SDP Annual Conference 2024

BRING ARTIFICIAL INTELLIGENCE INTO YOUR DECISION PROCESS WITH EXPLAINABLE AI

Patrick Elder, Director of the Data & AI Center of Excellence

WHAT WE WILL COVER

INTRODUCTIONS

ECS and Patrick Elder

ECS is a large systems integrator for the federal government. We serve Civilian, Defense, and Intelligence Community customers in wide variety of contracts. Our primary areas of expertise are Data & AI, Cybersecurity, Cloud, Application Services, and Management & Modernization.

Patrick Elder is the Director of the Data & AI Center of Excellence at ECS, with over 19 years of experience as a federal consultant working with data solutions. In this role, Patrick has led numerous community building initiatives, including hackathons, that have produced innovative concepts that ECS, and its parent company ASGN have funded as internal improvement projects focused on more efficient and impactful use of its data.
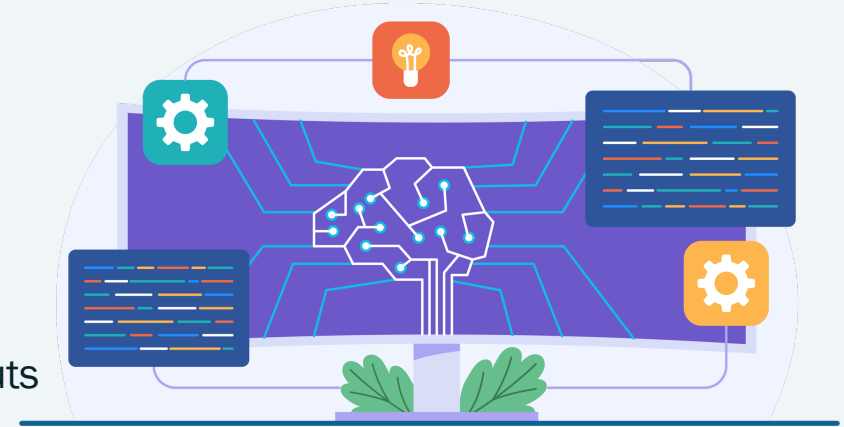
WHAT IS EXPLAINABLE AI (XAI)?

A QUICK INTRODUCTION TO AI

- ## What is AI?
  - AI refers to the simulation of human intelligence processes by machines
  - Examples range from simple to very complex
  - AI systems can analyze complex data inputs and give easily consumable outputs

- ## How does AI differ from traditional programming?
  - Traditional programming involves explicit instructions to achieve a specific task.
  - AI models learn from training data without being explicitly programmed.
  - AI models adapt and evolve based on the data used to train them
  - Traditional programming focuses on **deterministic** outcomes
  - AI models are generally focused on **probabilistic** outcomes

UNDERSTANDING RESULTS OF AI

- Why do we need to understand the results of AI?
  - Responsible, Ethical, and Appropriate Use
  - Use case can dictate how much we need to understand about AI
- Transparency vs. Opacity in AI Systems
  - Transparency: The ability to understand and interpret the inner workings of AI models.
  - Opacity: The lack of visibility into how AI models arrive at their decisions.
- Importance of Interpretability in Complex Models
  - Interpretability enables stakeholders to trust and validate AI outputs.
  - Interpretable models facilitate collaboration between data scientists and decision makers.
- Implications for Decision Making and Analysis
  - XAI enhances decision quality by providing insights into AI-driven decisions.
  - Improved interpretability leads to more confident decision-making processes.
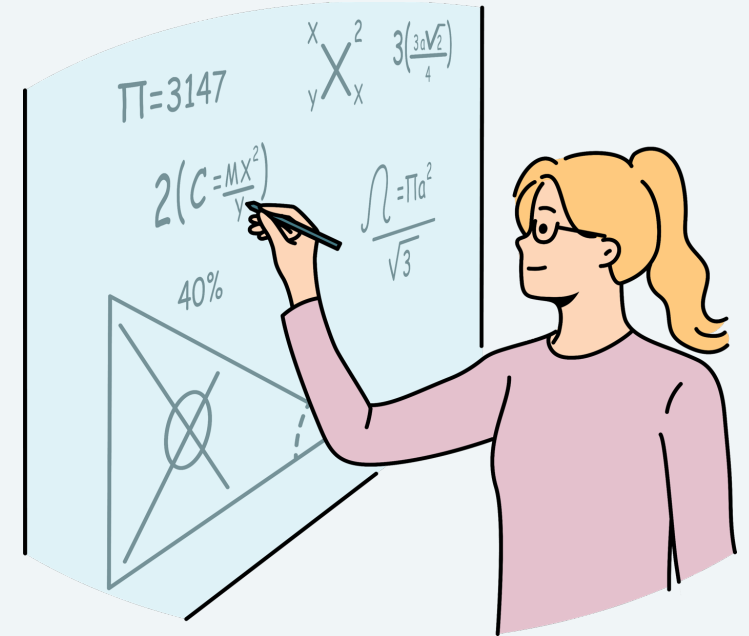
EXPLAINABLE VERSUS INTERPRETABLE

