Nil-Jana Akpinar

nakpinar.github.io/ niljana.akpinar@gmail.com linkedin.com/nil-jana-akpinar

Senior Applied Scientist at Microsoft working on AI safety and security, with a focus on evaluating the reliability and robustness of large language model systems. Previously worked on Responsible AI at Amazon and LinkedIn. PhD in Statistics and Machine Learning from Carnegie Mellon University.

Experience

Senior Applied Scientist – *Microsoft; Redmond, WA* Aug 2025 – ongoing

AI Safety and Security, AISP | Feb 2026 – ongoing

- Designing framework for estimating production accuracy of safety classifiers under distribution shift using taxonomy-based prevalence estimates and aggregate prediction signals.

E&D, GXP Research / Time and Places | Aug 2025 – Feb 2026

- Developed human-grounded evaluation methodologies and synthetic project-level benchmarks for measuring the performance of LLM-powered planning and productivity systems.

Postdoctoral Scientist – *Amazon AWS, Responsible AI; Seattle, WA* Sept 2023 – Aug 2025

- Evaluated LLM robustness in question-answering using inquiry persona-based perturbations; designed framework for automated evaluation with synthetic inquiry styles
- Led research on trust in LLM-generated summaries, including study design, participant recruitment, and mixed-methods analysis to uncover user perceptions and failure modes
- Developed techniques for LLM-assisted data augmentation and prompt tuning to evaluate and improve performance and fairness in downstream tasks; partly published in NeurIPs 2024 workshops
- Organized NeurIPs 2024 workshop on red-teaming GenAI, FAccT 2024 tutorial, and weekly Responsible AI science meetings, inviting cross-org and external researchers to present
- Mentored interns, contributed to internal tools and code review, and reviewed 15+ papers for FAccT, ICML, EMNLP etc.

AI – Machine Learning Engineering Intern – *LinkedIn, Responsible AI; Sunnyvale, CA* May 2021 – Aug 2021

- Led research on long-term fairness dynamics in connection recommender systems, developing a realistic simulation and theoretical framework to study feedback loops in statistical parity type interventions
- First-authored peer-reviewed paper 'Long-term Dynamics of Fairness Intervention in Connection Recommender Systems'; published at AIES 2022

Fairness and Privacy Research Engineering Intern – *LinkedIn, Anti-abuse; Sunnyvale, CA* May 2020 – Aug 2020

- Designed and implemented a named entity recognition pipeline to identify and redact privacy-sensitive information in user-generated content
- Collaborated with ML scientists and engineers to integrate models into internal tooling

Education

Carnegie Mellon University Aug 2018 - Aug 2023
Ph.D. in Statistics and Machine Learning (joint), GPA 4.0/4.0 *Pittsburgh, PA*

- **Thesis:** The Role of Noise, Proxies, and Dynamics in Algorithmic Fairness
- **Advisors:** Alexandra Chouldechova & Zachary C. Lipton

Carnegie Mellon University Aug 2018 - May 2020
M.Sc. in Statistics, GPA 4.0/4.0 *Pittsburgh, PA*

University of Freiburg Oct 2015 - Jul 2018
M.Sc. in Mathematics *Freiburg, Germany*

University of Freiburg Oct 2013 - Sept 2017
B.sc. in Economics *Freiburg, Germany*
University of Freiburg Oct 2012 - Sept 2015
B.Sc. in Mathematics *Freiburg, Germany*

Selected Publications (Google Scholar)

Under Review

Vivian Lai, Zana Bucinca, **Nil-Jana Akpinar** (2025) *Users Mispredict Their Own Preferences for AI Writing Assistance*. Under Review.

Sandeep Avula, **Nil-Jana Akpinar**, Brandon Dang, Kaza Razat, Vanessa Murdock, Pietro Perona (2025) *Behind the Labels: Content Annotator Perspectives on Resilience, Well-Being, and Work Design*. Under Review.

Chia-Jung Lee*, **Nil-Jana Akpinar***, Vanessa Murdock, Pietro Perona (2025) *Who's Asking? Evaluating LLM Robustness to Inquiry Personas in Factual Question Answering.* Under Review.

Published

Nil-Jana Akpinar, Sandeep Avula, Chia-Jung Lee, Brandon Dang, Kaza Razat, Vanessa Murdock (2025) *LLM or Human? Perceptions of Trust and Information Quality in Research Summaries*. CHI 2026.
Jaqueline Maasch, Violet Chen, Agni Orfanoudaki, **Nil-Jana Akpinar**, Kyra Gan (2025) *Local Causal Discovery for Structural Evidence of Direct Discrimination*. AAAI 2025.

Valeriia Cherepanova, Chia-Jung Lee, **Nil-Jana Akpinar**, Riccardo Fogliato, Martin Andres Bertran, Michael Kearns, James Zou (2025) *Improving LLM Group Fairness on Tabular Data via In-Context Learning*. Workshop paper: Neurips 2024 Safe Generative AI Workshop & Neurips 2024 Table Representation Learning Workshop. Full paper: AIES 2025.

Nil-Jana Akpinar*, Sina Fazelpour* (2025) *Authenticity and exclusion: social media recommendation algorithms and the dynamics of belonging in professional networks*. Workshop paper: MINT-Yale Workshop on Normative Philosophy of Computing 2024. Full paper: Synthese Journal 2025.

Vibhu Sharma, Shantanu Gupta, **Nil-Jana Akpinar**, Zachary Lipton, Liu Leqi (2024) *A Unified Causal Framework for Auditing Recommender Systems*. Workshop paper: RecSys 2024 FAccTRec. Preprint.

