AI: Facilitator of Change or Threat to Human Rights?

Ali Nikkhah
University of California, Irvine

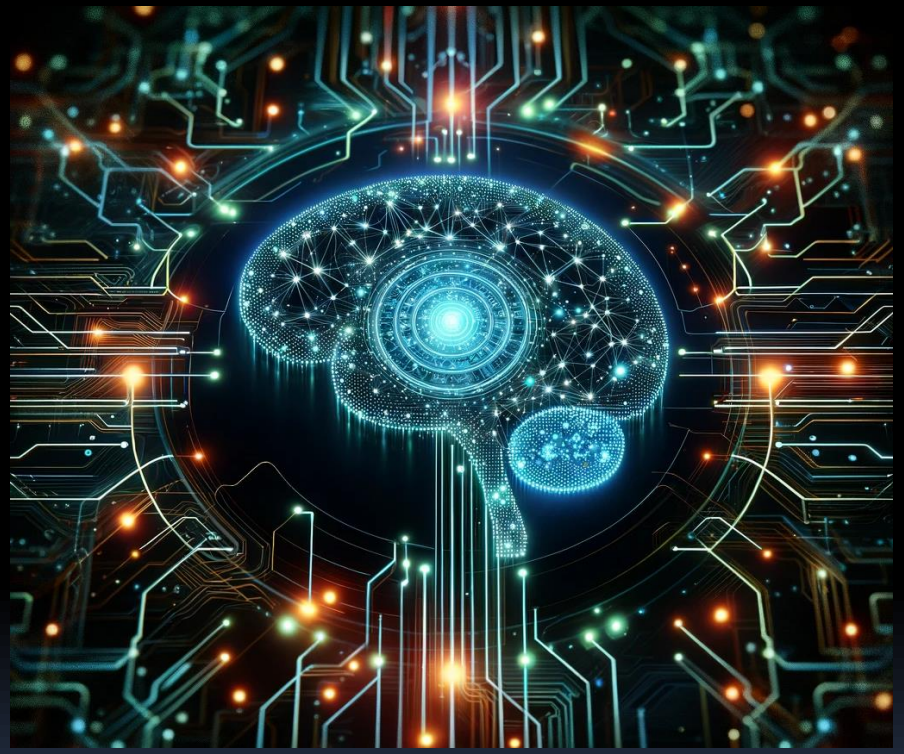
PTC'24
January 2024

Agenda

- Introduction and Background
- AI and Human Rights: Challenges and Impacts
- Navigating AI Challenges: Ethical Governance
- Building a Future with AI: Strategies and Vision

Introduction to AI

- AI's capabilities: Thinking, Learning, Problem-Solving
- The paradigm shift in human-technology interaction
- AI as a present reality and future shaper



DALL·E 3

The UN’s Global Digital Compact

- Goals: Safety, Empowerment, and Inclusion in the Digital Age
- Key Aspects: Digital Inclusion, Human Rights, Data Privacy, Ethical Technology
- Navigating the Future Digital Landscape

AI's Role in the Global Digital Era

- AI as a Driving Force in Digital Transformation
- Capabilities: Data Analytics, Automation, Decision-Making
- Impact on Industries, Governance, and Societal Norms
- Managing Data for a More Interconnected World

