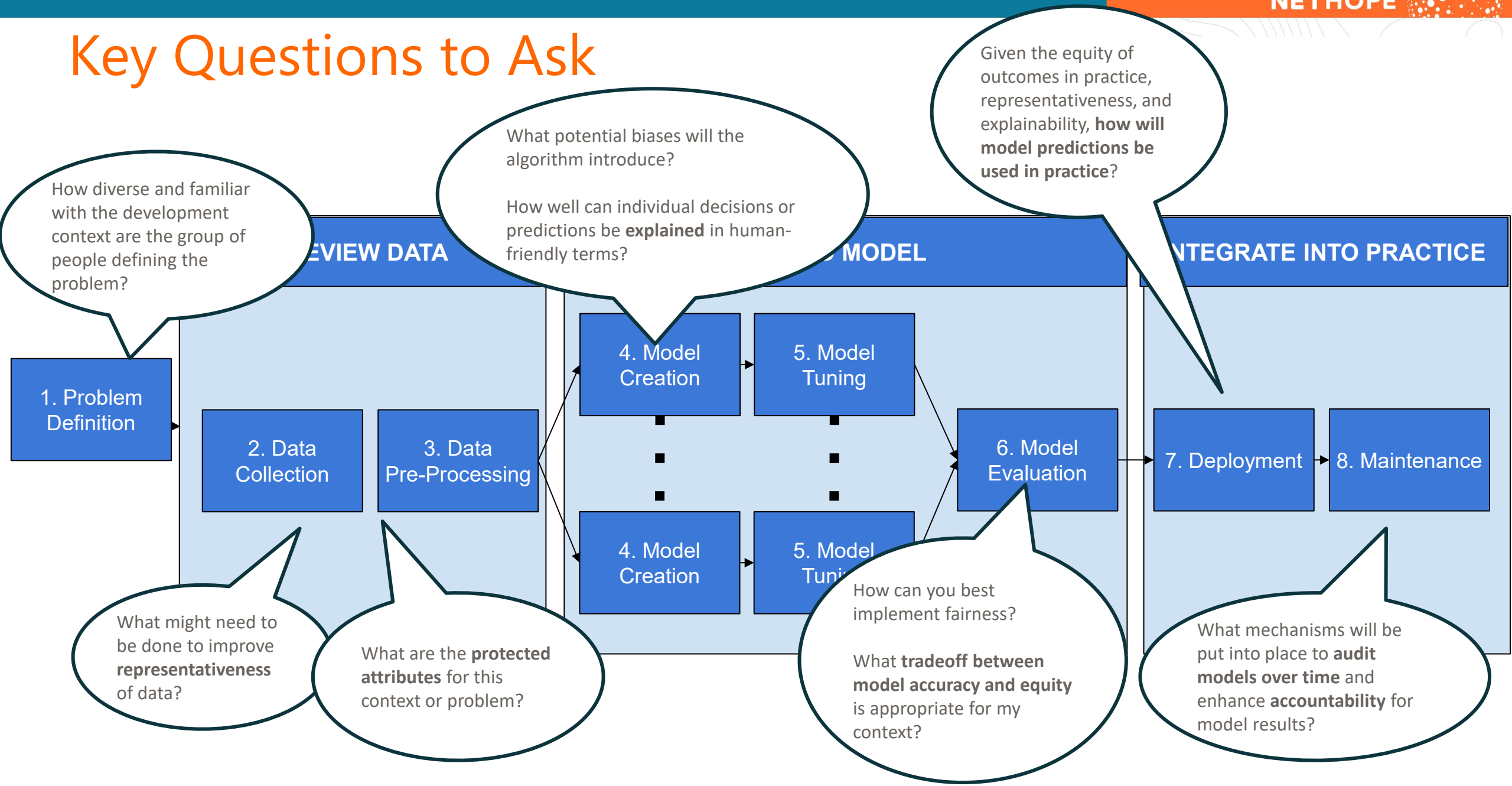
System Level

- Strengthening representativeness of data available for training ML models
- Support for auditing model outcomes, including consideration for open data and open algorithms
- Strengthening digital ecosystems and enabling environment for AI/ML
- Diversifying the workforce and organizations working in AI/ML

Machine Learning (ML) Project Overview



Key Questions to Ask



AI Suitability Framework for Nonprofits: https://nethope.app.box.com/file/643893297448

Protected Attributes

Traits that should not be used as a basis for decision-making in machine learning projects. Sometimes they are legally mandated. Your organization and data science team will need to define which traits to treat as protected in your context.

Typically, protected attributes include:

- race
- age
- gender
- sexual orientation
- religion
- socio-economic status

Other considerations

- Is the use of ML in your context solving a **relevant** problem?
- Is the application of ML technology adding **value** (eg informing more accurate, timely, actionable results?)
- Does your organization have sufficient **capacity** to implement the solution?
- Are there **other concerns (besides fairness**) you have about the proposed use of ML/AI?

