Risk identification questionnaire for unintended bias in machine learning development lifecycle

Michelle Seng Ah Lee and Jat Singh
Compliant & Accountable Systems Group, Cambridge Computer Lab

UNIVERSITY OF CAMBRIDGE
Department of Computer Science and Technology

AAAI / ACM conference on ARTIFICIAL INTELLIGENCE, ETHICS, AND SOCIETY

Summary

Developed and trialed a bias risk identification questionnaire with industry practitioners (excerpt to the right)

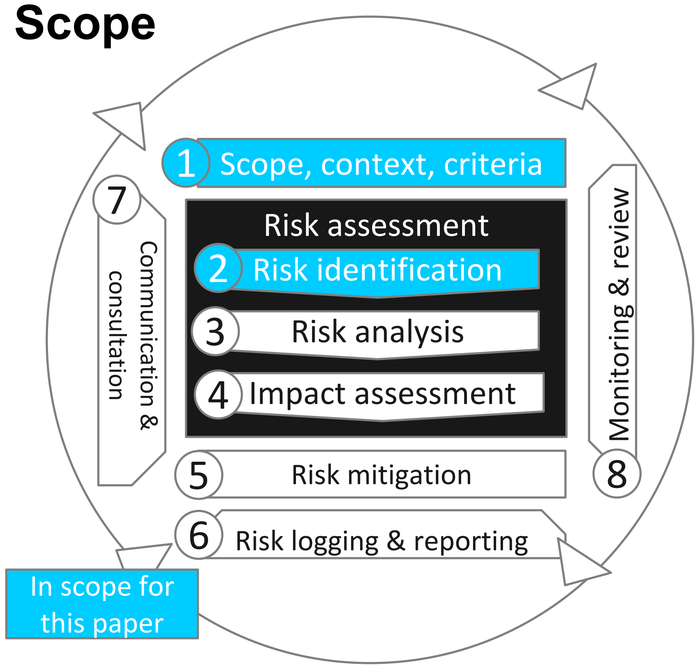
86% agree questionnaire can "proactively diagnose unexpected issues"

Full questionnaire:
https://github.com/michelleslee/bias_in_lifecycle

Questionnaire content

A. Background information
B. Design: historical / external bias
C. Data collection: representation bias
D. Feature selection: measurement bias
E. Model build: aggregation bias
F. Model evaluation: evaluation bias
G: Productionisation: deployment bias

Scope



1. Case studies
2. Bias risk identification questionnaire
3. Bias/fairness quantification, open source toolkits
4. Impact assessments, trade-offs of objectives
5. Technical "de-biasing" and/or non-technical mitigations (e.g. policy changes)
6. Logging: e.g. model cards, data sheets
7. Approval process, stakeholder engagement
8. Automated controls (e.g. anomalous data entry), checklists (e.g. open source toolkit license permissions)

E.g. insurance fraud

Historical bias: *Identification of potential criminal acts regularly accused of racial or faith based biases*

Representation bias: *lower data quality for claimants with poor English, "unknown unknowns" of missed fraud*

Measurement bias: *Attempts to locate geographical patterns of fraud can create unintended correlations with particular racial groups*

Deployment bias: *fraud investigators reinforce biases as they act as feedback mechanism*

Contact

Michelle Seng Ah Lee
University of Cambridge
sal87@cam.ac.uk