

Arjun Roy

Email: arjunroyhrpa@gmail.com
Ph: +49 15166421952
Add: Keltenstr. 2B, 82031 Grünwald
Git: github.com/arjunroyhrpa
LinkedIn: linkedin.com/in/arjun-roy-fub

Summary

AI/ML Scientist with 7+ years of experience in fairness-aware AI, robust ML systems, NLP, and multimodal learning. Led open-source projects like MMM-Fair and contributed to EU-funded initiatives across fintech and food safety. Specialized in bias mitigation, adversarial robustness, and responsible AI deployment. Acquired S/W engineering skills through an MCA, followed by advanced training in AI theory, mathematics, and applied machine learning during an M.Tech from IIT Patna.

Experience

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

- **MMM-Fair**: Lead developer of [mmm-fair](#), an open-source framework enabling compliance-ready model training & auditing, with LLM-powered explanations. Integrated into EU-funded projects for multi-attribute fairness evaluation.
- **EU Horizon project STELAR**:
 - **Adversarial Robustness**: Developed attacks increasing output damage by 69% 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**: Developing LLM-based pipeline for extracting food hazards (for food safety use-case) from incident reports; reached F1 > 0.72 on a highly imbalanced problem 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 fairness (SPD < 0.01) across demographics, 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**: 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 for FinTech use-case.

Research Scientist, L3S Research Center, Hannover, Germany Sep 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 Sep 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 Sep 2018 – Mar 2019

- 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 (chat, prompt, embeddings), LlamaIndex, OpenAI APIs.
- **ML/DL Frameworks:** PyTorch, TensorFlow, Keras, Scikit-learn, SpaCy, NLTK; experience building from scratch and integrating with modern stacks.
- **Distributed Training:** PyTorch DataParallel, DistributedDataParallel, torch.multiprocessing, FSDP; multi-GPU setups, experiment scaling.
- **MLOps & LLMOps:** Docker, MLflow, Weights & Biases, CI/CD (GitHub Actions).
- **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 communicator and creative problem-solver with a flexible, analytical mindset.
- **Languages:** English (Business Proficiency), German (A2)

Education

Freie Universität Berlin , Ph. D. in Computer Science (Expected Nov 2025)	Nov 2021 - Present
<ul style="list-style-type: none"> • Thesis: Multi-criteria of fairness-aware Machine Learning • Advisor: Prof. Eirini Ntoutsi 	
IIT Patna , Masters in Mathematics & Computing (9.16/10)	Jul 2017 – Aug 2019
<ul style="list-style-type: none"> • Thesis: Fake News Detection • Advisor: Prof. Pushpak Bhattacharyya, Prof. Asif Ekbal, Prof. Stefan Dietze 	
IGNOU , Masters in Computer Application (3/4)	Jul 2013 – Dec 2016
<ul style="list-style-type: none"> • Thesis: Online Journal Management System • Advisor: Assoc. Prof. Shalab Agarwal 	

Selected Publications

<ul style="list-style-type: none"> • Achieving Socio-Economic Parity through the Lens of EU AI Act A. Roy, S. Rizuo, S. Papadopoulos, E. Ntoutsi 	FACCT 2025
<ul style="list-style-type: none"> • ALMA: Aggregated Lipschitz Maximization Attack on Auto-encoders C. K. Ramanaik, A. Roy, E. Ntoutsi 	Under Review 2025
<ul style="list-style-type: none"> • FairBranch: Mitigating Bias Transfer in Fair Multi-task Learning A. Roy, C. Koutlis, S. Papadopoulos, E. Ntoutsi 	IJCNN 2024
<ul style="list-style-type: none"> • Adversarial Robustness of VAEs across Intersectional Subgroups C. K. Ramanaik, A. Roy, E. Ntoutsi 	BIAS 2024
<ul style="list-style-type: none"> • Exploring Fusion Techniques in Multimodal AI-Based Recruitment S. Swati, A. Roy, E. Ntoutsi 	EWAF 2024
<ul style="list-style-type: none"> • Multi-dimensional discrimination in law and machine learning A. Roy, J. Horstmann, E. Ntoutsi 	FACCT 2023
<ul style="list-style-type: none"> • Learning to teach fairness-aware deep multi-task learning A. Roy, E. Ntoutsi 	ECMLPKDD 2022
<ul style="list-style-type: none"> • MulCoB-MulFaV: Multimodal Content Based Multilingual Fact Verification A. Roy, A. Ekbal 	IJCNN 2021

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.



www.arjunroy.info