

Artificial Intelligence in Radiation Oncology

ICRO - PRODVANCE, Thiruvananthapuram

28 July 2024

Reducing Bias and responsible AI

Anupama Ananthasairam

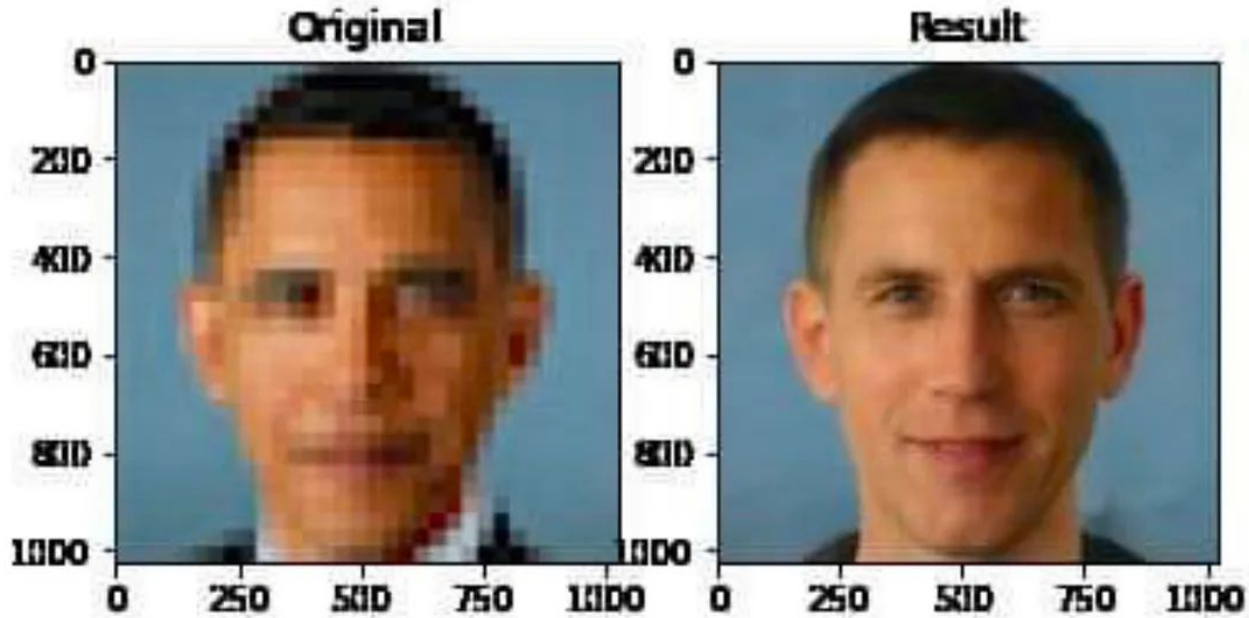
 KARKINOS



Objectives

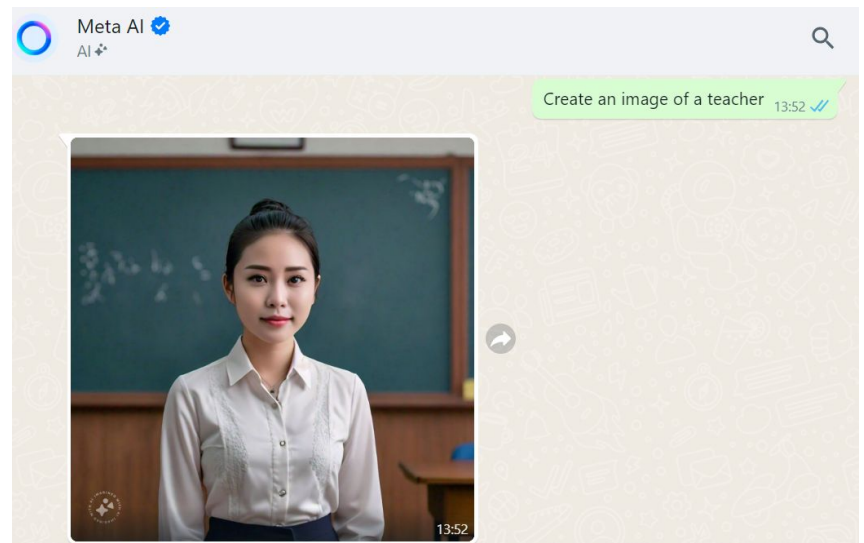
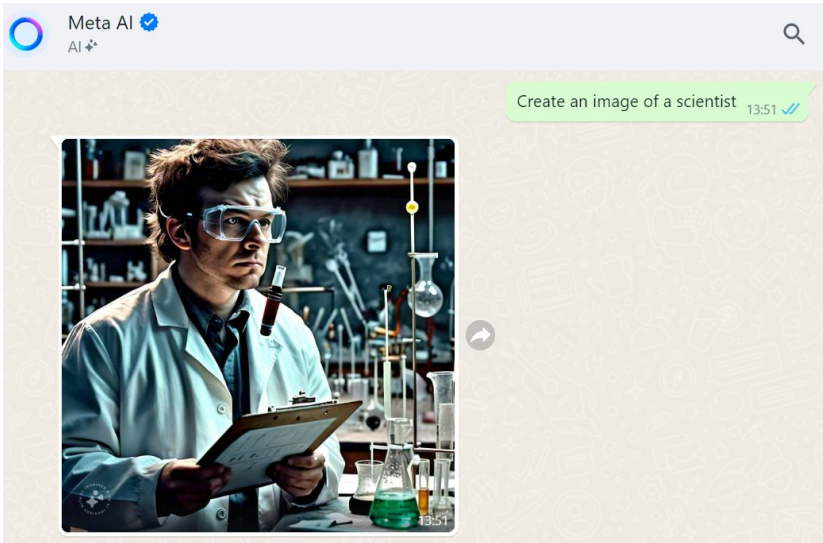
