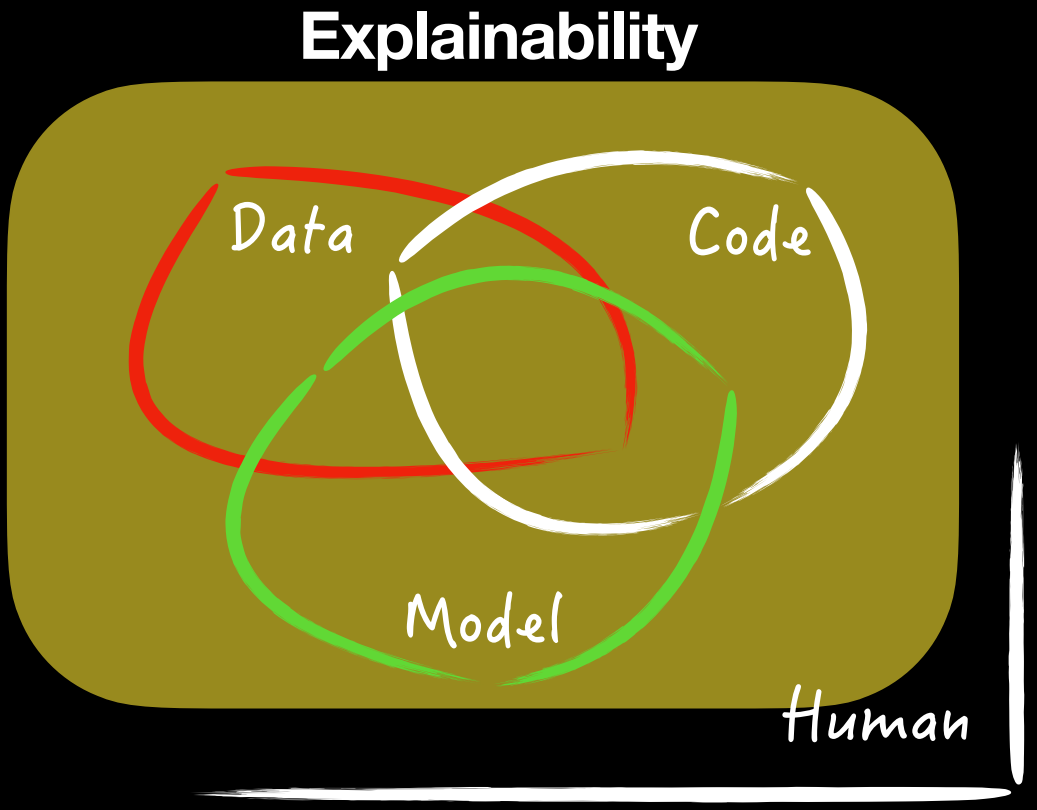
Accountability and Auditability

Jin Guo 12th Nov

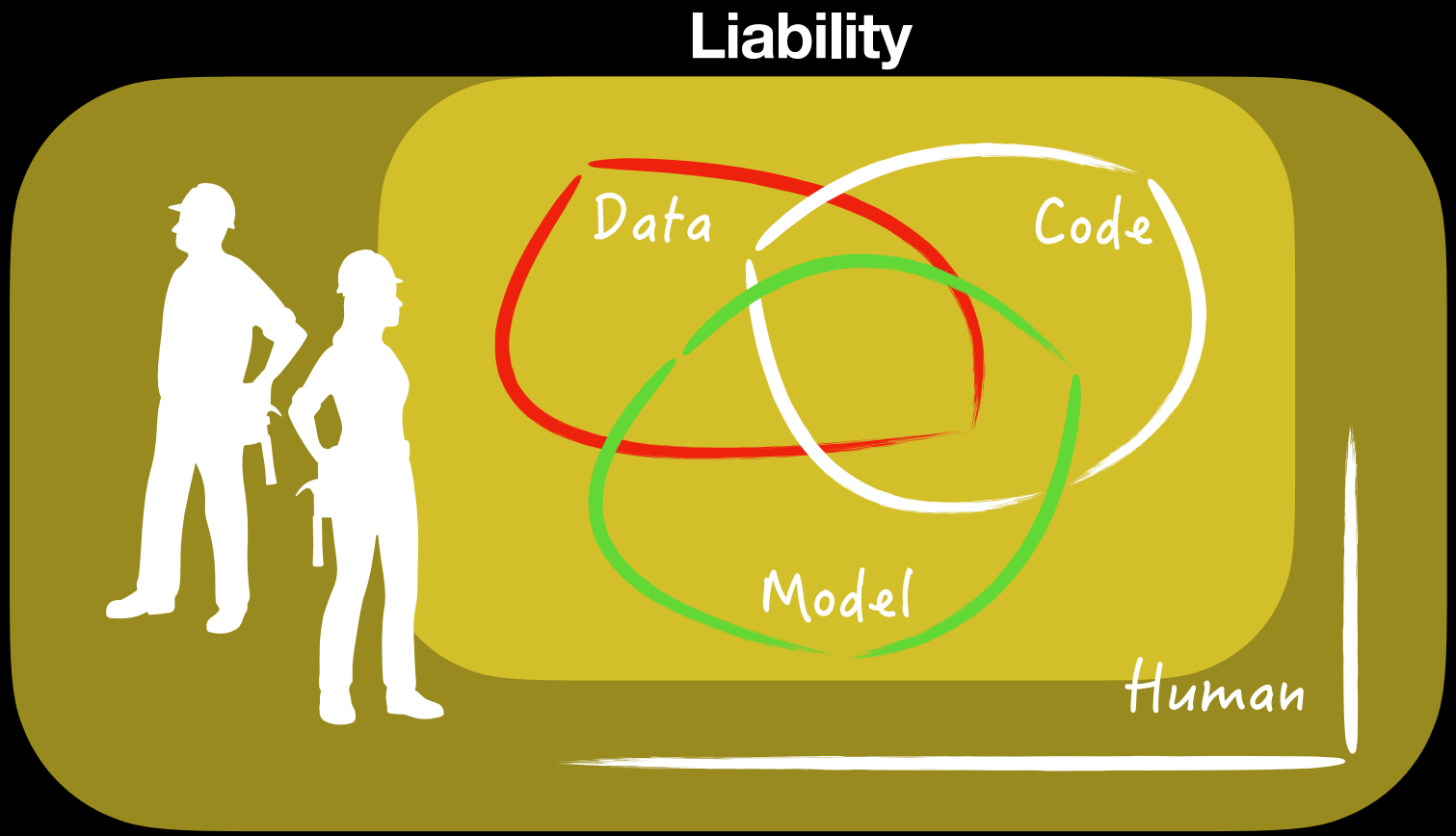
Accountability

A clear acknowledgement and assumption of responsibility and “answerability” for