

Name: Masood Ullah

Email: masoodullahansari@gmail.com

Resource: <https://ai.google/responsibilities/responsible-ai-practices/>

Assignment

Responsible Ai Practices

AI is creating new opportunities for improving people's lives and raising questions about building fairness, interpretability, privacy, and security into AI systems.

We should committed to responsible AI development and sharing knowledge and resources with the community by actively researching and developing ways to address AI's fairness, interpretability, privacy, and security concerns. Our current work and recommended practices are shared below and we will continue to adapt and learn as we progress.

General Recommended Ai Practices

1. Use a Human-Centered Design Approach

Design features with appropriate disclosures and consider augmentation and assistance to ensure a good user experience and assess the true impact of AI predictions, recommendations, and decisions. Model potential adverse feedback and engage with diverse users and use cases to incorporate feedback and increase the number of people who benefit from the technology.

2. Identify Multiple Metrics to Assess Training and Monitoring

Use multiple metrics to understand tradeoffs and consider metrics including user feedback, system performance, and false positive and false negative rates. Ensure that your metrics are appropriate for the context and goals of your system.

3. When Possible, Directly Examine Your Raw Data

Analyze raw data to understand it and, if necessary, understand input data as much as possible while respecting privacy. Identify and address training-serving skew during training and evaluation, use the simplest model that meets performance goals, and consider the relationship between data labels and items being predicted. Data bias is important to consider in relation to AI and fairness.

4. Understand the Limitations Of Your Dataset and Model

Avoid using a model trained to detect correlations to make causal inferences, and communicate the scope and coverage of the training data to clarify the capabilities and limitations of the model. Communicate limitations to users to improve feedback and better educate them.

5. Test Test Test

Learn from software engineering and quality engineering best practices to ensure the AI system is working as intended and can be trusted. Conduct unit and integration tests, detect input drift, use a gold standard dataset, conduct iterative user testing, and apply the poka-yoke principle of building quality checks into the system.

6. Continue to Monitor And Update the system After Deployment

Monitor your model and take real-world performance and user feedback into account. Allow time to address issues and consider both short- and long-term solutions. Analyze candidate and deployed models before updating to ensure the update will not negatively impact overall system quality and user experience.