AVIT AARUPADAI VEEDU INSTITUTE OF TECHNOLOGY Vinayaka Missions Chennai Campus | VINAYAKA MISSION'S RESEARCH FOUNDATION | AI NEXUS CLUB | INSTITUTION'S INNOVATION COUNCIL | INDIAai | AI IMPACT SUMMIT

Seminar / Session Report

1) Event Details

Title: ***Responsible AI: Risk Mitigation, Bias Reduction, and Model Transparency***

Date: 28.01.2026
Time: 10.00 AM – 11.00 AM
Venue: Intel Lab, AVIT
Organised by: AI Nexus Club / AVIT (CSE)
Resource Person: Dr. S. Pitchumani Angayarkanni, Professor, Department of CSE, AVIT

Student Strength: 44

2) Objective of the Session

The session aimed to introduce students to Responsible AI as an engineering discipline—where *risks, bias, and transparency are measurable and testable*, not just policy statements.

3) Session Summary (Topics Covered)

A. Why Responsible AI?

- Responsible AI was positioned as accountability in the age of automation, emphasizing that "accuracy alone isn't safety, fairness, or trust."
- Real-world examples were discussed to show how AI can create harm when trained on biased historical patterns or deployed without controls.

B. Risk in AI (Core Concept)

- AI Risk = Harm × Likelihood (harm: financial loss/denial of opportunity/safety/privacy; likelihood: data quality, drift, misuse, adversarial inputs).
- Need for a risk register + controls, similar to cybersecurity practice.

C. Three Pillars of Responsible AI

1. Risk Mitigation: prevention + detection + response
2. Bias Reduction: measure → mitigate → monitor
3. Model Transparency: explainability + documentation + traceability

D. Practical Tooling / Frameworks Introduced

- Risk mitigation checklist (data leakage, distribution shift, misuse, security/prompt injection, privacy inference) and the principle: "Every risk needs a test."
- Fairness concepts such as Demographic Parity gap and Equalized Odds gap with a scholarship-style example.

- Transparency methods: local vs global explanations, limitations, auditability; introduction to interactive explainability exploration.
- Model Card sections (intended use, data summary, metrics overall & per-group, fairness + mitigation, explainability approach, monitoring plan, ownership/versioning).
- Post-deployment monitoring risks: drift, bias re-emergence, feedback loops; monitor metrics by group + alert thresholds.
- Ownership clarity via RACI: Builder, Reviewer, Owner, Auditor.

4) Teaching–Learning Methods Used

- Concept explanation with real-world scenarios and responsible AI framing as an engineering workflow.
- Demonstration previews aligned to:
    - Data leakage trap (inflated accuracy through "cheating")
    - Fairness metrics + mitigation trade-offs
    - Explain a prediction (global + local explanation approaches)

5) Key Learning Outcomes (for Students)

By the end of the session, students were able to:

- Explain AI risk using the *Harm* × *Likelihood* framing and propose testable controls.
- Identify common AI failure modes: leakage, drift, misuse, privacy inference, and security risks in GenAI settings.
- Describe bias using fairness gaps (approval-rate gaps and error-rate gaps) and discuss mitigation trade-offs.
- Understand transparency artifacts: explanations, model cards, traceability, and monitoring responsibilities.

