



LUKE SCHUMACHER

AI Researcher, Data Engineer | ML Pipelines & Advanced Analytics

CONTACT

- Tallinn, Estonia
- +372 5556 8163
- luke.schumacher777@gmail.com
- luke-schumacher.github.io/Portfolio
- github.com/luke-schumacher
- linkedin.com/in/luke-schumacher

LANGUAGES

English	Native/C2
German	Native
French	A2 - Conversational

CORE SKILLS

Data Engineering

Python • PySpark • SQL • Pandas • NumPy • Airflow

ML & Analytics

TensorFlow • PyTorch • Scikit-learn • XGBoost • Predictive Modeling • Matplotlib

Cloud Platforms

AWS (Certified) • Azure • GCP • Databricks • Cloud-native Architecture

DevOps & MLOps

Docker • Kubernetes • CI/CD • MLflow • Weights & Biases • Terraform

Generative AI

LangChain • Transformers • Multi-Agent Systems • Neurosymbolic AI • LLM Applications

Software Engineering

Java • C# • JavaScript • Git • OOP • Clean Code

Collaboration

Agile/Scrum • Stakeholder Mgmt • Technical Presentations • JIRA • Cross-functional Teams

PROFESSIONAL PROFILE

I am a Data Engineer and AI Specialist with a strong international background spanning the Middle East, Europe, and Asia. I hold a bachelor's degree in CMD & IT and am completing a master's in AI for Sustainable Societies through the prestigious Erasmus Mundus program, with academic experiences in Tallinn, Porto, and Tampere. My expertise lies in designing scalable data pipelines for machine learning, deploying predictive models in production environments, and translating complex analytical findings into business impact. I combine deep proficiency in the Python/PySpark data stack with cloud-native engineering practices and hands-on MLOps experience. I thrive in cross-functional, high-performance teams and bring the technical rigour, client-facing communication skills, and relentless curiosity required to solve the hardest data problems at scale.

WORK EXPERIENCE

Working Student – DS & ML Engineering

Siemens Healthineers

Oct 2023 – Present

Germany (Remote)

Architecting and maintaining end-to-end data pipelines (Python, SQL, PySpark) that power predictive digital twin models for a global client base. Ensuring data quality, transforming operational data at scale, and deploying ML models within containerized environments using CI/CD best practices. Presenting analytical findings and strategic recommendations to senior management. Coordinating across data science, engineering, and business stakeholders in agile Scrum teams. Won **First Place at the ITT Summit** for innovative solution design.

AI Research Intern – Master's Thesis

Illigo, Deutsche Telekom & Siemens Healthineers

Dec 2024 – Present

Remote from Estonia

Conducting master's thesis on **Autonomous Protocols, Neurosymbolic Reliability, and Hierarchical Governance in Critical Infrastructure MAS**. Building distributed multi-agent systems in Python with generative AI components. Designing reproducible experiment pipelines and evaluating frameworks for robust autonomous decision-making. Defense planned May 2026.

Data Science & ML Engineering Intern

Infoteam Software AG

Jan 2022 – Oct 2023

Germany

Built and deployed ML-powered data pipelines for industrial clients, transforming raw production data into predictive models for process optimization and quality control. Applied clean code principles, conducted code reviews, and contributed to architectural decisions within agile teams. Translated business requirements into scalable, reproducible technical solutions.

Cloud Infrastructure Intern

Zero & One – AWS Partner

Jan 2021 – Jan 2022

UAE

Deployed and managed scalable AWS cloud services (EC2, S3, Lambda) for diverse client projects. Built automated CI/CD pipelines and infrastructure-as-code solutions using Terraform. Earned **AWS Cloud Practitioner Certification**, validating cloud-native engineering fundamentals.

CERTIFICATES & AWARDS

AWS Certified Cloud Practitioner

Amazon Web Services

ITT Summit – Siemens Healthineers

First Place – Innovative Solution Design

KEY HIGHLIGHTS

- Production ML pipelines (Python, PySpark, SQL)
- Cloud-native deployment (AWS, Azure, Docker, K8s)
- GenAI research (Multi-Agent Systems, LLMs)
- Native German & English; 3 continents, 5 countries

EDUCATION

Master of Science – AI for Sustainable Societies Aug 2024 - Present

University of Tallinn, Lusófona and Tampere

Tallinn, Porto, Tampere

Prestigious Erasmus Mundus program focusing on applying AI techniques to sustainable development and societal challenges. **Current focus:** Master's thesis on Autonomous Protocols and Neurosymbolic Reliability in Multi-Agent Systems for critical infrastructure, in collaboration with Deutsche Telekom/Illigo and Siemens Healthineers. Expected defense: May 2026.

Bachelor – Communication & Multimedia Design Aug 2019 - Aug 2024

Hanze University of Applied Sciences

Groningen, Netherlands

Specialization in interactive design, user experience, and digital technologies with focus on human-computer interaction and data-driven design solutions.

Minor – Artificial Intelligence & Machine Learning Feb - Sep 2022

Hanyang University

Seoul, South Korea

Top 5% of class in Machine Learning and AI+X: Introduction to Data Science courses. Completed final research project applying machine learning to understand patterns of stock market fluctuations across different industries.

PROFESSIONAL COMPETENCIES

Data Engineering: Scalable pipeline design • Data quality assurance • ETL/ELT workflows • Distributed computing • Cloud architecture • MLOps

Impact & Communication: Client-facing presentations • Stakeholder management • Translating technical complexity for non-technical audiences • Advisory mindset

Leadership: Scrum planning • Sprint facilitation • Team coordination • Mentoring • Project ownership • Competition winner

Growth Mindset: Resilience • Continuous learning • Intellectual curiosity • High-performance culture • Adaptability across 3 continents

REFERENCES

Georg Goertler – Project Manager, Siemens Healthineers

✉ georg.goertler@siemens-healthineers.com

“Available for a personal reference regarding: My contributions as a working student, data engineering and ML project delivery, cross-functional collaboration, and performance within the Siemens Healthineers team.”

Dr. Stefano Signoriello – Lead Data Scientist, Infoteam Software AG

✉ Stefano.Signoriello@infoteam.de • www.linkedin.com/in/dr-stefano-signoriello

“Reference regarding: My technical contributions to R&D and critical data infrastructure. Can attest to my rigorous engineering standards, expertise in complex data pipelines, and ability to handle high-stakes technical challenges with outstanding precision and endurance.”

Iolandá Vehlo – Assistant Professor, Universidade Lusófona

✉ iolandavelho@gmail.com • www.linkedin.com/in/iolandavelho-9a594a46

“Luke is one of my most outstanding students... clearly placing him among the top-tier Master’s students. Luke’s strong interest, combined with his solid programming skills, has contributed significantly to the success of his academic path. [He has] proven to be an outstanding student, delivering projects of a very high level.”

Casey Bennett – Associate Professor, Hanyang University & DePaul University

✉ cabennet@hanyang.ac.kr • www.caseybennett.com

“Luke finished among the top 5% of students in both my Machine Learning and Data Science courses. He stood out for his creative contributions to classroom discussions, often relating abstract concepts to everyday life experiences. Luke demonstrated strong technical skills, excellent leadership in group projects, and has consistently gone above and beyond expectations. I believe he represents immense potential as a future data science professional.”