Rıza Özçelik | Machine Learning Researcher

Eindhoven, The Netherlands
✉ r.ozcelik@tue.nl • github.com/rizaozcelik

Machine learning researcher with 7+ years of expertise in language models for natural sciences. Computer scientist. Recognized with 18 publications, 16+ selected talks, and 8 awards from industry and academia.

Employment

- 2022 – 2026: **Ph.D. Student**, *Molecular Machine Learning, Eindhoven University of Technology*.
  - Developed a new transfer learning strategy that unlocks zero-shot molecule design in ligand-based drug discovery.
  - Introduced state space models to chemistry. Improved models' capability to capture complex molecular properties by up to 15%.
  - Identified and solved widespread critical pitfalls in evaluating molecule designs. Defined new evaluation guidelines.
  - Collaborated with natural scientists and translated deep learning into real-world drug discovery applications in the lab.
  - Disseminated the results by publishing 12 manuscripts in top journals and presenting 16 talks and posters.

  Thesis supervisor: Francesca Grisoni
- 2021 – 2021: **Great Minds Research Intern**, *IBM Research Zurich.*
  - Selected for the competitive Great Minds Internship (<10% acceptance rate after pre-selection).
  - Created a link prediction model to discover otherwise hidden semantic relations in the application's knowledge graph.
  - Implemented a real-time system from scratch that translates natural language prompts of customers to internal workflows.
- 2019 – 2021: **Research & Teaching Assistant**, *Dept. Computer Engineering, Boğaziçi University*.
  - Co-led the development of the open-source Turkish language processing library, `nlptoolkit`.
  - Co-authored a research grant proposal on drug discovery. The proposal is funded $200K (≈10% acceptance rate).
  - Taught lectures for more than 200 hours. More than 15% of written feedback called me "the best TA ever."
- 2018 – 2018: **Data Scientist**, *ING Turkey.*
  - Collaborated with finance and operations experts and algorithmically formulate the idle cash minimization problem.
  - Prepared and presented data-driven reports to large and interdisciplinary teams.
  - Implemented machine learning solutions that saved more than $10K/day for the bank.

Education

- 2018–2021: **M.Sc.**, **Computer Engineering**, *Boğaziçi University, Turkey*.
  Thesis: Biomolecular Language Processing for Drug-Target Affinity Prediction. GPA: 3.94/4.00 (High Honor).
  Supervisor: Arzucan Özgür and Elif Ozkirimli (co-supervisor from Roche)
- 2014–2018: **B.Sc.**, **Computer Engineering**, *Boğaziçi University, Turkey.*
  Thesis: Disease Classification Based on Genomic Data with Machine Learning. GPA: 3.57/4.00 (High Honor).
  Supervisor: Arzucan Özgür

Selected Systems & Tools

- Real-time translator (IBM): Translated customer queries to internal pipelines with natural language processing in a real-time system.
- Cash Flow Optimization (ING): Saved ≈$10K/day by minimizing idle cash in ATMs.
- `s4dd`, `deepclp`, `pydebiaseddta`, `peptidy`: Created open-source Python libraries for deep learning on molecules, lowered entry barriers for interdisciplinary research.

Skills & Interests

- **Research:** Deep learning, drug discovery, predictive and generative modeling, language models.
- **Language modeling:** Transformers, SSMs, LSTMs, tokenization, finetuning, task arithmetic.
- **Software:** Python, PyTorch, Keras, XGBoost, RDKit, Git, Bash, HPC, Slurm, Anaconda, LaTeX.
- **Personal:** Teamwork, independence, communication, presentation, critical thinking, supervision.

Honors & Awards

1. Cover article on RSC Digital Discovery journal (2025).
2. Top-10 cited and top-10 read article certificates from ChemBioChem journal (2025).
3. Best Publication Award of 2024 at Institute of Complex Molecular Systems.
   Awarded for scientific impact and oral presentation quality by an interdisciplinary post-doc committee (18 contestants).
4. Google EMEA Academic Scholarship for students with disabilities.
   A highly competitive scholarship given to outstanding computer science students.
5. Best Abstract Award at Institute of Complex Molecular Systems.
   Awarded by an interdisciplinary post-doc committee for quality of written communication contest (20 contestants).
6. Three best bachelor thesis awards as a thesis supervisor.
   Two awarded by the Dept. of Computer Engineering, and one by Biomedical Engineering. Each contest had 50+ candidates.
7. Graduate student scholarship from Scientific Council of Turkey (TUBITAK).
   A highly competitive scholarship given to the top graduate students in Turkey.
8. Two High Honor Certificates.
   Awarded by Dept. of Computer Engineering of Boğaziçi University to graduates with an average grade over 3.50/4.00.

