Responsible AI using SciPy

Andrea Hobby

About Me

- First-year doctoral student at Johns Hopkins Bloomberg School of Public Health
  - Interested in developing responsible AI solutions to enhance patient safety and healthcare quality

Former Data Scientist at the Health Resources and Services Administration (HRSA)

- Former Data Scientist at Accenture Federal Services, focusing on data science solutions for state and federal governments
- Connect with me.
  - https://github.com/AndreaHobby
  - https://www.linkedin.com/in/andreahobby/
  - https://healthinnovation.substack.com/

Slides

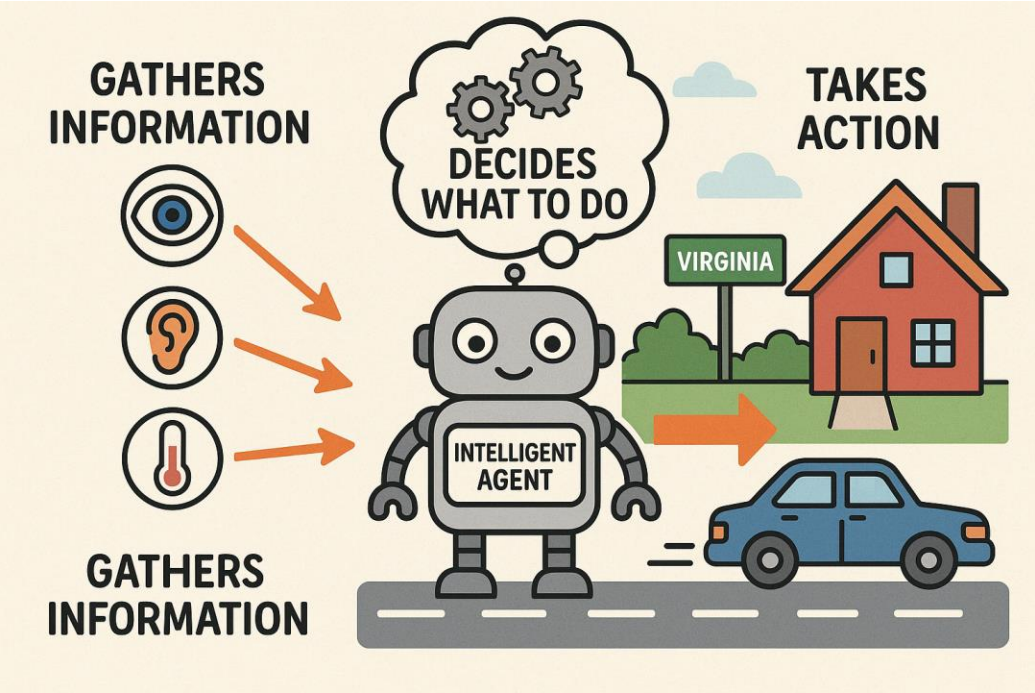
https://github.com/AndreaHobby/PyDataVirginia

What We'll Cover

- **Introduction**: Overview of AI, Responsible AI and its importance.
- **Principles of Responsible AI**: Fairness, Privacy and Security, Safety, Accountability, Transparency and Explainability.
- **SciPy Overview**: Key features.
- **3 Tutorials**: Real-world examples of Responsible AI using SciPy.
  - **Hands-on Tutorial 1**: Bias Detection and Mitigation using SciPy
  - **Hands-on Tutorial 2**: Sensitivity Analysis with SciPy and other python libraries
  - **Hands-on Tutorial 3**: Building Transparent and Explainable Machine Learning Models
- **Q&A**: Open the floor for questions.

Introduction

What is AI?

