

**The
Alan Turing
Institute**

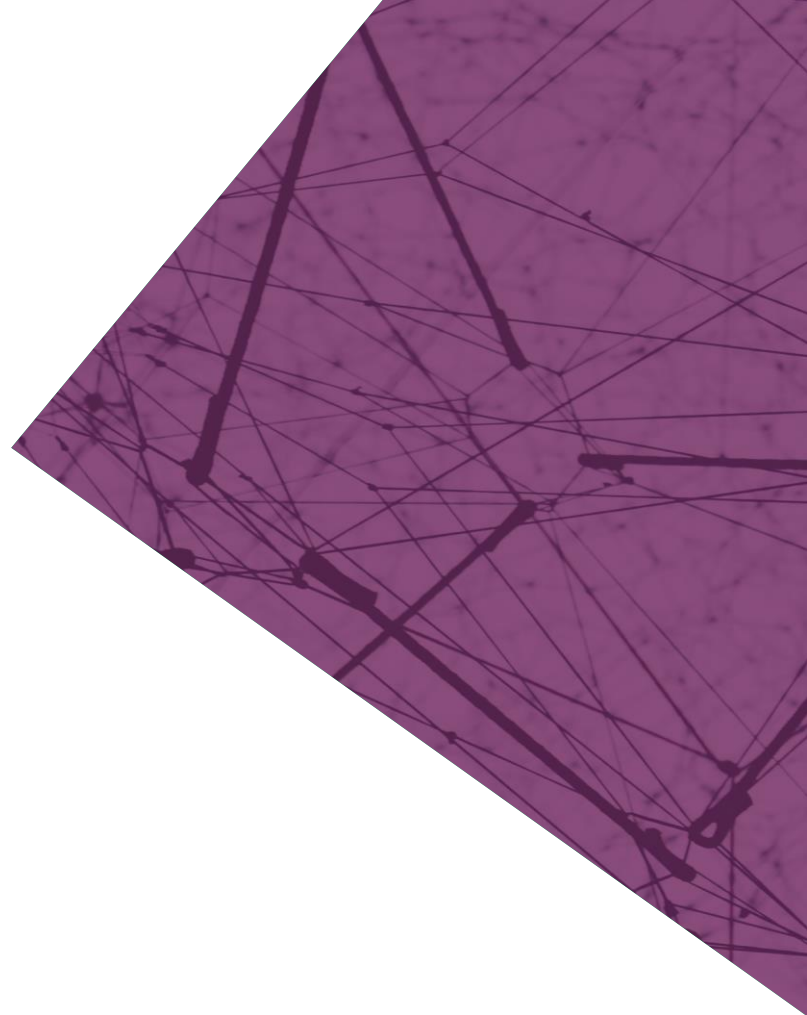
Intro to Trustworthy AI

**Milestone 5 : Trade-offs and Interactions
with other verticals in Trustworthy AI**

Roseline Polle

Postgraduate, UCL

roseline.polle.19@ucl.ac.uk



Course Overview

Part I: Ethical, Legal, and Governance

Milestone 1. Notions of Fairness

Milestone 2. Legal and Governance aspects of AI Fairness

Part II: Technical dimensions in Supervised ML

Milestone 3. Sources, Forms, and Quantification

Milestone 4. Mitigation

Milestone 5. Trade-offs and Interactions with other verticals in Trustworthy AI

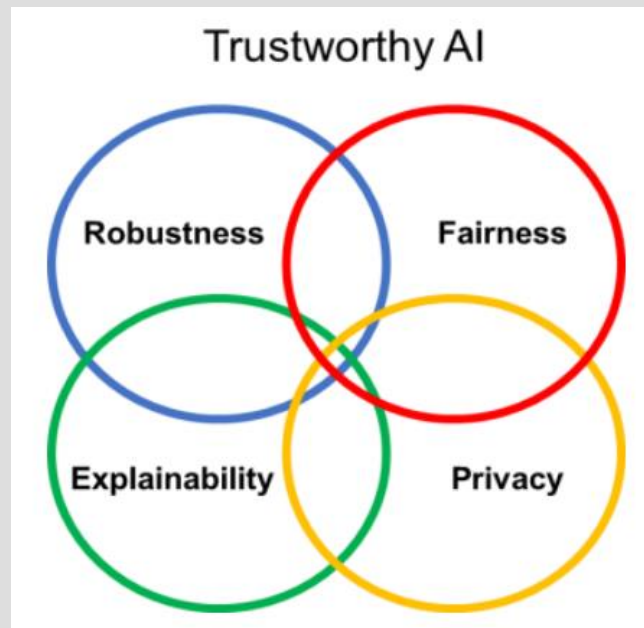
EC requirements for trustworthy AI

Ethics guidelines for trustworthy AI (EC): 7 requirements:

- Human agency and oversight
- Technical Robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

Zoom on four verticals

- **Privacy:** quality of a system to mitigate personal or critical data leakage.
- **Fairness:** quality of a system to avoid unfair treatment of individuals or organizations.
- **Explainability:** quality of a system to provide decisions or suggestions that can be understood by their users and developers.
- **Robustness:** quality of a system to be safe, not vulnerable to tampering



Milestone Overview

- **Part 1: Theory** → Slides + Supporting notebook
 1. Explainability
 2. Privacy
 3. Robustness
- **Part 2: Explainability Practice Notebook**