Agenda

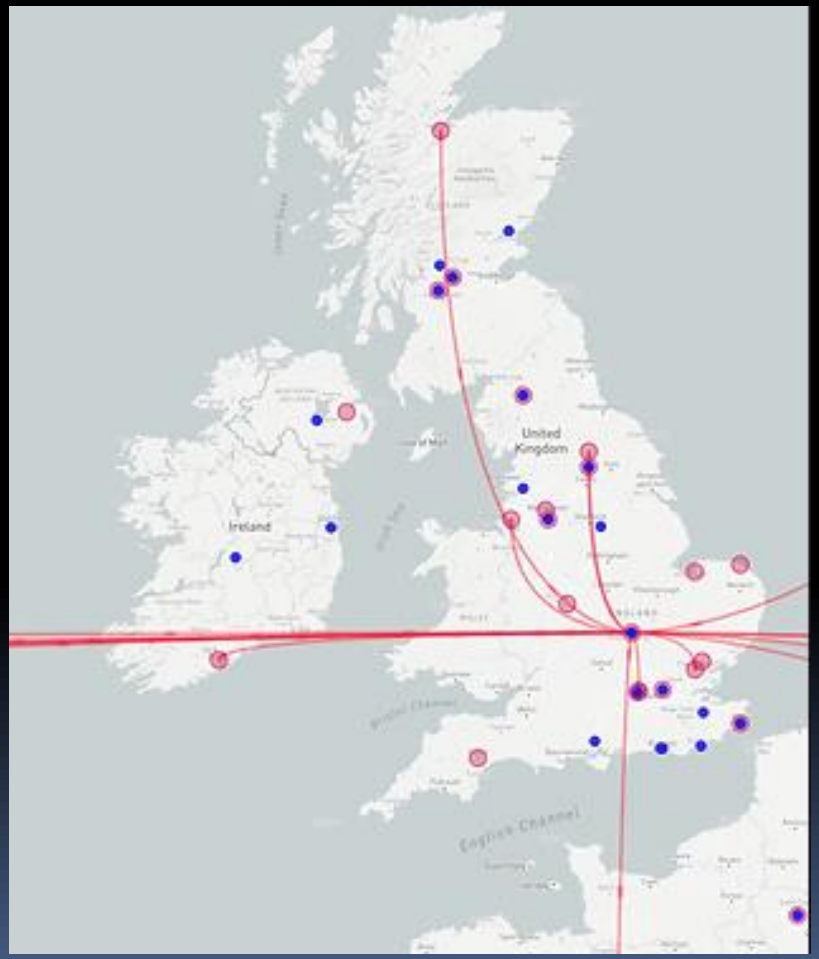
- Introduction and Background
- AI and Human Rights: Challenges and Impacts
- Navigating AI Challenges: Ethical Governance
- Building a Future with AI: Strategies and Vision

AI and Human Rights

- The Duality of AI in Human Rights
- AI as a Tool for Monitoring and Preventing Abuses
- Risks: Privacy Invasion, Bias in Decision-Making
- Aligning AI with Human Rights Principles

Case Study: AI's Positive Impact

- Case Study 1 - Trafficking and Exploitation Detection:
  - Example: Traffik Analysis Hub by IBM
  - Task: AI identifying patterns related to human trafficking and labor exploitation
  - Details:
    - Analysis of reports, online ads, financial transactions, and digital trails.
    - Heatmap
    - Track Journey with Routemaps
  - Impact:
    - Uncovering networks of trafficking and exploitation
    - Aiding law enforcement and NGOs.
    - Protecting victims' rights and dismantling trafficking networks.



Traffik Analysis Hub

Case Study: AI's Positive Impact

- Case Study 2 - Conflict Zone Monitoring:
  - Example: Monitor Violence in Myanmar By Human Rights Watch
  - Task: AI-powered satellite imagery analysis in conflict zones
  - Details:
    - Use AI to combine the thermal data with aerial images
    - Detection of destroyed buildings, troop movements, and refugee flows.
  - Impact:
    - Aid for human rights organizations in intervention strategies
    - Faster response times in high-risk areas.
    - More accurate assessments of human rights conditions.

blogs.nvidia

Case Study: AI's Positive Impact

- Case Study 3 - Online Hate Speech Detection:
  - Example: Rule By Example (RBE) by University of Michigan & Microsoft
  - Task: AI-enhanced detection of online hate speech
  - Details:
    - Combines deep learning and rule-based methods.
    - Improves transparency in automated content moderation.
  - Impact:
    - More effective identification of hate speech.
    - Balances free speech with protection from online harm.

Risks: AI's Negative Impact

- Case Study 1 - Bias and Discrimination:
  - Example: Amazon's AI Recruiting Tool
  - Task: Automated resume screening for job recruitment
  - Details:
    - The AI system was trained on historical data to identify top candidates.
    - It systematically biased against women
    - The bias stemmed from the historical data that reflected male dominance in the tech industry.
  - Impact:
    - Highlighted the risks of unconscious bias in AI, leading to broader awareness and discussion about ethical AI development.
    - Emphasized the need for diverse and unbiased training data in AI systems.