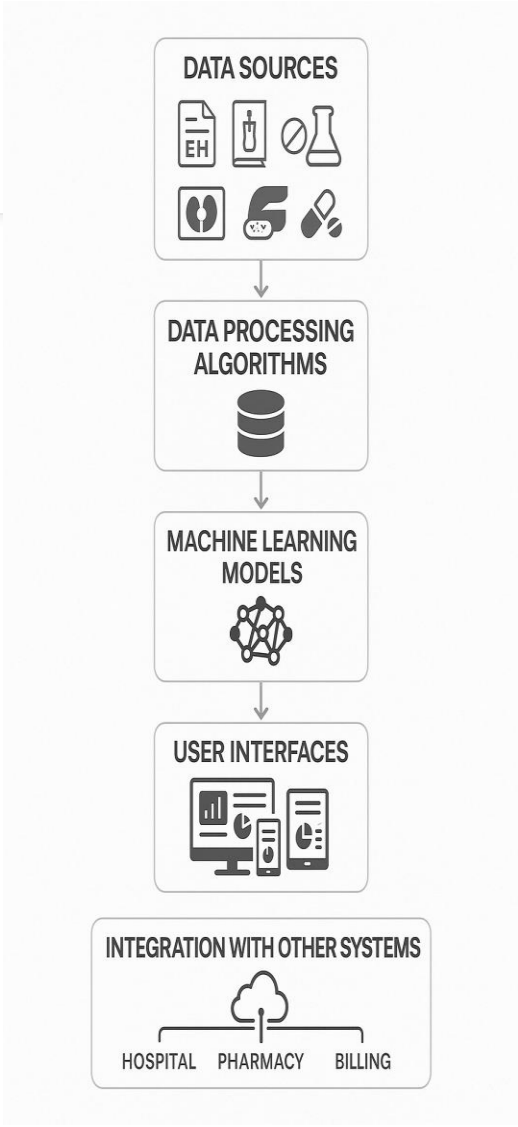


The goal of artificial intelligence (AI) is to create systems that can interact with their surroundings. These systems, referred to as "intelligent agents," are able to collect information from their surroundings, such as what they see or hear, and then use that knowledge to decide how to respond in a way that changes their surroundings. (Norvig 2021)

What is an AI System?

- An AI system is a tool that brings artificial intelligence to life. It's a computer program designed to do specific jobs or solve certain problems using AI algorithms.
- For example, AI systems can help with tasks like recognizing faces in photos, translating languages, or even driving cars. (Norvig 2021)

Example: AI System in a Hospital



What is Responsible AI?

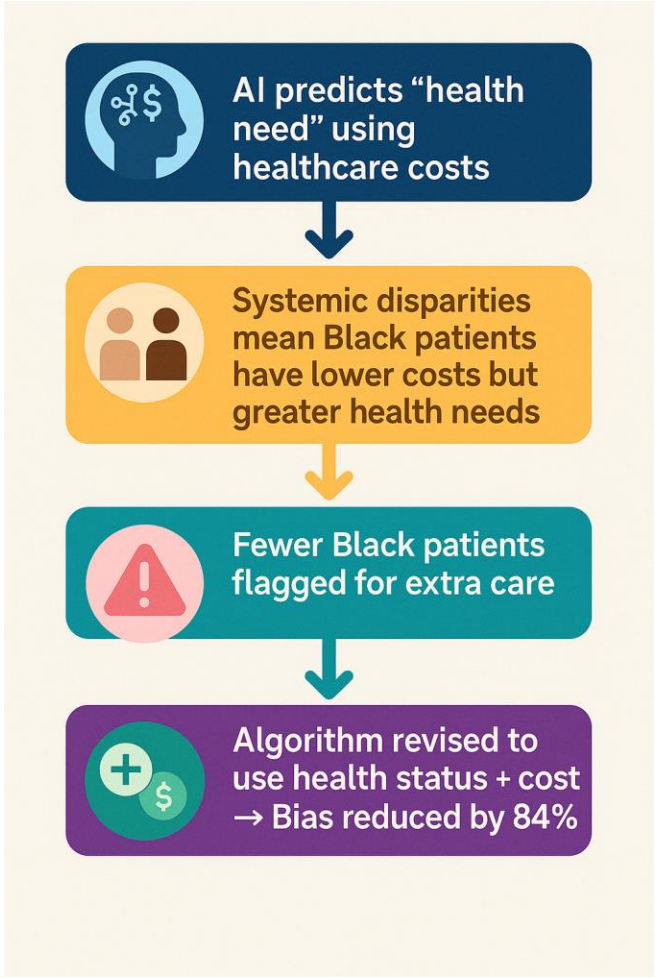
- Responsible AI ensures ethical, fair, and transparent use of artificial intelligence.
- Responsible AI is an approach to developing, deploying, and using AI systems that align with ethical principles and societal values. (IBM 2024)

Why is Responsible AI important?