- What is bias in AI and its impact in radiation oncology
- Identify sources of bias in AI
- Techniques to reduce the bias
- Importance of human in the loop
- What is Responsible AI?
- Importance of data governance in implementing responsible AI

What is Bias in AI?



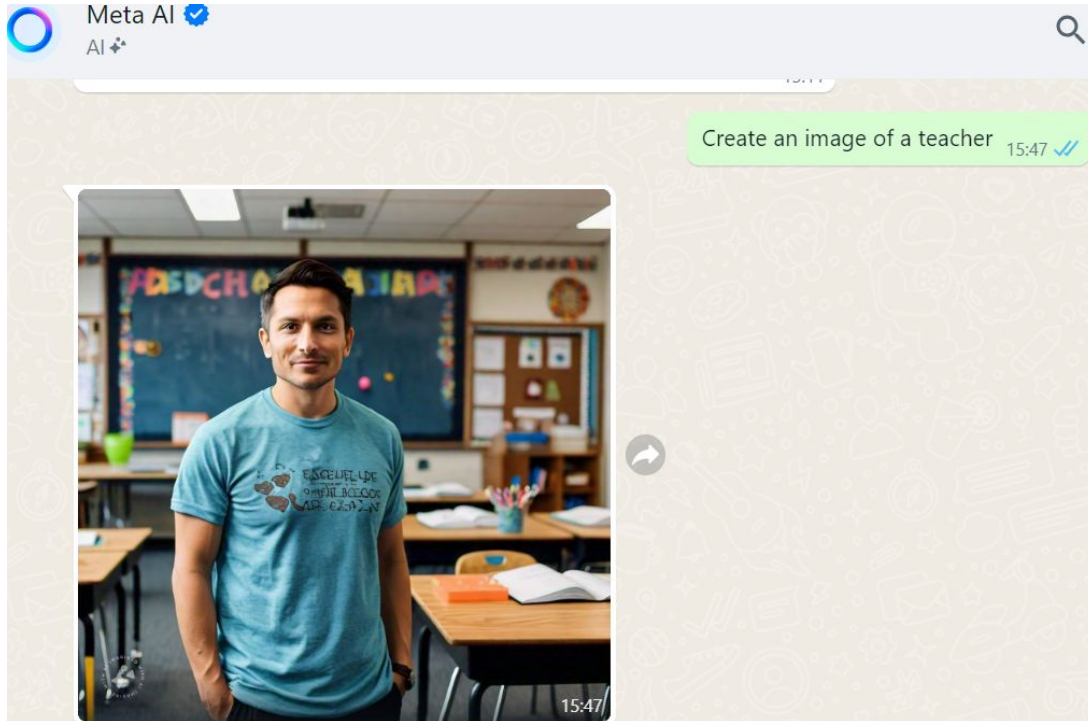
What's happening today?

