Reza Ghassemzadeh

MACHINE LEARNING & DATA ENGINEER · SAN DIEGO, CA

mrezaghassemzadeh@icloud.com rezaghassemzadeh.com github.com/MrezaGha linkedin.com/in/rezaghassemzadeh

SUMMARY

Machine learning and data engineer finishing an M.S. in Computer Science (Data Science) in December 2026. Builds the computer vision pipeline for a Caltrans-funded intersection safety study at SDSU, and fine-tuned Mistral 7B with LoRA to raise F1 from 0.33 to 0.91 on a product matching task. The M.S. thesis examines whether brain MRI classifiers learn meaningful features or exploit shortcuts in their training data. Before the degree, four years of part-time production engineering at Otto Group one.O and Prinovis (Bertelsmann): SQL pipelines on Oracle, validation checks, automation, monitoring, and Java and .NET applications.

EDUCATION

University of Hamburg — M.S. Computer Science (Data Science) Oct 2024 – Dec 2026 (expected)

San Diego State University — Visiting Graduate Student / Study Abroad Program · GPA 3.9 / 4.0 Jan 2026 – Dec 2026

University of Hamburg — B.S. Computer Science Oct 2020 – Sep 2024

EXPERIENCE

SDSU STAR Lab · Graduate Research Assistant Jun 2026 – Present

- Built the detection and tracking pipeline for a Caltrans-funded intersection safety study: YOLO-based vehicle detection and multi-object tracking across frames of aerial drone footage.
- Computed time-to-collision and post-encroachment safety measures from the extracted trajectories, and analyzed the cases where the two measures disagree.
- Work the drone recording sessions at the study's Southern California field sites, and own the footage from there on: preparation, validation, and the checks that catch unusable input before it reaches the pipeline.
- Document the pipeline and its data handling so other lab members can rerun it on new footage.

Otto Group one.O (Otto Group) · Data & Automation Engineer, Working Student Feb 2023 – Dec 2025

- Built and maintained the SQL views and scheduled data jobs on Oracle behind the team's recurring reporting, and checked results against business expectations before they shipped.
- Traced wrong figures in downstream dashboards back to the view or job that produced them, and fixed the cause rather than the number.
- Rebuilt the team's data definitions and technical documentation from scratch, which resolved inconsistencies between reports that had shown conflicting figures.
- Replaced a manual filing process with a REST API-based archiving service that ingests documents and extracts their metadata automatically.
- Built Power BI dashboards for business stakeholders and walked them through what the data could and could not tell them.
- Automated recurring data processing and monitoring tasks in Python and Power Automate.
- Helped design the company-wide Elastic Stack monitoring setup and rolled it out, covering log processing and alerting.
- Administered Linux and Windows servers, provisioned cloud databases on GCP, and ran server-side performance tests.

Prinovis (Bertelsmann) · Software Developer, Part-Time Sep 2021 – Jan 2023

- Cut version-related errors by roughly half, migrating core production applications from Java 8 to 17 and .NET 4.5 to 4.8, with tests written for each migrated component.
- Refactored the application architecture and the security layer so the codebase could take new features without breaking the old ones.
- Maintained the Gradle build and Jenkins CI pipelines, and supported a vendor selection including the technical interviews.

RESEARCH & PROJECTS

Brain Tumor MRI Classification with Explainable AI · M.S. Thesis, Univ. of Hamburg & SDSU — Jun 2026 – Present

- Investigates whether high accuracy on public brain MRI collections reflects meaningful learning or dataset-specific shortcuts.
- Built and compared CNN and Vision Transformer models in PyTorch under matched training and evaluation protocols, with cross-dataset evaluation.
- Tested Grad-CAM and HiResCAM saliency against expert tumor annotations, to check whether the maps fall on the regions the model actually uses.
- Manuscript in preparation.

Applied LLM Projects · SDSU, CS 668 — Jan 2026 – May 2026

- Fine-tuned Mistral 7B with LoRA for product matching and raised F1 from 0.33 to 0.91.
- Built a RAG assistant over the SDSU graduate bulletin, and a voice-based oral exam agent on the ElevenLabs API that asks the questions and scores the answers.

AI Tender Generation App · Lufthansa Industry Solutions — Sep 2025 – Jan 2026

- Cut manual tender drafting effort by about 70% with a web app on a Python backend and the OpenAI API.
- Added a self-correction loop and automatic scoring that catch weak output before it reaches the user.

Product Matching Framework · B.S. Thesis with Otto Group one.O — Feb 2024 – Aug 2024

- Built a two-stage pipeline: embedding-based retrieval with fine-tuned SentenceTransformers narrows the candidates, then a classifier decides the match. Matched over 75% of products correctly.

SKILLS

Languages: Python, SQL, Java, Bash; C and C++ (project experience); C# and Rust (basic)

Machine Learning: PyTorch, Hugging Face Transformers, scikit-learn, CNNs, Vision Transformers, LoRA fine-tuning, explainable AI (Grad-CAM, HiResCAM)

Computer Vision: YOLO, object detection, multi-object tracking, OpenCV, video pipelines

LLM Systems: OpenAI API, LangChain, RAG, embeddings and retrieval, agentic workflows, Claude Code

Evaluation: metric design, ablations, leakage detection, benchmark protocols, error analysis

Data Engineering: Oracle, PostgreSQL, SQL query optimization, ETL pipelines, scheduled jobs, Apache Spark, data validation, Power BI

Engineering & Platforms: Git, Jenkins CI, Gradle, unit testing, REST APIs, Linux, GCP, Azure, Elastic Stack

HONORS, CERTIFICATIONS & LANGUAGES

Honors: SDSU Global Campus Dean's List, Spring 2026

Certifications: Complete SQL Masterclass, Java Made Easy, Fundamentals of Machine Learning

Languages: English (fluent), German (native), Farsi (native), Spanish (conversational)