

Rinor Cakaj

PHD CANDIDATE · DEEP LEARNING

✉ cakajrinor@gmail.com | 📅 January 28st, 1998 | 🌐 www.rinorcakaj.com | 📷 rinorcakaj | 🌐 rinorcakaj

Summary

Researcher in Deep Learning with a focus on improving Convolutional Neural Networks (CNNs) performance and a deep understanding of advanced architectures, including Transformers like Vision Transformers (ViT) and Large Language Models (LLMs). I bring over 4 years of Python experience and more than 3 years in PyTorch, with a solid foundation in TensorFlow. With five published papers and five patent submissions, I am excited to contribute my expertise to impactful AI projects.

Education

University of Stuttgart

Stuttgart, Germany

PhD Candidate in Deep Learning | Institute of Signal Processing and System Theory

Sep. 2021 - Present

- Focused on improving CNN performance for computer vision.
- Developed the Mixture-of-Depths (MoD) technique for CNNs, optimizing computational efficiency by selectively processing relevant channels in feature maps (accepted at ACCV 2024). Key performance highlights:
 - Image Recognition (ImageNet): ResNet75-MoD matches ResNet50's accuracy with a 25% speed-up on CPU and 15% on GPU; ResNet86-MoD improves accuracy by 0.45%, with a 6% speed-up on CPU and 5% on GPU.
 - Semantic Segmentation (Cityscapes): FCN with ResNet86-MoD increases mIoU by 0.95% with similar computational costs.
 - Object Detection (Pascal VOC): Faster-RCNN with ResNet86-MoD raises mAP by 0.37% and AP50 by 0.4%, achieving a 10% CPU speed-up.
- Working with the Robert Bosch GmbH, under the supervision of Prof. Dr.-Eng. Bin Yang.

Technical University (TU) of Darmstadt

Darmstadt, Germany

M.Sc. Mathematics interdisciplinary Informatics

Oct. 2019 - Aug. 2021

- Advanced courses in mathematical optimization (Discrete Optimization, Nonlinear Optimization) and computer science (Statistical Machine Learning, Data Mining and Machine Learning, Artificial Intelligence).
- Minor in economics (Creating a Web Startup, Digital Transformations and Software & Digital Business).
- Master thesis: Exact Rule Learning via Boolean Compressed Sensing (1,3).
- GPA (German): 1,25

Technical University (TU) of Darmstadt

Darmstadt, Germany

B.Sc. Mathematics with Economics bilingual

Oct. 2016 - Sep. 2019

- Fundamentals in a broad range of mathematical areas (Analysis, Linear Algebra, Optimization, Stochastics, Numerical Analysis).
- Minor in economics (Micro- and Macroeconomics, Bookkeeping and Cost Accounting, Principles of Business Administration).
- Minor in computer science (Functional and Object-oriented Programming Concepts and Algorithms and Data Structures).
- Topic of bachelor thesis: Portfolio Optimization (1,3).
- GPA (German): 2,20

Work Experience

Quality Match GmbH

Heidelberg, Germany

Solution Architect

Mar. 2025 – Present

- Developing AI-driven techniques to automatically evaluate and improve data quality.
- Designing and building Nano-Task Trees and pipelines for efficient data annotation.
- Implementing a LLM-based solution that automatically designs Nano-Tasks.

Fletza

Stuttgart, Germany

Co-Founder & Full Stack Developer

Jan. 2025 – Present

- An AI-driven platform that transforms lecture materials (PDF, Word, PPTX) into flashcards, quizzes, and exam questions.
- The platform uses the SM-2 algorithm (similar to Anki) for effective learning and offers exam simulations with automatic feedback.
- Learn more at www.fletza.com.

Robert Bosch GmbH

Stuttgart, Germany

PhD Candidate in Deep Learning | Division Cross-Computing Solutions (XC) in the field for

Sep. 2021 - Feb. 2025

Advanced Driver Assistance Systems (ADAS)

- Gained a deep understanding of advanced deep learning architectures, including Convolutional Neural Networks (CNNs), Transformers (such as Vision Transformers (ViT) and Large Language Models (LLMs)), as well as Recurrent Neural Networks (RNNs) like LSTMs.
- Authored five patents related to the novel methods developed during my PhD, with one already published and four currently under review.
- Onboarded and provided technical guidance to two PhD students in their initial phases.
- Collaborated with the University of Stuttgart under the supervision of Dr. rer. nat. Jens Mehnert.

- Analyzed Kosovo's export and import data by product from 2010 to 2023, using Python for data processing.
- Developed recommendations to support Kosovo's economic policy, with insights drawn from tools like the BCG Matrix.

Segura & Jesberger GmbH

Frankfurt am Main, Germany

Working Student | Family Office

Feb. 2018 - Aug. 2021

- Digitized an annual capital market survey by coding a website using HTML, CSS, PHP, MySQL, and JavaScript, saving one month of work for a working student each year.
- Developed Python-based software for automatic portfolio verification, using OCR to extract tables from printed pages and compare them with Excel data, saving 16 hours of work each month.
- Developed automated data processing applications in Excel using VBA.

KPMG AG

Frankfurt am Main, Germany

Working Student | CIO Consulting - Digital Strategy

Nov. 2017 - Jan. 2018

- Researched technological trends to support digital strategy development, focusing on data-driven decision-making.
- Contributed to creating data and analytics training programs to enhance employee skills.