NETHOPE

PLAN INTERNATIONAL

YOUTH EMPLOYMENT SOLUTIONS

Case Study: TESSA Chatbot

What can go wrong? How do you address the issues?

PLAN INTERNATIONAL

YOUTH EMPLOYMENT SOLUTIONS

TESSA

Plan International's **T**raining, **E**mployment and **S**upport **S**ervices **A**ssistant

- **Problem:** Marginalized youth in Asia are unable to effectively communicate their skills and link to suitable economic opportunities
- **Current approach:** Community Development Facilitators support youth to 'formalize' their skills, then link them to opportunities
- **Limitations:** Quality not quantity
- **Solution:** AI-powered chatbot on Facebook Messenger - TESSA
- **Benefits:** Approachable, accessible, consistent quality

What can go wrong?

- TESSA **reinforces** the status quo in terms of **gender inequalities** in the labor market and **gender-stereotypical** skills, jobs, and careers through its engagement with and recommendations for the users

PLAN INTERNATIONAL

YOUTH EMPLOYMENT SOLUTIONS

Why?

The beast of bias

- Team
- Data
- Problem framing
- Implementation
- Model / Tools

→ Narrowed down options and opportunities for girls

PLAN INTERNATIONAL

YOUTH EMPLOYMENT SOLUTIONS

How to address these issues?

- **Team:** Diversify/restructure team to address power dynamics
- **Data:** Develop ML/AI for recognizing patterns in data, analyze, evaluate and flag.
- **Problem framing:** Principles and process to increase intentional inclusion
- **Model/Tools:** Assess for bias
- **Implementation:** Implementing an agile inclusion methodology.

PLAN INTERNATIONAL

YOUTH EMPLOYMENT SOLUTIONS

Unless we intentionally include, we will unintentionally exclude

Q&A

INSERT:

Photo of Speaker 1

Speaker 1
Name, Title, Organization, Email

INSERT:

Photo of Speaker 2

Speaker 2
Name, Title, Organization, Email

INSERT:

