Ethical Issues in the Development of Complex Machine Learning Algorithms

Dr. Sara R. Jordan

Policy Counsel, AI

Future of Privacy Forum

&

School of Public and International Affairs

Virginia Tech

Presentation Overview

- Principle Proliferation
- Recent Advice
  - World Economic Forum
  - General Services Administration
- Simplifying Principles: Accountability in ML Development
- Explicability in ML Deployment

Principle Proliferation

- > 230 statements on “AI ethics”, “Data ethics”, “ethics for machine learning”
  - Consistent 7 principles
- Increase in “targeted” applications of principles
  - “Ethically aligned design for ____” series
- Increase in attention to use of principles for audit functions
  - UK ICO, EU working groups

Recent Advice

US General Services Administration

- 7 Data Ethics Tenets
  1. Be aware of and uphold applicable statutes, regulations, professional practices, and ethical standards
  2. Be honest and act with integrity
  3. Be accountable and hold others accountable
  4. Be transparent
  5. Be informed of developments in the field of data science
  6. Be respectful of privacy and confidentiality
  7. Be respectful of the public, individuals and communities

World Economic Forum Applications

- Risk-benefit analysis based on 12 considerations
  1. Justify the choice of AI use
  2. Adopt a multi-stakeholder approach
  3. Consider relevant regulations and best practices
  4. Apply RBA across lifecycle
  5. User centered/ case based approach
  6. Lay out a risk prioritization schema
  7. Define performance metrics
  8. Define operational roles
  9. Specify data requirements and flows
  10. Specify lines of accountability
  11. Supporting a culture of experimentation
  12. Create educational resources

Simplifying Principles: Accountability

- In the absence of definitive regulation and enforcement authority*, the principle of accountability encourages
  - Attributability
  - Answerability
  - Action

*not explicitly encouraging this:

- Actions of accountability
  - Attributability: modular documentation in code, documentation about code for internal use, documentation about code for external consumption
  - Answerability: development leadership and sign off, building in redundant oversight
  - Action: authoritative and traceable go/ no-go choices

Simplifying Principles: Explicability

- Lessons from the A-Levels in the UK
  - Deformation of the term “algorithm” into a cudgel
    - Internal Explicability: 319 page Ofqual report
      - Careful methodology, clear pathways for decisions
      - Well-designed charts and graphs
        - Explicability for statisticians
    - External Explicability
      - Did not include “tweetable” tidbits
      - Clash of professional cultures and language
        - Was school testing history a corrective to grade inflation or was it taking context into account?
        - Not explicable for education community or lay public

Accountability and Explicability in Development and Deployment

Internal

- Attributability
  - This module contributes ___ to the end product
- Answerability
  - __ members of the team contributed ___ components to this module
- Action
  - ___ team lead made ___ go/no-go decision to use this

External

- Attributability
  - This ML technique was chosen from the following set to perform these tasks
- Answerability
  - This ML technique was chosen because it performed at this level, which is the appropriate level because
- Action
  - __ members of the team made the following choices for uses of this technique for these tasks at this expected performance level

Accountability and Explicability Communication

- Pay attention to and use ethics language
  - Correcting grade inflation or accounting for context
- Tweetable tidbits and simplified communications
- Use language carefully
  - Statistical model, machine learning, algorithm, decision-support system

Other components for Accountability

- Future of Privacy Forum Ethical Data Sharing Review Committee
  - developing risk assessment methods that help assure objectivity and independence in evaluation of projects
    - Data asset valuations
    - Data-use risk evaluation
    - Model-use risk evaluation

Thank you

- Contact: sjordan@fpf.org; srjordan@vt.edu