

ANNA HART

770-329-0761 – anna19@illinois.edu - [linkedin.com/in/ahart29/](https://www.linkedin.com/in/ahart29/)

EDUCATION

University of Illinois Urbana-Champaign

Fall 2023 - present

PhD Student in Computer Science

- Co-advised by Heng Ji (Natural Language Processing) and Huimin Zhao (Chemical Engineering)

Georgia Institute of Technology

December 2022

B.S. in Computer Science – 3.97 GPA

Minor in Chemistry and Biochemistry

- Concentration: Intelligence, Modeling & Simulation

RESEARCH EXPERIENCE

University of Illinois, Urbana-Champaign

Fall 2023 – current

- Research interests: AI for Science; Natural Language Processing for Bioengineering
- Research natural language processing algorithms for biochemical and bioengineering applications
- Research protein language model interpretability, analyzing attention specialization and performance differences between protein language models and general large language models
- Researched algorithms inferring tumor evolution and phylogenies

Summer Undergraduate Research Fellowship, Mayo Clinic

Summer 2021, 2022

- Designed scalable, reproducible pipeline for clustering and constructing phylogenetic trees of cancer mutations

Undergraduate Research, Laboratory for Synthetic Immunity

Spring, Summer, Fall 2019

- Conducted biomedical experiments involving immunotherapy and gene cloning

Vertically Integrated Projects, Global Health Device Innovation

Fall 2021 – Fall 2022

- Conducted biomedical experiments involving immunotherapy and gene cloning
- Researched solutions for creating an automated system for diagnosing soil transmitted helminth infections

PUBLICATIONS

- [In Preparation] **Anna Hart**, Jeonghwan Kim, Huimin Zhao, Heng Ji. “Rethinking Natural Language Processing for Proteins: How Protein Language Models Learn Differently” NeurIPS (2025)
- Leah Weber, **Anna Hart**, Idoia Ochoa and Mohammed El-Kebir. “Pharming: Joint Clonal Tree Reconstruction of SNV and CNA Evolution from Single-cell DNA Sequencing of Tumors.” 29th Annual International Conference on Research in Computational Molecular Biology (2025)
- Gamboa, Lena, Erick V. Phung, Haoxin Li, Jared P. Meyers, **Anna C. Hart**, Ian C. Miller, and Gabriel A. Kwong. “Heat-Triggered Remote Control of CRISPR-dCas9 for Tunable Transcriptional Modulation.” ACS Chemical Biology (2020)

ORAL PRESENTATIONS

- **Hart, Anna** “Rethinking NLP For Proteins: How Protein Language Models Learn Differently” AI For Basic Science Workshop, University of Illinois Urbana-Champaign (January 2025)
- **Hart, Anna** “Unraveling the 2024 Nobel Prize in Chemistry – AlphaFold” AI + Genomics Seminar, University of Illinois Urbana-Champaign (October 2024)

POSTER PRESENTATIONS

- **Hart, Anna**, Huimin Zhao and Heng Ji. “Predicting Rate-Limiting Enzymes with Multimodal Large Language Models.” Center for Advanced Biotechnology and Bioproducts Innovation Retreat (July 2024)
- **Hart, Anna**, Chuanyi Zhang, and Nicholas Chia. “Scalable Containerized Pipeline for Construction of Tumor Phylogeny from Mutation Frequencies.” Summer Undergraduate Research Fellowship Poster Session, Mayo Clinic (August 2022)
- **Hart, Anna**, Patricio Jeraldo, John Kalantari, and Nicholas Chia. “Building Phylogenetic Trees from Clustered Mutations in Colorectal Cancer with Probabilistic Machine Learning.” Summer Undergraduate Research Fellowship Poster Session, Mayo Clinic (August 2021)

WORK EXPERIENCE

Engineering Development Group Intern, MathWorks

Spring 2021

- Designed and implemented formal interface of tool in MATLAB to parse sections of text and check against standards
- Integrated tool and class-based unit tests into source control and automated building and testing environment

Computing for Engineers Teaching Assistant, Georgia Tech

Summer & Fall 2019 and 2020, Fall 2021

- Led software development team to produce and maintain software used in course activities
- Taught weekly recitation for ~40 students to learn introductory coding and held 3 office hours/week
- Developed exam and homework problems

STAR Tutors

Spring – Summer 2023

- Tutored high school and college students in chemistry, math, and biology

SKILLS AND RELEVANT COURSEWORK

Machine Learning • Advanced Natural Language Processing • Large Language Models • Protein Language Models • Graph Neural Networks • Computer Vision • Computational Modeling Algorithms • Generative AI

Languages and Workflows • Python (Scikit-learn, Pandas) • PyTorch • C++ • Git • Docker • Conda • Snakemake • GPU Acceleration • HuggingFace • Linux • Bash

Bioinformatics • Machine Learning for Bioinformatics • Computational Cancer Genomics • Systems Physiology • Protein Language Models • Antibody Design • TCGA • UniProt • Next Generation Sequencing • Single-Cell DNA Sequencing

Chemistry • Organic Chemistry I and II • Biochemistry I and II • Organic Chemistry Lab • Chemistry Communication

Mathematics • Linear Algebra • Differential Equations • Multivariable Calculus • Combinatorics • Discrete Math • Algorithms • Probability & Statistics

SERVICES

Reviewer – Language + Molecules ACL Conference Workshop

Summer 2024

Reviewer – AI4Research Workshop @ AAAI 2025

December 2024

OUTREACH

AI + Genomics Seminar Co-Organizer

Spring 2024 – current

- Leading and organizing interdisciplinary seminar connecting students in machine learning with students in biology-related fields

AI + Basic Science Planning Committee

Fall 2024 – current

- Institute-wide faculty-led initiative for strategizing the role of AI in basic science and facilitating interdisciplinary collaborations
- Planning mentorship program for graduate and undergraduate students

Molecule Maker Lab Institute Escape Room

Fall 2024

- Engaged public in learning about molecular synthesis through an escape-room activity

Imagine STEM Molecule Maker Lab Institute

Fall 2024

- Led group of 6-8th graders in creating an organic photovoltaic solar cell

AWARDS

- **Highest Honors, Georgia Institute of Technology** (2022)
- **Zell Miller Scholarship** (2018-2022)
- **National Merit Scholarship** (2018)