Denis Semkin

Saarbrücken, Saarland, Germany semkind001@gmail.com +4915119489367 in/denis-semkin

SUMMARY

M.Sc. Bioinformatics student specialising in computational drug discovery and ML-driven data pipelines. Built production workflows at Boehringer Ingelheim for small molecule analysis and automated energy management systems at Saarland University. Expertise in Python, cheminformatics (RDKit, molecular docking, MD simulations), machine learning, and scalable infrastructure (Docker, HPC, databases).

EXPERIENCE

Pharmaceutical Development Intern

Boehringer Ingelheim **July 2025 - January 2026, Biberach, Germany**

- Built automated data workflows for small molecule analysis in drug development, including data fetching, processing, and comprehensive dataset generation.
- Processed and analysed molecular structures using RDKit and cheminformatics tools.
- Executed computational workflows on high-performance computing clusters using Bash automation for scalable batch processing.
- Utilised MOE (Molecular Operating Environment) for molecular visualisation, property calculations, and structure preparation.
- Delivered production-ready datasets used for molecular dynamics simulation experiments.
- Master's thesis (starting Feb 2026): Building workflows to assess the developability of protein therapeutics using Boltz 2.

Student research assistant

Saarland University, AES **July 2023 - June 2025, Saarbrücken, Germany**

- Developed an automated modelling system for PV-powered battery energy management with ML-based prediction.
- Built and maintained Python data pipelines for continuous sensor data processing and analysis.
- Managed databases (MongoDB, InfluxDB) to ensure efficient data storage and retrieval.
- Containerised workflows with Docker and automated processes using Unix/Bash CronJobs.
- Delivered robust code for system optimisation, improving predictive accuracy and operational efficiency.

PROJECT

MolFormer Fine Tuning

Saarland University, Spoken Language Systems • January 2025 - March 2025

- Fine-tuned a pre-trained chemical language model on lipophilicity regression using SMILES data.
- Developed end-to-end workflow for tokenisation, dynamic padding, and regression head integration.
- Evaluated parameter-efficient strategies (BitFit, LoRA, iA3) through systematic ablation studies.
- Applied influence functions (LiSSA) for strategic external data selection.

Computer-Aided Drug Design (CADD) Pipeline

Saarland University, Volkamer " Data Driven Drug Design" Lab • October 2024 - March 2025

- Developed an end-to-end drug discovery pipeline integrating ML-based activity prediction, molecular filtering, docking, and MD simulations to identify novel EGFR inhibitors.
- Trained and evaluated machine learning models (Random Forest, SVM, Neural Networks) for compound activity prediction (pIC50) on 3,906 samples.
- Performed virtual screening of 1,051 compounds using molecular property filters (Lipinski's Rule of Five) and ML predictions, selecting the top 20 candidates.
- Executed molecular docking with AutoDock Smina and GNINA, achieving RMSD <2Å in validation studies.
- Conducted 100ns molecular dynamics simulations using GROMACS and CHARMM36 force field to assess ligand binding stability.
- Performed ADMETox analysis and off-target prediction using SwissADME, SwissTargetPrediction, and eMolTox.

EDUCATION

Master's degree in bioinformatics

Saarland University • Saarbrücken, Germany • 2026

Bachelor's degree in biotechnology

Lomonosov Moscow State University • Moscow, Russia • 2022

SKILLS

Python: NumPy, Pandas, Polars, MatPlotLib, Seaborn, Scikit-Learn, PyTorch, RDKit, PvLib, PyOmo

Other: PyMol, MOE, GROMACS, Boltz 2, Molecular Docking, Docker, Bash, HPC clusters, MongoDB, InfluxDB

Languages: English (C1), German (B1)