BENEDETTO CRISTOFANI/@SALZMANART

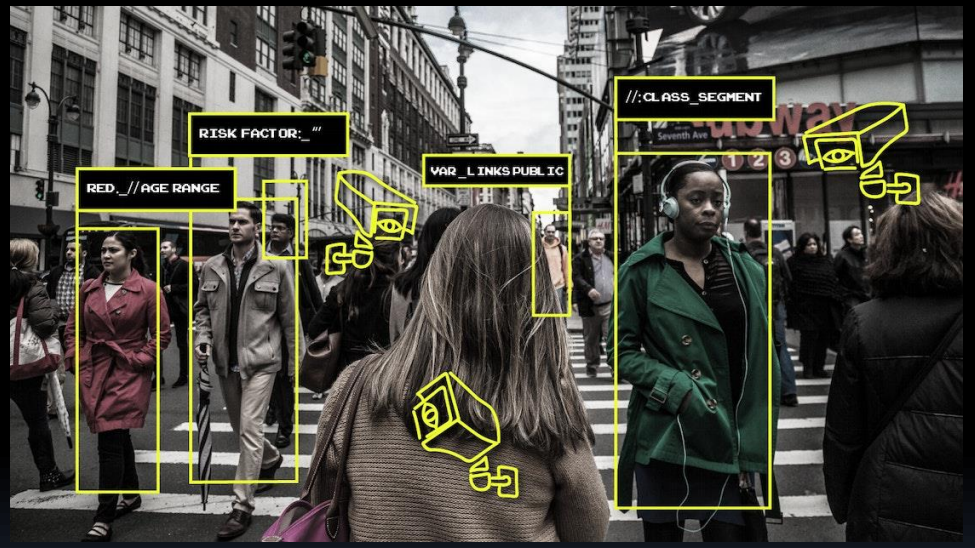
Risks: AI's Negative Impact

- Case Study 2- AI in Healthcare:
  - Example: IBM's 'AI Doctor' Project
  - Task: Developing an AI system for diagnosis and prescription
  - Details:
    - The AI was trained on a vast amount of medical data and aimed to assist healthcare professionals.
    - Despite high expectations, the AI system consistently failed to deliver accurate and reliable results.
  - Impact:
    - AI system trained predominantly on data from one racial group might be less accurate for other racial groups
    - Sensitive personal health information being breached or misused

DALL·E 3

Risks: AI's Negative Impact

- Case Study 1 - Privacy Invasion:
  - Example: Use of Facial Recognition for Surveillance
  - Task: Implementing facial recognition for public safety and security by Governments
  - Details:
    - Despite its intended use for security, the technology has raised significant privacy and ethical concerns
  - Impact:
    - There have been instances where facial recognition has led to misidentification, particularly among certain racial and ethnic groups, raising concerns about racial bias and discrimination.
    - The use of facial recognition has sparked debates about privacy invasion



Liberties.eu

Agenda

- Introduction and Background
- AI and Human Rights: Challenges and Impacts
- Navigating AI Challenges: Ethical Governance
- Building a Future with AI: Strategies and Vision

AI Governance Complexities

- Regulatory Challenges: Addressing the lag between AI advancements and policy development.
- Stakeholder Involvement: Importance of including diverse perspectives in governance.
- Balancing Act: Innovation vs. ethical and societal considerations.
- Global Standards: The need for international norms and guidelines.

AI and Ethical Governance

- Importance of Transparency: Ensuring AI decisions are explainable and understandable.
- Fairness and Non-Discrimination: Preventing AI from perpetuating biases or inequalities.
- Accountability and Responsibility: Establishing clear guidelines for AI developers and users.
- Privacy Protection: Safeguarding personal and sensitive data in AI operations.
- Human-Centric Approach: Keeping human welfare at the core of AI development.

Agenda

- Introduction and Background
- AI and Human Rights: Challenges and Impacts
- Navigating AI Challenges: Ethical Governance
- Building a Future with AI: Strategies and Vision

Global and Ethical Frameworks

- International Collaboration: Harmonizing AI standards and practices across borders.
- Universal Ethical Standards: Developing consensus on core ethical principles for AI.
- Data Governance Protocols: Ensuring responsible and secure handling of data worldwide.
- Sustainable AI Development: Aligning AI with global sustainability goals.
- Regulatory Alignment: Bridging the gap between various national regulatory frameworks.

Engaging Stakeholders in AI

- Involving Diverse Stakeholders in AI Discussions
- Contributions from Policymakers, Businesses, Academia, Public
- Enriching AI Discourse with Diverse Perspectives
- Ensuring Inclusive and Considerate AI Development

Vision for AI's Future

- AI as the Biggest Opportunity
- AI as the Biggest Threat
- Balancing Speed with Responsibility
- Raising Awareness Among Policymakers and Public

Conclusion

- AI's Dual Nature
- Ethical Governance
- Global Collaboration
- Stakeholder Engagement
- Raising Awareness


THANK YOU