

# Mustakim Shikalgar

+1 480-328-4635 | [✉ shikalgar.mustakim@gmail.com](mailto:shikalgar.mustakim@gmail.com) | [📄 /mustakim-shikalgar](https://github.com/mustakim-shikalgar) | [🌐 /MustakimFS](https://www.linkedin.com/in/mustakimshikalgar)

## EDUCATION

### Arizona State University

Tempe, AZ

Master of Science in Software Engineering; GPA: 3.75/4.0

Expected May 2026

Coursework: Data Processing at Scale, Machine Learning, Data Mining, Security, Algorithms, Semantic Web

## RESEARCH & PUBLICATIONS

### IEEE COMPSAC 2025 (27% acceptance rate)

Toronto, Canada

“Enhanced Tracking of Missing Persons using Knowledge Graph and Ontology Engineering”

Published

## PROFESSIONAL EXPERIENCE

### Software Engineering Lead

Jan 2026 – Present

METY Legal Chatbot (Industry Capstone) - MyEdmaster

Arizona State University

- Led architecture and engineering for a 6-person team building a full-stack AI legal assistant on Django, React, FastAPI, LangGraph, and MongoDB with LangSmith observability
- Designed a dynamic FSPR knowledge profiling system using async background LLM inference to build per-conversation user knowledge profiles, replacing shallow topic mastery tracking
- Rebuilt the LangGraph reasoning pipeline reducing redundant LLM calls and cutting per-query cost by over 80% through strategic routing between gpt-4o and gpt-4o-mini
- Implemented lawyer-style situational probing with gotcha insight surfacing for both parties, rolling conversation summarization, and a guided course system across 9 legal domains

## PROJECTS

### Distributed Key-Value Store | Java, Raft Consensus, gRPC

Jan 2025 – Present

- Architected fault-tolerant distributed KV store using Raft consensus across a 5-node cluster, ensuring partition tolerance
- Engineered gRPC communication with automatic failover, **sub-10ms latency** for strongly consistent reads
- Implemented tunable consistency (CP vs AP trade-offs) using configurable quorum-based reads

### Semiconductor Yield Prediction | Python, scikit-learn, XGBoost, Streamlit

Aug 2025 – Dec 2025

- Built binary classifier: **76% recall, 0.81 ROC-AUC** on imbalanced SECOM dataset (1,567 samples, 590 features, 14:1 ratio)
- Engineered L1 + Random Forest feature selection pipeline (590→113 features), outperforming XGBoost and SMOTE baselines
- Deployed Streamlit app with tunable decision threshold, optimizing for recall based on cost analysis

### Missing Persons Knowledge Graph | React, FastAPI, RDFLib, SPARQL

Aug 2024 – Dec 2024

- Architected knowledge graph integrating 10,000+ NamUs records using semantic web technologies (RDF, SPARQL)
- Replaced GraphDB/Azure backend (\$50/month) with FastAPI + RDFLib, eliminating costs (sub-100ms queries)

### De Bruijn Genome Assembler | Java, Spring Boot, React

Sep 2022 – May 2023

- Built genome assembler: **99.9% coverage** on phiX174 (5,386 bp) using de Bruijn graphs and Eulerian cycle
- Implemented error correction (tip removal, bubble detection) with multi-format support (FASTA, plain reads)

## TECHNICAL SKILLS

**Languages:** Java, Python, JavaScript/TypeScript, C++, SQL

**Frameworks:** Spring Boot, React, Django, FastAPI, PyTorch, TensorFlow, scikit-learn, LangChain, LangGraph

**Systems & Data:** gRPC, Raft, Docker, Kubernetes, MongoDB, PostgreSQL, Neo4j, Redis, SPARQL

**Tools:** AWS, Kafka, Git, Streamlit

## ACHIEVEMENTS

**LeetCode:** 815+ problems solved including 126 Hard (graph algorithms, dynamic programming, system design)

**Mathematics Olympiad:** Silver Medalist