Photo of Speaker 3

Speaker 3
Name, Title, Organization, Email

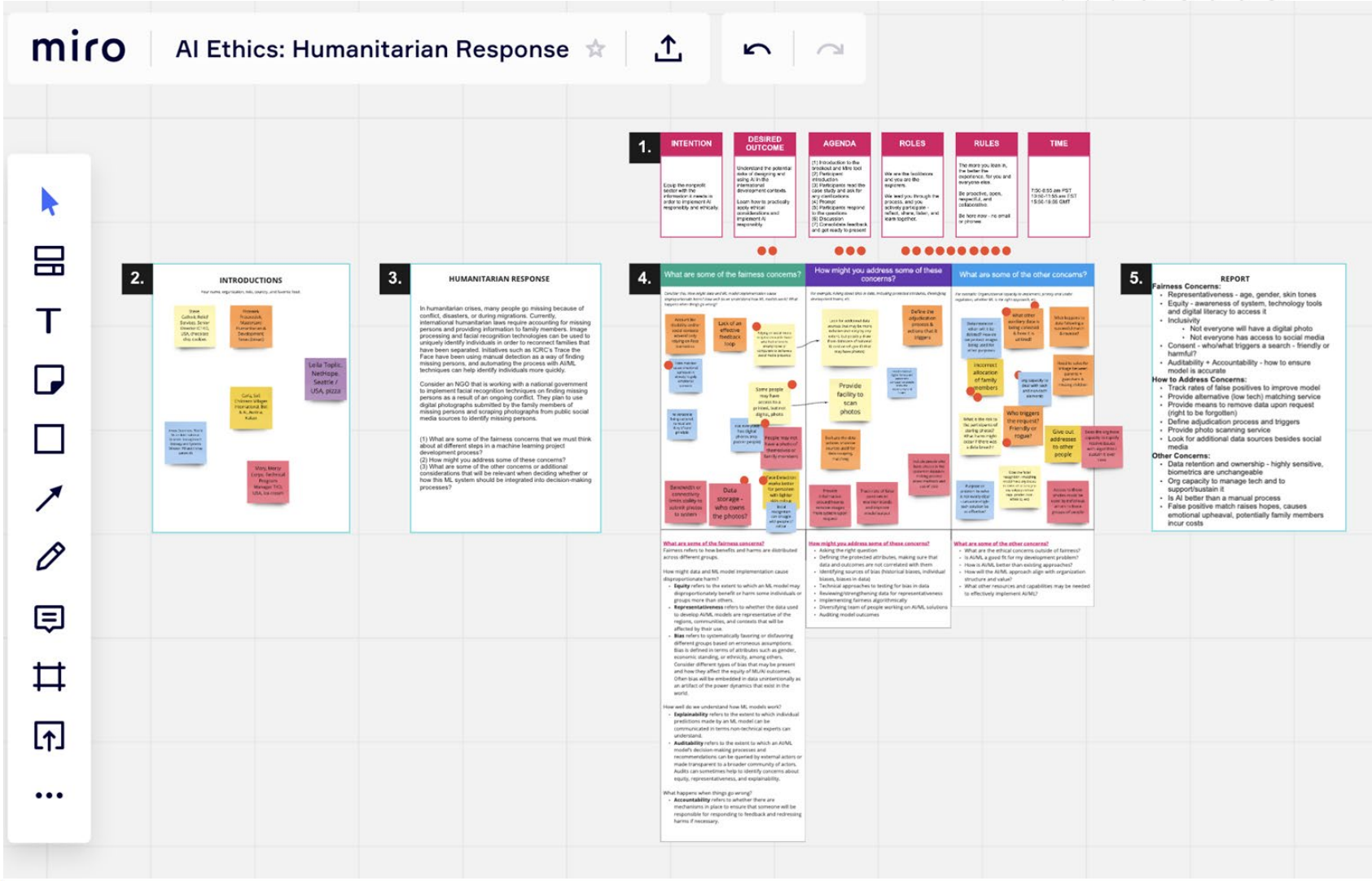
Short break

NETHOPE

Breakouts

Breakouts: Introduction

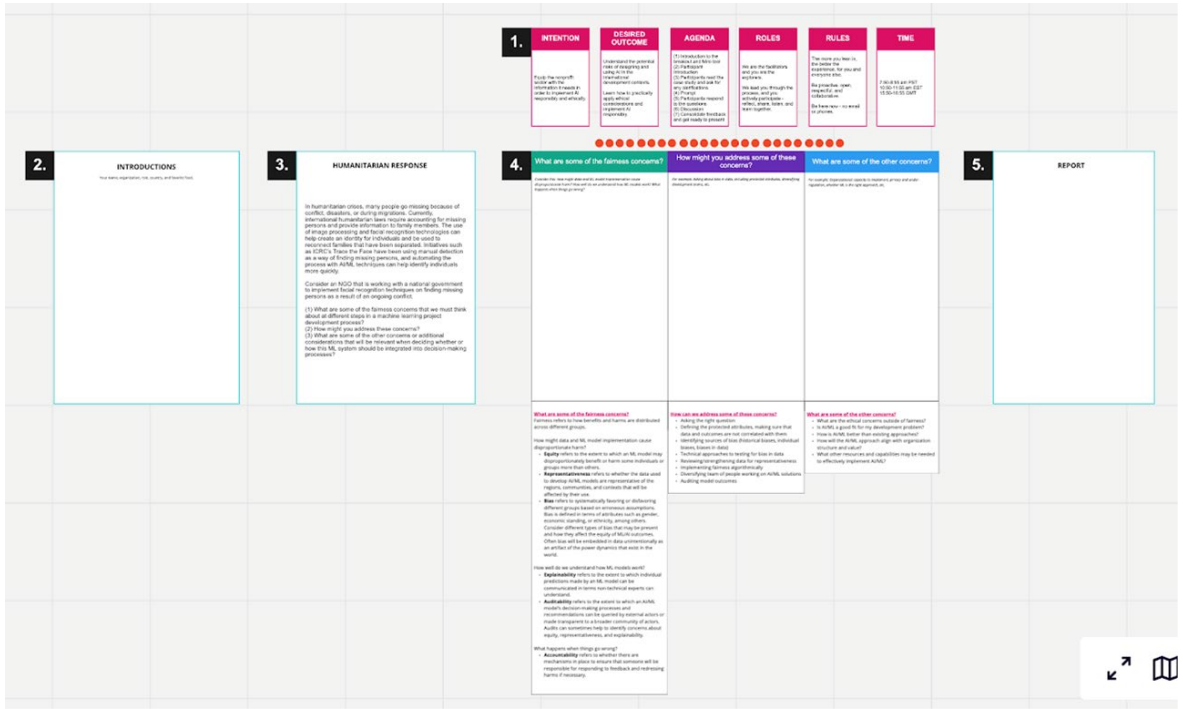
- 5 breakouts centered around use cases from humanitarian response, health, education, agriculture, workforce
- Hands-on, collaborative work using Miro and Zoom, followed by a report-out from each group

