DANIEL L. ZEIBERG, PhD

Cambridge, MA 02141 | 609-410-5967 | dzeiberg@me.com | https://www.linkedin.com/in/DanielZeiberg

PROFESSIONAL SUMMARY

Expert in machine learning (ML), artificial intelligence (AI), and computational biology with 7+ years of experience developing data-driven approaches to complex problems in biology and medicine. Demonstrated success in designing novel ML models for functional protein design and interpreting the effects of genetic variants.

KEY SKILLS

Protein design | Variant interpretation | Protein function optimization | Core ML | Deep learning | Probabilistic modeling | Generative modeling | Representation learning | Protein structure prediction | Python | PyTorch | TensorFlow | Git | Translational ML | Cloud computing | Technical presentation | Generative AI | Protein-protein interactions | Large language models | Knowledge graphs

PROFESSIONAL EXPERIENCE

INSTITUTE FOR EXPERIENTIAL AI | Boston, MA

Senior Machine Learning Engineer **Dec. 2025 – Present**

- Developed pipelines to predict protein-ligand binding affinities, enabling prioritization of enzyme candidates for experimental validation
- Deployed an interactive variant interpretation platform integrating knowledge graphs and LLMs to provide evidence-based clinical insights with interpretable predictions for genetic variants
- Trained flow-matching generative models for de novo protein design, optimizing sequences for target functions and structural constraints

Postdoctoral Research Fellow, AI + Life Sciences **Jan. 2025 – Dec. 2025**

- Develop advanced structure- and sequence-based models to identify and optimize enzyme activities for industrial biotechnology applications.
- Design and implement deep learning frameworks to identify catalytic sites, discover novel enzymatic activities, and accelerate enzyme discovery and candidate selection.
- Mentor internal teams on applying deep learning to molecular dynamics and protein conformational modeling, supporting functional protein analysis.

NORTHEASTERN UNIVERSITY | Boston, MA

Graduate Researcher, Khoury College of Computer Sciences **Sep. 2018 – Dec. 2024**

- Created core ML and probabilistic models to predict the phenotypic impact of genetic variants, enabling improved clinical interpretation.
- Collaborated with experimentalists to assess the performance of computational models in predicting the impact of genetic variants on human protein interaction networks
- Developed statistical methods to convert functional assay scores into probabilistic classifications for 40 rare disease-associated genes.
- Applied deep learning and gene-phenotype networks to identify diagnostic variants, achieving 80% accuracy in the Critical Assessment of Genome Interpretation’s Rare Genomes Project challenge.
- Published a comprehensive dataset predicting functional effects for over 73 million variants across 5,000 genes relevant to rare disease diagnosis.
- Improved drug-target interaction prediction accuracy by 12% through structure-based graphical models.

- Trained deep learning models linking genetic variants to rare disease phenotypes, increasing classification accuracy by 17%.
- Led an ML and bioinformatics bootcamp for 21 undergraduates, designing and delivering foundational curriculum and hands-on training.

EDUCATION

NORTHEASTERN UNIVERSITY | Boston, MA
Khoury College of Computer Sciences **Sep. 2018 – Dec. 2024**
Doctor of Philosophy, Computer Science
Thesis: Learning Calibrated Classifiers from Nonrepresentative Data

UNIVERSITY OF MICHIGAN | Ann Arbor, MI
Computer Science and Engineering Department, College of Engineering **Sep. 2014 – Apr. 2018**
Bachelor of Science in Engineering, Computer Science
Highlight: Developed a cutting-edge ML model to identify hospital patients at risk for acute respiratory distress syndrome, removing the need for manual, labor-intensive data review by clinicians.

PUBLICATIONS

Zeiberg D, Tejura M, McEwen AE, Fayer S, Pejaver V, Rubin AF, Starita LM, Fowler DM, O'Donnell-Luria A, Radivojac P. Gene-based calibration of high-throughput functional assays for clinical variant classification. *bioRxiv* [Preprint]. 2025 May 4:2025.04.29.651326.

Jain, S., Trinidad, M., Nguyen, T.B. *et al.* Evaluation of enzyme activity predictions for variants of unknown significance in Arylsulfatase A. *Human Genetics*. (2025) 144: 295-308.

Chen Y, Jain S, **Zeiberg D**, Iakoucheva LM, Mooney SD, Radivojac P, Pejaver V. Multi-objective prioritization of genes for high-throughput functional assays towards improved clinical variant interpretation. *Pacific Symposium on Biocomputing*. (2023) 28: 323-334.

Zeiberg D, Jain S, Radivojac P. Leveraging structure for improved classification of grouped biased data. *AAAI Conference on Artificial Intelligence*, AAAI 2023, pp.11113-11120, Washington, DC, U.S.A., February 2023.

The Critical Assessment of Genome Interpretation Consortium. Critical assessment of variant prioritization methods for rare disease diagnosis within the Rare Genomes Project. *Human Genomics* (2024) 18(1): 44.

Lugo-Martinez J, **Zeiberg D**, Gaudelet T, Malod-Dognin N, Przulj N, Radivojac P. Classification in biological networks with hypergraphlet kernels. *Bioinformatics* (2021) 37(7): 1000-1007.

IGVF Consortium. Deciphering the impact of genomic variation on function. *Nature* (2024) 633(8028): 47:57.

Zeiberg D, Jain S, Radivojac P. Fast nonparametric estimation of class proportions in the positive-unlabeled classification setting. *AAAI Conference on Artificial Intelligence*, AAAI 2020, pp. 6729-6736, New York, New York, U.S.A., February 2020.

Zeiberg D, Prahlad T, Nallamothu BK, Iwashyna TJ, Wiens J, Sjoding MW. Machine learning for patient risk stratification for acute respiratory distress syndrome. *PLOS ONE* (2019) 14(3): e0214465.