

Gökçe Uludoğan

Istanbul, Türkiye · gokce.uludogan@bogazici.edu.tr · Website · GitHub · Scholar

Research Profile

Machine-learning researcher working on biomolecular modeling, with a focus on protein representation learning, function prediction, molecular generation, and functional protein-unit discovery. Additional research experience includes cheminformatics and natural language processing, particularly drug design, Turkish language modeling, and hate speech detection.

Research areas: biomolecular machine learning; protein representation learning; protein function prediction; molecular generation; cheminformatics; natural language processing.

Education

Ph.D. in Computer Engineering *Boğaziçi University* 2021–Present

Istanbul, Türkiye; Expected 2027; co-advisors: Prof. Arzucan Özgür, Elif Ozkirimli.

M.S. in Computer Engineering *Boğaziçi University* 2018–2021

Istanbul, Türkiye; Research on biochemical language models for target-specific drug design.

B.S. in Computer Engineering *Boğaziçi University* 2013–2018

Istanbul, Türkiye; Erasmus exchange at KTH Royal Institute of Technology, 2016–2017.

Research and Teaching Experience

Researcher *Understanding the Language of Life (funded by the European Research Council (ERC) under grant 101089287) – Boğaziçi University* Nov. 2023–Present

- Lead research on function-aware protein unit discovery, ontology-informed function prediction, and systematic evaluation of Gene Ontology (GO)-based predictors.
- Develop representations spanning protein sequence, structure, function, and molecular dynamics within an European Research Council (ERC)-funded project.

Teaching Assistant *Department of Computer Engineering – Boğaziçi University* Feb. 2019–Present

- Conduct problem sessions, and laboratory sessions; prepare and grade quizzes, assignments, and course projects
- Support courses in bioinformatics, probability and statistics, information retrieval, databases, systems programming, computer organization, and introductory computing.

Applied NLP Researcher & Engineer *Utilizing Digital Technology for Social Cohesion (EuropeAid/170389/DD/ACT/Multi) – Boğaziçi University* Jun. 2023–Mar. 2025

- Created Turkish hate-speech datasets and models for social and print media, including target-aware and span-level analyses.
- Contributed to the development of pari, an AI-assisted tool that supports expert-led monitoring of hate speech in Turkish print and social media.
- Co-led a multilingual shared task on hate-speech detection in Turkish and Arabic, including task design, dataset preparation, participant coordination, and results analysis.

Molecular ML Researcher *Chemical Language Processing for Target-Based Drug Design – Boğaziçi University* Apr. 2020–Oct. 2023

- Designed biochemical language models, data-driven molecular tokenization methods, and latent-space optimization approaches for target-conditioned molecule generation. under Scientific and Technological Research Council of Türkiye (TÜBİTAK) grant 119E133.

NLP Software Developer *Using Turkish WordNet for Machine Translation – Boğaziçi University* Dec. 2018–Apr. 2019

- Tested, refactored, and documented an open-source Turkish natural language processing (NLP) toolkit and WordNet lexical database.
- Implemented user interfaces that improved access to the toolkit and Turkish lexical resources under Scientific and Technological Research Council of Türkiye (TÜBİTAK) grant 116E104.

Machine Learning Engineer *tazi.ai – Istanbul, Türkiye* Feb. 2018–Nov. 2018

- Engineered adaptive machine-learning components for real-time recommendation and anomaly-detection systems.