First prompt -



What's happening today?

Second Prompt -



What's happening today?



Finally 5th prompt...

Where can AI Bias occur?



Datasets

- Over/ Under representative sample data
- Inconsistent labeling



ML Algorithms

- Programming errors
- Unscientific weights allocation based on personal bias



Predictions

- Evaluation bias that chooses irrelevant metrics

Data Bias

- Response Bias
- Selection Bias
- Social Bias
- Automation Bias
- Group Attribution Bias



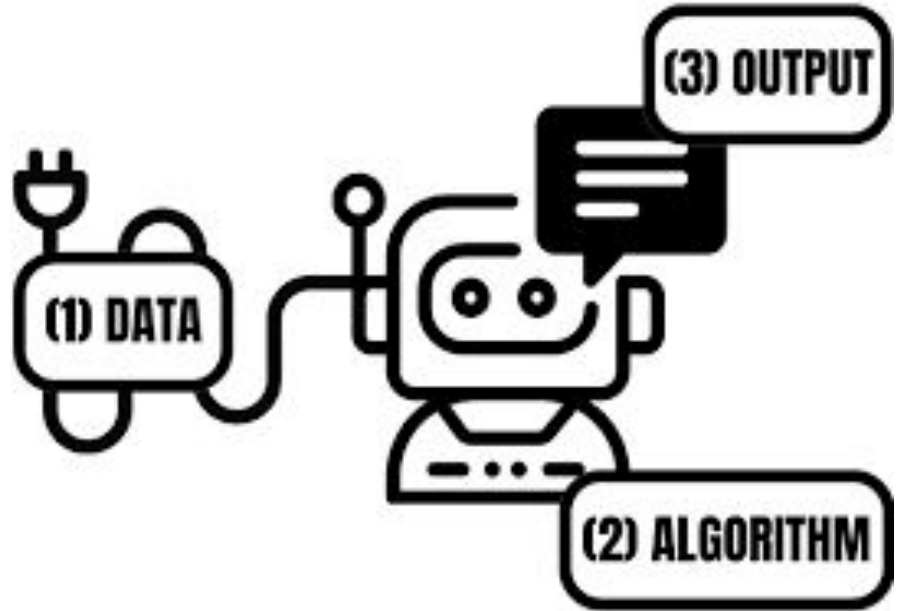
Algorithmic Bias

- Optimization Bias
- Model Design Bias
- Data Processing Bias
- Hyper parameter Bias
- Regularization Bias



Prediction Bias

- Confirmation Bias
- Anchoring Bias
- Overfitting Bias
- Omission Bias
- Evaluation Bias



Steps to reduce AI Bias in Radiation Oncology

- 1 Data Curation (Diverse, Representative, High quality datasets)
- 2 Data Annotation (Standardized guidelines, Multiple annotators)
- 3 Algorithmic Auditing (Continuous monitoring for drift, SHAP assessment)
- 4 Diverse development team (AI developer, SME, stakeholders)
- 5 Test Plan to have inclusive and diversity test cases
- 6 Encourage transparency (model interpretability, feature attribution)
- 7 Human - in - the -loop (Involve SMEs in every step)
- 8 Address Societal Biases (Review and validate Data Acquisition seriously)

What is Responsible AI?