actions, decisions, products and policies.



Accountability

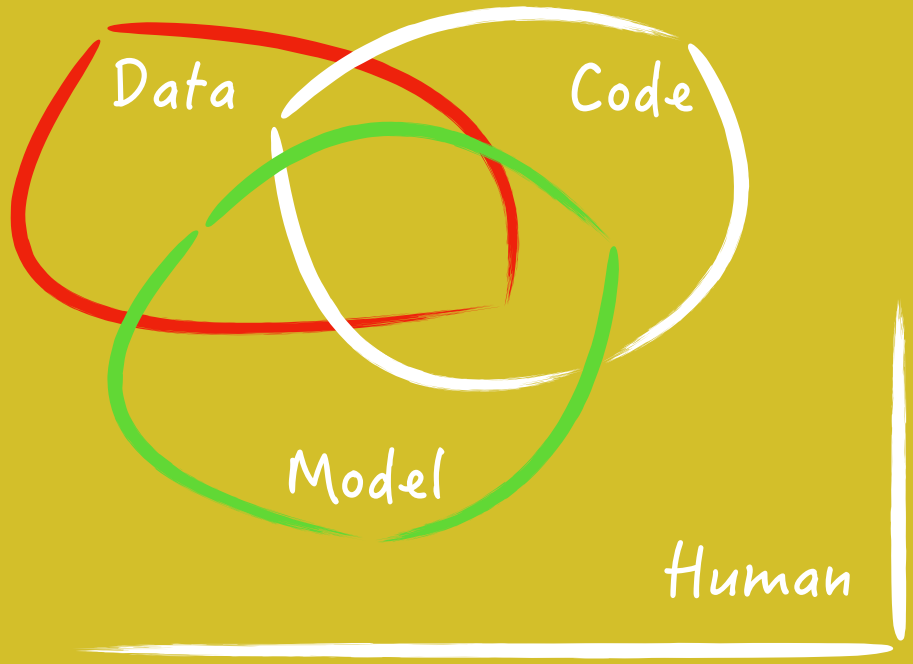
A clear acknowledgement and assumption of responsibility and “answerability” for actions, decisions, products and policies.



