DEPARTMENT OF APEX INSTITUTE OF TECHNOLOGY

PROJECT PROPOSAL

1. Project Title: - **Analyzing bias in AI models**

2. Project Scope: - (Max 500 words)

Artificial Intelligence (AI) plays an increasingly crucial role in decision-making across various domains, including healthcare, finance, recruitment, law enforcement, and education. However, AI models are susceptible to biases that arise from data, algorithms, and human intervention. Bias in AI can lead to unfair treatment, discrimination, and unethical outcomes, raising concerns about accountability and transparency. This project aims to systematically analyze AI bias, identify its sources, measure its impact, and explore effective mitigation strategies to ensure fairer and more equitable AI systems.

Objectives:

1. **Identify and Analyze Bias in AI Models:** Investigate different types of biases (data, algorithmic, and societal) and their impact on AI-driven decisions. Use real-world case studies to understand how bias affects sectors like hiring, healthcare, and law enforcement.

2. **Develop and Apply Bias Detection Techniques:** Implement statistical and computational methods to quantify bias in AI models. Use fairness metrics (e.g., demographic parity, disparate impact) and adversarial testing to assess model fairness across different demographic groups.

3. **Explore and Evaluate Bias Mitigation Strategies:** Study and compare three bias mitigation approaches: pre-processing (data balancing), in-processing (fairness-aware training), and post-processing (adjusting AI outputs). Analyze their effectiveness in reducing bias across various AI applications.

4. **Recommend Ethical and Policy Guidelines:** Assess existing AI regulations and propose best practices for ethical AI development. Provide recommendations for transparent, accountable, and fair AI systems based on industry standards and governance frameworks.

3. Requirements: -

- Hardware Requirements

1. Processor: Multi-core CPU
2. Storage: SSD with at least 256 GB
3. RAM: At least 8 GB

- Software Requirements

1. Machine Learning Frameworks
2. Data Processing Tools
3. Development Environment

STUDENTS DETAILS

Name	UID	Signature
Mustafiz Ahmed	21BCS6717	

APPROVAL AND AUTHORITY TO PROCEED

We approve the project as described above, and authorize the team to proceed.

Name	Title	Signature (With Date)