AI Ethics Webinar Series: Part II

June 4th, 2020
Amy Paul
USAID Center for Digital Development
Objectives

- Introduce relevant AI/ML applications in international development
- Highlight key ethics considerations for responsible use of AI/ML technology in international development
- Identify risks that may arise in the development and use of AI/ML technology with respect to ethics considerations
- Discuss approaches to safeguarding against risks
1. DATA
Text, Numeric, Audio, Image, Video

2. MACHINE LEARNING
Supervised, Unsupervised, Reinforcement, Deep Learning

3. ARTIFICIAL INTELLIGENCE
Computer Vision, Natural Language Processing, Chatbots, Robotics
Predictive Modeling with Machine Learning

Historical Data → Machine learning algorithms → Model
New Data → Model → Predictions
Key AI capabilities

**Computer vision** processes images or video in order to identify objects or interpret scenes or events.

**Natural language processing (NLP)** analyzes or synthesizes text of human languages such as English, Spanish, or Arabic.

MIT, Harvard: Identifying Infection in surgical scars
Key AI capabilities (...continued)

Speech or audio recognition analyzes audio files to recognize specific sounds or speech patterns. Speech recognition often relies on NLP to transcribe speech into written text.

Advanced Analytics carries out sophisticated analysis of multiple data sources, structures.

Content Generation and Interpretation creates new text, images, video from understanding of key patterns in training text, images, video.

Rainforest Connection: Detecting Illegal Deforestation via monitoring chainsaw sounds
Practical Examples: Chatbots

Users request information from a system, often using written or spoken queries.

Key capabilities: Natural language processing, speech recognition, conversational interfaces (chatbots)

Application areas: Health, Agriculture, Financial inclusion, humanitarian aid
Practical Examples: Text and Image Classification in Humanitarian response

Uses satellite imagery and other geographic information (e.g. geotagged user data) rapidly process information for situational awareness, disaster response planning, resource allocation, etc.

**Key capabilities:** Computer vision, geospatial analysis, NLP

**Application areas:** Humanitarian assistance, disaster response, law enforcement, policy planning

[Image of Ushahidi crisis mapping platform in Haiti]

[Image of CMU Software Engineering Institute: Disaster damage assessment with computer vision]
Practical Examples: Personalized predictions/ratings

Uses diverse types of data (behavioral, demographic, economic, etc.) to predict specific outcomes or behaviors.

Key capabilities: Advanced Analytics, NLP

Application areas: Employment, credit scoring, law enforcement, health

https://twitter.com/talamobile/status/671525771546988544
Practical Examples: Vision & Audio diagnostics

Analyzes images or sounds (often captured via smartphone) to diagnose disease.

Key capabilities: Computer vision, speech or audio recognition, NLP

Application areas: Health, Agriculture

MIT, Harvard: Identifying Infection in surgical scars

Ubenwa.ai uses machine learning to analyze baby cries in order to identify perinatal asphyxia at early stage

Makerere University AI Lab tests smartphone app for diagnosing malaria
Values for Machine Learning and AI -
Key considerations relevant to Fairness

- Equity
- Representativeness
- Explainability
- Auditability
- Transparency
- Suitability & Added Value
Equity

- Does an ML model disproportionately benefit or harm some individuals or groups more than others?
What does equity mean with respect to machine learning?

ML models can perform significantly better for one group than another, creating an uneven opportunity to utilize ML technology (e.g. language and image processing tools).

Equity

- Does an ML model disproportionately benefit or harm some individuals or groups more than others?

(See this article by Abdi Latif Dahir for more info on African languages in voice recognition tools)
Equity

Does an ML model disproportionately benefit or harm some individuals or groups more than others?

ML models can fail equally often across groups, but produce systematic differences in the type of error each group experiences (e.g. diagnostics, scoring/eligibility applications).
Equity

Does an ML model disproportionately benefit or harm some individuals or groups more than others?

What does equity mean with respect to machine learning?

ML models can be technically accurate, yet reinforce existing inequities and social bias (e.g., credit scoring, hiring, recommender apps)
Representativeness

Is the data used to train the ML models representative of the people who will be affected by the model’s application?

Image credit: Belle Demont
Representativeness

Is the data used to train the ML models representative of the people who will be affected by the model’s application?

Why does representativeness matter in machine learning?

If data aren’t representative of the real-world context in which model is used, ML models can produce misleading results that contribute to inequitable outcomes.

Explainability

Can individual predictions or decisions be explained in human-friendly terms?

Source: *Interpretable Machine Learning*, a book by Christopher Molnar
Explainability

Can individual predictions or decisions be explained in human-friendly terms?

Why is explainability important?

The choice of algorithm affects both model accuracy and our understanding of how predictions are made. If we can’t determine how a model is using input data, it is harder to identify when they produce unfair outcomes.
Auditability

Can the model’s decision-making processes and recommendations be queried by external actors?
Auditability: Can the model’s decision-making processes and recommendations be queried by external actors?

Opening up the model’s decision-making process for question and inspection increases likelihood of identifying potential harms and biases ahead of time.

Why is auditability important?
Accountability

Are there mechanisms in place to ensure that someone will be responsible for responding to feedback and redressing harms, if necessary?

Accountability

Are there mechanisms in place to ensure that someone will be responsible for responding to feedback and redressing harms, if necessary?

Without strong commitments to monitor outcomes, work collaboratively, and willingness to learn from failures, unintentional harms of machine-learning based tools may go unaddressed.
Other Considerations: Relevance & Added Value

- Is the use of ML in your context solving a relevant problem?
- Is the application of ML technology adding value (e.g. informing more accurate, timely, actionable results?)
Safeguards for Fairness in Machine Learning
Awareness Raising and Practical Tools and Resources

- Webinar and trainings
- Toolkits
- Tools that make it easier for non-coders to engage with and test machine learning based applications
Technical Approaches to Fairness in Machine Learning

- Addressing fairness consideration in the technical decisions of model development
  - data selection
  - choice of algorithm
  - model performance metrics

- Technical approaches to bias checks, greater interpretability
Capacity Strengthening and Diversification of Workforce

- Including people of different cultural, educational, economic, social, ethnic backgrounds to understand and address issues related to fairness

- Strengthening capacity for local innovation and technology development
Strengthening Digital Ecosystems

- Strengthening inclusive and secure data systems
- Supporting local innovation and diversity in start-up ecosystems
- Shaping policy frameworks for fair and accountable use of digital technology
END
Next Up!

June 18th, 2020
11am ET