PARSA GHAZVINIAN

PH.D. CANDIDATE, COMPUTER SCIENCE| EFFICIENT & TRUSTWORTHY AI

parsa.ghz10@gmail.com | github.com/parqz | linkedin.com/in/pghazvinian | Atlanta, GA, USA | Google Scholar

SUMMARY

Ph.D. candidate in Computer Science working on **efficient and trustworthy AI**: parameter-efficient fine-tuning (PEFT) of large language models via **low-rank adaptation (LoRA)**, including a controlled study showing that learned rank allocation was never detectably better than a fixed rank at the same parameter budget (under review, ACL Rolling Review); analyzing the evaluation standard used to judge **visual-token pruning in vision-language models**, showing that the commonly used upper bounds can be misleading (paper in preparation); reliability of LLM systems, including authority bias in LLM-based search engines and learning from noisy, subjective human annotations (**two EMNLP 2026 papers**); and memory-efficient encrypted inference via homomorphic encryption (IEEE S&P, ACNS, IEEE TPS). Seeking a **Summer 2027 research internship** in efficient LLMs/VLMs, model optimization, and trustworthy AI.

EDUCATION

PH.D., COMPUTER SCIENCE, GEORGIA STATE UNIVERSITY
Advisors: Dr. Rajshekhar Sunderraman, Dr. Yi Ding
Expected graduation: December 2027 — Spring 2022 – Present

M.SC., COMPUTER SCIENCE, GEORGIA STATE UNIVERSITY (AWARDED EN ROUTE TO PH.D.)
Spring 2022 – Fall 2024

M.SC., COMPUTER ENGINEERING (SOFTWARE), UNIVERSITY OF ZANJAN
Thesis: *FDI Attack Detection Using LSTMs in Smart Grids*| Ranked 1st in cohort — 2017 – 2020

B.SC., COMPUTER ENGINEERING (SOFTWARE), UNIVERSITY OF ZANJAN
Thesis: *Secure DBaaS Model Using NoSQL Databases*| Ranked 2nd in cohort — 2013 – 2017

PUBLICATIONS

- U. Jinadu, **P. Ghazvinian**, A. Budathoki, B. Ampel, R. Sunderraman, Y. Ding. "Authority Bias in Conversational Search Engines for Academic Paper Recommendation." *EMNLP 2026* (Main Conference).
- K. Jacob, **P. Ghazvinian**, U. Jinadu, N. Yang, Y. Ding. "When Disagreement Meets Noise: Noise-Robust Annotator Embeddings for Subjective NLP." *Findings of EMNLP 2026*.
- **P. Ghazvinian**, R. Podschwadt, M. GhasemiGol, D. Takabi. "MOFHEI: Model Optimization Framework for Fast and Efficient Homomorphically Encrypted Neural Network Inference." *IEEE TPS 2024*.
- R. Podschwadt, **P. Ghazvinian**, M. GhasemiGol, D. Takabi. "Memory-Efficient Privacy-Preserving Machine Learning Based on Homomorphic Encryption." *ACNS 2024*.
- **P. Ghazvinian**, R. Podschwadt, M. GhasemiGol, D. Takabi. "Packing-Aware Pruning for Efficient Private Inference Based on Fully Homomorphic Encryption." *IEEE S&P 2023* (extended abstract).
- **P. Ghazvinian**, M. Dhakal, U. Jinadu, Y. Ding, R. Sunderraman. "Does Learned Rank Allocation Help Shared and Standard Low-Rank Adapters? A Controlled Study at Equal Final Rank Budget." *Under review, ACL Rolling Review (Oct 2026 cycle)*.

RESEARCH EXPERIENCE

INSPIRE CENTER & CHAI CENTER, GEORGIA STATE UNIVERSITY — **GRADUATE RESEARCH ASSISTANT**
Atlanta, Georgia — Jan 2022 – Present

- Led a **controlled study of adaptive rank allocation in LoRA** (under review, ARR October 2026): built one masked rank-one framework covering LoRA, AdaLoRA, VB-LoRA and VeRA, and compared seven importance scores against random, uniform and shuffled-rank controls on RoBERTa (GLUE) and instruction-tuned LLaMA-3.1-8B and Qwen2.5-7B, with paired and equivalence (TOST) tests; learned allocation was **never detectably better than fixed rank** at the same budget (within ± 0.5 points on the GLUE average).

