

Romila Pradhan

School of Applied and Creative Computing, Purdue University
401 Grant St, West Lafayette, IN 47907, United States
rpradhan@purdue.edu; romilapradhan.github.io

Prepared June 1, 2026

RESEARCH INTERESTS

The design of trustworthy data-driven decision-making systems, and the application of data management principles to high quality, trusted end decisions. Research techniques draw on data quality, data integration, data science pipelines, and machine learning fairness and robustness.

EDUCATION

- 2018 **Purdue University**, West Lafayette, IN
Ph.D., Computer Science
Advisor: Sunil Prabhakar
Dissertation: *Guided Data Fusion*
- 2008 **Indian Institute of Technology**, Kharagpur, India
M.S., Mathematics and Computing
Advisor: M. P. Biswal
Thesis: *Multiple Choice Linear Programming – The Problem and the Solutions*
- 2008 **Indian Institute of Technology**, Kharagpur, India
B.S., Mathematics and Computing
Advisor: G. P. Rajasekhar
Dissertation: *Branch-Checking Algorithm for All-Pairs Shortest Paths*

PROFESSIONAL EXPERIENCE

- 2021–current **Purdue University**, West Lafayette, IN
Assistant Professor – School of Applied and Creative Computing
Affiliate Faculty Member – Center for Education and Research in Information Assurance and Security (CERIAS)
Core Faculty Member – Applied AI Research Center
- 2020–2021 **University of California San Diego**, La Jolla, CA
Postdoctoral Researcher – Halıcıoğlu Data Science Institute
- 2018–2020 **Purdue University**, West Lafayette, IN
Visiting Assistant Professor – Department of Computer Science
- 2011–2018 **Purdue University**, West Lafayette, IN
Ph.D. Student – Department of Computer Science
- 2008–2010 **Deutsche Bank Group**, Mumbai, India
Senior Analyst – Interest Rates Group

AWARDS

- 2026 EDBT Conference Best Reviewers Award
Outstanding Faculty in Discovery, School of Applied and Creative Computing
Exceptional Early Career Teaching Award, School of Applied and Creative Computing
- 2023 NSF CAREER Award
- 2022 Google Research Scholar Award

GRANTS

- 2026 NSF REU: CAREER: *Data Preparation for Trusted and Fair Data Science*
PI: Romila Pradhan
- 2023 NSF CAREER: *Data Preparation for Trusted and Fair Data Science*
PI: Romila Pradhan
- Center for Advancing Safety of Machine Intelligence (CASMI): *Diagnosing, Understanding, and Fixing Data Biases for Trusted Data Science*
PI: Romila Pradhan
- 2022 Google Research Scholar Award
PI: Romila Pradhan

PUBLICATIONS

Authorship Conventions

Conferences are the primary publication venue in databases and data management. Senior authors are typically listed last. Advisees are listed in **bold**. Romila Pradhan is underlined.

In Progress or In Review

- 2026 [1] **J. Hasan**, S. Jiang, **T. Singh**, S. Galhotra, R. Pradhan, D. Srivastava. *PipeLens: Identifying Interventions for Resolving Malfunctioning Data Science Pipelines*. In Review.
- [2] **A. Singh**, R. Pradhan. *Scalable Source Selection for Machine Learning Tasks*. In Review.
- [3] **A. Singh**, R. Pradhan. *Scalable Goal-Oriented Source Selection*. In Review.
- [4] **J. Hasan**, **J. Wang**, R. Pradhan. *Explanations and Recommendations for Machine Learning Pipelines under Data Drift*. In Progress.

Conference Proceedings

- 2026 [5] **J. Hasan**, R. Pradhan. *Selective Data Expansion for Model Performance*. EDBT 2026.
- [6] **J. Hasan**, **J. Wang**, R. Pradhan. *SENTINEL: Evaluating Pipeline Robustness to Distributional Shifts*. ICDE 2026. (Demonstration).
- [7] **A. Singh**, R. Pradhan. *SPLICE: An Efficient Framework for Selecting Sources for Machine Learning Tasks*. ICDE 2024. (Lightning talk).
- [8] **J. Hasan**, **J. Wang**, R. Pradhan. *Evaluating Pipeline Robustness Under Distributional Shifts*. ICDE 2024. (Lightning talk).

