VISHWAKARMA INSTITUTE OF TECHNOLOGY, PUNE

AI&DS Insights

Monthly Awareness Bulletin

ISSUE 09| May 2025

Vishwostav 2025

"You are stronger than you seem, braver than you believe, and smarter than you think." - Christopher Robin

Faculty Article

Responsible AI: Building Ethics into Intelligent Systems

Artificial Intelligence (AI) is no longer a futuristic concept—it's a present-day force transforming industries from finance to healthcare to logistics. However, as AI systems become increasingly autonomous and impactful, they raise critical questions about fairness, accountability, transparency, and safety. Responsible AI addresses these challenges by integrating ethical principles into the design, development, and deployment of AI technologies.

.

What is Responsible AI?

Responsible AI refers to a framework of practices and principles that guide the ethical and safe development of AI systems. It's not just about compliance—it's about building AI that earns trust and contributes positively to society.

The foundation of Responsible AI lies in five key pillars:

1. Fairness: AI should not discriminate. Yet, models trained on biased data can reinforce social inequalities. Fair AI practices involve identifying and mitigating algorithmic bias through representative datasets, fairness metrics, and ongoing audits.
2. Transparency: Many AI models, especially deep learning systems, are "black boxes." Responsible AI promotes explainability through interpretable models or post-hoc tools like SHAP or LIME, enabling users to understand how decisions are made.
3. Accountability: Who is responsible when AI goes wrong? Whether it's a flawed hiring algorithm or a misdiagnosis in healthcare, accountability requires clear governance structures, human oversight, and traceable system behavior.
4. Privacy: AI thrives on data, but not at the cost of user privacy. Techniques like federated learning and differential privacy are critical tools to ensure that personal data remains protected during training and inference.
5. Safety and Robustness: Responsible AI must function reliably under real-world conditions and adversarial environments. This includes rigorous testing, fail-safes, and continuous monitoring.

·Challenges in Practice

Despite growing awareness, several challenges hinder Responsible AI adoption:
Trade-offs between accuracy and fairness
Lack of standardized metrics for explainability or trustworthiness
Resource constraints for small organizations to conduct ethics reviews
Regulatory uncertainty in fast-changing markets
Nonetheless, industry leaders are establishing internal AI ethics boards, embedding ethics into DevOps pipelines (EthicsOps), and collaborating with regulators and academia to define best practices.

Case in Point: Responsible AI in Financial Services
In credit scoring, traditional models often disadvantage underrepresented groups due to historical bias in financial data. By adopting responsible AI, financial institutions can:

- **Use fairness-aware modeling techniques**
- **Audit models for disparate impact**
- **Provide explainable decisions to users and regulators**
-

Looking Ahead
Organizations that proactively embrace Responsible AI will gain a competitive edge. They won't just build better models—they'll build systems that align with human values.
In the end, Responsible AI is not just a technical imperative—it's a social contract. We must ensure that the systems we build reflect the society we want to create.

Dr Sunil M Sangve

Professor, AI&DS Department

Activities

NEP Orientation & Sensitization Programme under the Malaviya Mission Teacher Training Programme

The NEP Orientation and Sensitization Programme, organized by the AIDS Department in collaboration with Sant Gadge Baba Amravati University (SGBAU), was conducted under the Malaviya Mission Teacher Training Programme. This 10-day Faculty Development Programme (FDP), held from April 21 to 30, aimed to familiarize educators with the core principles and implementation strategies of the National Education Policy (NEP). The initiative focused on enhancing the pedagogical skills and policy awareness of faculty members, ensuring they are well-equipped to contribute effectively to the evolving educational landscape.

Faculty Publications

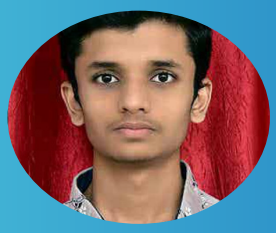
Name	Summary
Dr. Minal Barhate	Transformative integration of AI/ML technologies in law enforcement: A holistic approach for enhanced efficiency and public safety, Proceedings of the 1st International Conference on Applications of AI In 5G And IoT (ICAAI5GI2024), 9–10 August 2024, Greater Noida, India

Team - AI&DS INSIGHTS

Dr Minal Barhate

Prof. Lokesh Khedekar

Ishika Golecha
SY AIDS-A



Gaurang Gulhane
SY AIDS-A

Utkarsha Dhawle
SY AIDS-A

Janhavi Deo
SY AIDS-A