

Title: Trusted AI

Abstract: Artificial intelligence systems are increasingly being used to support human decision-making. While AI holds the promise of delivering valuable insights and knowledge across a multitude of applications, broad adoption of AI systems will rely heavily on the ability to trust their output. Human trust in technology is based on our understanding of how it works and our assessment of its safety and reliability. Towards this end, IBM Research AI is developing diverse approaches for how to achieve fairness, robustness, explainability, and provenance. In this talk, I will focus on fairness, robustness, and explainability and discuss some techniques for checking such desired behaviors in AI systems.

Bio: Dr Diptikalyan Saha is a Senior Researcher and Manager in Data & AI department at IBM Research India. He obtained his Ph.D. in Computer Science from the State University of New York at Stony Brook in 2006. His current research focus is to create trusted and scalable AI platforms and services. Earlier he has worked on various areas of Computer Science including logic programming, software engineering, natural language processing, security, and databases, and published in top conferences like ICLP, CCS, FSE, ICSE, VLDB, SIGMOD in these areas. He is an ACM senior member.