Talks & Workshops

AI Officer – Course

Stuttgart, Germany

Speaker

20 May 2025

- Delivered the same curriculum as the Hamburg session on 10 Apr. 2025 (see below).

AI Governance – Workshop

Erlangen, Germany

Technical Advisor

13 May 2025

- Provided technical input on the integration of AI systems into workplace environments with a focus on large language models and generative AI.
- Advised on key aspects relevant for a works council agreement, including model limitations, data usage, and monitoring mechanisms.
- Explained the operational principles of models like GPT, highlighting implications for fairness and interpretability.

AI Officer – Course

Hamburg, Germany

Speaker

10 Apr. 2025

- Introduced key topics in Generative AI, Machine Learning, and Deep Learning, with a focus on practical relevance.
- Explained the architecture and training process of GPT, including data collection, filtering, and pretraining, and showed how ChatGPT builds on this with instruction tuning, dialogue optimization, and user feedback.
- Presented key techniques such as Retrieval-Augmented Generation (RAG) and Reinforcement Learning from Human Feedback (RLHF).
- Explained prompt engineering strategies, including prompt structure, iterative refinement methods, and supporting tools.

Skills

Programming	Python	● ● ● ● ●	C++	● ● ● ○ ○
	PyTorch	● ● ● ● ●	Tensorflow	● ● ● ● ○
Deep Learning	Anaconda	● ● ● ● ○	TensorBoard	● ● ● ● ○
	CUDA	● ● ○ ○ ○		
Development & Documentation	Git	● ● ● ● ●	Docker	● ● ● ○ ○
	VS Code	● ● ● ● ○	Sphinx	● ● ● ○ ○
	Pytest	● ● ● ○ ○		
Data Science	Pandas	● ● ● ● ○	NumPy	● ● ● ● ○
	Matplotlib	● ● ● ● ○	SQL	● ● ● ○ ○
Languages	German	● ● ● ● ●	English	● ● ● ● ●
	French	● ● ○ ○ ○	Albanian	● ● ● ● ●

Publications

Conference Proceedings

- [1] **CNN Mixture-of-Depths**
Rinor Cakaj, Jens Mehnert, Bin Yang
2024 Asian Conference on Computer Vision (ACCV), Hanoi, Vietnam
- [2] **Spectral Wavelet Dropout: Regularization in the Wavelet Domain**
Rinor Cakaj, Jens Mehnert, Bin Yang
2024 International Conference on Machine Learning and Applications (ICMLA), Miami, Florida, USA

- [3] **Squeeze-and-Remember Block**
Rinor Cakaj, Jens Mehnert, Bin Yang
2024 *International Conference on Machine Learning and Applications (ICMLA)*, Miami, Florida, USA
- [4] **Spectral Batch Normalization: Normalization in the Frequency Domain**
Rinor Cakaj, Jens Mehnert, Bin Yang
2023 *International Joint Conference on Neural Networks (IJCNN)*, Goldcoast, Australia
- [5] **Weight Componder: A Simple Weight Reparameterization for Regularization**
Rinor Cakaj, Jens Mehnert, Bin Yang
2023 *International Joint Conference on Neural Networks (IJCNN)*, Goldcoast, Australia

Patents

Published Patents

- [1] **Method for Regularizing a Neural Network**
Rinor Cakaj, Jens Mehnert, Bin Yang
No.: EP4343619, Date: 27 Mar. 2024, Assignee: Robert Bosch GmbH

Projects & Competitions

Bosch Hackathon - Capture the Flag (CTF)

HTB Platform

Participant

7 Oct. 2024 - 10 Oct. 2024

- Collaborating with a colleague on CTF challenges in web, forensics, reversing, OSINT, hardware, and crypto.
- Solved tasks such as SQL injection, template injection, XSS, and buffer overflow.
- Gained hands-on experience with Python, JavaScript (client- and server-side), Linux command line, and a basic understanding of the OWASP Top 10 vulnerabilities.

Stock Index Trend Prediction Tool

Personal Project

Mar. 2023 - Present

- Developed an end-to-end deep learning pipeline for stock index trend prediction, including automated web scraping with data integrity checks, data processing with rolling statistics and augmentation.
- Built a deep learning model using multi-scale convolutions for pattern recognition and BiLSTM layers for temporal dependencies.
- Worked independently, using a variety of tools and libraries:
 - Development Environment: JupyterLab, VS Code, Anaconda
 - Data Processing & Visualization: Pandas, NumPy, Plotly, Matplotlib, Scipy
 - Automation & Testing: Selenium, Pytest, Optuna
 - Model Tracking & Deployment: Weights & Biases, TensorBoard, Docker, Git
 - Documentation: Sphinx

Volunteering

Erster Sindlinger Schwimm-Club 1901 e.V.

Frankfurt am Main, Germany

Member of the Executive Board, Treasurer and Trainer

Jun. 2018 - Jun. 2022

- Introduced and administered Office 365 to enhance collaboration between trainers and the board.
- Ordered and implemented electronic membership cards (RFID) with an access control system on a Raspberry Pi 4 using Python.
- Rewrote statutes and regulations for business, contributions, data protection, training, and honorary policies.
- Redesigned the website.
- Handled cash office responsibilities, including tax returns, bookkeeping, grant applications, collection of membership fees, compensation payments for exercise leaders, and donation processing.
- Taught children how to swim.