Pavel Dolin

pavel.a.dolin@gmail.com

linkedin.com/in/pavel-dolin
{github, gitlab}/izuminka
(908)590-2882

Selected Publications

Statistically Valid Post-Deployment Monitoring Should Be Standard for AI-Based Digital Health
P Dolin, W Li, G Dasarathy, V Berisha
Thirty-Ninth Annual Conference on Neural Information Processing Systems (**NeurIPS 2025**) — **8% acceptance**

ActionReasoningBench: Reasoning about Actions with and without Ramification Constraints
D. Handa*, **P Dolin***, S Kumbhar*, C Baral, T Son
13th International Conference on Learning Representations (**ICLR 2025**), *co-first author

The Art of Defending: A Systematic Evaluation and Analysis of LLM Defense Strategies on Safety and Over-Defensiveness
N Varshney, **P Dolin**, A Seth, C Baral
Findings of the Association for Computational Linguistics (ACL 2024)

Research Experience

Aisera May 2025 - August 2025
AI/ML Research Intern. Manager: Alexander Liss, PhD *Santa Clara, CA*

- Researched and architected a fast inference LLM routing pipeline that reduced baseline serving costs by 50% while maintaining the same accuracy. Led the engineering effort for a production-grade routing pipeline
- Fine-tuned transformer models in PyTorch using PEFT (LoRA/QLoRA, adapters) and selective fine-tuning (layer freezing, last-layer fine-tuning)

ASU ECEE, CHS, Berisha, Dasarathy Groups August 2024 - Present
RA. PIs: Prof. Visar Berisha, Prof. Gautam Dasarathy *Tempe, AZ*

- Investigated post-deployment reliability, robustness, and monitoring of AI models in healthcare applications. Proposed a framework for addressing it, leveraging statistical methods familiar to healthcare professionals, researchers, and regulators. First-author publication
- Investigated the theoretical foundations of data scaling laws. Designed a data- and compute-efficient estimation algorithm for learning curves, with applications to generative AI models. First-author publication in progress.

ASU SCAI, Cognition and Intelligence Group August 2022 - August 2024
RA. PI: Prof. Chitta Baral *Tempe, AZ*

- Investigated the capabilities of LLMs on reasoning about actions, effects, and change. Designed and implemented (Python) the experiments. Evaluated and fine-tuned multiple LLM models (LLaMa, Gemma). Analyzed the results of the experiments. Co-first author of the publication. ICLR 2025
- Evaluated, analyzed, and determined the best LLM defense strategies on safety and over-defensiveness. ACL 2024

Spiketrap (NLP startup acquired by Reddit) April 2021 - August 2022
Researcher. Mentor: Dr. Andrea Vatani *San Francisco, CA*

- Lead research on Twitch chat sentiment detection and neologisms (emotes) understanding. Created an unsupervised framework, outperforming the previous benchmark by 7.36% points. First author publication.
- Lead research on Twitch "Hateful Raids" (cyberharassment phenomenon) detection and resolution. Successfully created a dataset and an ML + rule-based classifier for raid detections and hate bot/user flagging

UCSB Computational Materials June 2017 - June 2019
Researcher. PI: Prof. Anton Van der Ven *Santa Barbara, CA*

- First-principles research of thermodynamics of 36 binary refractory alloys. Performed high-throughput quantum mechanical calculations, trained ML models, and ran Monte Carlo to generate phase diagrams. Journal Publication
- Investigated encoding of chemical and structural information of materials for prediction of properties via ML. Conducted research, data analytics, and statistical analysis

Engineering Experience

Spiketrap (NLP startup acquired by Reddit) Feb 2020 – Aug 2022
ML Engineer (Python + Major ML libraries, Ruby) *San Francisco, CA*

- Received ML Engineer III (IC3) offer from Reddit after Spiketrap's acquisition; declined to begin PhD at ASU.
- Served as a **lead engineer** for end-to-end development and deployment of listed ML projects:
- Topic Quality Project: Designed, trained, and deployed a topic quality model with custom metrics; built a retraining pipeline for drift handling and integrated into production ML infrastructure.
- Twitter Handles Ads: Designed, trained, and deployed an end-to-end ML pipeline using a lexicon- and embedding-based model to identify semantically similar Twitter handles from seed accounts and their followers
- Text Summarization Project: Researched and evaluated models for summarizing user reviews, balancing cost and quality to select optimal architectures. Built a cost-efficient end-to-end pipeline and owned the full lifecycle.

Software Engineer (Ruby on Rails, Python)

- **Lead engineer** for Notifications Project. Collaborated with multiple teams across the company to architect and ship a customizable, real-time notification service for clients. Designed the end-to-end pipeline, databases, REST APIs, architecture of the service, pipelines, fetching, storing, and processing of data. Fully implemented all the backend and was a supervisor for the frontend.

Education

Arizona State University August 2022 – Now
Ph.D. in Computer Science. ***Dean's Fellow****. GPA: 4.0* *Tempe, AZ*
Advised by Visar Berisha & Gautam Dasarathy

University of California, Santa Barbara September 2016 – June 2019
B.S. in Physics. GPA: 3.77 *Santa Barbara, CA*

Awards

NeurIPS 2025 Scholar Award 2025
Financial support for conference registration and hotel stay

Dean's Fellowship *ASU* 2022 – 2024
4% of PhD students receive the award. Provides tuition and $30,000 annually

Fulton Fellowship *ASU* 2022
An award of $1000 in "recognition of extraordinary achievements and support of future research and development."

Distinction in Major and High Honors *UCSB* 2019
Successful defense of honor's thesis and GPA within the 3.50 - 3.79 range

Technical Skills

Languages: Python (6+ yrs), Ruby (3 yrs), C++ (2 yrs), BASH (1+ yrs), SQL (2 yrs), Matlab (2 yrs)
Packages & Frameworks: PyTorch, TensorFlow, Scikit-learn, HuggingFace, NLTK, Pandas, NumPy, SciPy, Rails
Tools & Platforms: Linux, AWS, Git, Docker, MySQL, PostgreSQL, MongoDB, Redis
ML & AI: End-to-end development of predictive models and ML pipelines, large-scale data processing in production systems, training ML and DL models, LLM fine-tuning, PEFT (LoRA/QLoRA, adapters), prompt engineering, reinforcement learning, model evaluation and interpretability, experiment design, A/B testing, expertise in robustness
Mathematics: Strong foundation in linear algebra, probability, statistics, optimization, and numerical methods; expertise in algorithm design, statistical inference, and quantitative problem-solving grounded in first-principles thinking
Communication & Leadership: Strong written and verbal communication; experience in teaching and mentoring, cross-functional collaboration, technical leadership, and product management

Teaching Experience

CSE 310 Data Structures and Algorithms May 2024 – Present
Teaching Assistant (TA) *Tempe, AZ*

- Conducting recitation lectures for a class of 180 students.
- Helping students with HW and coding projects. Graded assignments. Administered Exams.