

Vaibhav KULKARNI

DATA & ML ENGINEERING LEAD (SYSTEM ARCHITECTURE | MACHINE LEARNING | INFORMATION SECURITY) · PHD (COMPUTER SCIENCE)

Switzerland (C Permit)

✉ vkulkarn@protonmail.com | 🏠 vaibhav90.github.io | 📄 vaibhav-kulkarni-b94891166 | 🎓 Vaibhav KULKARNI



Executive Summary

Leveraging over a decade of experience in data engineering, machine learning, & core infrastructure, I am a hands-on engineering lead adept at aligning technical strategies with business priorities. Currently spearheading a portfolio of data-centric projects at Debiopharm International, transforming data accessibility & automating knowledge extraction to boost efficiency & innovation. Proven track record in building & leading engineering teams to deliver data-driven products from the ground up across supply chain, public transport, medical devices & healthcare sectors. With a strong academic foundation & practical industry experience, I excel in collaborating with internal stakeholders to gather business requirements & managing external collaborators to deliver projects with strict deadlines ensuring platform scalability & driving digital transformation.

Work Experience

Debiopharm International

Lausanne, Switzerland

DATA ENGINEERING TEAM LEAD - RESPONSIBLE FOR THE ENTIRE ENTERPRISE DATA INFRASTRUCTURE & DATA-PLATFORMS

July 2022 - Present

- Led a team to architect & deploy a data infrastructure, integrating large clinical datasets daily from over 200 entities into central-data-warehouse
- Engineered an entity-recognition & search platform by mining scientific publication databases -Automated internal trend-monitoring functions
- Architected large-scale analytical applications & secure data solutions, enabling efficient workflows for clinical data science & biostatistics teams
- Productionised privacy-aware large language model-based applications to streamline document analysis and study protocol writing processes
- Supported compliance team via DevOps-based computerised system validation, infra qualification & security audits -Automated GxP workflows
- Responsible for technical due-diligence of healthcare-startups (technical-stack, software & ML strategy) to aid our VC funds investment decisions

SOPHiA Genetics

Lausanne, Switzerland

SENIOR DATA ENGINEER - RESPONSIBLE FOR NON-REGRESSION SYSTEM CRITICAL FOR NEW PRODUCT & FEATURE RELEASES

April 2021 - June 2022

- Implemented terabyte-scale ETL pipelines to transform raw genomics data in-to analytics-ready format -Significantly lowering processing time
- Developed API endpoints & CLI tools to expose database functionalities to internal teams- Facilitated data democratisation and access control
- Implemented regression testing mechanisms for variant detection bioinformatics pipelines & dashboards delivering real-time insights into KPIs
- Authored formal TLA+ specifications for ETL operations, enabling quantification and benchmarking of logical correctness across the data stack

GenLots

Lausanne, Switzerland

SOFTWARE ENGINEER | RESEARCH & DEVELOPMENT LEAD - RESPONSIBLE FOR ETL PIPELINES & DATABASE OPERATIONS

Sept. 2019 - Mar. 2021

- Developed a microservices architecture, streamlining ingestion, transformation, & delivery of client ERP data to production-grade ML services
- Deployed reinforcement learning-based solution, optimising end-to-end supply chain planning, warehouse management, and CO2 emissions
- Engineered a suite of tools for reporting & monitoring integrated into dashboards delivering real-time business performance metrics to clients
- Directed research projects in collaboration with Swiss watch manufacturing firms to predict inbound material needs & stabilise supply chains

Distributed Object Programming Lab

Lausanne, Switzerland

RESEARCH SCIENTIST

Nov. 2015 - Aug. 2019

- Partnered with public transport firms to architect & implement systems capturing high-frequency, real-time ridership data into a data warehouse
- Developed a Machine Learning-driven platform to proactively detect and reduce ticketless travel across the Lausanne public transport network
- Led platform development for a large-scale (>300 participants) spatiotemporal mobility-data collection project to facilitate ML-privacy research

ETH Zürich

Zürich, Switzerland

PROJECT ENGINEER

Jan. 2015 - Aug. 2015

- Implemented a software-defined-radio-based platform for reliable IoT product testing, generating interference patterns of wireless appliances
- Deployed the platform in production at ETH Zurich & TU Berlin with toolsets allowing for remote configuration & logging performance metrics

BOLT IoT

India

EMBEDDED SYSTEMS ENGINEER

Aug. 2011 - Aug. 2013

- Designed PCB layouts for embedded platforms used in robotics & IoT systems & collaborated with software teams to ensure firmware integration
- Implemented toolsets & applications to benchmark platform performance, validation tests, EMI simulations & industry compliance certifications

Education

University of Lausanne

Lausanne, Switzerland

PHD, COMPUTER SCIENCE (THESIS: INFORMATION SYSTEMS FOR PRIVACY-AWARE MACHINE LEARNING)

Nov. 2015 - Aug. 2019

Technische Universiteit Eindhoven

Eindhoven, Netherlands

MS, EMBEDDED SYSTEMS (GRADUATED WITH CUM LAUDE)

Aug. 2014 - Aug. 2015

Technische Universität Berlin

Berlin, Germany

MS, INFORMATION & COMMUNICATION TECHNOLOGY (GRADUATED WITH HONOURS)