- Interpretable
  - In general, we can understand what inputs factor into the model's output
  - Involves gaining insights into the relationships between input features, the model's internal representations, and its output
  - Useful to analyzing potential bias, feature engineering, and other ethical considerations
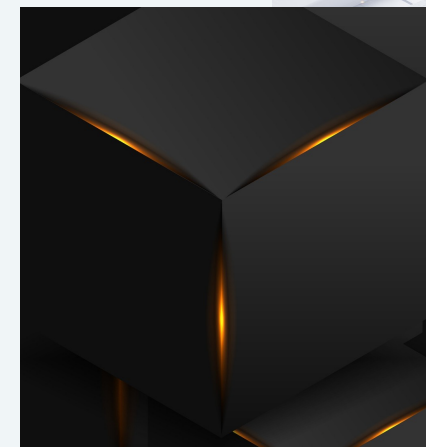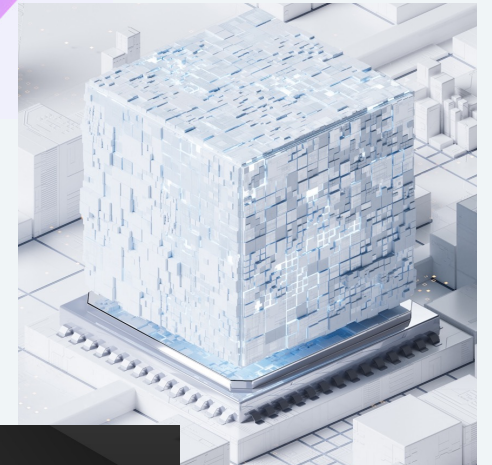
- Explainable
  - For an individual output of the model, we can identify specific inputs that were used
  - In some cases, how much specific inputs affected the model's output
  - Important for auditable tasks

TYPES OF MODELS, EXPLAINED

- Glass Box
  - Most transparent
  - Often are programmed or trained with instructions and decision points
  - Easily interpretable and explainable

- White Box
  - Somewhat transparent
  - Include well known methods such as linear regression, Naïve Bayes Classifiers, and Support Vector Machines
  - Easily interpretable, but present more challenges for explainability

- Black Box
  - Least transparent
  - Large Language Models (LLMs), diffusion models, and recommendation engines
  - Difficult if not impossible to achieve interpretability and explainability

METHODS TO ACHIEVE XAI

- Feature Visualization

- Rule Based Explanations

- LIME and SHAP

- Layer-wise Relevance Propagation (LRP)

- Contrastive Explanations

CASE STUDY IN EXPLAINABLE AI

In a research paper published on ScienceDirect, "Explainable AI and Interpretable Machine Learning: A Case Study in Perspective", the researchers used a medical dataset to predict a patient's probability of being diabetic or non-diabetic.