- It is important for keeping people impacted by AI safe and those who create AI systems accountable.
- Examples in the healthcare industry
  - Improving accuracy and speed of Breast Cancer Diagnosis (McKinney 2020)
  - Reduce bias in machine learning models for Heart Disease (Li 2023)

Real-World Example: Biased Healthcare Risk Prediction Algorithm

Optum which is a part of UnitedHealth Group developed a risk prediction algorithm. In 2019, an error was found in this algorithm. (Weber 2023)



Principles of Responsible AI

General Principles of Responsible AI

- Fairness
- Privacy and Security
- Safety
- Accountability
- Transparency and Explainability

Fairness

The fairness principle is that AI systems should not make decisions that discriminate against individuals or groups based on their gender, race, ethnicity, religion, or other personal characteristics. (Dignum 2019)

Privacy and Security

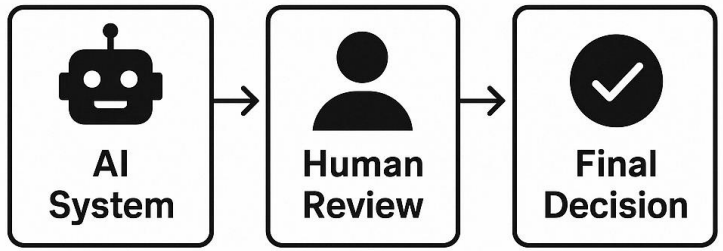
The privacy and security principle is about AI systems handling personal and sensitive data in ways that respect individual rights, maintain confidentiality, and protect against unauthorized access or misuse. (Dignum 2019)

Safety

The safety principle is so that AI systems operate reliably and as intended, without causing harm to people, property, or the environment. (Dignum 2019)

Accountability

The accountability principle focuses on the process by which people and institutions that design, develop, implement, and use AI systems are held responsible for the results and operations of these systems. This involves maintaining human oversight, guaranteeing decision traceability, adhering to legal requirements, and putting in place explicit procedures for observation, error or bias correction, and safety protection. (Dignum 2019)



Transparency and Explainability

- Transparency is being clear and transparent about how AI systems are developed, how they operate, what data they use, and how decisions are made. (Dignum 2019)
- Explainability is an AI system's ability to produce coherent, human-readable explanations for its decisions and outputs. (Dignum 2019)

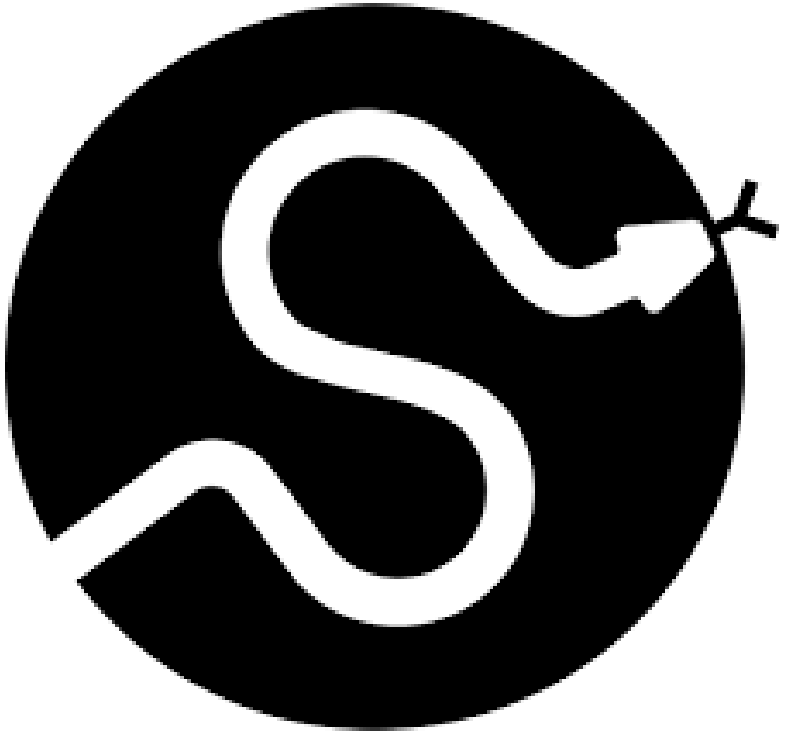
Best Practices for Implementing Responsible AI

