Erik Miehling

mail: emiehling@gmail.com

Work Experience

- **IBM Research** *Dublin, Ireland*
  Research Scientist (AI) **Aug 2022 - present**
  - Developed a benchmark for evaluating the steerability of large language models, leveraging stochastic control theory to determine how prompting, fine-tuning, and decoding strategies shape model behavior.
  - Created novel black-box explainability methods for large language models by developing intelligent search algorithms to generate contrastive explanations; applied these methods to explainability in open-text generation, automated red-teaming, and conversational AI (under review at EMNLP 2024).
  - Designed and implemented a light-weight transformer-based classifier for detecting conversational degradation in human-AI conversations; designed custom algorithms for generating synthetic training data using large language models (under review at EMNLP 2024).
  - Developed recommender system simulators to study dynamic fairness in online advertising domains; demonstrated that informational consent decisions have disparate impact on the recommendation accuracy of specific user groups (published at NeurIPS 2023; featured in BBC Science Focus).
- **University of Illinois at Urbana-Champaign** *Urbana, IL*
  Postdoctoral Research Associate **Feb 2018 - Aug 2022**
  - Advised 10 Ph.D. students and published 15 peer-reviewed articles in multi-agent reinforcement learning, stochastic control, and machine learning spanning five federally funded projects (total research funding: $39 500 000 USD).
  - Made foundational contributions to multi-agent reinforcement learning in both cooperative domains (RNN-based information embeddings) and adversarial domains (online attacker intent inference for defense).
  - Co-wrote a successful NSF grant valued at $500 000 USD which funded a three year research program on modeling, learning, and control of epidemic processes.

Education

- **University of Michigan** *Ann Arbor, MI*
  Ph.D. – Electrical Engineering & Computer Science (advisor: Demos Teneketzis) **Sept 2011 – Dec 2017**
- **University of British Columbia** *Vancouver, Canada*
  M.A.Sc. – Electrical & Computer Engineering **Sept 2009 – Aug 2011**
  B.A.Sc. – Electrical Engineering **Sept 2006 – May 2009**

Selected Community Engagements

- **Dublin City University** *Dublin, Ireland*
  Instructor: Hosted a 3-hr coding tutorial/lab on transformer-based models and applications to master's students in the Deep Learning for NLP program. **Apr 2024**
- **Trinity College Dublin** *Dublin, Ireland*
  Mentor: Mentored a team of 11 undergraduate computer science students on building a compliance assistant for the EU AI Act; won the "public choice" award (3rd of 33 teams). **Jan 2024 – Apr 2024**

Internships

- **Oak Ridge National Laboratory** *Oak Ridge, TN*
  Intern: Designed efficient pricing mechanisms for coordinated charging of plug-in electric vehicles **Summer 2013**
- **Defense Research & Development Canada** *Ottawa, Canada*
  Intern: Developed GPU parallelized radar resource management algorithms using Monte-Carlo methods **Summer 2010**
- **Broadcom Canada** *Richmond, Canada*
  Intern: Designed DSP algorithms for voice recognition and filtering **Summer 2008**

Skills

Programming: Python (torch, transformers, langchain), GitHub, MLOps
Theory: probability & statistics, machine learning, RL, optimization & control, game theory
Documentation, presentation, & graphics: LaTeX, Keynote, Adobe Illustrator