Training Course on Bias Detection and Mitigation in Machine Learning

Professional Development Training Course Outline

Category	Data Science	**Duration**	10 Days
Base Fee	$3,000 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

Training Course on Bias Detection & Mitigation in Machine Learning delves into the critical and evolving field of **Bias Detection & Mitigation in Machine Learning**. As **AI systems** become increasingly prevalent and influential in various sectors, from healthcare to finance, ensuring their **fairness, transparency, and accountability** is paramount. This program equips participants with the essential knowledge and practical skills to identify, analyze, and effectively address **algorithmic bias** throughout the entire **machine learning lifecycle**. By mastering advanced **ethical AI frameworks** and **responsible AI practices**, participants will learn to build and deploy robust, equitable, and trustworthy AI solutions that mitigate potential societal harms and comply with emerging **AI regulations**.

The course will explore diverse sources of bias, from **data collection** and **preprocessing** to **model training** and **deployment**. Through **hands-on exercises** and **real-world case studies**, attendees will gain proficiency in utilizing **fairness metrics**, implementing **bias mitigation techniques**, and establishing **ethical AI governance**. This training is crucial for organizations committed to developing **responsible AI**, fostering **data ethics**, and building public trust in their **AI-driven innovations**. It addresses the urgent need for professionals who can navigate the complex ethical landscape of AI and champion the development of **equitable algorithms** in a rapidly advancing technological world.

Programme Curriculum

Training Course on Bias Detection & Mitigation in Machine Learning: Identifying and Addressing Algorithmic Bias

Introduction

Training Course on Bias Detection & Mitigation in Machine Learning delves into the critical and evolving field of **Bias Detection & Mitigation in Machine Learning**. As **AI systems** become increasingly prevalent and influential in various sectors, from healthcare to finance, ensuring their **fairness, transparency, and accountability** is paramount. This program equips participants with the essential knowledge and practical skills to identify, analyze, and effectively address **algorithmic bias** throughout the entire **machine learning lifecycle**. By mastering advanced **ethical AI frameworks** and **responsible AI practices**, participants will learn to build and deploy robust, equitable, and trustworthy AI solutions that mitigate potential societal harms and comply with emerging **AI regulations**.

The course will explore diverse sources of bias, from **data collection** and **preprocessing** to **model training** and **deployment**. Through **hands-on exercises** and **real-world case studies**, attendees will gain proficiency in utilizing **fairness metrics**, implementing **bias mitigation techniques**, and establishing **ethical AI governance**. This training is crucial for organizations committed to developing **responsible AI**, fostering **data ethics**, and building public trust in their **AI-driven innovations**. It addresses the urgent need for professionals who can navigate the complex ethical landscape of AI and champion the development of **equitable algorithms** in a rapidly advancing technological world.

Course Duration

10 days

Course Objectives

Upon completion of this training, participants will be able to:

1. Comprehend the ethical and societal implications of algorithmic bias in cutting-edge AI systems.
2. Identify diverse sources of bias in machine learning pipelines, including historical bias, representation bias, and measurement bias.
3. Apply advanced data preprocessing techniques to reduce bias in training datasets and enhance data fairness.
4. Implement various bias detection metrics and fairness evaluation methodologies (e.g., demographic parity, equalized odds).
5. Master in-processing and post-processing bias mitigation strategies to build more equitable AI models.
6. Utilize explainable AI (XAI) techniques to understand model decision-making and uncover hidden biases, promoting AI interpretability.
7. Conduct comprehensive fairness audits of machine learning models across different demographic groups.
8. Design and develop debiased machine learning algorithms for diverse applications, including natural language processing (NLP) and computer vision.
9. Navigate emerging AI ethics regulations and compliance frameworks (e.g., EU AI Act, GDPR) relevant to bias.
10. Establish robust AI governance frameworks and responsible AI development practices within organizations.
11. Assess and mitigate potential harms associated with biased AI systems in real-world scenarios.
12. Develop strategies for continuous monitoring and retraining of AI models to address evolving biases.

13. Foster a culture of ethical AI innovation and promote interdisciplinary collaboration for responsible AI deployment.

Organizational Benefits

- Demonstrates commitment to ethical AI, building public and stakeholder confidence.
- Ensures compliance with evolving AI ethics regulations and minimizes potential legal liabilities.
- Addressing bias often leads to more robust, generalizable, and accurate models.
- Promotes social responsibility and prevents discriminatory impacts.
- Fosters a responsible innovation ecosystem, ensuring long-term viability of AI initiatives.
- Appeals to professionals seeking to work for ethically conscious and forward-thinking organizations.
- Positions the organization as a leader in responsible AI development and deployment.
- AI systems free from bias lead to more reliable and trustworthy insights for strategic decisions.

