Elyas Heidari

PhD Researcher
AI for Life Sciences
DKFZ & EMBL Heidelberg
Heidelberg, Germany

✉ elyas.heidari@dkfz-heidelberg.de
github.com/EliHei2
LinkedIn: tinyurl.com/lnkdns
Google Scholar: tinyurl.com/gglscl

PROFILE

I build scalable learning systems for biomedical data that transform ill-posed real-world structure into robust, high-performance implementations spanning concrete modeling, optimized systems design, and reliable deployment. I operate across disciplines and institutions, connecting efficient computation with biomedical domain insight through close collaboration with experimentalists, mathematicians, and engineers.

RESEARCH EXPERIENCE

DKFZ & EMBL Heidelberg Heidelberg, DE
PhD Researcher (Stegle & Gerstung Labs) Aug 2022 – Present

- Led **end-to-end ML pipelines**, **scalable graph learning**, and **distributed multi-GPU systems** for **multimodal spatial omics** at **10M+ scale**; current focus on **multi-agent systems** and **automation**.
- Delivered production-grade segmentation and spatial inference systems underpinning multiple large-scale analyses and publications; enabled routine processing of datasets previously considered computationally intractable.

Cancer Research UK & EMBL-EBI Cambridge, UK
Research Trainee (MSc Thesis; Marioni Lab) Apr 2021 – Apr 2022

- Developed a **supervised representation learning** framework using **graph attention** and **transformers** for spatial inference from single-cell data.

University of Zurich Zurich, CH
Research Assistant (Robinson Lab) Sep 2019 – Apr 2022

- Engineered **single-cell analysis pipelines** (QC, integration, cell typing) and developed **immune repertoire** analysis.

EMBL Heidelberg Heidelberg, DE
Research Trainee (BSc Thesis; Huber Group) Jul 2018 – Sep 2018

- Developed an early **graph-based representation learning** method for single-cell data; explored **probabilistic** and **random-walk** graphical models and their **scalability limits**.

Sharif University of Technology Tehran, IR
Research Assistant 2017 – 2018

- Developed a **population-scale graphical modeling** toolkit for **high-dimensional mixed-type multivariate data**.

AWARDS & RECOGNITION

- **ETH Medal** (2023) — Outstanding Master's thesis, ETH Zurich (top <0.1%)
- **EMBL-EBI Research Fellowship** (2021–2022) — Competitive 12-month traineeship
- **EMBL Research Fellowship** (2018) — Undergraduate research traineeship

EDUCATION

Heidelberg University Heidelberg, DE
PhD, Faculty of Biosciences 2022 – 2026 (expected)

- Thesis: *Structured Representation Learning for Large-Scale Spatial Omics Data*

ETH Zurich Zurich, CH
MSc, Computational Biology & Bioinformatics 2019 – 2022

- GPA: 5.76/6.0 (Top 3 in cohort); Thesis: *Supervised spatial inference of dissociated single-cell data* (**ETH Medal**)

Sharif University of Technology Tehran, IR
BSc, Computer Engineering & Applied Mathematics 2014 – 2019

- GPA: 16.8/20; emphasis on algorithms, optimization, and mathematical modeling. Founder, **Sharif DataDays**. Head TA for Advanced Programming and Probability & Statistics.

TECHNICAL EXPERTISE

- **Machine Learning:** Graph neural networks (heterogeneous, message-passing, attentional), transformer architectures, representation learning, variational inference, generative models
- **Agentic AI & Developer Tooling:** LLM-assisted development workflows for code generation, refactoring, and debugging; agent-based research engineering pipelines (e.g. Claude Code, Cursor)
- **Frameworks:** PyTorch, PyTorch Geometric, PyTorch Lightning, Pyro; JAX/XLA (developing proficiency)
- **Distributed Computing:** Multi-GPU training and inference pipelines, Dask, RAPIDS, HPC cluster management (SLURM)
- **Data Infrastructure:** Zarr, XArray, HDF5, TIFF; experience with datasets exceeding 10^7 observations
- **Software Engineering:** CI/CD pipelines, API design, documentation standards, reproducible research practices
- **Programming:** Python (primary), R (extensive), Java, C, Bash · **Languages:** English (fluent), German (B2), Persian (native)

SELECTED SOFTWARE

- **Segger (Python)** — large-scale GNN-based cell segmentation for spatial transcriptomics (30M transcripts in 10 minutes, multi-GPU). github.com/EliHei2/segger_dev
- **SageNet (Python)** — graph-attention-based spatial inference for dissociated single-cell data. sagenet.readthedocs.io github.com/MarioniLab/sagenet
- **scGCN (Python)** — early Graph Convolutional Modeling on gene regulatory networks for single-cell annotation. github.com/EliHei2/scPotter
- **MUVis (R)** — structured dependency modeling and visualization for mixed-type multivariate data. baio-lab.github.io/muvis

