

## **Profile**

Experienced data scientist with a proven track record in mining open-source software repositories. Extensive experience in business intelligence consulting, financial and investment analysis, and project management across various industries. Familiar with applying machine learning, deep learning, and natural language processing techniques, with proficiency in statistics and operations research, to extract insights and drive data-informed decisions. Recognized for key competencies including:

- Advanced analytical and problem-solving capabilities
- Ability to tackle complex, intellectually demanding problems with minimal guidance
- Rapid learning and adaptability in dynamic environments
- Leadership in cross-functional teams, with strong communication and organizational skills

## **Professional Experience**

### **Graduate Research Assistant**

#### **University of Tennessee Knoxville**

**May 2021 - Present**

- Leveraged natural language processing (NLP) and machine learning (ML) to research open source software supply chains, focusing on copy-based code reuse and its implications for software reliability, and development efficiency.
- Designed an automated source code autocuration technique to improve the quality of pre-training datasets for large language models (LLMs), addressing issues related to bugs, vulnerabilities, and license violations.
- Served as a key contributor to the [World of Code](#) infrastructure, leveraging Bash scripting to enable scalable data curation, augmentation, and repository analysis for over 130M open source software projects.

### **Senior Data Scientist**

#### **Mobile Communications Company of Iran**

**May 2019 - Apr 2020**

- At the ERP department, data warehousing & business intelligence office, bridged the gap between business units and data warehouse teams by acting as a liaison, translating business needs into technical requirements for the design, testing, and development of data warehouse solutions.
- Developed and maintained reports, dashboards, and analyses using oracle business intelligence enterprise.

### **Strategic Investments Lead**

#### **Mobile Communications Company of Iran**

**Feb 2018 - May 2019**

- At the mergers and acquisitions department, feasibility studies office, led a team of five professionals in managing complex investment projects, including business plan development, feasibility analysis, valuation of acquisition targets and due diligence.
- Collaborated with C-level management and the board of directors to negotiate contract terms and cooperation models with potential counterparts, such as strategic partners or business sellers.

### **International Investment Analyst**

#### **Mobile Communications Company of Iran**

**Feb 2016 - Feb 2018**

- At the mergers and acquisitions department, international investments office, developed financial models to project earnings and assess the viability of acquisition opportunities in international markets.
- Screened markets for potential investment targets, interpreting financial statements and other data to evaluate risk and profitability.

## **Languages**

English (Fluent), German (Working Knowledge), Persian (Native)

## Skills

- **Programming Languages** | Bash (Scripting), R, Python, C
- **AI/ML** | TensorFlow, Keras, PyTorch, Scikit-learn, NumPy, Pandas, NLTK
- **Generative AI** | Large Language Models (LLMs), Transformer Architectures, Text Generation, GPT
- **Data Analysis** | SQL, MongoDB, Tableau, Power BI, MATLAB, Matplotlib, Seaborn, ggplot2
- **Big Data** | Hadoop, Apache Spark
- **DevOps & Productivity** | Git, Docker, Jira, MS Project, COMFAR

## Education

### **Ph.D. in Computer Science | May 2021 - May 2025 (Expected)**

University of Tennessee, Knoxville, USA

- Dissertation: Copy-Based Reuse and its Implications in Open Source Software Supply Chains.
- Supervisor: [Dr. Audris Mockus](#)
- GPA: 4.00/4

### **Master of Science in Industrial Engineering | Sep 2011 - Sep 2013**

Sharif University of Technology, Tehran, Iran

- Thesis: The Influence of Information Presentation and Risk Attitude on Asset Allocation in Financial Markets.
- Supervisor: [Dr. S. T. Akhavan Niaki](#)

### **Bachelor of Science in Industrial Engineering | Jan 2007 - Jul 2011**

Mazandaran Institute of Technology, Babol, Iran

- Ranked 6th nationwide in the Industrial Engineering Graduate Admissions Exam.

## Certificates

### **[Deep Learning Specialization](#) by DeepLearning.AI on Coursera | Dec 2024**

1. Structuring Machine Learning Projects, 2. Neural Networks and Deep Learning, 3. Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization, 4. Convolutional Neural Networks, 5. Sequence Models

## Publications

- **Jahanshahi M., Mockus A.** "Cracks in The Stack: Hidden Vulnerabilities and Licensing Risks in LLM Pre-Training Datasets". *Accepted in the Second International Workshop on Large Language Models for Code (LLM4Code)*. 2025. **Won the LLM4Code Best Paper Award.**
- **Jahanshahi, M., Reid, D., & Mockus, A.** "Beyond Dependencies: The Role of Copy-Based Reuse in Open Source Software Development". *Accepted in ACM Transactions on Software Engineering and Methodology (TOSEM)*. 2025.
- **Jahanshahi, M., Reid, D., McDaniel, A., & Mockus, A.** "OSS License Identification at Scale: A Comprehensive Dataset Using World of Code". *Accepted in 2025 IEEE/ACM 22st International Conference on Mining Software Repositories (MSR)*. IEEE, 2025.
- Miller, C., **Jahanshahi, M., Mockus, A., Vasilescu, B., & Kästner, C.** "Understanding the Response to Open-Source Dependency Abandonment in the npm Ecosystem". *Accepted in the 47th International Conference on Software Engineering (ICSE)*. 2025. **Won the ACM SIGSOFT Distinguished Paper Award.**
- Thakur, A., Milewicz, R., **Jahanshahi, M., Paganini, L., Vasilescu, B., & Mockus, A.** "Scientific Open-Source Software Is More Sustainable Than One Might Think!". *Accepted in The ACM International Conference on the Foundations of Software Engineering (FSE)*. 2025.
- **Jahanshahi, M. & Mockus, A.** "Dataset: Copy-based Reuse in Open Source Software". *2024 IEEE/ACM 21st International Conference on Mining Software Repositories (MSR)* (pp. 42-47). IEEE, 2024.
- Reid, D., **Jahanshahi, M., & Mockus, A.** "The extent of orphan vulnerabilities from code reuse in open source software". *Proceedings of the 44th International Conference on Software Engineering (ICSE)* (pp. 2104-2115). 2022. **Nominated for the ACM SIGSOFT Distinguished Paper Award.**
- Lyulina, E., & **Jahanshahi, M.** "Building the collaboration graph of open-source software ecosystem". *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)* (pp. 618-620). IEEE, 2021.