ACCOUNTABILITY IN AI DECISION-MAKING

Artificial Intelligence System Engineering

INTRODUCTION

- AI accountability refers to the idea that artificial intelligence should be developed, deployed, and utilized such that responsibility for bad outcomes can be assigned to liable parties.
- XAI – Explainable AI
- Improtance:
  - Trust in AI systems
  - Compliance with laws and ethical norms
  - Mitigation of bias and harm

EXPLAINABLE AI AND INTERPRETABILITY

- Trained models consist of hundreds or thousands of formulas and parameters
- XAI tries to help with:

  The model

  Justifications and documentations of model behavior for ethical and regulatory needs.

  Debugging and optimization

LOCAL EXPLAINABILITY

- Understand the parameters which influence a single prediction
- Create and analyze a simple and explainable model
- Steps:

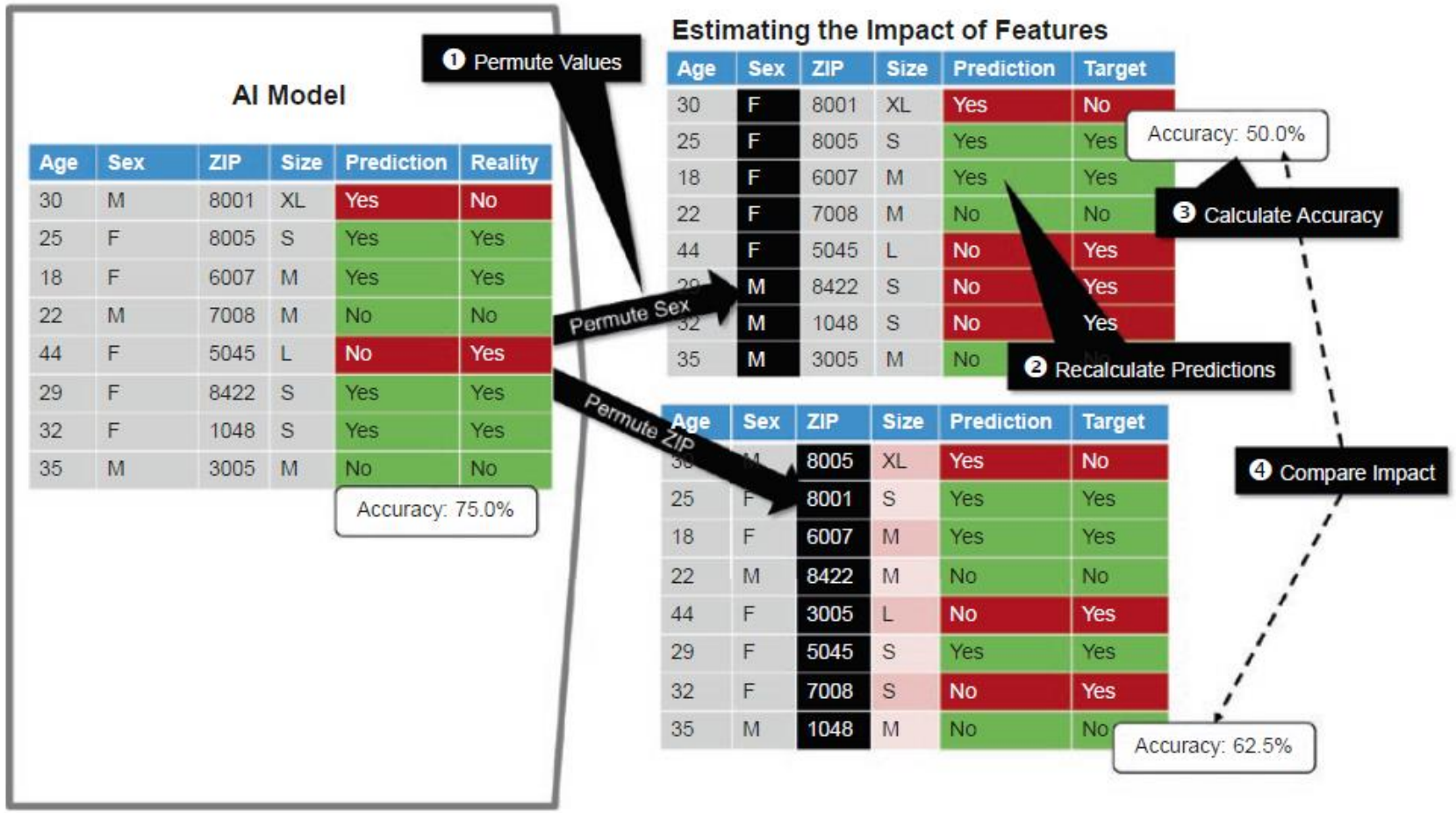
  Probing

  Constructing an explainable model

  Understanding and explaining

Important: It is essential not to calculate just the gradient but to probe around the points of interest

GLOBAL EXPLAINABILITY



- Global explainability does not look at single predictions
- Permutation importance determines the impact of the various input features on a given model's predictions.

