Microsoft

Proudly supported by SAP

edunet foundation

Lab Manual

Responsible AI in Practice: Case Studies

Case Study 1: Amazon's AI Recruitment Tool

Theme: AI Bias and Fairness in Hiring

Objective

To understand how machine learning systems can unintentionally reinforce historical discrimination in recruitment processes, and the importance of ethical design in HR-related AI tools.

Prerequisites

- Understanding of supervised learning
- Familiarity with classification models

Problem Statement

In 2014, Amazon developed an experimental AI-powered recruitment tool intended to automate the resume screening process. The system was trained on 10 years of past hiring data, which reflected a male-dominated tech industry. The AI model began penalizing resumes that included the word "women," such as "women's chess club," and favored terms more commonly associated with male candidates.

Solution

After internal audits, Amazon identified that the model was not gender-neutral and reinforced sexist biases. The project was quietly scrapped in 2018. The incident highlighted the need for:

- Diverse datasets
- Pre-launch bias audits
- Inclusion of ethical oversight in model design
- Human-AI hybrid decision-making for fairness

Questions

1. Could anonymizing resumes have reduced bias in this AI system? Why or why not?
2. What protocols should be mandated before deploying AI in sensitive domains like hiring?
3. How can HR departments balance efficiency with fairness when using AI?
4. Should governments enforce AI bias testing in all employment-related AI tools?

Case Study 2: COMPAS Algorithm and U.S. Criminal Justice

Theme: Historical Bias in Predictive Policing

Objective

To explore how biased training data in judicial AI tools can perpetuate social inequalities and examine accountability and transparency in high-stakes decision-making.

Prerequisites

- Understanding of classification algorithms
- Risk scoring systems in criminal justice
- Basics of systemic bias and racial discrimination

Problem Statement

COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is an AI tool used to assess a defendant's risk of reoffending. A 2016 investigation by ProPublica revealed that COMPAS consistently gave Black defendants higher risk scores than white defendants with similar histories. This resulted from historical bias embedded in policing and court data.

Solution

The study led to widespread criticism and prompted efforts to:

- Review and correct biased training data
- Introduce “right to explanation” in sentencing algorithms
- Limit the use of AI tools unless thoroughly audited for bias
- Increase human oversight in AI-assisted decisions

Questions

1. What safeguards can prevent historical injustices from entering AI tools?
2. How can transparency and explainability protect individual rights in legal systems?
3. Is it ethical to use risk prediction tools in courts if biases persist?
4. How should AI accountability be assigned when an unfair decision affects a defendant’s future?

Case Study 3: Midjourney's Cultural Stereotyping

Theme: Generative AI and Cultural Representation

Objective

To analyze how generative AI systems reflect and reinforce cultural stereotypes and explore the need for inclusive training datasets.

Equipment Required

- Midjourney or similar image-generation platform
- Prompt datasets by region/culture

Prerequisites

- Understanding of generative AI models
- Knowledge of prompt engineering and diffusion models
- Awareness of cultural representation and media bias

Problem Statement

A study by Rest of World in 2023 analyzed 3,000 AI-generated images across 6 countries. For prompts like "a woman" or "a house," Midjourney often produced stereotyped images—for example, an "Indian person" was always shown as an elderly man in a turban, ignoring women or youth. Similarly, Mexican individuals were portrayed only in sombreros, and Nigerian food or clothing was shown in clichéd visuals.

Solution

The findings exposed how AI can distort real-world cultural diversity when trained on narrow or biased datasets. The study prompted discussions about:

- Dataset diversification
- Human review of generative outputs
- Culturally sensitive prompt design
- Regulation around AI-generated media representations

Questions

1. Should global generative AI platforms be required to localize their datasets?
2. How can AI reflect diversity without amplifying stereotypes?
3. Who is responsible for the outputs of generative AI—developers or users?
4. Can AI ever truly "understand" culture, or is it limited to patterns?

Case Study 4: GDPR's Right to Explanation

Theme: Explainability and Data Privacy

Objective

To examine how data privacy laws like the EU's GDPR enforce transparency and explainability in AI-powered decision systems.

Equipment Required

- GDPR policy documents
- AI model for credit approval (black-box and explainable versions)

Prerequisites

- Basics of EU data protection laws
- Model interpretability concepts
- Regulatory and compliance frameworks for AI

Problem Statement

AI systems increasingly make impactful decisions—loan approvals, job selection, insurance pricing—but users often don't know how or why decisions were made. GDPR addresses this through the "Right to Explanation," enabling individuals to understand how automated decisions affect them.

Solution

Companies must now:

- Provide model explanations
- Collect informed user consent
- Log data usage for audit trails
- Modify model architecture to ensure interpretability
  This ensures fairness and improves consumer trust.

Questions

1. Should "right to explanation" apply to all AI systems, regardless of risk level?
2. Can technical explanations be simplified for non-expert users?
3. How can regulators audit XAI implementations for compliance?
4. What happens if the explanation itself introduces bias?

Case Study 5: California Consumer Privacy Act (CCPA)

Theme: Consumer Data Rights and AI Governance

Objective

To explore how regional laws like the CCPA reshape the way organizations build AI systems that rely on consumer data.

Equipment Required

- CCPA legal text
- Customer data pipeline in an AI application (e.g., recommendation engine)

Prerequisites

- Knowledge of consumer data lifecycle
- Understanding of targeted advertising and personalization

Problem Statement

Before the CCPA, users had little control over how their data was collected, stored, or sold by AI-powered platforms. Companies used browsing behavior, purchases, and preferences to target users without transparent consent.

Solution

Under CCPA:

- Consumers can know, delete, or restrict use of their personal data
- Businesses must offer opt-out mechanisms
- Companies are incentivized to adopt data minimization and transparency
  This forces AI developers to rethink data pipelines and retrain models with compliant practices.

Higher-Order Thinking Questions

1. How can AI systems be redesigned to respect “privacy by default”?
2. Should AI explain its data usage and recommendation logic to users?
3. How do privacy laws influence data-driven business models?
4. Can stricter privacy lead to more ethical AI or hinder innovation?