

---

# Mohamad Zamini

---

Laramie, WY 82070

7012157110 | mohamad.zamini@gmail.com

## Summary

---

As a PhD student in Machine Learning and Natural Language Processing, I specialize in visual and text reasoning. I also excel in deploying large language models (LLMs) effectively. My solid foundation in software engineering has empowered me to make meaningful contributions to notable projects such as **Pytorch-Geometric**, **LangChain**, **Open-Assistant**, **Gymnasium**, and **Gorilla**. I excel in collaborative teams and driving projects to success.

05/2025 - 08/2025

**Microsoft** – One Microsoft Way, Redmond, WA

### Data Science Intern – Bing Map Team

- Analyze the mobile telemetry dataset to uncover usage insights.
- Make recommendations to improve user retention, engagement, and satisfaction.
- Convert insights into measurable product metrics.

07/2024 - 09/2024

**Numenta** – 889 Winslow Street, Redwood City, CA

### Machine Learning Intern

- Fine-tuned LLM models, including Mistral, LLaMA, and GPT, leveraging techniques such as activation sparsity and attention sparsity to optimize performance.
- Developed techniques such as KWTa, dynamic context pruning, and KV caching.

06/2022 - 08/2024

**Petrolern** - 5460 Bridge Pointe Dr, Atlanta, GA

### Digital Innovation Intern

- During my internship at Petrolern as a Digital Innovation Intern, I gained experience in both machine learning and data compression techniques
- I developed a semantic compression technique using a deep autoencoder to effectively map data tuples into a lower-dimensional representation
- As a machine learning engineer, I built models for analyzing geothermal data and improved their performance through algorithmic optimization

## Jobs

06/2018 - 08/2019

**Lifeweb** – Tehran, Iran

### NLP Engineer

- Fine-tune models like BART for summarization on Persian text data.
- Implementing Matrix Factorization for topic modeling.
- BiLSTM-CRF Models for sequential tagging.

## Education and Training

---

01/2021-current

**University of Wyoming** Laramie, WY (transferred from University of North Dakota)

Ph.D.: Computer Science

Area of research: Advancing efficient token compression and hierarchical cross-modal attention to boost reasoning accuracy in Multimodal Large Language Models.

Created a new benchmark dataset and developed PDF-LLaVA, a model that outperforms the original LLaVA using only 25% of the tokens. Currently focusing on precise decomposition of occluded shapes to enhance visual question answering within the dataset.

09/2016 - 09/2018	<b>University of Tarbiat Modares</b> Tehran, Iran Master of Science: Information Technology Engineering
09/2014 - 09/2016	<b>University of Science And Culture</b> Tehran, Iran Bachelor of Engineering: Computer Engineering

## Papers

---

- Zamini, M., Shukla, D. (KDD 25 Submission). Q-Route: Quaternion based Actor-Critic for KG Reasoning
- Zamini, M., Shukla, D. (ICCV 25 Submission). PDF: Prune deep and fast LLaVA
- Zamini, M., Reza, H. & Rabiei, M. (2022). A review of knowledge graph completion. Information, 13(8), 396.
- Zamini, M., Montazer, G. Credit card fraud detection using autoencoder based clustering. In 2018 9th International Symposium on Telecommunications (IST), 486-491. IEEE.

## Skills

---

- |   |   |
|---|---|
| <ul style="list-style-type: none"> <li>• Machine Learning</li> <li>• Python</li> <li>• LLM/VLM</li> </ul> | <ul style="list-style-type: none"> <li>• PyTorch</li> <li>• Codebase</li> <li>• ONNX</li> </ul> |
|---|---|

## Teaching Assistant Experiences

---

- Introduction to Programming
- Introduction to Artificial intelligence
- Software Design
- Machine Learning