Accountability

A clear acknowledgement and assumption of responsibility and “answerability” for actions, decisions, products and policies.

Fairness, Explainability, Auditability, Responsibility, Accuracy, etc.



Challenges

- Indicator of Accountability at different level
- Need to establish clarity and/or agreement on the roles of various actors and stakeholders in ensuring accountability in AI.

Auditability

The process should be understandable by people apart from process participants, who can check that process standards are being followed and make suggestions for process improvement.

External and Internal

Gender Shade — Example of External Auditing

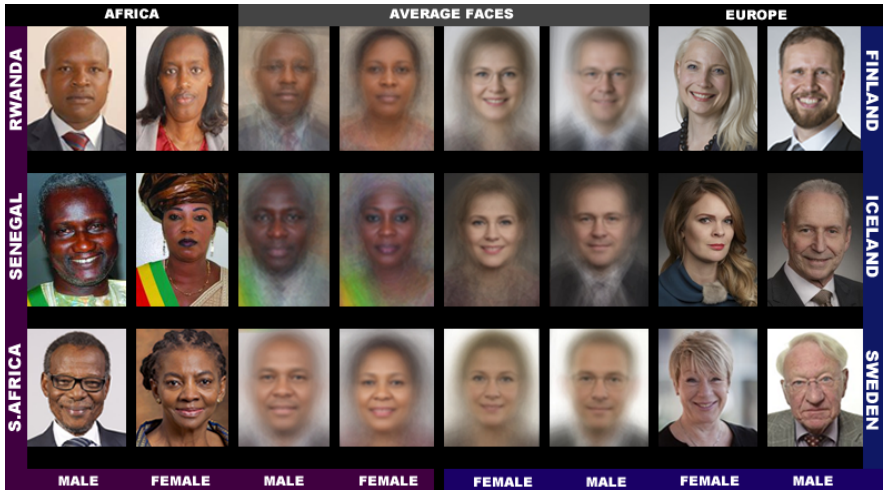


Figure 1: Example images and average faces from the new Pilot Parliaments Benchmark (PPB). As the examples show, the images are constrained with relatively little variation in pose. The subjects are composed of male and female parliamentarians from 6 countries. On average, Senegalese subjects are the darkest skinned while those from Finland and Iceland are the lightest skinned.

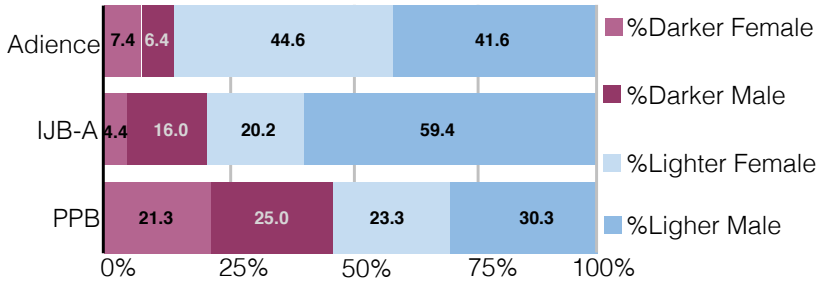


Figure 3: The percentage of darker female, lighter female, darker male, and lighter male subjects in PPB, IJB-A and Adience. Only 4.4% of subjects in Adience are darker-skinned and female in comparison to 21.3% in PPB.

Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." In *Conference on fairness, accountability and transparency*, pp. 77-91. 2018.

Gender Shade — Example of External Auditing

Microsoft, IBM, Face++

“using advanced statistical algorithms that ‘may not always be 100% precise’”

- All classifiers perform better on male faces than female faces (8.1% – 20.6% difference in error rate)
- All classifiers perform better on lighter faces than darker faces (11.8% – 19.2% difference in error rate)
- All classifiers perform worst on darker female faces (20.8% – 34.7% error rate)
- Microsoft and IBM classifiers perform best on lighter male faces (error rates of 0.0% and 0.3% respectively)
- Face++ classifiers perform best on darker male faces (0.7% error rate)
- The maximum difference in error rate between the best and worst classified groups is 34.4%

Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." In *Conference on fairness, accountability and transparency*, pp. 77-91. 2018.

SMACTR — Example of Internal Auditing

- Draw lessons from other domains:
  - Aerospace
  - Medical devices
  - Finance

Design Checklist

Traceability

FMEA (Failure Modes and Effects Analasis)

Raji, Inioluwa Deborah, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. "Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 33-44. 2020.

