



Serhii Hamotskyi

✉ serhii@serhii.net | 🌐 https://serhii.net | 📷 pchr8

Experience

Anhalt University of Applied Sciences (Köthen, Germany)

Jan. 2022 – current

PROJECT: CONSTRAINED RETRIEVAL-AUGMENTED LANGUAGE MODELS (CORAL)

Jan. 2025 – current

- LLM training data acquisition, preprocessing
- LLM-as-a-judge evaluation of LLM-generated text

PROJECT: MAGNETIC RESONANCE FOR THE DIGITIZATION OF PLANTS (50% POSITION)

May 2024 – Dec. 2024

- Medical Imaging / Computer Vision, 4D/MRI data
- 3DSlicer extension development (Python)

RESEARCH ASSOCIATE (50% POSITION)

Apr. 2023 – Apr. 2024

- Development of prototypes for research projects and grant proposals
- Establishment of CI/CD pipelines for ML projects (Docker, GitLab, Kubernetes)
- Assistance in the preparation of teaching materials

PROJECT: ELIGIBILITY EXAMINATION OF SECURITIES, GERMAN CENTRAL BANK (50% POSITION FROM JULY 2022)

Jan. 2022 – Dec. 2022

- NLP for automatic document (financial security) analysis (transformers, pytorch, spacy)
- Layout parsing, checkbox item matching, text flow reconstruction
- Language detection; annotation mapping on textflow changes

ExB Research & Development GmbH (Leipzig, Germany)

May 2019 – Dec. 2021

JUNIOR RESEARCHER → MACHINE LEARNING ENGINEER

July 2020 – Dec. 2021

- Vision-based layout detection and document parsing (pytorch, Detectron2)
- Improved comparison/matching of scanned addresses (record linking, string similarity, OCR errors simulation, ...)
- Rapid prototyping of training/inference scripts, REST APIs, converters; making them production-ready
- Adapted, visualized, presented results to technical and non-technical audiences
- Improved documentation; created a curriculum with tasks of increasing complexity to teach internal tool usage

WORKING STUDENT IN RESEARCH TEAM (50% POSITION)

May 2019 – June 2020

- Natural language processing, especially Named entity recognition
- Implemented custom Tensorflow layers, metrics, losses
- Implemented, trained and evaluated Language Models based on BERT, investigated improvements on downstream tasks
- Ran experiments on existing models, implementing various features and changes, improved stability and reproducibility

Israel Electric Corporation (Tel Aviv, Israel)

Aug. 2018 – Sep. 2018

MACHINE LEARNING INTERN (IAESTE)

- Used ML and statistical approaches for anomaly detection to find electrical meters at risk of failure (Keras, scikit-learn)
- Implemented and tested various LSTM-based multi-step load forecasting algorithms (Keras, pandas)
- Created and implemented an algorithm for filling gaps in sensor data, preserving daily/weekly periodicity (Python, pandas)
- Exploratory data analysis on black-box database containing sensor data without any available context or documentation

Education

Anhalt University of Applied Sciences

Köthen (Germany)

M.Sc. IN DATA SCIENCE

Oct. 2022 – May 2024

Thesis: “Eval-UA-tion: Benchmark for Evaluation of Ukrainian Language Models” (LLM, Evaluation, Morphology)

Merseburg University of Applied Sciences (HoMe)

Merseburg (Germany)

B.Sc. IN COMPUTER SCIENCE

Sep. 2015 – Nov. 2019

Thesis: “Natural language identification in user-generated data via ensemble learning” (POS-tagging; SVM, RF, DNN; pandas, nltk)

National Technical University of Ukraine “Kyiv Polytechnic Institute”

Kyiv (Ukraine)

B.Sc. IN SOFTWARE ENGINEERING

Sep. 2013 – June 2017

Thesis: “A system of encryption based on steganography using HTML as medium” (WebExtensions, Django; Shamir secret sharing)

Selected Publications

Eval-UA-tion 1.0: benchmark for evaluating Ukrainian (large) language models

UNLP @ LREC-COLING 2024

S. HAMOTSKYI, A.-I. LEVBARG, AND C. HÄNIG

3 different tasks and 9 novel datasets in total, with human baselines. Includes UA-CBT, a fill-in-the-gaps task on LLM-generated fairy tales.

Development and evaluation of a German language model for the financial domain

FinNLP @ LREC-COLING 2024

N. KOZAEVA, S. HAMOTSKYI, AND C. HÄNIG

Fine-tuning a LLM on the FinCorpus-DE10k corpus and evaluating it on financial NER and text classification tasks.

FinCorpus-DE10k: A corpus for the German financial domain

LREC-COLING 2024

S. HAMOTSKYI, N. KOZAEVA, AND C. HÄNIG

A corpus of German and bilingual PDF documents from the financial domain.

NLP-based decision support system for examination of eligibility criteria from securities prospectuses at the German central bank

AAAI-23

C. HÄNIG, M. SCHLÖSSER, S. HAMOTSKYI, G. ZAMBAKU, AND J. BLANKENBURG,

NLP-based decision support system for examining securities eligibility at the German Central Bank.

Automatized generation of alphabets of symbols

FedCSIS 2017

S. HAMOTSKYI, A. ROJBI, S. STIRENKO, AND Y. GORDIENKO

The paper is an intersection of computer science, cognitive science and linguistics and proposes a framework for creating symbols and alphabets according to given criteria and medium (touchpads, joysticks, eye-tracking, ...)

Open Source Projects

pymorphy-spacy-disambiguation

<https://github.com/pchr8/pymorphy-spacy-disambiguation/>

Disambiguate between pymorphy morphologies using spacy morphology data

ukr_numbers

https://github.com/pchr8/ukr_numbers/

Convert integers to Ukrainian numerals and adjectives, target declination shown using natural language:

(124, “першою”) → “сто двадцять четвертою”

Tools, libraries, etc.

ML / DS Pandas, numpy, sklearn; HF transformers/datasets

Operating Systems Linux (Debian, Arch), basic server administration

Other Data visualization (matplotlib/seaborn); LaTeX, HTML & CSS fundamentals

(Natural) languages **Ukrainian, Russian:** Native speaker; **English:** Proficient; **German:** Upper intermediate; **Italian:** Intermediate