Peng Wang

Linkedin: https://www.linkedin.com/in/pengw96/
Website: https://peter-peng-w.github.io/
Email : pw7nc@virginia.edu
Mobile : +1-434-328-9834

Education

- **University of Virginia** — Charlottesville, VA
  Ph.D. Student in Computer Science, Advisor: Jing Yang and Cong Shen — *Aug. 2022 – Present*
- **University of Virginia** — Charlottesville, VA
  Master of Science in Computer Science, Advisor: Hongning Wang — *Aug. 2019 – Dec. 2021*
- **Tsinghua University** — Beijing, China
  Bachelor of Engineering in Computer Science and Technology — *Sept. 2014 – Jun. 2018*

Publications

[1] **P. Wang**, Z. Chu, C. Shi, M. Poloczek, C. Shen, and J. Yang, "Skill-coupled policy optimization with calibrated group-wise advantage estimation," *In Submission*,

[2] R. Liu, **P. Wang**, D. Li, C. Shen, and J. Yang, "A shared low-rank adaptation approach to personalized rlhf," *International Conference on Artificial Intelligence and Statistics*, 2025.

[3] L. Fan, **P. Wang**, J. Yang, and C. Shen, *Chain-of-thought enhanced shallow transformers for wireless symbol detection*, 2025. arXiv: 2506.21093 [cs.LG].

[4] S. Wang*, **P. Wang***, T. Zhou, Y. Dong, Z. Tan, and J. Li, "Ceb: Compositional evaluation benchmark for fairness in large language models," *The Thirteenth International Conference on Learning Representations*, 2025, Spotlight Paper.

[5] S. Wang*, **P. Wang***, T. Zhou, *et al.*, "On demonstration selection for improving fairness in language models," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, *Workshop on Socially Responsible Language Modelling Research*, Spotlight Paper, 2024.

[6] **P. Wang**, R. Cai, and H. Wang, "Graph-based extractive explainer for recommendations," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 2163–2171.

RESEARCH INTEREST

- My research interests span various topics in machine learning, including information retrieval, reinforcement learning, and trustworthy AI. Recently, I have been particularly interested in exploring **Reinforcement Post-training** techniques to improve models' reasoning abilities by designing variants of GRPO and building skill-targeted agentic pipelines. Furthermore, I am interested in the **trustworthiness of LLMs**, including their robustness against malicious attacks during instruction tuning and fairness issues in both training-free evaluation (e.g., through in-context learning) and alignment tuning.

Skills Summary

- **Programming Languages**: Adept at Python, C/C++, familiar with Linux, Java, R, SQL
- **Machine Learning**: Adept at PyTorch, familiar with TensorFlow

Technical Research

- **LLM Reasoning** — Charlottesville, USA
  Research Assistant, Directed by Prof. Jing Yang and Prof. Cong Shen, University of Virginia — *Sep. 2024 - Present*

  Compositional GRPO
  - Developed a skill-aware grouping strategy that structures GRPO updates over semantically aligned prompt sets, yielding more reliable optimization and better generalization across diverse reasoning skills.
  - Implemented variance- and coverage-oriented training components, including a James-Stein advantage estimator for robust group baselines and a maximum-entropy adaptive prompt sampler to balance exploration across skill groups for efficient and stable variance-reduced policy optimization.

  PSR Recursive Reasoning Model
  - Formulated stepwise reasoning as a POMDP and introduced a compact belief/state abstraction to shorten trajectories and reduce error propagation from noisy intermediate steps.
  - Distilled structured state representations via integrated forward planning and retrospective self-reflection of the teacher model, then optimized the resulting abstraction with PPO/GRPO to improve reasoning robustness and interpretability.
- **Alignment of LLM** — Charlottesville, USA
  Research Assistant, Directed by Prof. Hongning Wang, University of Virginia/Tsinghua University — *Sep. 2023 - Sep. 2024*

  - Introduced Reward/Advantage-weighted Regression to promote model's alignment performance during both SFT and DPO.
  - Investigating data selection and generation methods that integrate trajectory rewards to enhance multi-step reasoning in formal mathematical proof generation.

- **Fairness in LLM** Charlottesville, USA
  Research Assistant, Directed by Prof. Yangfeng Ji and Prof. Jundong Li, University of Virginia *Jan. 2024 - Present*
  - Developed a synthesized benchmark to assess LLMs' zero-shot and few-shot fairness across various tasks, including stereotype recognition/classification, toxic content generation, and decision-making based on sensitive attributes.
  - Proposed multi-stage clustering strategies to adaptively select in-context demonstrations, improving LLMs/VLMs' group fairness in decision-making tasks in various domains such as EHR.

- **Explainable Recommendation (XRec)** Charlottesville, USA
  Research Assistant, Directed by Prof. Hongning Wang, University of Virginia *Sep. 2020 - May. 2023*
  - Reimplemented baseline models including NRT and Att2Seq and evaluated them on datasets including Yelp and TripAdvisor.
  - Proposed to use graph structure to model the relationship between user, item, attributes and candidate explanations.
  - Leveraged on Graph Attention Network to predict the relevance score of each candidate sentences to form explanations.
  - Conducted data poisoning attacks on matrix-based and neural network-based XRec methods to investigate their robustness.

- **Continual Reinforcement Learning** Los Angeles, USA
  Research Assistant, Directed by Prof. Yan Liu, University of Southern California *Jul. 2018 - Oct. 2018*
  - Reproduced DQN, Double DQN, Duel DQN and Prioritized Experience Replay and evaluated them on Atari games.
  - Implemented various unsupervised representation learning methods to improve the training speed of the current DQN method.
  - Combined DQN with a novel expandable neural network structure to achieve continual RL.

Work Experience

- **Amazon** NYC, USA
  Research Scientist Intern *Jun. 2025 - Sep. 2025*
  - Proposed and developed **CAMPAIGN**, a collaborative multi-agent LLM framework for advertising analytics, enabling tool-integrated multi-step reasoning over large-scale relational advertising databases.
  - Designed SQL/Python-driven tool-use pipelines that join heterogeneous advertiser-, campaign-, and keyword-level tables to construct structured KPI features for causal and counterfactual strategy evaluation.
  - Formulated LLM-based strategy recommendation as a reasoning-over-database problem, leveraging advertiser historical trajectories to generate interpretable bidding and targeting strategies beyond nearest-neighbor heuristics.
  - Built a **diagnosis then factuality verification** pipeline to mitigate numerical, aggregation, attribution, and specification hallucinations in tool-augmented reasoning chains, and benchmarked tool-integrated reasoning across multiple foundation models to characterize failure modes and accuracy–quality trade-offs.

- **Zhipu AI (Z.ai)** Beijing, China
  Machine Learning Engineer Intern, RLHF Group *Jun. 2024 - Aug. 2024*
  - Worked on LLM post-training for Automatic Theorem Proving in Lean.
  - Implemented multiple search strategies including whole-proof sampling, per-step tactic generation via best-first search, and MCTS, which were then used to synthesize theorem proofs to scale up the SFT dataset.
  - Performed both SFT and step-DPO to enhance the base model's reasoning ability on theorem proving.

Service

- Reviewer of ACM TIST, IEEE TCCN, ICLR, KDD, WWW, and AAAI

Teaching

- Guest Lecture, ECE 6501 6501 Reinforcement Learning and LLM, University of Virginia (Fall 2025)
- Graduate Teaching Assistant, CS 6501 Natural Language Processing, University of Virginia (Spring 2025)
- Graduate Teaching Assistant, CS 6762 Signal Processing, Machine Learning and Control, University of Virginia (Fall 2024)