ARTIFICIAL INTELLIGENCE | BEGINNER | 3 DAYS

Responsible AI & AI Safety Foundations

A practical, low-barrier introduction to Responsible AI, AI ethics, governance, safety principles, and emerging risks without technical prerequisites.

Why Responsible AI?

Reduce product and reputational risk

Prevent biased or unsafe AI behavior before it reaches users.

Enable compliance readiness

Translate NIST and EU AI Act concepts into practical actions.

Improve customer trust

Build transparency and accountability into AI-powered features.

Safer LLM adoption

Recognize hallucinations, harmful content, and overtrust, and apply guardrails.

What you'll be able to do

- ✓ Understand ethical risks and harmful failure modes of AI systems.
- ✓ Identify common sources of bias in data, labeling, and product design.
- ✓ Apply fairness, transparency, explainability, and safety principles in practical scenarios.
- ✓ Explain major governance and regulatory frameworks in simple business language.
- ✓ Recognize unsafe LLM outputs and implement beginner-friendly guardrails.
- ✓ Participate confidently in AI governance and Responsible AI discussions.

Who this is for

- Students and early-career practitioners
- Product Managers and Business Analysts
- HR, Operations, and Policy teams
- Junior Data Scientists
- Anyone new to Responsible AI and AI safety

Curriculum

01 Module 1: Introduction to Responsible AI and Ethical Foundations

- What Responsible AI means in real-world systems
- How AI systems make decisions (non-technical explanation)
- Real-world AI failures and ethical breakdowns
- Sources of risk across data, design, and deployment
- **Hands-on:** Evaluate an AI decision for fairness and bias

02 Module 2: Understanding Bias, Fairness and Inclusivity

- Types of bias in AI systems
- Historical, sampling, labeling, and representation bias
- Fairness concepts explained without mathematics
- Designing inclusive AI products
- **Hands-on:** Identify bias sources in a dataset example

03 Module 3: Safe Use of Large Language Models

- How LLMs generate responses
- Common LLM risks including hallucinations and harmful outputs
- Overtrust and misuse of generative AI
- Introduction to prompt safety and red-teaming
- **Hands-on:** Detect unsafe outputs and rewrite safer prompts

04 Module 4: Transparency, Explainability and Accountability

- What transparency means in AI products
- Explainability vs interpretability with simple examples
- Model documentation basics
- Accountability structures inside organizations
- **Hands-on:** Create a simple Model Card for a fictional AI system

05 Module 5: Global Frameworks and Regulations (Beginner Edition)

- Why AI governance frameworks exist
- NIST AI RMF explained for non-technical teams
- EU AI Act risk tiers simplified
- OECD and emerging regional AI guidelines

06 Module 6: Designing Safe and Responsible AI Products

- Safety-by-design principles
- Non-technical guardrails for AI systems
- Human oversight and escalation thresholds
- Monitoring misuse and harmful interactions
- **Hands-on:** Redesign an AI feature with safety controls

07 Module 7: Capstone – Responsible AI Review

- Ethical risk identification
- Fairness and bias assessment
- Lightweight governance documentation
- Safety improvement recommendations