Leonardo Kuffo

Amsterdam, The Netherlands | lxkr@cwi.nl | lkuffo.com | github.com/lkuffo

I am a PhD researcher in the Database Architectures group @ CWI. I am working on improving the efficiency of vector indexing and (filtered) search in data systems. My other research interests are data compression and SIMD programming.

Experience

Centrum Wiskunde & Informatica — Amsterdam, The Netherlands　　　　　　2023 – Present
PhD. Researcher @ Database Architectures Group
· Created ALP, an algorithm to compress floats, now used by DuckDB, Vortex, and (soon) Parquet.

· Created PDX, a data layout for vector embeddings that accelerates vector search by up to 20x.

· Created SuperKMeans to accelerate vector embeddings indexing by up to 10x.

Google — Zurich, Switzerland　　　　　　Jun'25–Dec'25
Research Student @ Information Retrieval Team
· Created Semantic Recall, a new metric to assess vector search retrieval quality.

Amazon — Luxembourg　　　　　　Jun'22–Nov'22
Data Engineer Intern @ Global Transportation Team
· Engineered an architecture in AWS Lambda to monitor dozens of Redshift clusters.

· Added batch processing functionalities on ETL pipelines running in EMR Spark.

Shippify Inc. — Guayaquil, Ecuador　　　　　　Oct'18–Oct'21
Back-End Engineer & Lead Engineer
· Scaled systems to process millions of deliveries and thousands of route creations per day.

CERN — Geneva, Switzerland　　　　　　Jul'18–Sep'18
Reserch Intern @ OpenLab
· Developed ML models to find workload similarities using CPU performance metrics.

Publications

2026

A Super Fast K-means for Vector Embeddings. **VLDB'26**. *Under review*
Kuffo, L., Hepkema, S., & Boncz, P.

Semantic Recall for Vector Search. **SIGIR'26**. *Under review*
Kuffo, L., Tsakalidou, I., De Viti, R., Angel A., Jiri I., & Lendhardt R.

2025

Bang for the Buck: Vector Search on Cloud CPUs. **DaMoN'25.** 🏆 *Best Paper Award*
Kuffo, L., & Boncz, P.

PDX: A Data Layout for Vector Similarity Search. **SIGMOD'25**.
Kuffo, L., Krippner, E., & Boncz, P.

Towards Efficient Vector Similarity Search in Analytical Databases.. **DBDBD'25**.
van Noort, S., Kuffo, L., & Boncz, P.

2024

ALP: Adaptive Lossless floating-Point Compression. **SIGMOD'24.** 🏆 *Best Artifact Award*
Afroozeh, A., Kuffo, L., & Boncz, P.

Towards a Vertical Layout for Vector Similarity Search. **DBDBD'24**.
Kuffo, L., & Boncz, P.

Earlier Carreer

Know your customer: Detection of CX in Social Platforms using Text Categorization. **IEEE Big Data '18.**
Kuffo, L., Vaca, C., Izquierdo, E., & Bustamante, J. C.

Mining Worldwide Entrepreneurs Psycholinguistic Dimensions from Twitter. **ICEDEG'18**.
Kuffo, L., Vaca, C., Izquierdo, E., & Bustamante, J. C.

Back to #6d: Predicting Venezuelan states political election results through Twitter. **ICEDEG'17**.
Castro, R., Kuffo, L., & Vaca, C.

Talks

Juggling High-Dimensional Vectors in the AI-era.
@ Bytedance.

ALP: Adaptive Lossless floating-Point Compression.
@ CWI, Meta.

Education

PhD Computer Science — *Vrije Universiteit Amsterdam.* Amsterdam, NL.		2023–Present
MSc Computer Science — *Vrije Universiteit Amsterdam.* Amsterdam, NL.		2021–2023
BSc Computer Science — *Escuela Superior Politécnica del Litoral.* Guayaquil, EC.		2014 –2019

Beyond my Experience

Volunteering: CIDR'23, ICDE'24, CIDR'25.

Organization: The Dutch Belgian Database Day '24, The Dutch Seminar on Data Systems Design (2024, 2025, 2026).

TA: Data Preparation @ Universiteit van Amsterdam (2026), Exploratory Data Analysis @ ESPOL (2017), Programming Fundamentals @ ESPOL (2015, 2016, 2017)

Languages: Spanish (Native), English (Fluent), Dutch (B1).

Udemy professor with over 12K students. I teach Web Scraping, Data Analysis, and Data Visualization techniques.

Computer Science professor on YouTube with over 1.5M views.