6) Participation Details

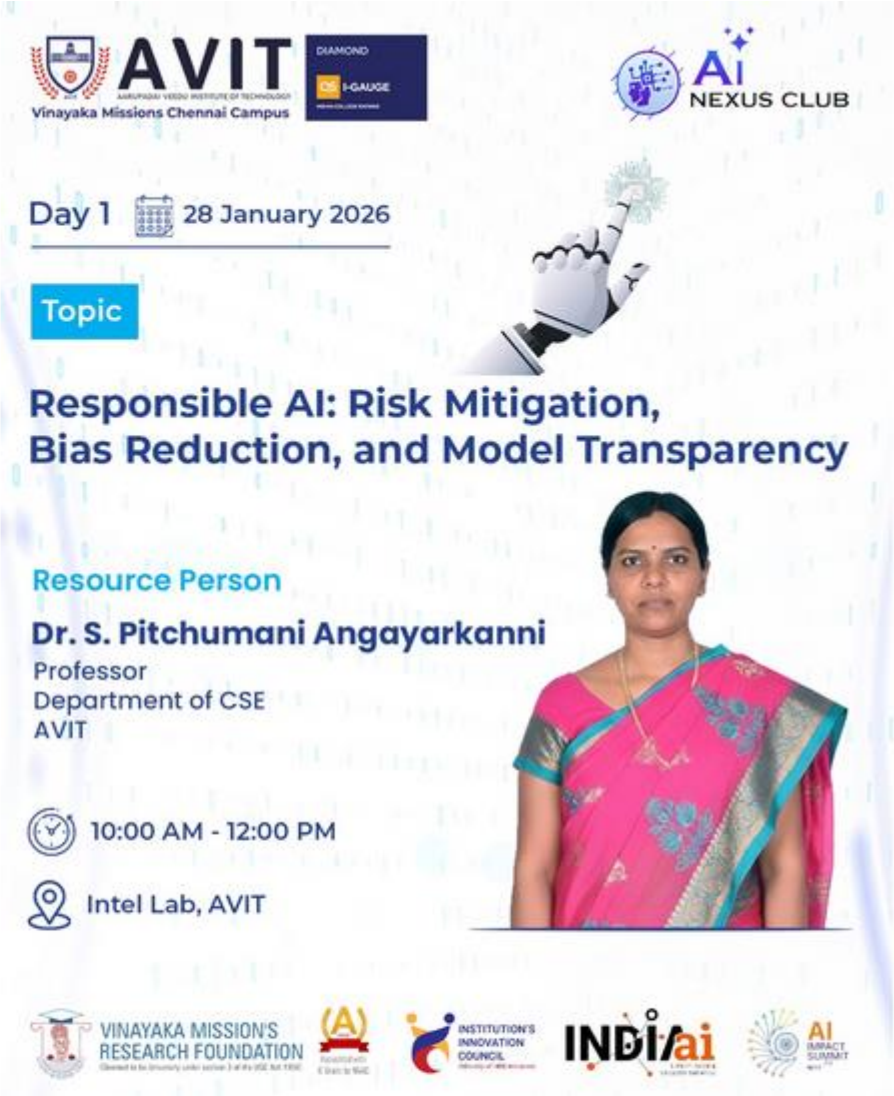
- Total Participants (Students): 44
- Active interaction during examples on risk, fairness, and accountability workflow.

7) Conclusion & Follow-up Suggestions

The session successfully reinforced that Responsible AI is operational—implemented through tests, documentation, and monitoring, supported by clear ownership roles.

APPENDIX

Event Poster



AVIT
Vinayaka Missions Chennai Campus

DIAMOND
QS I-GAUGE

Ai NEXUS CLUB

Day 1 28 January 2026

Topic

Responsible AI: Risk Mitigation, Bias Reduction, and Model Transparency

Resource Person

Dr. S. Pitchumani Angayarkanni
Professor
Department of CSE
AVIT

10:00 AM - 12:00 PM

Intel Lab, AVIT

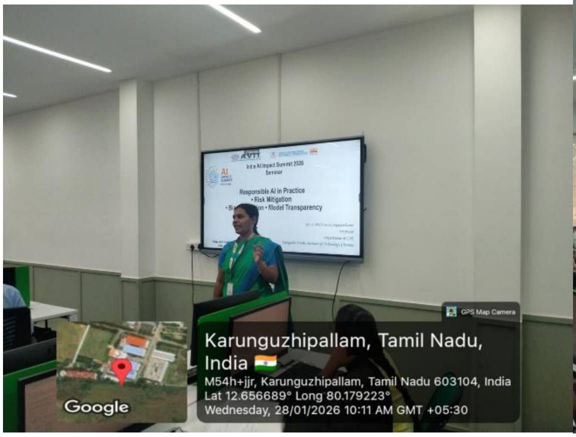
VINAYAKA MISSION'S RESEARCH FOUNDATION

INSTITUTION'S INNOVATION COUNCIL

INDIAai

AI IMPACT SUMMIT

Photographs of the Session


Responsible AI in Practice
• Risk Mitigation
• Model Transparency
GPS Map Camera
Google
Karunguzhipallam, Tamil Nadu, India 🇮🇳
M54h+jjr, Karunguzhipallam, Tamil Nadu 603104, India
Lat 12.656689° Long 80.179223°
Wednesday, 28/01/2026 10:11 AM GMT +05:30

GPS Map Camera
Google
Karunguzhipallam, Tamil Nadu, India 🇮🇳
M54h+wmm, Karunguzhipallam, Tamil Nadu 603104, India
Lat 12.65743° Long 80.178789°
Wednesday, 28/01/2026 10:11 AM GMT +05:30