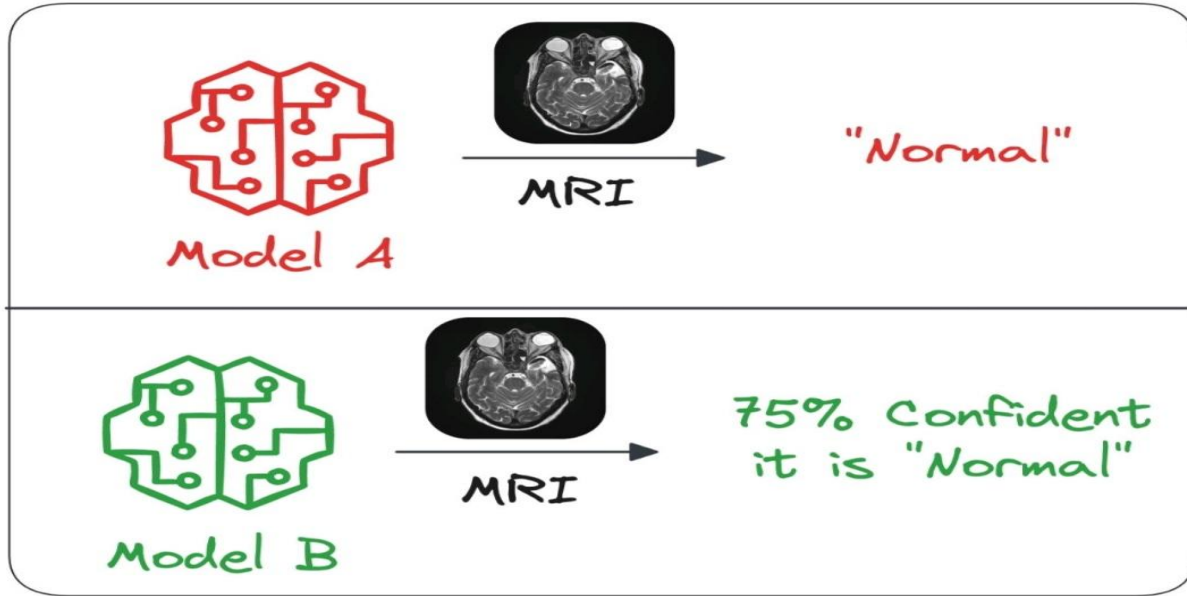


Responsible AI

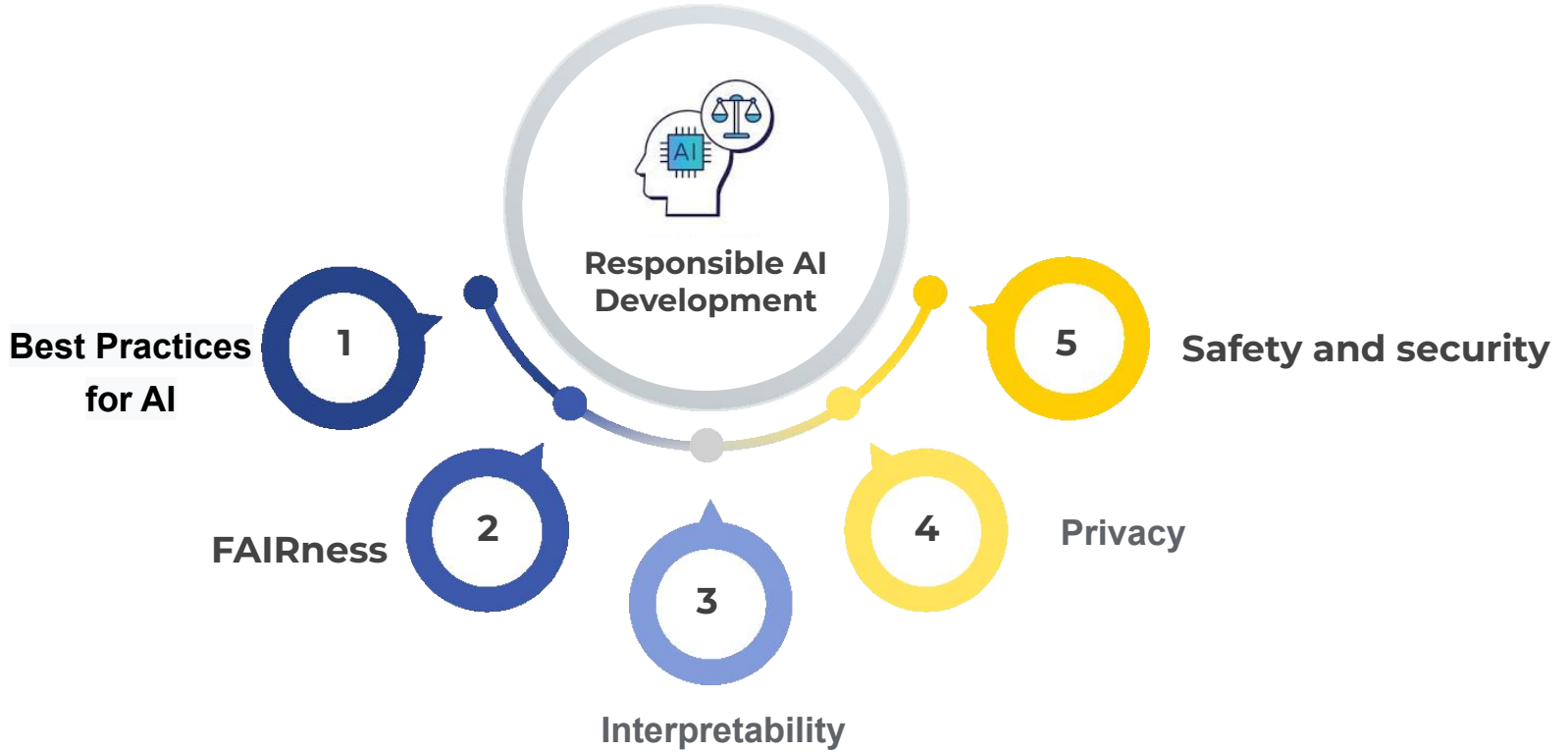
(Re-**spaan**-suh-bl **AY**-eye)

