Arjun Roy

Email: arjunroyihrpa@gmail.com
Ph: +49 15166421952
Git: github.com/arjunroyihrpa
Site: sites.google.com/view/arjunroy
LinkedIn: linkedin.com/in/arjun-roy-fub

Summary

Applied AI/ML Scientist with 7+ years of hands-on experience in fairness-aware AI, robust ML systems, and multi-modal learning. Proven record in leading open-source toolkits (e.g., MMM-Fair), designing SOTA solutions for real-world use-cases, and spearheading AI efforts in EU-funded projects. Specialize in bias mitigation, adversarial robustness, and scaling trustworthy ML from research to deployment.

Experience

Research Scientist, Research Institute CODE, Munich, Germany June 2023 – Present

- **MMM-Fair:** Lead developer of mmm-fair, an open-source framework enabling compliance-ready model training & auditing. Integrated in EU-funded projects for multi-attribute fairness evaluation.
- **EU Horizon project STELAR:**
  - **Adversarial Robustness:** Developed attacks increasing output damage by +68.6% over latent space attack SOTA on CelebA/FFHQ under Lipschitz bounds; designed defenses reducing reconstruction error by up to 95% in diffusion and NVAE models.
  - **Fair Multimodal Recruitment AI:** Built text-tabular-vision fusion model improving MAE by 80% over baselines; achieved MAE < 0.1 and reduced KLD bias by 55% across ethnicity groups.
  - **Multi-task Hazard Classification:** Leading LLM-based classifier for food hazard & product category detection; achieved F1 > 0.72 on highly imbalanced, 10-class dataset using LoRA and focal loss.
- **EU Horizon project MAMMOth:**
  - **AI Ethics & EU AI Act Compliance:** Co-led a multidisciplinary study on algorithmic fairness under the EU AI Act. Introduced a novel Socio-Economic Parity (SEP) fairness notion; constrained optimization using SEP achieved SPD $<$ 0.01 across demographics and improved identification of underprivileged, high-potential individuals by +66% over prior gold standard (CDP).
  - **Bias Transfer in Multi-task Learning:** Defined and tackled the novel problem of Bias Transfer in MTL alongside Negative Transfer and gradient conflict issues. On tabular (5 tasks) and visual (31 tasks) benchmarks, achieved SOTA fairness on 6/6 metrics and top-2 accuracy on 3/4 metrics.
  - **Open-source Contribution – MAI-BIAS Toolkit:** Core contributor to MAI-BIAS , an open-source framework for bias exploration and mitigation. Developed 5 of 27 core technical modules, on fairness evaluation and debiasing.

Research Scientist, L3S Research Center, Hannover, Germany Sept 2019 – May 2023

- **Volkswagen Foundation project BIAS:**
  - **Fair MTL via RL (L2T-FMT):** Designed a student-teacher MTL framework using deep reinforcement learning to dynamically optimize task objectives; achieved 12–19% fairness gains and up to 2% accuracy improvement over SOTA on tabular and visual tasks.
  - **AI Ethics & Anti-discrimination Law:** Co-led seminal interdisciplinary research bridging fairness in ML and anti-discrimination law, providing the first in-depth comparative analysis of multi-discrimination concepts across disciplines.
  - **Multi-attribute Fairness under Imbalance:** Outperformed IBM-Fairlearn (SOTA) on 3/4 datasets with up to 11% better performance under class imbalance.

Visiting Researcher, ITI-CERTH, Thessaloniki, Greece Sept 2022 – Nov 2022

- Designed a kernel-alignment-based algorithm to cluster tasks for efficient branching in Multi-task Learning.

Researcher Intern, L3S Research Center, Hannover, Germany Sept 2022 – Nov 2022

- Built a pipeline for stance detection in news headlines to support automated fact-checking.

Skills

- **Fairness & Responsible AI**: Bias-aware learning (tabular, CV, NLP), fairness in MTL, multi-attribute bias mitigation, model explainability, legal-alignment of fairness.
- **NLP & Multimodal AI**: Pre-LLM to LLM-era NLP, multilingual and low-resource modeling, fake news and stance detection, cross-lingual transfer, multimodal fusion (text + tabular + vision).
- **LLMs & Fine-Tuning**: LoRA, PEFT, AdapterHub, prompt-tuning, custom head training; experience with LLaMA, GPT, Mistral, PaLM, LaBSE, etc.

- **LLM Frameworks**: Hugging Face Transformers, LangChain, LlamaIndex, OpenAI APIs (chat + embeddings), RAG pipelines.
- **ML/DL Frameworks**: PyTorch, TensorFlow, Keras, Scikit-learn, SpaCy, NLTK; experience building from scratch and integrating with modern stacks.
- **MLOps & LLMOps**: Docker, MLflow, Weights & Biases, CI/CD (GitHub Actions), experiment tracking, reproducible pipelines.
- **Cloud & Tooling**: AWS (EC2, S3, Lambda), GCP, REST APIs, Flask, bash scripting, data pipelines.
- **Programming Languages**: Python (primary), C, C++, Java, R, SQL, JavaScript, HTML/CSS.
- **Teamwork & Leadership**: Led interdisciplinary teams, managed EU project WPs, strong communication across research, legal, and policy stakeholders.

Education

Freie Universität Berlin, Ph. D. in Computer Science — Expected 2025
- **Thesis:** Multi-criteria of fairness-aware Machine Learning
- **Advisor:** Prof. Eirini Ntoutsi

IIT Patna, Masters in Mathematics & Computing (9.16/10) — 2017 – 2019
- **Thesis:** Fake News Detection
- **Advisor:** Prof. Pushpak Bhattacharyya, Prof. Asif Ekbal, Prof. Stefan Dietze

IGNOU, Masters in Computer Application (3/4) — 2013 – 2016
- **Thesis:** Online Journal Management System
- **Advisor:** Assoc. Prof. Shalab Agarwal

Selected Publications

- Achieving Socio-Economic Parity through the Lens of EU AI Act — FAccT 2025
  A. Roy, S. Rizuo, S. Papadopoulos, E. Ntoutsi
- ALMA: Aggregated Lipschitz Maximization Attack on Auto-encoders — Under Review 2025
  C. K. Ramanaik , **A. Roy**, E. Ntoutsi
- FairBranch: Mitigating Bias Transfer in Fair Multi-task Learning — IJCNN 2024
  A. Roy, C. Koutlis, S. Papadopoulos, E. Ntoutsi
- Adversarial Robustness of VAEs across Intersectional Subgroups — BIAS 2024
  C. K. Ramanaik, **A. Roy**, E. Ntoutsi
- Exploring Fusion Techniques in Multimodal AI-Based Recruitment — EWAF 2024
  S. Swati, **A. Roy**, E. Ntoutsi
- Multi-dimensional discrimination in law and machine learning — FAccT 2023
  A. Roy, J. Horstmann, E. Ntoutsi
- Learning to teach fairness-aware deep multi-task learning — ECMLPKDD 2022
  A. Roy, E. Ntoutsi
- MulCoB-MulFaV: Multimodal Content Based Multilingual Fact Verification — IJCNN 2021
  A. Roy, A. Ekbal

Google Scholar: scholar.google.de/citations?user=HJ0FBh4AAAAJ

Honours

DAAD Scholarship 2018 (Master's Sandwich Program): Awarded to top 10% within IITs, placing in the top 0.016% of all postgraduate students nationwide in India.

GATE 2017 Top 3%: Among 100,000+ engineering/CS graduates nationwide; qualified for IIT M.Tech admission and received government-funded scholarship.

View CV online

Last updated on 09.04.2025