Feb. 2013 - Aug. 2014

Goa University

Goa, India

BS, ELECTRONICS & TELECOMMUNICATION ENGINEERING (GRADUATED WITH HONOURS)

June 2008 - May 2012

Skills & Competencies

- **software & design:** Python, Java, Scala and Rust. Skills: object-oriented-programming, test-driven development & software design patterns
- **database management:** PostgreSQL, MySQL, MongoDB, Cosmos DB, Redis. skills: data modeling, query optimisation, replication & migration
- **data engineering:** architecting ETL pipelines (Azure Data Factory, Azure Synapse Analytics, Data Build Tool (DBT), Apache Airflow, & Databricks)
- **infrastructure & services:** Docker, Kubernetes, Kafka, Azure and AWS. skills: containerisation, orchestration, on-prem monitoring with Zabbix
- **machine learning & data science:** NumPy, scikit-learn, TensorFlow, PyTorch. skills: data cleaning & quality patterns, DL-models and validation
- **observability & monitoring:** Elasticsearch-Kibana-Logstash, Prometheus, Grafana. skills: logging, traceability, alerting and anomaly detection
- **devops & versioning:** Azure DevOps, GitHub Actions, Terraform and Ansible. skills: release management, pipeline automation (CI/CD) and IaC
- **data visualization:** matplotlib, Tableau, Dash, Power BI, SAS visual analytics. skills: interactive dashboards, data wrangling and transformation
- **security:** Azure DevOps security, Darktrace and Firewall configuration. skills: security policies, application & infrastructure security procedures
- **LLM (GPT)-based app development:** OpenAI APIs, Vector Databases & Langchain. skills: fine-tuning & training techniques, prompt engineering
- **project management:** Atlassian Suite (JIRA, Confluence and Bitbucket). skills: agile methodologies, sprint planning and backlog prioritization
- **languages:** English (fluent), French (working proficiency, B2), German (elementary proficiency, A2), Hindi (Fluent), Marathi and Konkani (Native)

Publications, Open Source Projects, & Datasets

- **Generative models for simulating mobility trajectories** (collaboration with Statistics Department at Columbia University)
V. KULKARNI, N. TAGASOVSKA, T. VATTER, B. GARBINATO (published at: *Neural Information Processing Systems*, Montreal, 2018)
Impact: Advanced state-of-the-art results for synthetic data generation using ML (Ranked 1 Conference in AI)
- **Breadcrumbs: A mobility dataset with point-of-interest annotations** (collaboration with Information Security-Privacy Lab)
A. MORO, V. KULKARNI, PA GHIRINGHELLI, B. CHAPUIS, K. HUGUENIN, B. GARBINATO (published at: *SigSpatial*, Chicago, 2019)
Impact: Large-scale geospatial dataset to facilitate ML research (Ranked 1 Conference in Geo Information Systems)
- **Controlled interference generation for wireless coexistence research** (ETH Zurich - Distributed Systems Research Group)
A. HITHNAWI, V. KULKARNI, S. LI, H. SHAFAGH (published at: *Implementation-Mobile Computing and Networking*, Paris, 2015)
Impact: open source project for a complete software-defined-radio platform for controllable interference generation
- **Information Disclosure in Location-based services: An Extended Privacy Calculus Model** (collaboration with BISA Lab)
D. NAOUS, V. KULKARNI, C. LEGNER, B. GARBINATO (published at: *International Conference on Information Systems*, 2019)
Impact: Integrate user information disclosure behaviours in Apps (Ranked 1 Conference in Information Systems)
- **Privacy-Preserving Location-Based Services by using Intel Secure Guard Extensions** V. KULKARNI, B. CHAPUIS, B. GARBINATO
Open-source software project, (published at: *International Conference on Embedded Networked Systems*, Delft, 2017)
Impact: Demonstrated the ability of hardware-based execution environments to deploy privacy-aware ML applications

Continuous Learning & Certifications

Data Engineering with Rust

Duke University, June 2024

Generative AI with Large Language Models

DeepLearning.AI & AWS, July 2023

Data Engineering with Azure

Udemy, Jan. 2023

Applied Machine Learning & Data Science

University of Michigan, Apr 2021

Amazon Web Services

AWS, May 2020

Information Security & Privacy

University College London, Jan 2019

Fundamentals of Machine Learning

Stanford University, Jan 2017

Talks, Grants & Awards

- Data Repository: Making Informed Technology Choices - BioTechX Europe (Premier Healthcare Tech Conference)
- Expert Viewpoint on AI's Role in Data Space (Swiss Association of Actuaries)
- Generative AI Report for Healthcare (AI for Health)
- HEC-research grant for independent doctoral research by Swiss National Fund
- Zeno Karl Schindler award for master thesis at ETH Zürich
- EIT-ICT Mobility award best business plan in IoT domain

Extracurricular Activities

- Launched a technology outreach initiative aimed at fostering a passion for science and technology among rural Indian students. Personally visited 25 schools, engaging over 400 students with hands-on demonstrations and presentations on real-world scientific applications
- Competed in multiple endurance races - Lausanne Marathon ('18, '19), Kilomètre Vertical de Fully, & 20KM de Lausanne ('16-'24)
- Played for 'Flyhigh Lausanne' in Ultimate Frisbee (2018, 2019) and actively participated in Padel Tennis tournaments in Switzerland