- 2025 [9] **T. Surve**, R. Pradhan. *Explaining Fairness Violations using Machine Unlearning*. EDBT 2025.
- [10] **S. Thandri**, R. Pradhan. *Label Flipping for Group Fairness*. QDB Workshop @VLDB 2025.
- [11] **J. Hasan**, R. Pradhan. *Explanations for Machine Learning Pipelines under Data Drift*. HILDA Workshop @SIGMOD 2025.
- 2024 [12] **Ekta**, R. Pradhan. *Valuation-based Data Acquisition for Machine Learning Fairness*. QDB Workshop @VLDB 2024.
- 2022 [13] R. Pradhan, T. Li. *Human-in-the-loop Bias Mitigation in Data Science*. HCAI Workshop @ NeurIPS 2022.
- [14] R. Pradhan, J. Zhu, B. Glavic, B. Salimi. *Interpretable Data-based Explanations for Fairness Debugging*. SIGMOD 2022.
- [15] J. Zhu, R. Pradhan, B. Glavic, B. Salimi. *Generating Interpretable Data-Based Explanations for Fairness Debugging using Gopher*. SIGMOD 2022 (Demonstration).
- [16] R. Pradhan, A. Lahiri, S. Galhotra, B. Salimi. *Explainable AI: Foundations, Applications, Opportunities for Data Management Research*. SIGMOD 2022 (Tutorial).
- [17] R. Pradhan, A. Lahiri, S. Galhotra, B. Salimi. *Explainable AI: Foundations, Applications, Opportunities for Data Management Research*. ICDE 2022 (Tutorial).
- 2021 [18] S. Galhotra*, R. Pradhan*, B. Salimi. *Explaining Black-Box Algorithms Using Probabilistic Contrastive Counterfactuals*. SIGMOD 2021. (*=Equal contribution)
- [19] S. Galhotra, R. Pradhan, B. Salimi. *Feature Attribution and Recourse via Probabilistic Contrastive Counterfactuals*. Algorithmic Recourse Workshop @ICML 2021.
- [20] P. Y. Wang, S. Galhotra, R. Pradhan, B. Salimi. *Demonstration of Generating Explanations for Black-Box Algorithms Using Lewis*. VLDB 2021 (Demonstration).
- 2019 [21] R. Pradhan, S. Prabhakar. *AuthIntegrate: Toward Combating False Data on the Internet*. Truth Discovery and Fact Checking Workshop @SIGKDD 2019.
- 2018 [22] R. Pradhan, S. Bykau, S. Prabhakar. *A Framework to Integrate User Feedback for Rapid Conflict Resolution*. ICDE 2018 (Demonstration).
- [23] R. Pradhan, W. G. Aref, S. Prabhakar. *Leveraging Data Relationships to Resolve Conflicts from Disparate Data Sources*. DEXA 2018.
- 2017 [24] R. Pradhan, S. Bykau, S. Prabhakar. *Staging User Feedback toward Rapid Conflict Resolution in Data Fusion*. SIGMOD 2017.

TEACHING EXPERIENCE

Course Development

- 2022 Responsible Data Management, Purdue CNIT graduate course
- 2019 Large-scale Data Analytics, Purdue CS undergraduate course

Teaching

- 2026 Spring Instructor, *Enterprise Data Management* (CNIT 39200, 81 students)
- 2025 Fall Instructor, *Responsible Data Science* (CNIT 57100, 19 students)
- Instructor, *Database Administration* (CNIT 48700, 7 students)
- 2025 Spring Instructor, *Enterprise Data Management* (CNIT 39200, 81 students)

2024 Fall	Instructor, <i>Responsible Data Science</i> (CNIT 57100, 7 students)
	Instructor, <i>Database Administration</i> (CNIT 48700, 14 students)
2024 Spring	Instructor, <i>Enterprise Data Management</i> (CNIT 39200, 61 students)
	Instructor, <i>Responsible Data Management</i> (CNIT 58100-RDM, 13 students)
2023 Fall	Instructor, <i>Responsible Data Science</i> (CNIT 58100-RDM, 19 students)
	Instructor, <i>Database Administration</i> (CNIT 48700, 7 students)
2023 Spring	Instructor, <i>Enterprise Data Management</i> (CNIT 39200, 56 students)
	Instructor, <i>Responsible Data Management</i> (CNIT 58100-RDM, 12 students)
2022 Fall	Instructor, <i>Responsible Data Science</i> (CNIT 58100-RDM, 5 students)
	Instructor, <i>Database Administration</i> (CNIT 48700, 5 students)
2022 Spring	Instructor, <i>Enterprise Data Management</i> (CNIT 39200, 61 students)
2022 Fall	Instructor, <i>Database Administration</i> (CNIT 48700, 6 students)

