Red flags and checklist: Outline

Part 1: (what questions and why)

Accuracy and appropriate use– the tool's ability to accurately predict recidivism, crime, etc.– i.e. a higher rate of predictive accuracy.

§ Is the data used in the algorithm appropriate, consistent and compatible for the purpose and use of the technology?
§ Does the technology reveal details about accuracy rates, margin of error and measures of uncertainty?

Training Bias and Fairness– the tool's ability to treat all users/communities fairly without racial bias, or discrimination the underlying influence of the data and the people who built the algorithm

§ How accurate and recent is the data used? How accurately does the data represent real-world situations?
§ Is the data representative? does it include equal sample population from diverse voices/communities and context?
§ Is the data local and representative of the community?
§ Does the algorithm make final assessments and recommendations without an interference by a person?
§ Which data and variables were used in the algorithms?
§ Are variables such as race, gender, zip codes, age that are proxies for discrimination included in the algorithm?
§ Is the technology reinforcing or amplifying existing bias?

Impact – examine the algorithm in terms of effect it will have on the community members/users

§ What is the overall impact of the technology?
§ What is the type of impact? i.e. access to benefits or services, financial, privacy, freedom, rights.
§ Who will be impacted by the technology?
§ Will the technology provide positive changes to all community members?
§ What are the potential negative impacts of the technology on community/ community members based on age, race, religion, national origin, gender, disability?
§ Will the data follow users throughout their lifetime, affect their reputations and impact future opportunities?

Transparency – the extent to which the algorithms (codes, data) used are available to users.

§ Are the data, models and algorithms available for reviewing and understanding?
§ Does the technology explain the goal, purpose and intent of the algorithm?
§ Is there any information about the data used in terms of accuracy, uncertainty, limitations, assumptions and the sample population?
§ How was the data collected, transformed and analyzed?
§ Was the data used private or public, are there any identifying aspects that could cause privacy harm to the individuals.?
§ Which tools were used to model the data? What features or variables are used in the algorithm?

Accountability –

§ Who or what made the decisions about the data used?
§ How were the decisions made?
§ Can you review or audit those decisions?
§ How was the algorithm tested before being used?
§ How is it tracked, or modified?

Interpretability – the extent to which the tool can be interpreted by users, government agencies, officials, stakeholders and community organizations

§ Can users correctly interpret the outputs of the technology?
§ Does the technology provide clear documentation on how to interpret the outputs?

Operability - the extent to which the tool can be administered by government agencies and officials.

§ Will the agencies and/or organizations operating the technologies be trained how to use them?
§ Is the algorithm straightforward and uses non-technical terms that describe the technology and explain its use, inputs and outcomes?
§ Who will have access to the technology?

Security & privacy –

§ Will the data collected be used only for the purpose intended?
§ How will personal information be protected from data breaches?
§ Does the technology provide comparable privacy protections to different groups?

Part 2:

**Add 2 -3 specific examples of technologies and their harms.

Part 3 additional features/resources:

AI Now. 2018. Algorithm accountability policy toolkit.

Bavitz, Christopher, Sam Bookman, Jonathan Eubank, Kira Hessekiel, and Vivek Krishnamurthy. 2018. Assessing the Assessments: Lessons from Early State Experiences in the
Procurement and Implementation of Risk Assessment Tools. *Berkman Klein Center for Internet & Society research publication.*

Center for Government Excellence. Ethics and Algorithms Toolkit https://ethicstoolkit.ai/

Diakopoulos N. 2016. Accountability in Algorithmic Decision Making. *Communications of the ACM, Vol. 59(2).*
https://cacm.acm.org/magazines/2016/2/197421-accountability-in-algorithmic-decision-making/fulltext