Recent Talks & Presentations

- 5 talks and 5 posters in 2024 and 2025:
  - "New Approaches in AI-Driven Drug Discovery: From Design to Evaluation." @ Institute of Complex Molecular Systems Symposium (upcoming). Invited for winning the best publication award of 2024.
  - "Jungle of Generative Drug Discovery: Traps, Treasures, and Ways Out." and 2 posters @ ELLIS ML4Molecules Workshop – a highly-competitive venue for machine learning for drug discovery and materials science research;
  - 3 posters @ Machine Learning for Life and Material Science Workshop of International Conference on Machine Learning (ICML) – a top workshop for AI for life sciences researchers;
  - "Chemical Language Modeling with Structured State Spaces." @ Applied Machine Learning Days (AMLD) – the largest machine learning event of Switzerland;
  - "Chemical Language Modeling with Structured State Spaces for De Novo Design." @ M2D2 talk series – organized by Valence.AI, this talk series invites prestigious research on AI for drug discovery.
- 5 talks and 1 poster in 2023:
  - "Structured State Space Sequence Models for De Novo Drug Design." @ ELLIS ML4Molecules Workshop;
  - "AI-driven Molecule Design with S4 Models." @ World Chemistry Congress (IUPAC-CHAINS);
  - 1 poster @ IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology – a prestigious computational biology conference;
  - "Revitalizing Artificial Intelligence for Molecule Design." Institute of Complex Molecular Systems Symposium – an event that brings together the interdisciplinary scientists at the institute. I was invited for winning the Best Abstract Award;
  - "Accelerating Drug Discovery with Artificial Intelligence." Invited talk at Oncode Cancer Institute – an event organized to foster application of AI for cancer treatment;
  - "DebiasedDTA: A Framework for Improving the Generalizability of Drug-Target Affinity Prediction Models." @ International Conference on Research in Computational Molecular Biology (RECOMB) – a highly prestigious computational biology conference.

Publications

*Shared first authorship

Peer Reviewed

1. **Özçelik, Rıza**, and Francesca Grisoni. “A Hitchhiker’s Guide to Deep Chemical Language Processing for Bioactivity Prediction.” *Digital Discovery*, 2025. (Cover article)

2. **Özçelik, Rıza**, Laura van Weesep, Sarah de Ruiter, and Francesca Grisoni. “peptidy: A Light-weight Python Library for Peptide Representation in Machine Learning.” *Bioinformatics Advances*, 2025.

3. **Özçelik, Rıza**, Francesca Grisoni. “De Novo Drug Design by Chemical Language Modeling.” Book chapter in *An Introduction to Generative Drug Discovery*, 2025.

4. Birolo, Rebecca, **Rıza Özçelik**, Andrea Aramini, Roberto Gobetto, Michele R. Chierotti, Francesca Grisoni. “Deep Supramolecular Language Processing for Co-crystal Prediction.” *Angewandte Chemie*, 2025.

5. **Özçelik, Rıza**, Sarah de Ruiter, Emanuele Crusciolo, and Francesca Grisoni. “Chemical Language Modeling with Structured State Space Sequence Models.” *Nature Communications*, 2024.

6. van Tilborg, Derek, Helena Brinkmann, Emanuele Criscuolo, Luke Rossen, **Rıza Özçelik***, and Francesca Grisoni. “Deep learning for Low-data Drug Discovery: Hurdles and Opportunities." *Current Opinion in Structural Biology*, 2024.

7. Multari, Silvia, **Rıza Özçelik**, Angelica Mazzolari, Salvatore Nobile, and Francesca Grisoni. “Predicting Metabolic Reactions with a Molecular Transformer for Drug Design Optimization.” *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2024.

8. Büşra Temizer, Asu, Gökçe Uludoğan, **Rıza Özçelik***, Taha Koulani, Elif Ozkirimli, Kutlu Ulgen, Nilgun Karalı, and Arzucan Özgür. “Exploring Data-Driven Chemical SMILES Tokenization Approaches to Identify Key Protein-Ligand Binding Moieties.” *Molecular Informatics*, 2024.

9. **Özçelik, Rıza***, Derek van Tilborg, Jose Jiménez-Luna, and Francesca Grisoni. “Structure-based Drug Discovery with Deep Learning.” *ChemBioChem*, 2023. (Top-10 cited article of the journal in 2023)

10. **Özçelik, Rıza**, Alperen Bağ, Berk Atıl, Melih Barsbey, Arzucan Özgür, and Elif Özkırımlı. “A Framework for Improving the Generalizability of Drug-Target Affinity Prediction Models.” *Journal of Computational Biology*, 2023.