Klaus Haller. Managing AI in the Enterprise

ACCOUNTABILITY

- Often defined too imprecisely
- High-Level Expert Group (HLEG) reports, the GDPR and the Artificial Intelligence Act (AIA)
- Accountability can be understood as answerability, requiring three key conditions:

  Authority recognition

  Interrogation

  Limitation of power

NOVELLI, C., TADDEO, M. & FLORIDI, L. ACCOUNTABILITY IN ARTIFICIAL INTELLIGENCE: WHAT IT IS AND HOW IT WORKS. AI & SOC 39, 1871–1882 (2024). HTTPS://DOI.ORG/10.1007/S00146-023-01635-Y

Features	Explanations
1. Context (what for?)	Fields in which an accountability relation is established
2. Range (about what?)	Tasks, like actions, services, decisions, and assessments taken by the accountable agent
3. Agent (who?)	The entity who exercises the delegated powers, accepting to be blamed or praised
4. Forum (to whom?)	The entity engaged in actual interrogation and supervision and/or the bearer of the interests served through delegation of tasks (*principal*)
5. Standard (according to what?)	Principles, rules, and benchmarks against which the conduct of the accountable agent is assessed
6. Process (how?)	Procedures through which the agent is called to account
7. Implications (what follows?)	Consequences, formal or informal, triggered by the accountability assessment

GOALS OF ACCOUNTABILITY

ACCOUNTABILITY IN AI

- Context - field of use, function, the level of autonomy of the AIs in question
- Range - design, development or deployment
- Agents – all the steps in the range are performed by someone. They can be individuals, corporations, etc.
- Standards – all tasks in the range must be assessed
- Process - rules, metrics and procedures. Performed by developers of AI, third parties, ...
- Implications – consequences of AI

PRACTICE TASK

- Pabandyti pagal straipsnyje siūlomą struktūrą, įvertinti savo MBD.

INTELECTUAL PROPERTY IN AI

INTELECTUAL PROPERTY IN AI

- The global artificial intelligence market size is expected to grow twentyfold between 2021 and 2030 (Statista 2023)
- Emergence of user-friendly ‘generative AI’ tools as ChatGPT, Midjourney, Speechify, Synthesia, and Amper AI
- Which components of AI require IP laws? WIPO (2023) distinguishes four definitions of AI inventions.
- WIPO - World Intellectual Property Organization

DIFFERENT AI INVENTION CONCEPTS

- AI models or algorithms. Inventions on core AI technology itself
- AI-based inventions / creations
- AI-generated inventions / creations
- AI-assisted inventions / creations

AI ALGORITHMS AND MODELS

- In many jurisdictions, abstract ideas, including mathematical algorithms, are not patentable
- Technical point of view
- Thoroughly describe the AI model, including its architecture, training data, and specific applications, to meet the enablement and written description requirements

AI-GENERATED OUTPUTS

- US – no human involvment, no patent
- Human crated elements may qualify for copyright protection
- Inventors must be human
- Use of existing works in AI training – each situation is different and still there is no single consensus
- Generative AI Copyright Disclosure Act

TRAINING DATA

- Similar issues as in previous slide

AI SOFTWARE

- The code that implements AI systems is protected under software copyright laws
- Innovative AI algorithms and models may be eligible for patent protection if they meet criteria such as novelty, non-obviousness, and utility

AI HARDWARE

- Hardware innovations designed to optimize AI operations, such as specialized processors (e.g., GPUs or TPUs), can be patented

AI-BASED PRODUCTS OR SERVICES

- Novelty, non-obviousness, and utility

AI DEVELOPMENT TOOLS AND PLATFORMS

- Proprietary tools and platforms used to develop AI systems are protected under copyright law.

AI-SUPPORTED BUSINESS PROCESSES

- Proprietary tools and platforms used to develop AI systems are protected under copyright law.

LEGAL DOCUMENTS

- US Patent and Trademark Office (USPTO)
- European Patent Office (EPO)