Welcome!

We will get started at the top of the hour!

>>Please activate the Zoom Chat window for panelist and attendee interactions<<

YOU WILL EXPERIENCE SILENCE UNTIL WE START
WELCOME!
NetHope
Solutions Center

AI ETHICS WEBINAR SERIES
PART ONE
Housekeeping

• Let’s keep this interactive:
  Post questions in Zoom Chat window for the Q&A and discussion session

• Look for a follow-up email with link to recording and collateral on NetHope Solutions Center

• Please respond to webinar satisfaction poll presented after the webinar
Introduction to AI Ethics Webinar Series

Leila Toplic
Lead for Emerging Technologies Initiative, NetHope
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 problems before they spread and escalate
  - Prevent loss of life and resources
- Challenge: Responsibly design and use AI technology - maximizing its benefits while minimizing risks and protecting human rights

https://solutionscenter.nethope.org/et
Nora Lindstrom
Global Lead for Digital Development, Plan International

Amy Paul
Technical Advisor, USAID Center for Digital Development

Kendra Leith
Associate Director for Research, MIT D-Lab

Amit Gandhi
Instructor, MIT D-Lab

Leila Topic
Lead for Emerging Technologies Initiative, NetHope

NetHope
USAID
PLAN International
MIT D-Lab
AI Ethics Webinar Series

- **May 21**: Brief introduction to ethics and an example of what can go wrong with AI and how to address the issues.
- **June 4**: How can we use AI/ML tools in international development? What are some illustrative values that should guide responsible use of AI in practice? What risks might be encountered with respect to implementing ML/AI tools in accordance with those values? How do you mitigate risks and engage in responsible development and use of AI?
- **June 18**: Fairness in Machine Learning projects
Brief Ethics Primer

If you’d like additional training on ethics, please let us know that via the webinar survey.
Today

- What is ethics?
- What is responsible innovation?
- What are some of the ethical issues surrounding technology use today?
- A practical example that highlights risks, issues, and path to achieving ethical AI.
Ethics is a set of values and principles that guide the way in which we behave as individuals and as a group.
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 technology solution?

Technology 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.
Ethical Principles, Guidelines, Frameworks

Source: The Humanitarian Data Science and Ethics Group’s Ethical Framework
Ethical Principles, Guidelines, Frameworks

“Principles Artificial Intelligence Project” by Berkman Klein’s Cyberlaw Clinic
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/values such as surveillance.
- Unfair outcomes like discrimination and prejudice stemming from bias.
Bias

AI systems can:

- “reproduce, reinforce, and amplify the patterns of marginalisation, inequality, and discrimination”
- “replicate their designers’ preconceptions and biases”
- use data that is “insufficiently representative of the populations from which they are drawing inferences.”

*The Alan Turing Institute*
Example of what can go wrong with AI and how to address the issues

Hycinth Umaran
Product Owner, Plan International

Nora Lindstrom
Global Lead for Digital Development, Plan International
TESSA

Plan International’s Training, Employment and Support Services Assistant
What problem are you trying to solve?

Marginalized youth in Asia, particularly young women, are unable to effectively express and communicate their skills and link to suitable economic opportunities.
Hi, I’m Tessa! Your virtual “Ate” here to support your journey into a future YOU choose.

Build resume  Mentor & Coach

Identify Skills  Follow & Grow

Connect to Opportunity

https://www.youtube.com/watch?v=suB_XdPVdH0
**What can go wrong?**

<table>
<thead>
<tr>
<th>Team</th>
<th>Data</th>
<th>Problem Framing</th>
<th>Model/ Tools</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did not ensure team was diverse/balanced</td>
<td>Data collected was bias and then reused.</td>
<td>In solving one problem statement, we unintentionally created another.</td>
<td>3rd party tools were bias blind</td>
<td>We didn’t have contingencies or corrections on hand.</td>
</tr>
</tbody>
</table>
## How do we address the issues?

