Ioannis Mouratidis

Austin, Texas
imourati@gmail.com | LinkedIn | Google Scholar

Computational biology researcher focused on AI safety and dual-use risks, with 35 publications (10 first or senior author) and 3 patents in large-scale data analysis and ML applications in genomics. Co-founded startup in cancer diagnostics, authored grants securing $4M+ in competitive funding and currently lead a 5-member research team.

Experience

University of Texas at Austin **Austin, TX**
Senior Research Engineer Associate (Research Engineer/Scientist Associate V) 08/2025 – Present

- Direct 12 research initiatives (4 independently, 8 collaboratively) resulting in 2 peer-reviewed publications with an additional 4 under review and 6 in preparation.
- Lead 5-member team focused on safety profiles and capability assessments of biological foundation models.
- Lead red-teaming of agentic AI scientists (Biomni, Kosmos, Google Co-Scientist) in collaboration with SecureBio, assessing susceptibility to adversarial attacks.
- Evaluate adversarial robustness of genomic foundation models, including susceptibility to data poisoning attacks that are very difficult to detect in large omics datasets but can subtly compromise safety and performance.
- Direct benchmarking of genomic language models against real genomes, evaluating capabilities in synthetic biology and identifying systematic discrepancies that can be used to identify synthetically generated sequences.
- Own lab's scientific computing infra: migrated 150 TB from Penn State; manage 60K node-hours/25 users.

The Pennsylvania State University **Hershey, PA**
Research Associate & Research Assistant 08/2023 – 08/2025

- Engineered novel algorithms in Python and Bash for the analysis of short DNA and protein sequences; 12 peer-reviewed publications (7 as first/corresponding author).
- Built scalable High-Performance Computing (HPC) pipeline in Slurm enabling the identification of pathogen biomarkers across 45,000 genome assemblies; led scaling to 500,000 assemblies.
- Developed the *kmerDB* database, consolidating information on short DNA and protein sequences across over 60K genomes and proteomes from RefSeq and UniProt; co-developed 4 additional public scientific databases.
- Created open-source *Zseeker* Python tool; co-developed 5 additional open-source bioinformatics tools.
- Delivered oral presentations at 4+ national and international conferences, ranging from 150-300 attendees.
- Organized introductory workshops on Python and ML for experimental biologists (40+ student and faculty attendees), demonstrating ability to communicate complex technical concepts to non-expert audiences.
- Supervised 8+ researchers and coordinated collaborative projects across multiple research groups; mentored 5 junior researchers with no prior bioinformatics experience to their first first-author publications.
- Designed research projects and drafted grant applications resulting in $3M+ in competitive funding.

Neomer Diagnostics **Brussels, Belgium**
Co-founder and Chief Technical Officer 01/2022 – 05/2023

- Co-founded diagnostics startup and coordinated across technical, regulatory, and investor stakeholders to translate patented nullomer research into clinical platform.
- Developed ML pipeline for cancer detection from liquid biopsies, achieving AUC ranging from 0.89 to 0.94.
- Designed and delivered investor demos for non-technical audiences, securing $850K.

Independent Contractor **Brussels, Belgium**
Software Engineering Consultant 02/2020 – 08/2023

- Optimized ML pipeline using Julia; 50x speedup, saving >3 months of computational time and $50K+ in costs.
- Created distributed computing pipelines for analysis of 100+ TB of genomic data.
- Deployed and managed genomic analysis pipeline to GCP, enabling processing of 14 TB of cfDNA sequencing data.
- Consulted on study design and grant writing, securing $400K+ in awarded funding.

Education

The Pennsylvania State University **Hershey, PA**
Ph.D. in Bioinformatics and Genomics (Cumulative GPA: 4.0) 08/2023 – 02/2026
Doctoral thesis: Computational and machine learning methods for k-mer sequence analysis in large-scale omics data

KU Leuven **Leuven, Belgium**
M.Sc. in Artificial Intelligence (Honours: Cum Laude) 09/2020 – 07/2023
Master's thesis: Frequentmers – a k-mer based approach for the detection of liver cirrhosis gut microbiome mNGS

Aristotle University of Thessaloniki **Thessaloniki, Greece**
B.Sc. in Mathematics (Honours: Very Good) 10/2013 – 03/2018

Skills

Technical skills:

- Programming Languages: Python (Pandas, NumPy, scikit-learn, PyTorch, Matplotlib, FastMCP), Bash, Julia
- Tools & Platforms: High-Performance Computing (Slurm, SGE, LSF), Google Cloud Platform, Git, Docker
- Computational Methods: AI evals, genomic language models, empirical ML research, algorithm design and optimization, large-scale data analysis, machine learning pipelines
- AI Safety: Dual-use risk assessment, red-teaming, threat modeling
- Research Skills: Study design, scientific writing, grant writing

Languages: English (Fluent), French (Fluent), German (Fluent), Japanese (Elementary), Greek (Native)

Honors and Awards

- Center for Molecular Carcinogenesis and Toxicology Symposium Abstract Award (1st place)
- Alumni Society Award, Penn State College of Medicine, 2024 (5% acceptance rate)
- University Graduate Fellow & Huck Distinguished Fellow, Penn State, 2023 (1% acceptance rate)

Patents

- Georgakopoulos-Soares, I., **Mouratidis, I.**, Provatas, K. (2025). System and Method for Extracting Neomers. U.S. Provisional Patent.
- Ahituv, N., Yizhar-Barnea, O., Georgakopoulos-Soares, I., **Mouratidis, I.**, Hemberg, M. (2023). Systems for mutation caller and methods of using the same. WO2024103003A3.
- Ahituv, N., Yizhar-Barnea, O., Georgakopoulos-Soares, I., **Mouratidis, I.**, Hemberg, M. (2022). Compositions comprising nullomers and methods of using the same for cancer detection and diagnosis. WO2022235718A3.

Selected Publications

1. Tzanakakis, A., **Mouratidis, I.**[+], & Georgakopoulos-Soares, I. (2026). Fundamental limitations of genomic language models for realistic sequence generation. *bioRxiv*.
2. Georgakopoulos-Soares, I.*, Yizhar-Barnea, O.*, **Mouratidis, I.***, et al. (2025). Leveraging sequences missing from the human genome to diagnose cancer. *Communications Medicine*, *5*(1).
3. Wang, G.*, **Mouratidis, I.***, Provatas, K., et al. (2025). ZSeeker: an optimized algorithm for Z-DNA detection in genomic sequences. *Briefings in Bioinformatics*, 26(3).
4. **Mouratidis, I.***[+], Konnaris, M. A.*, Chantzi, N.*, Chan, C. S. Y.*, et al. (2025). Identification of the shortest species-specific oligonucleotide sequences. *Genome Research*, *35*(2).
5. **Mouratidis, I.***, Baltoumas, F. A.*, Chantzi, N.*, et al. (2024). kmerDB: A database encompassing the set of genomic and proteomic sequence information for each species. *Computational and Structural Biotechnology Journal*, *23*.
6. **Mouratidis, I.**, Chantzi, N., Khan, U., et al. (2023). Frequentmers - a novel way to look at metagenomic next generation sequencing data and an application in detecting liver cirrhosis. *BMC Genomics*, *24*(1).

Complete list of published work can be found in Google Scholar.

* Equally contributing authors [+] Co-corresponding (supervising) author