Piotr (Peter) Mardziel piotrm@gmail.com, (+1)5088735439,
Citizenship: USA, Poland (EU) piotr.mardziel.com, linkedin, google scholar

Summary: Research software engineer with a PhD in CS from top 15 CS program; 10 years of research experience in programming languages (PL), security, machine learning (ML) safety, privacy, interpretability, and fairness with 5 years of that in post-doctoral and faculty roles at Carnegie Mellon University featuring teaching and advising at the graduate level; and 3 years of professional software engineering experience at a Silicon Valley startup.

2021–now Research & Software Engineer, Truera (Series B Startup)

- Research, design, and implementation of
  - Metrics and software for model monitoring and drift evaluation (Java, Scala, proprietary).
  - Explanations of deep neural networks (Python, Tensorflow, Pytorch, Keras, Trulens on Github); algorithms and visualizations for natural language processing (NLP) explanations and robustness (Python).
    - Leveraged PL and research expertise to apply state-of-the-art approaches for deep neural network interpretabiity to neural vision and NLP models.
  - Library for instrumenting and monitoring LLM (large language model) apps (Python, LangChain, LlamaIndex. Trulens-eval on Github *** 1.5k github stars);
    - Leveraged PL and python expertise to achieve “minimal time to value” – 2 lines of code gets you LLM app monitoring and evaluation.
- Developed and instructed workshop sessions on model monitoring and courses on ML safety.
- Research community service: served on program committee and/or as a peer reviewer for many ML/NLP venues including: AAAI, NeurIPS, ICLR, ACL, EMNLP.

2016–2020 Systems Scientist (special faculty), prior Post-doc, Carnegie Mellon University

- Designed, taught, and/or advised in courses on privacy, security, and fairness in ML. Security and Fairness of Deep Learning ‘20, Foundations of Privacy ‘18.
- Research: AI Safety (see publications below). Advised MS and PhD research projects.
- Research community service: similar venues to the above plus security/privacy venues including: CSF, PETS, S&P/Oakland, CCS. Chair of “Programming Languages and Analysis for Security” 2019.

2015 (obtained) Doctor of Philosophy (PhD), Computer Science, University of Maryland, College Park

Research (links are publications, see also google scholar)

AI Safety: Explaining machine learning methods; defining, discovering, and remedying privacy, security, and fairness violations. NeurIPS’21, NeurIPS’20, ACL’20, AAAI’20, CCS’17. Measuring and alleviating gender bias in natural language models. Chapter’20

Programming languages and static analysis: Secure coding contest: Design and deployment of the Build It, Break It, Fix It contest. CCS’16, CSET’15. Java static analysis: Abstract interpreter and related techniques for java bytecode analysis, integrating numeric and aliasing properties. ESOP’18

PL with security and privacy: Information flow in software systems: with dynamic secrets: POST’17, CSF’15, S&P’14; in secure computation: PLAS’13, PLAS’12; static analysis: BookChapter’20, JCS’13, CSF’11.