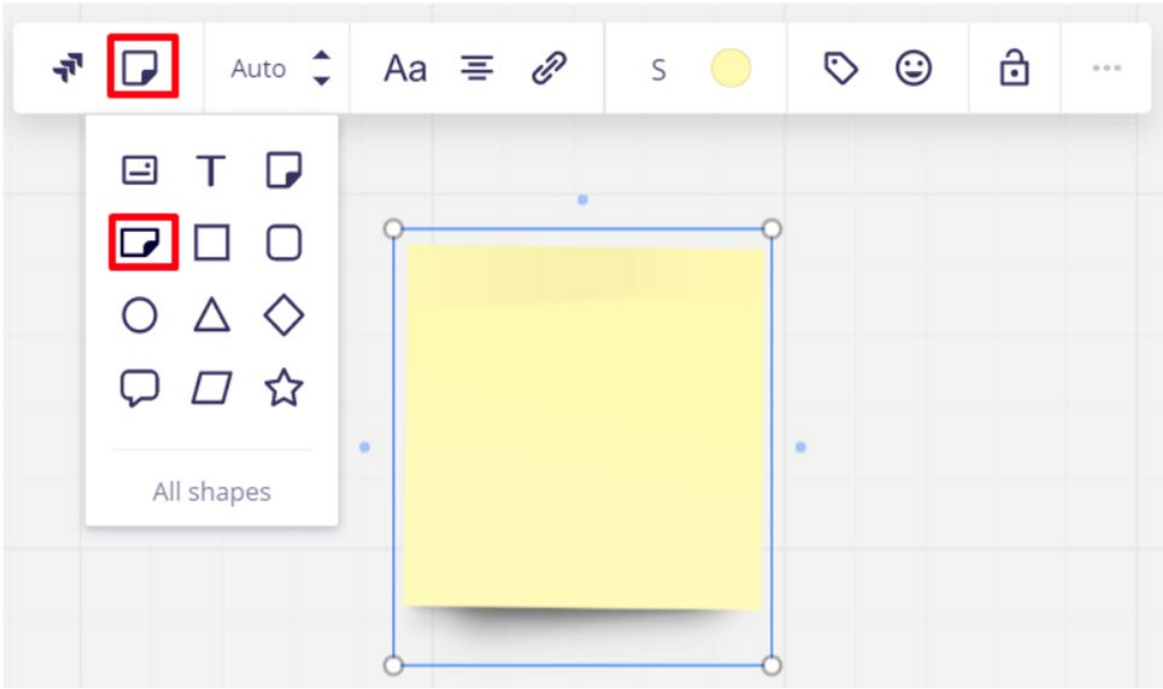


Breakouts: Facilitators

PHOTO	PHOTO	PHOTO	PHOTO	PHOTO	PHOTO
Name Organization	**Name** Organization	**Name** Organization	**Name** Organization	**Name** Organization	**Name** Organization
Humanitarian Response	**Health**	**Education**	**Agriculture**	**Workforce**	**Main room**

Breakouts: Miro demo





Welcome back, hope you enjoyed the breakout!

Please get ready to share.

Report-out and Discussion

Each group will have 2 minutes to share their key insights, using their Miro board as a backdrop.

- Fairness concerns
- How to address fairness-related concerns
- Other concerns or additional considerations

Few minutes at the end to share observations, feedback, additional insights on the use cases.

Order: workforce, health, education, agriculture, humanitarian response

Breakouts: Use Cases

Workforce

Machine learning is used to screen resumes from job applicants and determine which ones should be offered interviews.

Health

Machine learning is used to better predict which people living with HIV/AIDS are most at risk of being "lost to follow up" in the first 12 months of their treatment.

Education

Machine learning is used to evaluate the quality of student writing among high school students across India, with the goal to make tailored recommendations of support needed to improve writing performance.

Agriculture

Machine learning is used to improve farmer income by helping them determine where and when to purchase inputs and sell crops as well as how to connect with the appropriate markets.

Humanitarian Response

A facial recognition system designed to help identify and find missing persons (missing due to conflict, disaster or migration) with the goal to reconnect them with their families.

Next steps:

- Continue learning about ethical, responsible development and use of AI.

  See resources on the next slide.
- If you are a NetHope Member, join NetHope's AI Working Group:

  http://bit.ly/ET_WorkingGroup

Resources

- Key AI Ethics concepts
- Exploring Fairness in Machine Learning for International Development by MIT D-Lab, with the support from USAID
- AI Ethics: 5 Considerations for Nonprofits
- NetHope's AI Ethics webinars:
    - Part I (recording, slides)
    - Part II (recording, slides)
    - Part III (recording, slides)
- NetHope AI Suitability Toolkit for Nonprofits
- USAID: Reflecting the Past, Shaping the Future: Making AI Work for International Development
- AI Primer (recording, slides)

Thank you

for participating in the **AI Ethics for Nonprofits** Workshop!

NETHOPE