11. Barsbey, Melih, **Rıza Özçelik***, Alperen Bağ, Berk Atıl, Arzucan Özgür, and Elif Özkırımlı. “A Computational Software for Training Robust Drug-Target Affinity Prediction Models: pydebiaseddta.” *Journal of Computational Biology*, 2023.

12. Budur, Emrah, **Rıza Özçelik**, Dilara Soylu, Omar Khattab, Tunga Güngör, and Christopher Potts. “Building Efficient and Effective OpenQA Systems for Low-Resource Languages." *Knowledge-Based Systems*, 2023.

13. **Özçelik, Rıza***, Hakime Öztürk, Arzucan Özgür, and Elif Ozkirimli. “ChemBoost: A chemical language based approach for the prediction of protein - ligand binding affinity.” *Molecular Informatics*, 2021.

14. Budur, Emrah, **Rıza Özçelik**, Tunga Güngör, and Christopher Potts. “Data and Representation for Turkish Language Inference.” *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.

15. Köksal, Abdullatif, Hilal Dönmez, **Rıza Özçelik**, Elif Ozkirimli, and Arzucan Özgür. “Vapur: A Search Engine to Find Related Protein–Compound Pairs in COVID-19 Literature.” *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP*, 2020.

16. **Özçelik, Rıza**, Gökçe Uludoğan, Selen Parlar, Özge Bakay, Özlem Ergelen and Olcay Taner Yıldız. "User Interface for Turkish Word Network KeNet." *Signal Processing Applications*, 2019.

17. Uludoğan, Gökçe, **Rıza Özçelik**, Selen Parlar, Gökhan Ercan and Olcay Taner Yıldız. "User Interfaces for Turkish Natural Language Processing." *Signal Processing Applications*, 2019.

18. Ercan, Gökhan, Orçun Erkek, Onur Açıkgöz, **Rıza Özçelik**, Selen Parlar and Olcay Taner Yıldız. "Data Set Generation for Analysing of Turkish Semantic and Sentence Similarity." *3rd International Conference on Computer Science and Engineering (UBMK)*, 2018.

Preprints

1. **Özçelik, Rıza**, Francesca Grisoni. "The Jungle of Generative Drug Discovery: Traps, Treasures, and Ways Out." *ArXiv doi: 10.48550/arXiv.2501.05457 (Under review)*, 2025.

2. **Özçelik, Rıza***, Helena Brinkmann, Emanuele Criscuolo, Francesca Grisoni. "Generative Deep Learning for De Novo Drug Design – A Chemical Space Odyssey." *ChemRxiv doi: 10.26434/chemrxiv-2025-1zs8j (Under review)*, 2025.

3. van Weesep, Laura, **Rıza Özçelik***, Marloes Pennings, Emanuele Criscuolo, Christian Ottmann, Luc Brunsveld, Francesca Grisoni. "Identifying 14-3-3 Interactome Binding Sites with Deep Learning", *ChemRxiv doi: 10.26434/chemrxiv-2025-18tn6 (Under review)* 2025.

4. Criscuolo, Emanuele, **Rıza Özçelik**, Derek van Tilborg, Francesca Grisoni. "The Surprising Ineffectiveness of Molecular Dynamics Coordinates for Predicting Bioactivity with Machine Learning." *ChemRxiv doi: 10.26434/chemrxiv-2024-rp81v*, 2024.

In Preparation

1. **Özçelik, Rıza**, Sarah de Ruiter, Francesca Grisoni. "Molecular Task Arithmetic: A New Transfer Learning Approach for Molecule Design in Low-Data", *Submitted to NeurIPS 2025*.

2. **Özçelik, Rıza**, Francesca Grisoni. "An Efficient and Effective Framework to Evaluate De Novo Design Models", 2025.

Medium Blogposts

1. Different Methods in Geometric Deep Learning

2. An Intuitive Explanation of DeepWalk

3. An Intuitive Explanation of NeoDTI

4. An Intuitive Explanation of EmbedS

5. A*: Adding Heuristics to Dijkstra

6. Solving Minimum Path Cover on a DAG

7. EDA on Graphs via networkx

8. Using GraphSAGE to Learn Paper Embeddings in CORA

9. Regression Trees from Scratch in 30 lines of Python

10. A Very Friendly Starting Guide for Terminal and Git