STUDENTS

Current	Jahid Hasan. 2023–current
Ph.D.s	<i>Intelligent Data-Centric Systems for Fair, Robust, and Adaptive Data Science Pipelines</i>
	Ambarish Singh. 2023–current
	<i>Efficient Source Selection for Machine Learning</i>
	Ikechukwu Andrew Ben Obi, (co-chaired with B. C. Min). Expected 2026
	<i>Neurosymbolic Robot Learning for Safe Language-Controlled Robot Systems</i>
	Jingya Wang. 2025–current, <i>Andrews Fellowship</i>
	<i>Toward Practical Machine Unlearning Techniques</i>
M.S. Advising	Omkar Pote. 2024–current
	<i>Optimizing Machine Learning Pipelines Under Data Drift</i>
	Ananya Uppal. 2024–current
	<i>Role of Data Valuation Techniques in Detecting and Repairing Data Quality Issues in Machine Learning Pipelines</i>
	Harshita Rathee. 2024–2026
	<i>The Role of Data Selection on the Fairness of Large Language Models</i>
	Kevin Chittilapilly. 2023–2024
	<i>Estimating Fairness in Machine Learning Systems through Data Characteristics</i>
	Tejendra Singh. 2023–2024
	<i>Leveraging Data Characteristics to Optimize Machine Learning Pipelines</i>
	Shashank Thandri. 2023–2024
	<i>Improving Fairness in Machine Learning Systems through Label Flipping</i>
	Ekta 2022–2024.
	<i>Using Data Acquisition to Improve Machine Learning Model Fairness</i>
	Sathvika Kotha. 2022–2024
	<i>Studying the Impact of Data Quality and Preprocessing on ML Model Fairness</i>
	Tanmay Surve. 2022–2023
	<i>Data-Based Explanations of Random Forests using Machine Unlearning</i>

- Ph.D. Weizhang Wang. 2025–current
 Committees *Reinforcement Learning from Human Feedback and Large Language Model Logical Reasoning*
- Srujana Kaddevarmuth. 2024–current
Comprehensive Ethical Frameworks for Developing and Scaling AI Systems
- James Prater. Defended 2025
Effects of Automation Bias on Human-AI Teaming in High-stress Environments
- Tung Vu. Defended 2023
A Statistical Approach to Detecting Generative Adversarial Network Overfitting
- M.S. Dhairya Dhuvad. 2025–current
 Committees *Fairness of Locally Differentially Private Models*
- Meghana Gorripati. 2025-2026
Explainable AI for Enhancing Cloud Autoscaling
- Rohit Bankar. 2025
Resource Optimization for Cloud-native ML Models
- Athreyan Mohana. 2024-2025
Ranked Cluster Influence: A Defense against Adversarial Labeling Attacks on Fairness
- Shridhar Ayush. 2023-2024
Intelligent Auto-scaling of Cloud-native Machine Learning Workloads
- Anuj Dubey. 2023-2024
Capturing Style through Large Language Models: An Authorship Perspective
- Yiqun Zhang. 2023
Predicting Answer Acceptance on Stack Overflow Using Asker’s Participation in Answer Comments
- Kanwardeep Singh Walia. 2021-2022
Using Temporal Networks to Find the Influencing Buggy Site in the Code Communities
- External Haakon Robinson, Ph.D., Norwegian University of Science and Technology. Defended 2023
 Committees *Trustworthy Machine Learning for Controlled Dynamical Systems*
- Aksel Vaaler, M.S., Norwegian University of Science and Technology. Defended 2023
Design and Simulation of Predictive Safety Filter for Autonomous Ship
- Undergrad Bo-Cheng Lu. 2026-current
 Research *Ensuring Fairness in Machine Learning Pipelines*
 Advising
- Ritwik Jayaraman. 2025-current
Zero-shot Machine Unlearning for Tabular Data
- Animesh Joshi. 2024-2025
Debugging Machine Learning Pipelines
- Ahana Bhattacharyya (University of Buffalo). 2023-2024, Ph.D. at Brown University
Data Validity for Machine Learning Model Fairness
- Liming Liu. 2023
- Anusha Sarraf. 2023
Investigating Explainability of Machine Learning Model Decisions for Injury Surveillance
- Manas Paranjape. 2022
Investigating Explainability of Machine Learning Model Decisions for Injury Surveillance

- Andrea Sanchez. 2022
Detecting Bias in ML-based Credit Risk Algorithms
- Jiawei Chang. 2022
Detecting Bias in ML-based Income Prediction
- Jiongli Zhu. 2022
Fairness Debugging via Interpretable Data-based Explanations
- Rohith Ravindranath. 2019
Identifying Direct and Indirect Causal Effects on Outcomes in Healthcare