- Using a black box model, called a Random Forest classifier, to make the prediction
- While it is easy to measure how well the model performs, it is hard to explain its results
- Using LIME and SHAP, the researchers gained both interpretability and explainability
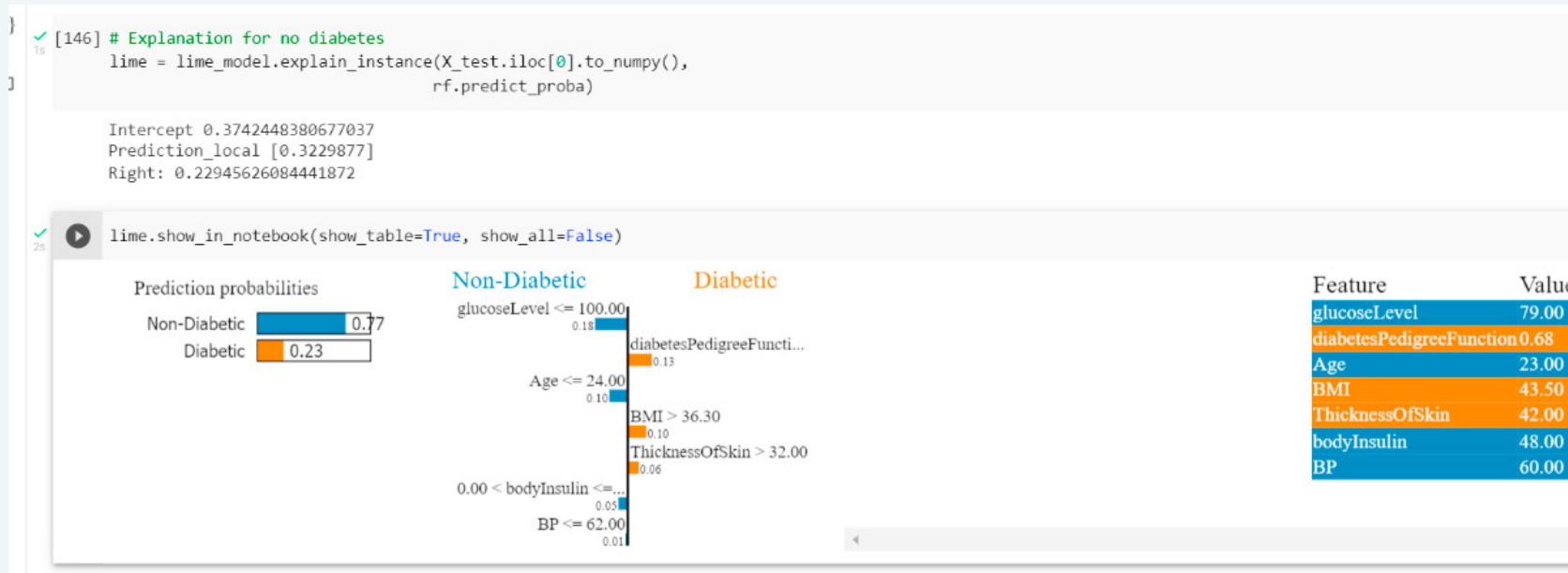
Source: https://www.sciencedirect.com/science/article/pii/S1877050922008432

INTEGRATING AI INTO DECISION PROCESSES – WHY XAI MATTERS

WHY WE SHOULD INCLUDE AI IN DECISION PROCESSES

- Empowering Decision Makers and Analysts
  - Rapid advancements in AI and machine learning have revolutionized decision-making processes across industries.
  - AI technologies offer unprecedented capabilities for data analysis, prediction, and automation.
  - Decision makers are increasingly relying on AI models to inform critical decisions.
- Dilemma Faced by Decision Makers and Analysts
  - Decision makers encounter challenges due to the opacity of AI models.
  - Lack of understanding and interpretability of AI outputs can hinder trust.
  - Decision analysts struggle to justify and explain decisions based on AI algorithms.
- The Promise of Explainable AI (XAI)
  - Explainable AI (XAI) aims to provide transparency and interpretability in AI.
  - XAI bridges the gap between AI models and end-users, empowering decision makers.
  - With XAI, decision makers can understand and trust AI-driven decisions, improving overall effectiveness.