Responsible AI ensures that the development and deployment of AI systems aligns with ethical principles and values, including transparency and accountability. RAI principles and best practices help to reduce the potential negative impacts of AI caused by machine bias.

What do you prefer?



Responsible AI



Best Practices for Responsible AI Development

1. Embrace human-centered design
2. Use multiple metrics for evaluation
3. Examine raw data directly when possible
4. Understand dataset and model limitations
5. Test thoroughly
6. Continuously monitor and update post-deployment



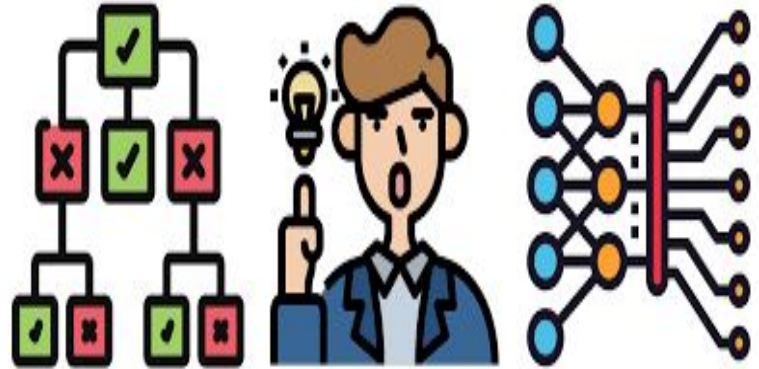
FAIRness

1. Foster an inclusive workforce with diverse knowledge and perspectives
2. Seek input from communities early in the research and development process
3. Assess training datasets for potential sources of unfair bias
4. Train models to remove or correct problematic biases
5. Evaluate models for disparities in performance
6. Continuously test final AI systems for unfair outcomes through adversarial testing
7. Utilize ML models to identify and address conscious and unconscious human biases and barriers to inclusion



Interpretability

- Develop a roadmap for achieving interpretability in your AI system.
- Prioritize interpretability as a key aspect of the user experience.
- Design AI models with interpretability in mind from the outset.
- Select metrics that align with the desired outcomes and tasks.
- Gain a deep understanding of the trained model's behavior and decision-making processes.
- Effectively communicate explanations of the model's outputs to users.
- Rigorously test and validate the AI system



Privacy, Safety and Security

- Data minimization
- Anonymization
- Encryption
- Access controls
- Data protection policies
- Transparency
- Consent
- Data storage
- Regular audits
- Adhere to regulations
- Privacy by design
- Human oversight





References

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, [online] 54(6), pp.1–35. doi:<https://doi.org/10.1145/3457607>.