Explainable AI (XAI) for Research Decisions Training Course

Professional Development Training Course Outline

Category	Research and Data Analysis	**Duration**	5 Days
Base Fee	$1,500 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

As artificial intelligence continues to shape global research landscapes, the need for transparency and accountability in AI models—especially in sensitive domains like healthcare, criminal justice, human rights, and social policy—has become crucial. Explainable AI (XAI) for Research Decisions, equips researchers, data scientists, policy-makers, and ethics professionals with practical tools and critical insights to leverage Explainable AI for ethically sound and socially responsible research outcomes.

Through an in-depth exploration of XAI frameworks, interpretability techniques, bias detection methods, and case-driven analysis, this course empowers learners to enhance algorithmic transparency, promote research integrity, and make evidence-based decisions when addressing vulnerable populations or high-impact societal challenges. The training integrates cutting-edge methods, real-world datasets, and sensitivity-aware modeling practices to ensure responsible use of AI in delicate research contexts.

Programme Curriculum

Explainable AI (XAI) for Research Decisions Training Course

Introduction

As artificial intelligence continues to shape global research landscapes, the need for transparency and accountability in AI models—especially in sensitive domains like healthcare, criminal justice, human rights, and social policy—has become crucial. Explainable AI (XAI) for Research Decisions, equips researchers, data scientists, policy-makers, and ethics professionals with practical tools and critical insights to leverage Explainable AI for ethically sound and socially responsible research outcomes.

Through an in-depth exploration of XAI frameworks, interpretability techniques, bias detection methods, and case-driven analysis, this course empowers learners to enhance algorithmic transparency, promote research integrity, and make evidence-based decisions when addressing vulnerable populations or high-impact societal challenges. The training integrates cutting-edge methods, real-world datasets, and sensitivity-aware modeling practices to ensure responsible use of AI in delicate research contexts.

Course Objectives

1. Understand the core principles of Explainable AI (XAI) in sensitive research contexts.
2. Identify and analyze bias, fairness, and accountability issues in AI models.
3. Apply interpretable machine learning techniques to complex datasets.
4. Evaluate ethical and social implications of AI in high-stakes research decisions.
5. Develop risk mitigation strategies in sensitive data environments.
6. Implement post-hoc explanation methods for AI transparency.
7. Use model-agnostic XAI tools for sensitive topic research.
8. Understand privacy-preserving machine learning in ethical research contexts.
9. Incorporate sociotechnical perspectives in AI model interpretation.
10. Explore causality and counterfactual reasoning in XAI.
11. Examine regulatory frameworks (e.g., GDPR, AI Act) and their implications.
12. Conduct impact assessments using explainable AI metrics.
13. Design human-centered research workflows with XAI integration.

Target Audiences

1. Academic Researchers
2. Data Scientists
3. Policy Makers
4. Ethics Officers
5. Human Rights Analysts
6. Public Health Researchers
7. Journalists Investigating AI
8. NGO & Social Impact Professionals

Course Duration: 5 days

Course Modules

Module 1: Introduction to Explainable AI in Sensitive Research

- Definition and scope of XAI
- Key challenges in researching sensitive topics
- Why explainability matters in vulnerable contexts
- Overview of AI decision-making in social research
- Ethics and accountability foundations
- **Case Study:** XAI in predictive policing and human rights audits

Module 2: Understanding Bias and Fairness in AI

- Types of bias in datasets and models
- Fairness-aware machine learning algorithms
- Intersectionality in model outcomes
- Measuring and mitigating bias
- Stakeholder inclusion in fairness evaluations
- **Case Study:** Fairness analysis in gender-based violence prediction models

Module 3: Interpretable Machine Learning Techniques

- Glass-box models: Decision trees, linear models
- Model simplification strategies
- Global vs local interpretability
- Trade-offs between performance and transparency
- Visualization of interpretable outputs
- **Case Study:** Transparent health-risk modeling for vulnerable populations

Module 4: Post-hoc Explanation Methods

- SHAP (SHapley Additive Explanations)
- LIME (Local Interpretable Model-agnostic Explanations)
- Counterfactual and contrastive explanations
- Anchors and feature attribution
- Explanation fidelity vs comprehensibility
- **Case Study:** Using SHAP in refugee data modeling

Module 5: Privacy and Data Protection in Sensitive Research

- Overview of data protection regulations (GDPR, HIPAA)
- De-identification and anonymization techniques
- Federated learning for privacy preservation
- Consent frameworks for sensitive data use
- Secure multi-party computation basics
- **Case Study:** Federated XAI in mental health studies

Module 6: Causality and Counterfactual Reasoning

- Introduction to causal inference in AI
- DAGs and structural causal models
- Identifying confounders in sensitive topics
- Counterfactual queries for explanation
- Causal discovery with limited data
- **Case Study:** Causal reasoning in child welfare policy impact analysis

Module 7: Regulatory and Ethical Frameworks

- Understanding AI governance (AI Act, OECD principles)
- Human rights-based approaches to AI research
- Ethics checklists and audit tools
- Documentation and transparency standards
- Building trust through accountability
- **Case Study:** Ethics audit of AI in migration policy

Module 8: Designing Human-Centered Research Workflows

- Co-designing with communities
- Participatory research approaches
- Communicating AI decisions to lay users
- Inclusive interface design for explanations
- Feedback loops in human-AI collaboration
- **Case Study:** Human-centered XAI in indigenous population health research

Training Methodology

- Interactive lectures with real-world applications

- Hands-on coding exercises using Python, SHAP, LIME
- Group discussions on ethical dilemmas in XAI
- Case-based learning from global research initiatives
- Guided projects using sensitive datasets

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call *+254769199797*

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	27 Nov 2026	Online	$1,000
23 Nov 2026	27 Nov 2026	Physical	$1,500
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$6,000
23 Nov 2026	27 Nov 2026	Physical	$4,500
23 Nov 2026	27 Nov 2026	Physical	$6,000

Start Date	End Date	Location	Price
23 Nov 2026	27 Nov 2026	Physical	$6,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com