PRACTICAL APPLICATION OF AI IS HARDER THAN IT SEEMS

Multiple studies have shown that 70% of AI project fail to deliver value, with common challenges including:

- Technological
  - The right data is not available, large enough, or of high enough quality to deliver useful AI models
  - Lack of well-defined data dictionary leads to foundational misunderstanding of data
  - Costs and timelines are easily underestimated
- Organizational
  - Resistance to change within the organization to support or make use of AI
  - Business problem definition lacks detail needed once project starts
  - Construction of the team is too distributed across the organization, especially data expertise
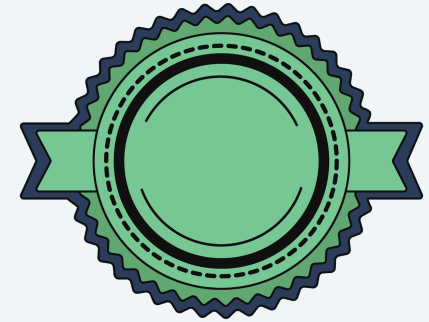- Cultural
  - Lack of Trust in AI
  - Fear of Unknown Bias in AI
  - Decision makers continued use of "private knowledge"

ECS

WHY XAI MATTERS– ADDRESSING THE CHALLENGES

- Enhancing Decision Quality
  - XAI helps identify biases and errors in AI models, leading to more accurate decisions.
  - Transparent AI outputs enable decision makers to make informed judgments based on evidence.

- Empowering Decision Makers and Analysts
  - XAI empowers decision makers to understand the rationale behind AI-driven recommendations.
  - Decision analysts can effectively communicate AI outputs to stakeholders, fostering trust and collaboration.

- Ensuring Traceability and Accountability
  - XAI promotes accountability by providing a clear audit trail of decision-making processes.
  - Transparent AI models enable stakeholders to trace decisions back to their underlying factors.

ECS

KEY CONSIDERATIONS

- Ensuring Traceability and Accountability

- Data Quality and Bias

- Interpretability and Transparency

- Human Oversight and Collaboration

- Robustness and Reliability

- Ethical and Legal Compliance

- Risk Management and Contingency Planning

- User Education and Training

- Scalability and Integration

- Cost-Benefit Analysis

- Continuous Improvement and Innovation
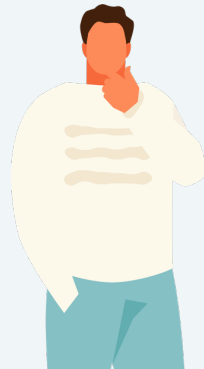
UNDERSTANDING LIMITATIONS AND RISKS

LIMITATIONS OF XAI

- Trade-off Between Interpretability and Performance

- Evolutionary Nature of AI Systems

- Limited Coverage of Explanations

- Interpretation Bias and Subjectivity

- Lack of Standardization and Consistency

- Challenges in Human Interpretation

RISKS WHEN USING AI DECISION PROCESSES

- Algorithmic Bias and Discrimination
  - Models are only as good as the data they are trained on
  - Existing biases can be perpetuated by models
- Unintended Consequences and Error Propagation
  - Best intentions do not guarantee best results
  - Models chained together in application can propagate errors
- Data Privacy and Security Risks
  - Data used for AI model development can sensitive
  - Adversarial Attacks and Manipulation
  - System Vulnerabilities and Cybersecurity Threats
- Overreliance on Automation
  - Gradual degradation of expertise needed for "human in the loop"
  - Loss of Human Expertise and Decision Diversity
- Ethical Dilemmas and Accountability
  - Can AI represent the ethics of the organization that uses it?
  - Regulatory and Compliance Risks

CONCLUSION

CONCLUSION

- The Importance of Explainable AI
  - XAI is an important tool that can promote transparency and interpretability in AI systems.
  - Leveraging XAI enables decision makers and analysts to take advantage of AI with reduced risk.

- Responsible AI Adoption in Decision Processes
  - Ethical considerations and responsible practices in AI deployment.
  - Prioritize transparency and accountability in AI initiatives.

- Further Exploration and Implementation
  - Continue exploring XAI techniques and integrating them into their decision-making processes.
  - Use XAI to drive positive change and innovation across industries.

ECS