Target Audience

1. Machine Learning Engineers.
2. Data Scientists & Analysts
3. AI Product Managers.
4. Software Developers
5. Compliance & Legal Professionals.
6. Business Leaders & Executives.
7. Researchers & Academics.
8. Policy Makers & Regulators.

Course Outline

Module 1: Foundations of Algorithmic Bias and Ethical AI

- Understanding the rise of **AI ethics** and its importance in modern society.
- Defining **algorithmic bias** and its various manifestations (historical, representation, measurement, aggregation).
- Exploring the societal and individual impacts of biased AI systems.
- Introduction to core **fairness definitions** (e.g., disparate impact, individual fairness).
- **Case Study:** COMPAS Recidivism Algorithm bias against Black defendants.

Module 2: Sources of Bias in the Machine Learning Lifecycle

- Bias in **data collection**: sampling bias, societal biases reflected in data.
- Bias in **data preprocessing**: feature engineering, imputation, data labeling.
- Bias in **model training**: algorithmic choices, optimization objectives.
- Bias in **model evaluation** and **deployment**: human feedback loops, interpretation biases.
- **Case Study:** Amazon's biased hiring algorithm and its implications.

Module 3: Data-Level Bias Detection and Mitigation

- Techniques for **auditing datasets** for bias: demographic analysis, statistical parity checks.
- **Data rebalancing** strategies: oversampling, undersampling, synthetic data generation.
- **Fairness-aware data preprocessing** methods: reweighing, disparate impact remover.
- Measuring and visualizing **data fairness metrics**.

- **Case Study:** Gender bias in image datasets (e.g., facial recognition training data).

Module 4: In-Processing Bias Mitigation Techniques

- Integrating **fairness constraints** into model optimization objectives.
- **Adversarial debiasing** methods for training fair models.
- Regularization techniques to promote fairness and prevent disparate impact.
- Exploring fair machine learning libraries and frameworks (e.g., AIF360, Fairlearn).
- **Case Study:** Bias mitigation in loan approval prediction models.

Module 5: Post-Processing Bias Mitigation and Explainable AI (XAI)

- **Threshold adjustment** and **calibration techniques** for fair outcomes.
- **Reject option classification** and other post-prediction debiasing strategies.
- Introduction to **XAI methodologies**: LIME, SHAP, feature importance.
- Using XAI to identify and understand sources of bias in black-box models.
- **Case Study:** Explaining biased medical diagnostic AI systems to build trust.

Module 6: Fairness Metrics and Auditing AI Systems

- Deep dive into various **fairness metrics**: equality of opportunity, predictive parity, treatment equality.
- **Model card** and **data card** documentation for transparency.
- Developing and implementing a **fairness audit framework** for AI systems.
- Continuous monitoring of fairness and performance in deployed models.
- **Case Study:** Auditing facial recognition systems for racial and gender bias.

Module 7: Ethical AI Governance and Responsible AI Development

- Principles of **responsible AI**: accountability, transparency, human oversight.
- Establishing **AI ethics committees** and **governance structures**.
- Developing **organizational policies** and best practices for ethical AI.
- Integrating **ethical considerations** throughout the entire **MLOps pipeline**.
- **Case Study:** Google's AI ethics challenges and efforts to establish responsible AI guidelines.

Module 8: Emerging Trends, Legal Landscape, and Future of Ethical AI

- Overview of **global AI regulations** (e.g., EU AI Act, NIST AI Risk Management Framework).
- The role of **AI literacy** and public awareness in promoting ethical AI.
- Addressing **intersectionality** in bias detection and mitigation.
- Future challenges and opportunities in **ethical AI research** and development.
- **Case Study:** The use of AI in criminal justice and the push for algorithmic accountability.

Training Methodology

This course employs a participatory and hands-on approach to ensure practical learning, including:

- Interactive lectures and presentations.
- Group discussions and brainstorming sessions.
- Hands-on exercises using real-world datasets.
- Role-playing and scenario-based simulations.
- Analysis of case studies to bridge theory and practice.

- Peer-to-peer learning and networking.
- Expert-led Q&A sessions.
- Continuous feedback and personalized guidance.

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	04 Dec 2026	Online	$2,000
23 Nov 2026	04 Dec 2026	Physical	$3,000

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$9,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$12,000

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

Register Online Now

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