

# Besim Shala

Munich, Germany - [besim\\_shala@outlook.de](mailto:besim_shala@outlook.de) - [linkedin.com/in/besim-shala-b32211250](https://www.linkedin.com/in/besim-shala-b32211250)

## PROFILE

---

AI and cloud-focused Information Systems student with a passion for building solutions that drive real business value. With 3+ years of enterprise experience, I combine hands-on engineering in intelligent automation and AWS architectures with a focus on tangible ROI. Recognized for leading internal AI initiatives. I thrive on translating complex business problems into deployable, scalable systems that deliver measurable efficiency gains.

## EDUCATION

---

**Technical University of Munich (TUM)** Munich, Germany  
Master of Science in Information Systems (120 ECTS) April 2026 – Present  
Current Grade: /

**Munich University of Applied Sciences (HM)** Munich, Germany  
Bachelor of Science in Business Informatics - Information Technology (210 ECTS) Oct 2022 – Feb 2026  
Thesis: Contextual LLM-Based Early Detection System — summa cum laude (1.0)

## EXPERIENCE

---

**Amazon** Munich, Germany  
**Business Analyst - AI & Data Automation** (*Working Student*) April 2026 – Present

- Identification and evaluation of AI use cases within internal applications to optimize operational workflows
- Automation of end-to-end processes within business applications to increase efficiency and reduce manual effort
- Co-leading an internal AI community to promote knowledge sharing and adoption of AI-driven solutions
- Partnering with cross-functional teams to translate operational data into scalable process improvements
- Developing data-driven dashboards and performance monitoring frameworks to support decision-making
- Conducting root cause analyses and contributing to the design of automation-based solutions

**BMW Group** Munich, Germany  
**Applied AI Engineer** (*Working Student*) Feb 2025 – Feb 2026

- Supported a board-level AI transformation initiative targeting multi-million euro operational savings
- Engineered AI pipelines using Python, OpenAI APIs, vector databases, and scalable chunking strategies
- Automated enterprise workflows via n8n orchestration and AI-driven process integration
- Translated business requirements into deployable AI systems in collaboration with technical and domain teams
- Presented AI solutions to senior management and contributed to strategic adoption decisions
- Co-led a 200+ employee AI community, driving internal knowledge transfer and AI best-practice adoption

**BMW Group** Munich, Germany  
**Software & Cloud Engineer** (*Intern*) Sep 2024 – Jan 2025

- Developed a full-stack web application for TP-4 project reporting to centralize and visualize key business KPIs across multiple data sources
- Designed and integrated AWS QuickSight dashboards into an Angular-based frontend and Spring Boot backend for interactive data visualization
- Containerized the application with Docker and deployed it on AWS ECS for scalable and reliable access
- Utilized AWS S3 for data storage and implemented secure data flows between services
- Engineered a private AWS API Gateway in a VPC using Terraform, implementing VPC endpoint architecture, Route 53 custom domain routing, and certificate-based secure access

## Lösch & Partner GmbH (External at BMW Group)

Munich, Germany

### Software Developer (*Working Student*)

Jan 2024 – Aug 2024

- Enhanced the enterprise application Codebeamer by developing a custom Java-based data structure for user management task processing
- Implemented a job queue system that sequentially executes background tasks, replacing error-prone nightly batch triggers
- Improved system stability and reduced job failures by optimizing process synchronization and log handling
- Collaborated within an agile SCRUM team using Jira and Confluence for sprint planning and documentation

## Munich University of Applied Sciences

Munich, Germany

### Software Developer (*Student Assistant*)

Feb 2024 – Mar 2024

- Developed a web-based nursing support application to assist the newly established Faculty of Nursing in managing and locating nearby care institutions
- Implemented a secure user authentication system using Flask and Python
- Built and containerized the application with Docker for portable and scalable deployment
- Integrated external APIs to calculate user distances and automatically identify the nearest nursing institution through a custom distance algorithm

## metafinanz Informationssysteme GmbH

Munich, Germany

### Business Analyst (*Working Student*)

Jan 2023 – Dec 2023

- Developed and maintained web-based applications (JavaScript, CSS, HTML)
- Resolved support requests for digital applications
- Implemented custom code adaptations (JavaScript, CSS)

## PROJECTS

---

### Containerized Microservices Planning Platform

2025

- Designed and implemented a cloud-native microservices architecture for a Jira-like planning system using Angular, Spring Boot, Docker, and MySQL
- Built two independent vertical microservices (Kanban + Settings) with isolated frontend, backend, and database containers orchestrated via Docker Compose
- Implemented secure Single Sign-On authentication with Keycloak and OAuth2 across services
- Developed REST-based horizontal service communication enabling dynamic cross-service data exchange
- Engineered a scalable container architecture supporting independent deployment and extension of services
- Demonstrated domain-driven microservice decomposition and production-ready DevOps setup

### Contextual LLM-Based Sequential Detection System

2026

- Architected an LLM-driven pipeline for long-context discourse modeling and early signal detection
- Implemented hierarchical chunk aggregation for scalable transformer inference on extended sequences
- Designed sliding-window temporal analysis for progressive early-warning classification
- Converted qualitative language signals into quantitative metrics via LLM-as-a-Judge scoring
- Built prompt-engineered inference framework with deterministic JSON evaluation outputs
- Optimized decision thresholds via grid search under precision-recall constraints
- Developed reproducible experimental evaluation protocol for AI system benchmarking

## SKILLS

---

**AI & LLM Engineering:** Prompt engineering, long-context LLM pipelines, chunk-based inference, LLM evaluation frameworks, model calibration, experimental AI benchmarking

**Programming:** Python, Java, TypeScript, JavaScript

**Frameworks:** Spring Boot, Angular, Flask, Docker

**Cloud:** AWS (ECS, Lambda, API Gateway, S3, VPC, Route 53), Terraform, containerized architectures

**Tools:** Git, Jira, Confluence, VS Code, IntelliJ IDEA

**Languages:** German (Native), Albanian (Native), English (C1)