PUBLICATIONS AND PREPRINTS

- *Segger: Fast and accurate cell segmentation of imaging-based spatial transcriptomics data.*
  Heidari, E.*, Moorman, A.*, Unyi, D., et al. bioRxiv, 2025 (under revision *Nature Methods*).
  Lead developer: formulated cell segmentation as a large-scale heterogeneous GNN problem; designed multi-GPU pipeline (10–100M nodes) achieving 1000× speedup and production-grade robustness.
- *Integrated in vivo combinatorial functional genomics and spatial transcriptomics of tumours to decode genotype-to-phenotype relationships.*
  Breinig, M.*, Lomakin, A.*, **Heidari, E.***, et al. *Nature Biomedical Engineering*, 2025.
  Co-lead developer: built end-to-end spatial phenotyping pipeline integrating spatial transcriptomics and scRNA-seq to link genetic perturbations to tumor microenvironments.
- *SpatialData: an open and universal data framework for spatial omics.*
  Marconato, L.*, Palla, G.*, Yamauchi, K.A.*, Virshup, I.*, **Heidari, E.**, et al. *Nature Methods* 22(1):58–62, 2025.
  Contributor: designed and implemented the flagship multi-layer breast cancer analysis demonstrating a universal grammar for multimodal spatial omics; heavy debugging and system maturation.
- *Supervised spatial inference of dissociated single-cell data with SageNet.*
  Heidari, E., Lohoff, T., Tyser, R.C.V., Marioni, J.C., Robinson, M.D., Ghazanfar, S. bioRxiv, 2022.
  Lead developer: introduced graph-attention-based spatial inference; repurposed GNNs for neighborhood-aware representation learning; outperformed Tangram and NovoSpaRc.
- *An end-to-end workflow for high-throughput discovery of clinically relevant insights from large biomedical datasets.*
  Heidari, E., Sadeghi, M.A., Meresht, V.B., et al. bioRxiv, 2020.
  Lead developer: designed a scalable, modular analytics pipeline for population-scale biomedical data; emphasized reproducibility and automation.
- *Pin1 regulatory miRNAs as novel candidates for Alzheimer's disease treatment.*
  Heidari, E., Siavashani, E.S., Rasooli, M., et al. bioRxiv, 2018.
  Lead analyst: statistical- and meta-analysis identifying candidate regulatory miRNAs.

- *snRNA-seq stratifies multiple sclerosis patients into distinct white matter glial responses.*
  Macnair, W., Calini, D., Agirre, E., **Heidari, E.**, et al. *Neuron* 113(3):396–410.e9, 2025.
  Contributor: architected and drafted large-scale sc/snRNA-seq pipelines (200+ patients, 1M cells), including integration, cell type annotation, and downstream statistical analysis.
- *Meta-analysis of single-cell method benchmarks reveals the need for extensibility and interoperability.*
  Sonrel, A., Luetge, A., Soneson, C., Mallona, I., Germain, P.L., **Heidari, E.**, et al. *Genome Biology* 24(1):119, 2023.
  Contributor: analysis and synthesis of benchmarking results; advocated to software design principles.

* equal contribution

TALKS, POSTERS & COMMUNITY

Talks:

- **Segger: Fast and accurate cell segmentation in spatial transcriptomics** — ISMB 2025 (Liverpool); ESSB 2025 (Heidelberg); scverse 2024 (Munich); MOPITAS 2024 (Copenhagen)
- **Supervised spatial inference of dissociated single-cell data with SageNet** — CSHL Genome Informatics 2021 (online)
- **scGCN: A geometric deep learning framework on single-cell gene networks** — EuroBioC 2020 (online)

Posters:

- **Segger** — Single-Cell Genomics 2025 (Stockholm); ESSB 2024 (Berlin)
- **SageNet** — Single Cell Genomics 2023 (Bern)
- **scGCN** — ISMB/ECCB 2019 (Basel)

Hackathons:

- **scverse × Owkin Hackathon 2025** — Paris (project lead)
- **SpaceHack Germany 2024** — Berlin (project lead)
- **SpaceHack Germany 2022** Lutherstadt-Wittenberg (project lead)
- **CSAMA 2019** — Brixen, Italy

REFERENCES

- Dr. Oliver Stegle (EMBL and DKFZ Heidelberg) — oliver.stegle@embl.de
- Dr. Moritz Gerstung (DKFZ Heidelberg) — moritz.gerstung@dkfz-heidelberg.de
- Prof. Mark Robinson (University of Zurich) — mark.robinson@mls.uzh.ch
- Dr. John Marioni (Genentech) — marioni@ebi.ac.uk