AI Ethics & Governance

AI Ethics & Governance Internships

Internship Categories

Contact +91-8977624748 for joining process

Visit ASCEND Portal for more details →

Browse AI Ethics & Governance internship programs at NTHRYS Biotech Labs — academic, professional, industrial and research tracks across the categories below.

CATEGORIES

Algorithmic Bias Detection in Computer Vision
Research methods to identify and measure demographic biases in image classification and object detection systems across different population groups.

Fair Machine Learning Model Evaluation Metrics
Develop and test new fairness-aware evaluation metrics that balance accuracy with equitable performance across protected attributes.

Explainability Techniques for Neural Networks
Implement and compare interpretability methods like LIME, SHAP, and attention visualization to understand deep learning model decisions.

Adversarial Robustness Testing Frameworks
Design and execute systematic tests to evaluate how AI systems respond to adversarial examples and malicious inputs.

Privacy-Preserving Machine Learning Implementation
Research and implement federated learning, differential privacy, and encrypted machine learning techniques for sensitive data.

Algorithmic Transparency Auditing Standards
Develop audit frameworks and documentation standards to assess transparency requirements in deployed AI systems.

Natural Language Processing Bias Analysis
Investigate gender, racial, and cultural biases embedded in language models and text processing systems.

Fairness in Automated Decision-Making Systems
Study fairness implications in hiring algorithms, credit scoring, and criminal justice risk assessment tools.

AI Policy Compliance and Legal Analysis
Analyze and document AI system compliance with regulations like GDPR, CCPA, and emerging AI governance frameworks.

Stakeholder Impact Assessment Methodologies

Design research protocols to identify and evaluate impacts of AI systems on different stakeholder communities.

Data Governance and Curation Best Practices

Research methods for responsible data collection, annotation, and management to reduce bias in training datasets.

Algorithmic Accountability Through Traceability

Develop systems and documentation approaches to trace AI decision-making processes from input to output for accountability.

Facial Recognition Ethics and Regulation

Investigate accuracy disparities, privacy concerns, and governance models for facial recognition technology deployment.

Recommendation System Fairness Research

Study bias in content recommendation algorithms and develop fair ranking methods across user demographics.

AI Transparency in Criminal Justice Systems

Evaluate explainability and fairness of predictive policing and risk assessment algorithms in law enforcement.

Human-AI Collaboration Ethics Study

Research how humans and AI systems can collaborate ethically with appropriate oversight, transparency, and human control.

Synthetic Data Generation for Bias Mitigation

Develop and evaluate synthetic data generation techniques to balance datasets and reduce historical bias representation.

Responsible AI Governance Framework Development

Design comprehensive governance frameworks including policies, processes, and oversight mechanisms for organizational AI deployment.

Environmental Impact of AI Systems

Research energy consumption, carbon footprint, and sustainability considerations in training and deploying large AI models.

AI Ethics Curriculum and Education Design

Develop educational materials and training programs to build AI ethics literacy among technologists and stakeholders.

Autonomous Vehicle Safety and Ethical Testing

Research ethical dilemmas, safety validation methods, and governance approaches for autonomous vehicle deployment.

AI Misinformation and Deepfake Detection

Develop detection systems and evaluate AI''s role in generating and identifying synthetic media and misinformation.

Algorithmic Discrimination in E-commerce Pricing

Study price discrimination algorithms and their differential impact on consumer groups and market fairness.

Healthcare AI Fairness and Equity Research

Investigate fairness in medical diagnosis algorithms, treatment recommendations, and resource allocation systems.

Consent and Data Rights Management Systems

Design systems and mechanisms for managing user consent, data ownership, and rights in AI-driven applications.

Interpretable Machine Learning Feature Analysis

Research methods to understand which features drive AI model decisions and validate their fairness implications.

AI Surveillance and Privacy Impact Assessment

Analyze privacy risks and ethical implications of surveillance AI systems in public and private contexts.

Bias in AI-Generated Content and Outputs

Evaluate biases in generative AI systems producing text, images, and audio for fairness and representation issues.

AI Incident Reporting and Documentation Standards

Develop standardized protocols for documenting, analyzing, and reporting failures and harms from AI systems.

Participatory Design in AI System Development

Research co-design approaches that include affected communities in developing and governing AI systems.

Model Card and Documentation Best Practices

Create comprehensive documentation frameworks that transparently communicate AI model capabilities, limitations, and ethical considerations.

Algorithmic Auditing Tools and Automation

Develop automated audit tools to continuously monitor deployed AI systems for bias, drift, and ethical violations.

Disability Rights and AI Accessibility

Research accessibility issues, digital equity impacts, and inclusive design principles for AI-powered applications.

Financial Services AI Discrimination Detection

Investigate bias in loan approval, credit scoring, and insurance algorithms that may discriminate against protected groups.

Explainable AI for Non-Technical Stakeholders

Develop communication methods and visualization techniques to explain AI decisions to policymakers and affected communities.

AI System Lifecycle Governance and Oversight

Research governance structures and oversight mechanisms throughout AI system development, deployment, and maintenance phases.

Whistleblower Protection in AI Organizations

Study mechanisms and policies to enable and protect ethical reporting of AI safety and fairness concerns.

Multi-Stakeholder AI Governance Models

Examine collaborative governance approaches involving technologists, regulators, civil society, and affected communities.

AI Transparency in Hiring and Recruitment

Analyze bias in resume screening algorithms and evaluate fairness interventions in employment decision systems.

Interpretability vs. Accuracy Trade-off Research

Investigate the balance between model explainability and predictive performance in practical fairness-aware applications.

Vulnerable Population Protection in AI Systems

Research safeguards and ethical guidelines for AI systems affecting vulnerable groups including minors and marginalized communities.

AI Regulatory Sandbox Implementation Studies

Analyze and evaluate regulatory sandbox programs that test AI systems in controlled environments before wider deployment.

Algorithmic Recourse and User Redress Systems

Design mechanisms that allow individuals to understand, challenge, and appeal AI-based decisions affecting them.

International AI Ethics Standards Comparison

Analyze and compare AI ethics guidelines and governance approaches across different countries and regions.

AI Reproducibility and Scientific Integrity

Research methodologies for ensuring reproducibility, documentation, and scientific integrity in AI research and development.

Consumer Protection and AI-Driven Marketing

Investigate manipulative AI marketing practices, dark patterns, and consumer protection mechanisms in algorithmic systems.

AI Impact on Labor Markets and Employment

Study societal and economic implications of AI automation on workforce displacement and job quality.

Fairness-Aware Data Pre-processing Techniques

Research and implement data cleaning and preparation methods that reduce historical bias before model training.

Third-Party AI Auditing Infrastructure

Develop independent auditing mechanisms and standards for external evaluation of AI systems'' fairness and safety.

AI Ethics in Developing Countries Context

Research unique ethical challenges and governance needs for AI deployment in resource-constrained developing country contexts.