AI Ethics and Explainability

Duration: 5 Days

Language: en

Course Code: PI2-111

Objective

By the end of this course, participants will be able to:

- Understand key ethical issues in AI development and deployment.
- Identify risks related to bias, discrimination, and privacy.
- Apply frameworks and tools for assessing ethical AI practices.
- Explore methods and tools for achieving explainability in AI models.
- Balance model performance with transparency and interpretability.
- Navigate AI governance, compliance, and regulatory standards.
- Build or evaluate AI systems that are socially responsible and auditable.

Audience

This course is ideal for:

- Data scientists, AI engineers, and machine learning practitioners.
- Product managers and tech leads developing AI-based services.
- Compliance officers and legal professionals in tech organizations.
- Policymakers and regulators overseeing AI governance.
- Academics and researchers in AI, ethics, or digital rights.
- NGO representatives and advocates focused on AI fairness.
- IT consultants implementing AI solutions in sensitive domains.

Training Methodology

This course uses a mix of instructor-led sessions, ethical dilemma discussions, real-world case analysis, and hands-on explainability exercises. Participants will explore AI audit tools, compare black-box vs. interpretable models, and evaluate case studies through ethical frameworks. The course promotes critical thinking, dialogue, and practical implementation.

Summary

As artificial intelligence continues to influence decision-making in healthcare, finance, justice, and beyond, questions around fairness, accountability, and transparency have moved to the forefront of AI development. This course introduces the ethical challenges posed by AI systems and focuses on explainable AI (XAI) as a critical component for trust, compliance, and responsible innovation.

Participants will explore the intersection of ethics, regulation, and technical solutions designed to make AI systems understandable and fair. The course offers a balance of conceptual foundations, real-world case studies, and practical tools to assess and improve the transparency of AI models—especially in high-stakes or regulated environments.

Course Content & Outline

Section 1: Foundations of AI Ethics

- Why AI ethics matters: trust, safety, and social responsibility.
- Key ethical principles: fairness, accountability, transparency, autonomy.
- Common risks: bias, discrimination, opacity, surveillance.
- Real-world case studies: when AI fails or harms.
- Ethical frameworks: consequentialism, deontology, virtue ethics in AI.
- Stakeholder impact and ethical design considerations.
- Role of human oversight and moral reasoning.

Section 2: Understanding Bias and Fairness in AI

- How bias enters AI systems (data, design, deployment).
- Types of bias: historical, algorithmic, selection, measurement.
- Evaluating fairness: individual vs. group fairness metrics.
- Tools for bias detection and mitigation (Fairlearn, AI Fairness 360, etc.).
- Fairness trade-offs: performance vs. equity
- Inclusive dataset design and demographic parity.
- Guidelines for equitable model development.

Section 3: Explainable AI (XAI) Principles and Techniques

- What is XAI and why it matters.
- Black-box vs. white-box models: strengths and limitations.
- Post-hoc explanation methods: SHAP, LIME, counterfactuals.
- Interpretable models: decision trees, rule-based systems, linear models.
- Domain-specific challenges in explainability (e.g., healthcare, finance).
- Communicating AI decisions to non-technical stakeholders.
- Hands-on walkthrough: explaining a black-box classifier.

Section 4: Governance, Compliance, and Regulation

- Overview of global AI regulations and guidelines (EU AI Act, OECD, UNESCO).
- Industry standards: ISO/IEC 24028, NIST AI Risk Management Framework.
- Ethics review boards, model documentation, and audit trails.
- AI in high-risk sectors: health, law, public services.
- Data protection, consent, and the right to explanation.
- Building internal governance structures for ethical AI.
- Preparing for regulatory audits and external evaluations.

Section 5: Designing Ethical and Transparent AI Systems

- Ethical design thinking in AI product development.
- Balancing accuracy, explainability, and user trust.
- Human-in-the-loop systems and oversight protocols.
- Transparency by design: UI, feedback, documentation.
- Building ethical AI cultures inside organizations.
- Scenario planning for ethical decision-making.
- Action planning: applying XAI and ethics to your projects.

Certificate Description

Upon successful completion of this training course, delegates will be awarded a Holistique Training Certificate of Completion. For those who attend and complete the online training course, a Holistique Training e-Certificate will be provided.

Holistique Training Certificates are accredited by the British Accreditation Council (BAC) and The CPD Certification Service (CPD), and are certified under ISO 9001, ISO 21001, and ISO 29993 standards.

CPD credits for this course are granted by our Certificates and will be reflected on the Holistique Training Certificate of Completion. In accordance with the standards of The CPD Certification Service, one CPD credit is awarded per hour of course attendance. A maximum of 50 CPD credits can be claimed for any single course we currently offer.

Categories

AI, Data and Visualisation, IT & Computer Application, Technology

Tags

AI Ethics

Related Articles

Top 15 Skills Every Data Scientist Needs in 2025

Discover the top 15 data scientist skills you need to succeed in 2025, including technical, communication, cloud, and soft skills—plus expert tips on how to master them.