- Leading an **analysis of visual-token pruning in vision-language models** (paper in preparation): showed that the answer-searched "upper bounds" used to judge pruning methods can make models report objects absent from the image, and built a held-out evaluation on Qwen2.5-VL, Qwen3-VL and LLaVA-1.5 across GQA, POPE and DocVQA, finding that these bounds **overstate the remaining headroom by about 30 points**.
- Built **MOFHEI**, an end-to-end framework that converts pretrained models into HE-compatible form and applies iterative packing-aware block pruning; reached a **98% pruning ratio** on LeNet, cutting encrypted-inference latency **9.63**× and memory **4.04**× at negligible accuracy loss; released as open source (github.com/inspire-lab/MOFHEI).
- Co-developed a caching and dynamic weight-loading scheme for batch-packed FHE convolutional networks, reordering intra-layer operations to maximize reuse of recurring values; cut peak memory up to **50**× for 13% runtime overhead and ran **4.5**× **faster** than OS-level swapping, enabling encrypted inference on low-cost hardware (ACNS 2024).
- Implemented the Robust InfoNCE (RINCE) contrastive objective behind **NRA-Embed**, after showing that SotA multi-annotator models degrade sharply under label noise; preserves minority perspectives instead of majority-vote aggregation and holds clean-setting accuracy while degrading gracefully as noise rises (code).

DATA MINING LAB, UNIVERSITY OF ZANJAN **GRADUATE RESEARCH ASSISTANT & TEACHING ASSISTANT**

Zanjan, Iran — Sep 2018 – Dec 2020

- Built an LSTM-based false-data-injection (FDI) detection model for smart-grid telemetry, reaching **96.7% detection accuracy** on the IEEE 118-bus test system.
- Teaching Assistant, Machine Learning (45 students/semester); ran lab sessions and authored assignments.

TECHNICAL SKILLS

- **Languages:** Python, C++, Java, SQL, Bash
- **ML Frameworks:** PyTorch, Hugging Face (Transformers, PEFT, Accelerate, Datasets), TensorFlow, Keras, scikit-learn, NumPy
- **LLM / Model Efficiency:** LoRA, AdaLoRA, VB-LoRA, VeRA, adapters, structured & unstructured pruning, quantization, knowledge distillation, KV-cache and token sparsity, instruction tuning, contrastive learning
- **Models:** LLaMA-3.1, Qwen-2.5, LLaVA-1.5, Qwen2.5-VL, Qwen3-VL, RoBERTa, BERT, ViT
- **Infrastructure:** Multi-GPU training (A100, RTX A6000), CUDA, Slurm, Docker, Linux, Git, Weights & Biases
- **Privacy-Preserving ML:** Homomorphic encryption schemes CKKS and BGV (Microsoft SEAL, OpenFHE, IBM HELayers, Concrete), secure model inference
- **Evaluation:** GLUE, GQA, POPE, DocVQA, RefCOCO, ROUGE, SacreBLEU, ablation, significance and equivalence testing (paired t-tests, Holm correction, TOST)

ACADEMIC SERVICE

- **Reviewer**, *IEEE Transactions on Information Forensics & Security* (T-IFS) — reviewed 2 submissions in cryptography and systems security, 2025.
- **Artifact Evaluation Committee**, *Proceedings on Privacy Enhancing Technologies* (PoPETs) 2026 — evaluated research artifacts for functionality and reproducibility across multiple review rounds.
- **Reviewer**, *International Joint Conference on Neural Networks* (IJCNN) 2026 — reviewed 3 submissions.
- **Secondary Reviewer**, *ACL Rolling Review* (ARR) 2026 — co-reviewed 2 NLP submissions with primary reviewer across full iterative review cycles.

HONORS & AWARDS

Student Scholarship Award, Professional Organization of Iranian Americans (POOIA) 2025

Student Travel Grants — IEEE S&P (2023), IEEE TPS (2024) 2023 – 2024