<table>
<thead>
<tr>
<th>Team</th>
<th>Data</th>
<th>Problem Framing</th>
<th>Model/ Tools</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did not ensure team was diverse/balanced</td>
<td>Data collected was bias and then reused.</td>
<td>In solving one problem statement, we unintentionally created another.</td>
<td>3rd party tools were bias blind</td>
<td>We didn’t have contingencies or corrections on hand.</td>
</tr>
<tr>
<td>Restructured team to switch power dynamics.</td>
<td>Developing ML/AI for recognizing patterns in data and flagging.</td>
<td>Principles and process to increase intentional inclusion</td>
<td>Gender transformative assessment for TESSA (and others)</td>
<td>Implementing an agile inclusion methodology.</td>
</tr>
</tbody>
</table>
The beast of bias.

- Bias existed from the beginning
- We were not intentional about our design
- It expressed itself in our conversation
- It expressed itself in the job recommendations
- It was reinforced by the recommendation engine
- It narrowed down options for girls
Gender Transformative Assessment

- Gender Norms
- Agency
- Working With Boys & Men
- Condition & Position
- Diversity
- Enabling Environment
Gender & inclusion requirements gather

DECEMBER 2019

JANUARY 2020
Development of inclusion backlog

APRIL 2020

SEPTEMBER 2020
Open source of findings/features

MAY 2020

AUGUST 2020
Agile integration of inclusion backlog
“Unless we intentionally include, we will unintentionally exclude”
Is Artificial Intelligence Doomed To Reflect The Worst In All Of Us?

Hey Siri, stop perpetuating sexist stereotypes, UN says

AI robots are sexist and racist, experts warn

Amazon’s sexist AI recruiting tool: how did it go so wrong?
Almost 90% of Men/Women Globally Are Biased Against Women

New Analysis Provides Clues to “Glass Ceiling”; Tools to Shatter It

TABLE 3

Bias against gender equality is on the rise

<table>
<thead>
<tr>
<th>Index</th>
<th>Description</th>
<th>Group</th>
<th>2004–2009</th>
<th>2010–2014</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSNI</td>
<td>With some bias</td>
<td>Women</td>
<td>83.4</td>
<td>84.6</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Men</td>
<td>89.4</td>
<td>89.9</td>
<td>0.5</td>
</tr>
<tr>
<td>GSNI2</td>
<td>With moderate to intense biases</td>
<td>Women</td>
<td>56.6</td>
<td>59.7</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Men</td>
<td>70.0</td>
<td>70.8</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Source: Human Development Report Office calculations based on World Values Survey for 31 countries with time series data, representing 59 percent of the global population.
Gender data gaps in economic opportunities

- Access to Financial Services
- Access to ICT
- Asset Ownership
- Informal Employment
- Productivity in Agriculture
- Unpaid Work

- Access to Childcare
- Employment Mobility
- Entrepreneurship
- Migrant Worker Conditions
- Pay Gaps & Opportunity Costs

- Aspirations & Expectations
- Decent Work

*Compared to Data2X's 2014 mapping exercise
The Gender Imbalance in AI Research Across 23 Countries

*Among 4000 researchers who have been published at the leading conferences NIPS, ICML or ICLR in 2017*

Lack of diversity in AI
**Ethical, humane technology** is technology that supports individual and collective well-being and enhances our ability to tackle global challenges.

**ML/Al as ally** in breaking down harmful gender social norms

**Equality Tech:**
Technology that by design advances gender equality
hän on presidentti
hän on koodari
hän on lastenhoitaja

he is president
he is a coder
she is a nanny
hän on presidentti

Translations are gender-specific. LEARN MORE

she is the president (feminine)

he is the president (masculine)
Key questions

- Is your team diverse?
- Do you have and/or can you get lots of data for this problem? Does the data you need exist?
- Is there bias in your data?
- What are potential biases that AI may introduce or amplify in your context? Can use of your model reinforce existing bias in the data?
- What are potential negative outcomes and harmful use cases for the solution you are creating?
- Does your AI maintain the status quo or does it advance equality?

Resources on AI Ethics

- Exploring Fairness in Machine Learning for International Development by MIT D-Lab, with the support from USAID
- Understanding artificial intelligence ethics and safety, by The Alan Turing Institute
- The Humanitarian Data Science and Ethics Group’s Ethical Framework
- Data Ethics, AI and Responsible Innovation edX course by The University of Edinburgh
- Principled AI, by the Berkman Klein Center at Harvard University
- Ethics Guidelines for Trustworthy AI