

# Tatiana Malygina

Deep Learning | Protein Language Models | ML for Drug Discovery

Deep Learning Researcher with 5+ years in applied ML (CV, NLP, bioinformatics).  
Focus on protein language models and biosynthetic gene clusters.  
Publications at ICML, ISMB, MIDL.

 [merlettaia@gmail.com](mailto:merlettaia@gmail.com)  [latticetower](https://github.com/latticetower)  [Google Scholar](https://scholar.google.com/citations?user=...)  [LinkedIn](https://www.linkedin.com/in/...)

## SKILLS

- ▶ Programming: Python, C++, Julia, SQL
- ▶ Deep Learning, NLP, Computer Vision
- ▶ Algorithms, Structural Bioinformatics, Pipelines
- ▶ Libs: pytorch, lightning, pytorch-geometric



## EDUCATION

- ▶ PhD (2021-current time)  
Helmholtz-Institute for Pharmaceutical Research Saarland (Germany)
- ▶ MSc in Algorithmic Bioinformatics (2013-2015)  
St. Petersburg Academic University (Russia)
- ▶ BSc in Applied Mathematics (2003-2010)  
Northern (Arctic) Federal University (Russia)

## WORK EXPERIENCE

- ▶ 2021 - current time - Scientist (bioinformatics, deep learning) at **Helmholtz-Institute for Pharmaceutical Research Saarland**
- ▶ 2017-2021 - Deep learning researcher (medical imaging, bioinformatics) at **BotkinAI**
- ▶ 2016 - QA intern at **Debian Med**
- ▶ 2014-2015, 2016 - Bioinformatician (intern, junior) at **BIOCAD**
- ▶ 2010-2013 - Software developer at **Innovation Center LLC**

## AWARDS

- ▶ Adaptyv Translation Prize at Bio x ML Hackathon (2024)
- ▶ 2nd place at AI Hackathon with the project on small molecule generation and properties prediction (St.Petersburg, 2017)
- ▶ 21st place at ICFCP 2011 (functional programming contest)

## SELECTED RESEARCH

- ▶ Malygina et al. (2019) Data Augmentation with GAN: Improving Chest X-Ray Pathologies Prediction on Class-Imbalanced Cases. Best paper award (AIST 2019); presented at MIDL 2019.  
[https://doi.org/10.1007/978-3-030-37334-4\\_29](https://doi.org/10.1007/978-3-030-37334-4_29)
- ▶ Knyazev et al. (6th author) (2021) Accurate assembly of minority viral haplotypes from next-generation sequencing through efficient noise reduction, Published at NAR.  
<https://doi.org/10.1093/nar/gkab576>
- ▶ Papillon et al. (2023) ICML 2023 topological deep learning challenge: Design and results. Published at Proceedings of Machine Learning Research  
<https://proceedings.mlr.press/v221/papillon23a.html>
- ▶ Malygina and Kalinina (2024) Exploring sequence landscape of biosynthetic gene clusters with protein language models. Presented at ISMB 2023, VAAM 2023, ICML 2024  
<https://openreview.net/forum?id=UguakxTKG5>

KAGGLE  
EXPERT

OPEN SOURCE  
CONTRIBUTOR

Scan the QR code for  
an online version of this  
CV with clickable links