- Implement Responsible AI principles
  - Diverse dataset
  - Human review
  - Secure the data
  - Transparent decision-making
- Use Responsible AI tools (e.g. Google's What-If Tool, Model Cards, Microsoft Azure Responsible AI Dashboard)
- Responsible AI python libraries (fairlearn, responsibleai, shap and others)
- Monitor AI Systems for abnormal or incorrect output

SciPy Overview

Introduction to SciPy

- SciPy is an open-source python library for scientific computing.



Key SciPy modules

- `scipy.optimize`,
- `scipy.integrate`,
- `scipy.stats`
- `scipy.signal`

Tutorials

- See GitHub for data and notebooks.

- https://github.com/AndreaHobby/PyDataVirginia/tree/main/notebooks

Q&A

References

- Abràmoff, M. D., Tarver, M. E., Loyo-Berrios, N., Trujillo, S., Char, D., Obermeyer, Z., ... & Maisel, W. H. (2023). Considerations for addressing bias in artificial intelligence for health equity. *NPJ Digital Medicine, 6*(1), 170.
- Coeckelbergh, M. (2020). *AI ethics*. MIT Press.
- Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way* (Vol. 2156). Cham: Springer.
- Google Cloud. (n.d.). Explainable AI. Retrieved April 16, 2025, from https://cloud.google.com/explainable-ai1
- How to implement AI – responsibly. *Harvard Business Review*. Accessed April 13, 2025. https://hbr.org/2024/05/how-to-implement-ai-responsibly
- IBM. (n.d.). Responsible AI. IBM Think. Retrieved April 17, 2025, from https://www.ibm.com/think/topics/responsible-ai
- Kim, J. Y., Hasan, A., Kellogg, K. C., Ratliff, W., Murray, S. G., Suresh, H., ... & Sendak, M. P. (2024). Development and preliminary testing of Health Equity Across the AI Lifecycle (HEAAL): A framework for healthcare delivery organizations to mitigate the risk of AI solutions worsening health inequities. *PLOS Digital Health, 3*(5), e0000390.

References Cont.

Li, F., Wu, P., Ong, H. H., Peterson, J. F., Wei, W. Q., & Zhao, J. (2023). Evaluating and mitigating bias in machine learning models for cardiovascular disease prediction. Journal of Biomedical Informatics, 138, 104294.

Masood, A. (2023). Responsible AI in the enterprise: Practical AI risk management for explainable, auditable, and safe models with hyperscalers and Azure OpenAI. Packt Publishing, Limited.

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. Nature, 577(7788), 89-94.

Molnar, C. (n.d.). Interpretable machine learning: A guide for making black box models explainable. Retrieved April 16, 2025, from https://christophm.github.io/interpretable-ml-book/

Peter, N., & Intelligence, R. S. A. (2021). A modern approach. Pearson Education, USA.

scikit-learn developers. (n.d.). Permutation feature importance. Retrieved April 16, 2025, from https://scikit-learn.org/stable/modules/permutation_importance.html

Weber, T. (2023). Rooting out AI's biases. Hopkins Bloomberg Public Health Magazine. https://magazine.publichealth.jhu.edu/2023/rooting-out-ais-biases