

Evaluating AI Systems

MoreThanFair's guide to creating automated systems responsibly

Guiding Principles

1

AI Should Be Safe and Effective

Developers of AI systems for lending should proactively consider the ethical implications of their work and ensure that it doesn't infringe upon existing laws, human rights, or democratic values.

2

AI Should Be Explainable

AI systems for lending should be capable of explaining their purpose, subject to oversight ensuring that they function correctly, and communicated in plain language to users to promote accountability and trust with the public.

3

AI Should Be Equitable and Fair

AI systems for lending should be developed in a way that mitigates any form of discrimination and considers the needs of all users, regardless of race, gender, or socioeconomic status.

4

AI Should Be Secure

AI systems for lending should be developed and used in ways that prioritize security and establish safeguards to protect the privacy and personal data of individuals.

5

AI Should Be Accountable

AI systems for lending should be evaluated based on outcomes they engender, including with regard to any societal harms that may arise from the use of the systems.

Practical Questions to Ask

Know Your Purpose

Before you begin building an AI system, speak to key stakeholders and collaborators about the impetus behind a project. What business problem(s) are they trying to solve? What technologies will be needed to achieve these objectives? What potential risks do we foresee for consumers, institutions, and society at large?

Collect the Right Data

Building an AI system likely involves training a model based on some amount of data. So what data is available? Is it appropriate for this use case? Does the data allow the model to operate accurately across groups defined by race, sex, and gender (when relevant)? Addressing issues with data quality, data paucity, and data insufficiency can dramatically reduce risks down the line.

Assemble Diverse Teams

AI systems are built by humans, so it's important to focus on team composition. Who are the people responsible for making decisions at each step in the process? Is that group of people sufficiently diverse in their backgrounds and perspectives? Diverse and well trained teams can often spot problems and confront challenges more rigorously.

Assess Outcomes Holistically

Evaluating an AI system involves more than just seeing how well it performs a designated task. You have to consider the broader picture and potential externalities. How well is the system serving the complete set of users? Are the results of the model clear and explainable? What are the circumstances in which the model shouldn't be used?

Continue to Monitor and Validate

Launching an AI system into production isn't the end of the road. Far from it. Stakeholders need to continue to monitor the performance of the system to see if it is sufficiently valid and robust in the real world. Did the model rest on any assumptions that have been proven incorrect? Is the model protected from cyber attacks or other bad actors?

1

2

3

4

5