

Leo J. Paggen

Maastricht, Netherlands · +32 .. · l.paggen@student.maastrichtuniversity.nl

linkedin.com/in/lpaggen · github.com/lpaggen

Summary

Research-oriented Data Scientist and Engineer with experience extending Amazon production forecasting systems, developing AI infrastructure, and building symbolic verification tools. Interests include statistical machine learning, reliable AI, time-series forecasting, and autonomous intelligent systems.

Education

M.Sc. Data Science

Maastricht University, Netherlands

Sep 2024 – Sep 2026 (Expected)

- Thesis: Designing an LLM-Assisted Maritime Autonomy Platform for Autonomous Surface Vessels
- Machine Learning, Statistical Modeling, Time Series Analysis, Optimization, Big Data Systems

Exchange Semester Economics

Sungkyunkwan University, South Korea

Feb 2023 – Jun 2023

Technical Skills

- **Programming:** Python, Java, SQL, Rust
- **Machine Learning:** Time Series Forecasting, Statistical Modeling, Large Language Models, NLP, Supervised Fine-Tuning, Uncertainty Quantification
- **AI and Systems:** vLLM, Z3 SMT Solver, Static Program Analysis, ReAct Agents

Experience

Data Scientist Intern, QuBIT Midas

Amazon, Seattle, WA

Jul 2025 – Oct 2025

- Extended a production demand forecasting system supporting approximately \$48M in annual cost savings, improving inbound fulfillment center volume forecasting for North American operations.
- Independently owned three forecasting research initiatives spanning uncertainty estimation, recurring event modeling, and operational disruption modeling, designing and experimentally evaluating statistical improvements to minimize Weighted Absolute Percentage Error (WAPE).
- Developed a bootstrap-based uncertainty estimation framework and implemented a multiprocessing pipeline achieving up to $4.7\times$ speedup over sequential execution, enabling practical large-scale statistical inference for production forecasting.
- Designed and validated statistically distinct forecasting overlays for recurring holidays and fulfillment yard closures, reducing WAPE by up to 6.8% on evaluation datasets through configurable historical weighting and warehouse-specific calibration.
- Extended Amazon Redshift data pipelines and SQL workflows supporting new forecasting methodologies, integrating additional operational data into the existing production forecasting framework.
- Authored a 10-page internal research paper documenting methodology, experimental evaluation, and implementation recommendations, and presented findings to a multidisciplinary team of scientists and engineers.

Data Consultant Intern

PwC, Brussels, Belgium

Nov 2025 – Jan 2026

- Developed an AI-assisted classification pipeline supporting compliance with the EU Deforestation Regulation (EUDR) by mapping client product descriptions to 4-digit Combined Nomenclature (CN) commodity codes.
- Experimentally evaluated multiple retrieval and language-model approaches, pivoting from embedding-based semantic retrieval to frontier LLMs after comparative performance analysis.
- Designed a two-stage inference pipeline in which an independent language model validated initial classifications, improving recall from 80% to 88% on internal evaluation datasets.
- Delivered a proof-of-concept solution while navigating enterprise infrastructure and hardware constraints, documenting recommendations for future deployment.

Teaching Assistant

Maastricht University

Jan 2025 – Present

- Assisted teaching in Natural Language Processing, Principles of Programming Languages, and Linear Programming.
- Mentored students in machine learning, optimization, debugging, and software engineering concepts.

- Guided students through programming assignments, debugging sessions, and optimization exercises.

Selected Research and Software Engineering Projects

Static Verification Framework for PyTorch Tensor Programs — In Development Python, Rust

- Developing a cross-language static verification framework for PyTorch tensor programs using symbolic constraint solving and SMT-based reasoning.
- Built a Python frontend that lowers the Python AST into a semantic intermediate representation with source-span metadata for precise diagnostics, serialized through a custom Protobuf (Google Protocol Buffers) schema.
- Developing a Rust analysis engine with cross-module import resolution and Z3-based constraint propagation across tensor operations.
- Propagates symbolic tensor constraints across program execution, reporting shape inconsistencies only when accumulated symbolic constraints become unsatisfiable.
- Previously developed a complete compiler pipeline—including lexer, parser, semantic analysis, and type checking—for a custom programming language, providing the foundation for the current PyTorch verification framework.

M.S. Thesis: LLM-Based Maritime Autonomy Platform - In Development Python, C++

- Designing an AI-assisted autonomy layer for MOOS-IvP that enables natural-language reasoning while preserving deterministic vessel control.
- Developed a custom C++/Python interface that streams real-time vessel state—including AIS contacts, radar information, and navigation context—into an LLM reasoning pipeline.
- Implemented a ReAct-style navigation agent that interprets vessel state and generates COLREG-compliant navigation actions through a constrained tool interface.
- Generated synthetic maritime navigation data using simulation and fine-tuned open-weight language models (Mistral-7B and Qwen-8B) with supervised fine-tuning (SFT), significantly improving the reliability of structured JSON command generation.
- Architected the platform for future expansion into modular navigation, communication, and mission-planning agents.
- Supports concurrent execution of multiple local language models through vLLM to enable modular agent orchestration.

Cisco ACIBILIZER (Open Source) Python

- Developed an open-source tool converting Cisco ACI configurations into reproducible Ansible automation.

Quantum Technology Forecasting and Collaboration Analysis (Under Peer Review)

- Co-authored a journal manuscript currently under peer review in forecasting global quantum technology trends using publications, patents, funding, and collaboration networks. Journal: Technological Forecasting and Social Change.
- Designed and implemented the collaboration network analysis, constructing research graphs from 25,000+ scientific publications and 9,000+ standardized institutions.
- Developed an institution canonicalization pipeline combining LLM-assisted normalization, rule-based processing, and manual validation to resolve heterogeneous affiliation records into consistent organizational identities.
- Applied network science techniques—including weighted collaboration graphs, betweenness centrality, Newman edge weighting, and network efficiency analysis—to identify structural patterns in the global quantum research ecosystem.
- Co-authored the methodology and interpretation of the collaboration analysis under faculty supervision.