INVITED TALKS

- 2025 Explainability, Robustness, and Fairness of Data Science Applications
School of Applied and Creative Computing Research Seminar Series, Purdue University
- Debugging, Explaining, and Mitigating Unfairness in Machine Learning
Cornell University
- 2024 Explaining Black-box Probabilistic Attacker-Defender Models
Sandia National Laboratories
- Debugging and Explaining Unfairness in Machine Learning Models
Applied Artificial Intelligence Research Center, Purdue University
- Debugging and Explaining Unfairness in Machine Learning Models
CERIAS Symposium, Purdue University
- Diagnosing, Understanding, and Fixing Data Biases for Trusted Data Science
Center for Advancing Safety in Machine Intelligence
- 2023 Explainability and Fairness in Algorithmic Systems
Honors College, Purdue University
- Toward Trustworthy and Responsible Data Science
Brandeis University
- Interpretable Data-based Explanations for Fairness Debugging
Industrial Engineering, Purdue University
- Causality-based Post hoc Explanations for Fairness Debugging
Computer Science & Artificial Intelligence Laboratory, MIT
- 2022 Data Preparation for Fair and Trusted Machine Learning
Cisco Research
- 2021 Explaining Black-box Algorithms Using Probabilistic Contrastive Counterfactuals
Computer Science & Artificial Intelligence Laboratory, MIT
- 2021 Explaining Black-box Algorithms Using Probabilistic Contrastive Counterfactuals
Department of Computer Science, University of California San Diego

UNIVERSITY SERVICE

- 2021-current Data and Software Development Curriculum Sub-Committee
School of Applied and Creative Computing
- 2022-2026 Department Faculty Search Committee
School of Applied and Creative Computing

- 2025-2026 Co-chair, Research Synergies Working Group
School of Applied and Creative Computing
- 2025 Organizer, Applied AI Day
Purdue Polytechnic Institute
- 2022-current Library Committee, Purdue University

COMMUNITY SERVICE

- 2026 SIGMOD PC
VLDB PC
EDBT PC
ICDE PC
TKDE Review Board
TWeb Review Board
TDSC Review Board
TII Review Board
HILDA PC and Mentor
QDB PC
- 2025 SIGMOD PC
VLDB PC
EDBT PC
ICDE PC
TaDA PC
TKDE Review Board
TWeb Review Board
TDSC Review Board
NASEM Rapid Expert Consultation Reviewer
- 2024 EDBT PC
SDM PC
CIKM PC
TKDE Review Board
TWeb Review Board
TDSC Review Board
NSF IIS Panelist
- 2023 EDBT PC
SIGKDD PC
FAccT PC
CIKM PC
TKDE Review Board
TWeb Review Board
TDSC Review Board
ICDE PhD Symposium PC
NSF IIS Panelist
US-Israel Binational Science Foundation Panelist
- 2022 SIGMOD PC
EDBT PC
CIKM PC
SIGKDD PC
TKDE Review Board
TDSC Review Board

- 2021 TKDE Review Board
TDSC Review Board
- 2020 TKDE Review Board
TDSC Review Board
- 2019 TDSC Review Board

OUTREACH

- 2024 VISION Summer camp course on Bias in Machine Learning
Purdue Polytechnic Institute
Discovering Opportunities in Technology (DOiT) camp course on Bias in Machine Learning
Purdue University
- 2022 Hands-on activities on Machine Learning and Bias to students from Purdue Polytechnic High School
Purdue University
Summer camp data science module on Machine Learning and Bias
Purdue University
- 2019 Summer camp data science module on exploratory data analysis and classification to Global Engineering Program participants from Taiwan
Purdue University

Glossary

- SIGMOD ACM SIGMOD/PODS International Conference on Management of Data
- VLDB International Conference on Very Large Data Bases
- ICDE IEEE International Conference on Data Engineering
- EDBT International Conference on Extending Database Technology
- HILDA Human-in-the-Loop Workshop at SIGMOD
- QDB Quality in Databases Workshop at VLDB
- TaDA Tabular Data Analysis Workshop at VLDB
- TKDE IEEE Transactions on Knowledge and Data Engineering
- TWeb IEEE Transactions on Web
- TDSC IEEE Transactions on Dependable and Secure Computing
- TII IEEE Transactions on Industrial Informatics
- NSF National Science Foundation
- III Division of Information and Intelligent Systems at NSF
- REU Research Experience for Undergraduates
- NASEM National Academies of Sciences, Engineering, and Medicine