SMACTR — Example of Internal Auditing

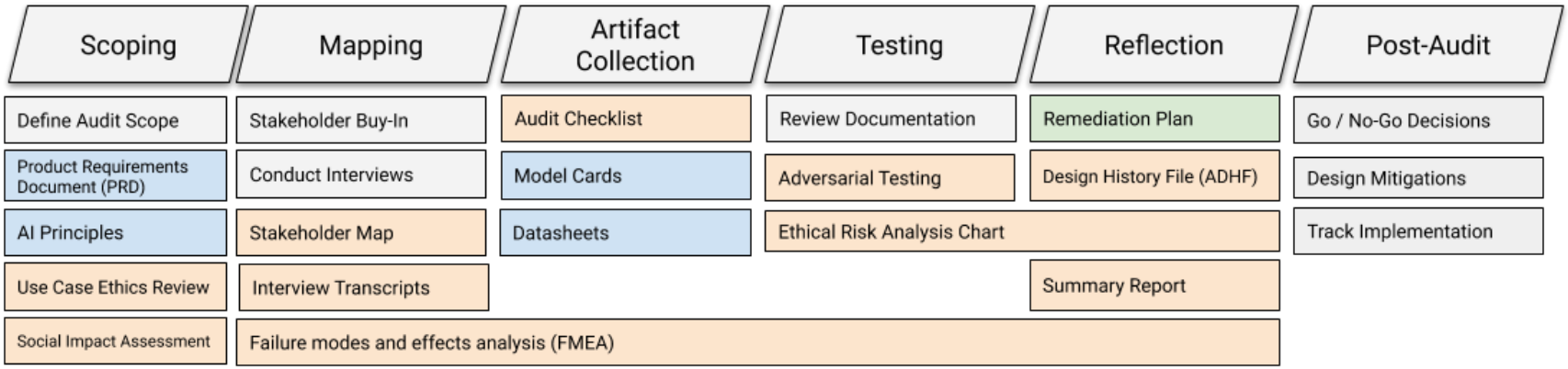
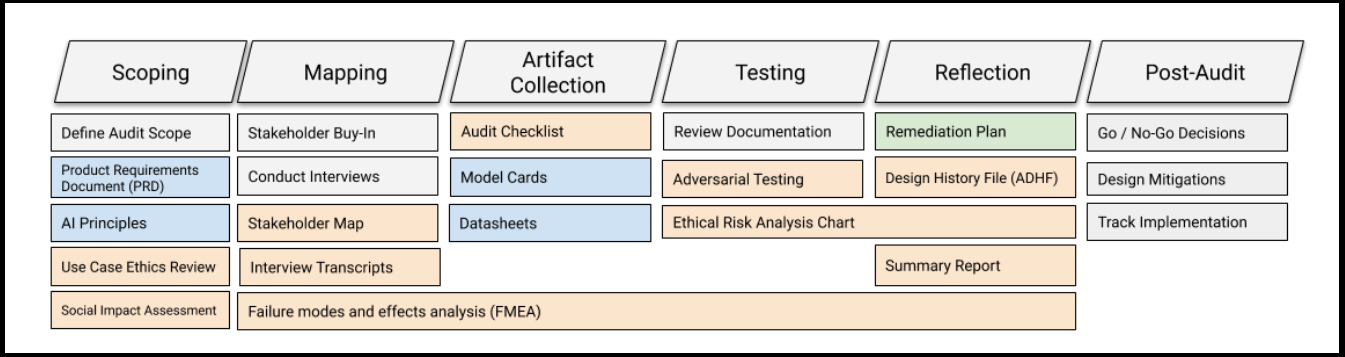


Figure 2: Overview of Internal Audit Framework. Gray indicates a process, and the colored sections represent documents. Documents in orange are produced by the auditors, blue documents are produced by the engineering and product teams and green outputs are jointly developed.

Raji, Inioluwa Deborah, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. "Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 33-44. 2020.

Activity



- Each room is assigned with one stage from the above framework.
  - (Individually) Read the related section from the paper “Closing the AI accountability gap” Section 4.
  - (Individually) Choose one artifact that might applicable to your case study project form the course assignments.
  - (Individually) Brainstorm the content of the artifacts if possible. If not, consider how to generate the artifacts (plan).
  - (Group) Share within the group: your project context, the artifact plan or draft from this activity.