Publications

1. **PUFFIN: Protein Unit Discovery with Functional Supervision.** G. Uludoğan, B. Giledereli, E. Özkırmı, A. Özgür. *Bioinformatics, Intelligent Systems for Molecular Biology (ISMB) Proceedings*, 2026. [view]
2. **STAR-GO: Improving Protein Function Prediction by Learning to Hierarchically Integrate Ontology-Informed Semantic Embeddings.** M. E. Akça, G. Uludoğan, A. Özgür, İ. M. Baytaş. *Bioinformatics*, 42(4), btg146, 2026. [view]
3. **ConGen: Targeted Molecule Generation Through Contrastive Learning and Latent Optimization.** C. Koban, G. Uludoğan, E. Özkırmı, A. Özgür. *Molecular Informatics*, 45(5), e70038, 2026.
4. **HATECAT-TR: A Hate Speech Span Detection and Categorization Dataset for Turkish.** H. K. Şeker, G. Uludoğan, P. Önal, A. Özgür. *Findings of Empirical Methods in Natural Language Processing (EMNLP) 2025*, 2025.
5. **TURNA: A Turkish Encoder-Decoder Language Model for Enhanced Understanding and Generation.** G. Uludoğan, Z. Y. Balal, F. Akkurt, M. Türker, O. Güngör, S. Üsküdarlı. *Findings of the Association for Computational Linguistics (ACL) 2024*, 10103-10117, 2024. [view]
6. **A Concise Report of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-Political Events from Text.** A. Hürriyetoğlu, S. Thapa, G. Uludoğan, S. Dehghan, H. Tanev. *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-Political Events from Text*, 2024.
7. **Overview of the Hate Speech Detection in Turkish and Arabic Tweets (HSD-2Lang) Shared Task at CASE 2024.** G. Uludoğan, S. Dehghan, İ. Arın, E. Erol, B. Yanıkoğlu, A. Özgür. *Proceedings of Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE) 2024*, 229-233, 2024.
8. **Detecting Hate Speech in Turkish Print Media: A Corpus and a Hybrid Approach with Target-Oriented Linguistic Knowledge.** G. Uludoğan, A. E. Yüksel, Ü. Tunçer, B. Işık, Y. Korkmaz, D. Akar, A. Özgür. *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-Political Events from Text*, 2024.
9. **Exploring Data-Driven Chemical SMILES Tokenization Approaches to Identify Key Protein-Ligand Binding Moieties.** A. B. Temizer, G. Uludoğan, R. Özçelik, T. Koulani, E. Özkırmı, K. O. Ülgen, N. L. Karalı, A. Özgür. *Molecular Informatics*, 43(3), e202300249, 2024. [view]
10. **BOUN-TABI@SMM4H'22: Text-to-Text Adverse Drug Event Extraction with Data Balancing and Prompting.** G. Uludoğan, Z. Yirmibeşoğlu. *Social Media Mining for Health (SMM4H) 2022*, 2022.
11. **Exploiting Pretrained Biochemical Language Models for Targeted Drug Design.** G. Uludoğan, E. Özkırmı, K. O. Ülgen, N. L. Karalı, A. Özgür. *Bioinformatics*, 38(Suppl. 2), ii155-ii161, 2022. [view]
12. **User Interface for Turkish Word Network KeNet.** R. Özçelik, G. Uludoğan, S. Parlar, Ö. Bakay, Ö. Ergelen, O. T. Yıldız. *IEEE Signal Processing and Communications Applications Conference (SIU)*, 2019.

Talks

PUFFIN: Protein Unit Discovery with Functional Supervision <i>Intelligent Systems for Molecular Biology (ISMB) Proceedings Talk</i>	2026 — Washington, D.C., United States
STAR-GO: Ontology-Informed Semantic Embeddings for Protein Function <i>Intelligent Systems for Molecular Biology (ISMB) Highlights Talk</i>	2026 — Washington, D.C., United States
Detecting Hate Speech in Turkish Print Media: A Corpus and a Hybrid Approach with Target-Oriented Linguistic Knowledge <i>7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE) 2024</i>	2024 — St. Julian's, Malta
Overview of the Hate Speech Detection in Turkish and Arabic Tweets (HSD-2Lang) Shared Task at Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE) 2024 <i>7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE) 2024</i>	2024 — St. Julian's, Malta
Exploiting Pretrained Biochemical Language Models for Targeted Drug Design <i>European Conference on Computational Biology (ECCB) 2022 & Health Informatics and Bioinformatics Symposium (HIBIT) 2022</i>	2022 — Sitges, Spain & Mersin, Türkiye

Teaching

CMPE 150 <i>Introduction to Computing</i>	CMPE 344 <i>Computer Organization</i>
CMPE 230 <i>Systems Programming</i>	CMPE 493 <i>Information Retrieval</i>
CMPE 321 <i>Introduction to Database Systems</i>	CMPE 549 <i>Bioinformatics</i>
CMPE 343 <i>Introduction to Probability and Statistics</i>	

Awards

Intelligent Systems for Molecular Biology (ISMB) 2026 Travel Fellowship *International Society for Computational Biology (ISCB)* 2026

Graduate Scholarships Performance Program (2250), four award periods *Scientific and Technological Research Council of Türkiye (TÜBİTAK) Science Fellowships and Grant Programmes Department (BİDEB)* 2022–2024

European Conference on Computational Biology (ECCB) 2022 Travel Fellowship *International Society for Computational Biology (ISCB)* 2022

National M.S. Scholarship (2210-A) *Scientific and Technological Research Council of Türkiye (TÜBİTAK) Science Fellowships and Grant Programmes Department (BİDEB)* 2018–2020

Academic Service

Reviewing — Conferences *ACL Rolling Review, multiple cycles, 2025–2026; Association for Computational Linguistics Student Research Workshop (ACL SRW) 2026; International Workshop on Pattern Recognition in Healthcare Analytics 2023–2026; Signal Processing and Communications Applications Conference (SIU) 2022–2026*

Reviewing — Journals *Journal of Computational Biology; Turkish Journal of Mathematics and Computer Science; Nature Communications (secondary); PLOS Computational Biology (secondary)*

Organization *HSD-2Lang Shared Task on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE), European Chapter of the Association for Computational Linguistics (EACL) 2024*

Editor *the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE), European Chapter of the Association for Computational Linguistics (EACL) 2024*

Volunteering *European Chapter of the Association for Computational Linguistics (EACL) 2024; Intelligent Systems for Molecular Biology (ISMB)/European Conference on Computational Biology (ECCB) 2023; Research in Computational Molecular Biology (RECOMB) 2023; Women in Machine Learning (WiML) Workshop at NeurIPS 2022*