Riccardo Fogliato, Pratik Patil, **Nil-Jana Akpinar**, Mathew Monfort (2024) *Precise Model Benchmarking with Only a Few Observations*. EMNLP 2024.

Nil-Jana Akpinar, Zachary C. Lipton, Alexandra Chouldechova (2024) *The Impact of Differential Feature Under-reporting on Algorithmic Fairness.* FAccT 2024.

Nil-Jana Akpinar, Manish Nagireddy, Logan Stapelton, Hao-Fei Cheng, Haiyi Zhu, Steven Wu, Hoda Heidari (2022) *A Sandbox Tool to Bias(Stress)-Test Fairness Algorithms.* Poster: EAAMO 2022 poster. Preprint.

Nil-Jana Akpinar*, Leqi Liu*, Dylan Hadfield-Menell, Zachary C. Lipton (2022) *Counterfactual Metrics for Auditing Black-Box Recommender Systems for Ethical Concerns.* Workshop paper: ICML 2022 Responsible Decision Making in Dynamic Environments.

Nil-Jana Akpinar, Cyrus DiCiccio, Preetam Nandu, Kinjal Basu (2022) *Long-term Dynamics of Fairness Intervention in Connection Recommender Systems.* AIES 2022.

Nil-Jana Akpinar, Maria De-Arteaga, Alexandra Chouldechova (2021) *The effect of differential victim crime reporting on predictive policing systems.* FAccT 2021.

Nil-Jana Akpinar, Aaditya Ramdas, Umut Acar (2020) *Analyzing Student Strategies In Blended Courses Using Clickstream Data.* Conference on Educational Data Mining (EDM) 2020.

Nil-Jana Akpinar, Bernhard Kratzwald, Stefan Feuerriegel (2020) *Sample Complexity Bounds for Recurrent Neural Networks with Application to Combinatorial Graph Problems.* AAAI 2020 student abstract. Preprint. Poster.

Nil-Jana Akpinar, Simon Alfano, Gregory Kersten, Bo Yu (2017) *The Role of Sentiment and Cultural Differences in the Communication Process of e-Negotiations.* Conference on Group Decision and Negotiation (GDN) 2017.

* Equal contribution

Talks, Workshops & Tutorials

Invited talks

- Keynote at BIAS workshop SIGIR 2024, website (2024)
- MILA x Vector Institute DEFirst group, recording (2023)
- INFORMS annual meeting, Session on Advances in Responsible AI: From Theory to Applications (2022)

- LinkedIn Responsible AI team (2022)
- AMS Sectional Meeting on Social Change in and through Mathematics and Education (2021)
- Guest lecture on Word Embeddings in Text Analysis class, CMU (2019)

Accepted Tutorials

Alicia Sagae, **Nil-Jana Akpinar**, Riccardo Fogliato, Mia Mayer, Michael Kearns (2024) *Responsible AI in the Generative Era: Science and Practice*. Tutorial at FAccT 2024.

Accepted Workshops

Valeriia Cherepanova, Bo Li, Niv Cohen, Yifei Wang, Yisen Wang, Avital Shafran, **Nil-Jana Akpinar**, James Zou (2024) *Red Teaming GenAI: What Can We Learn From Adversaries?* Workshop at Neurips 2024.

Honors and Awards

- Amazon Graduate Research Fellowhip (2021)
- Invited to Doctoral Consortium: FAccT (2021, 2022), EAAMO (2022)
- AAAI best three minute student presentation award (2020)
- German National Academic Foundation Scholarship (2013-2018) & Research Visit Grant (2016)
- Economics Award by Südwestmetall (2012)
- Mathematics Award by the German Mathematical Society (2012)

Selected Press

- *Long-term Dynamics of Fairness Intervention in Connection Recommender Systems*, ML@CMU blog, link (2022)
- *Algoritmos de predicción policial: para qué se usan y por qué se ensañan con los maś pobres*, El País, link (2021)
- *Predictive policing is still racist – whatever data it uses*, MIT Technology Review, link (2021)

Teaching and Mentorship

Teaching Assistant, Carnegie Mellon University
- Statistical Graphics and Visualization (Spring 2020), Special Topics: Text Analysis (Fall 2019), Advances Methods for Data Analysis (Spring 2019), Probability Theory for Computer Scientist (Fall 2018)

Data Science Initiative Fellow, Carnegie Mellon University
- Mentored groups of undergraduate students in corporate data science consulting projects.
- Penguin Random House (2019), Giant Eagle (2020)

Teaching Assistant, University of Freiburg
- Management Information Science / Introduction to Programming in R (Spring 2016), Mathematics for Natural Scientists (Fall 2015), Introduction to Programming in C (Spring 2014, 2018), Linear Algebra (Fall 2014)

Service

- **Reviewer and Area Chair:** TMLR (2023), AIES (2023, 2024), ACM Transactions on Recommender Systems (2023), FAccT (2022, 2023, 2024, 2025, 2026), JMLR (2022), Neurips (2021), ICML ML4Data workshop (2021), ICML ethics review (2024, 2025), ACL rolling review (2024, 2025), IRRJ (2024), ICLR Responsible AI workshop (2021, area chair), AAAI (2026)
- **CMU:** Editor ML@CMU blog (2021 – 2023), Organizer Fairness, Ethics, Accountability, Transparency in ML reading group (2021), Board member CMQ+ (2019 – 2020)
- **University of Freiburg:** Three times elected member of faculty council (2015 – 2018), elected member of senate (2015), student council member (2013 – 2018), member of examination board and faculty appointment committees

Last updated: March 2026