Artificial Intelligence System Engineering

Ethics, its relevance and relations to AI. Ethical Theories And Frameworks. International, national, and industry-specific AI regulations

Responsible AI

- Responsible AI is the practice of designing, developing, and deploying AI systems with good intention and sufficient awareness to empower users, to engender trust, and to ensure fair and positive impact to society. It consists of areas like fairness, privacy, transparency, and accountability. (Designing Machine Learning systems)
- Responsible AI is a broad and complex topic that usually requires quite a lot of time to understand
- Discussions about ethical IT systems (including AI) have been discussed for a long time, but now the rules are being applied.

Risks

Bias and Discrimination:	• AI systems can inherit biases from data or developers.
Transparency:	• AI decisions often lack explainability (black-box models).
Privacy and Surveillance:	• Data collection raises concerns about user privacy.
Accountability:	• Who is responsible for AI's decisions and their consequences?
Autonomy:	• Balancing human decision-making with AI recommendations.

Risks

Unethical use cases

- Discrimination due to data bias
- Violations of privacy
- False information and manipulation

Unethical engineering practices

- Deliberately ignoring bias
- Inadequate testing and verification practices
- Irresponsible data management
- Unethical work organization

Unethical AI models

How to handle dilemmas?

Top-down approach - ethical guidelines guide through critical situations

- requires close collaboration of data scientists or engineers with ethics specialists

Bottom-up approach - AI system learns by observation.

- Rare and might take years for such situation to happen
- Might learn not the best practices

Ethical AI Models

Bias

What is the training data and it's distribution? Does it reflect reality?

Fairness - trained AI models behave similarly well for all relevant subgroups, even for small ones

Is accuracy similar in all groups?

Ethical Theories

- **Deontology**:
  - Focus on rules and duties.
- **Utilitarianism**:
  - Aim to maximize overall happiness and minimize harm.
- **Virtue Ethics**:
  - Emphasis on the moral character of the user.
- **Justice and Fairness**:
  - Focus on equity and avoiding discrimination.

Ethical Frameworks for AI

- **Asilomar AI Principles**
  - AI should be beneficial, transparent, and accountable
  - https://futureoflife.org/open-letter/ai-principles/
- **EU Guidelines for Trustworthy AI**:
  - Seven principles: Accountability, transparency, diversity, fairness, etc.
  - https://digital-strategy.ec.europa.eu/en/policies/artificial-intelligence
- **UNESCO AI Ethics Recommendation**:
  - A global framework for ethical AI implementation
  - https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence
- **IEEE Ethics in Autonomous and Intelligent Systems**:
  - Standards for ethical AI development
  - https://standards.ieee.org/industry-connections/activities/ieee-global-initiative/

International AI Regulations

- **European Union**:
  - **AI Act**: Proposes a risk-based approach to AI regulation (prohibited, high-risk, and low-risk applications).
  - **GDPR**: Data privacy regulation with implications for AI.
- **United Nations**:
  - Focus on global AI ethics and sustainable development goals (SDGs).
- **OECD AI Principles**:
  - Recommendations for trustworthy AI: Human rights, fairness, transparency.

Industry-Specific AI Regulations

- **Healthcare**:
  - **HIPAA (USA)**: Protects health data used in AI.
  - **MDR (EU)**: Medical Device Regulation for AI in healthcare.
- **Finance**:
  - AI must comply with anti-discrimination laws (e.g., Fair Lending Regulations in the US).
- **Autonomous Vehicles**:
  - Compliance with safety standards (e.g., ISO 26262 for functional safety in automotive systems).
- **Technology**:
  - Big Tech self-regulations (e.g., Google's AI Principles).

Summary

- Adopt transparency and explainability practices.
- Conduct bias audits and diversity testing.
- Establish accountability for AI decisions.
- Engage diverse stakeholders in AI design.
- Create policies for data privacy and security.

National AI Regulations

- Lithuania
  - Lithuanian Artificial Intelligence Strategy: A Vision of the Future
  - Action Plan for the Development of Artificial Intelligence Technologies in Lithuania for 2023–2026
  - Alignment with EU Regulations
- Poland:
  - Policy for the Development of Artificial Intelligence in Poland from 2020
  - Alignment with EU Regulations

PRACTICE TASKS

Analyzing Industry specific regulations related to Your topic

Responsible AI (Reasearch task)

- NIST Special Publication 1270: Towards a Standard for Identifying and Managing Bias in Artificial Intelligence
- ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) publications
- Trustworthy ML - Resources
- FATE/CV - Schedule

Responsible AI (Research task)

- Analyze Regulations related to AI in these countries:
  - USA
  - China
  - India
  - Taiwan
  - Russia