Hamza Tahboub

linkedin.com/in/hamzatahboub
Availability: June–December, 2024

tahboub.h@northeastern.edu
+1 (858) 371-8866
Boston, MA

Education

- **Northeastern University, Khoury College of Computer Sciences** — Boston, United States
  Bachelor of Science — Computer Science and Mathematics; GPA: 3.93, Dean's List — *2021 – 2025*

  Relevant Courses: Adv. Programming with Data, Practical Neural Networks, Adv. Linear Algebra, Statistics and Stochastic Processes, Intro to Data Management and Processing (graduate), Information Retrieval, Programming in C++, Adv. Technical Writing

- **King's Academy** — Madaba, Jordan
  High School Degree — AP Program, Honors Roll — *2017 – 2021*

  Completed all offered advanced math courses, including Multivariable Calculus, Number Theory, AP Statistics, and Linear Algebra.

Skills Summary

- **Languages**: Python, Java, C++, SQL, LaTeX, Racket (Lisp dialect), MATLAB, HTML/CSS, C, JavaScript
- **Tools**: Numpy, Pandas, Git, PyTorch, TensorFlow, Apache Spark, Hadoop, Plotly, Slurm, Matplotlib, Jupyter, MySQL, SQLite, MongoDB, XPath, Azure, Google Cloud (GCP), Amazon Web Services (AWS)
- **Concepts**: Model Distillation, Transformers & Attention, Fine-tuning, MLE, Language Models, MapReduce

Experience

- **Genentech, Subsidiary of Roche** — San Francisco, CA
  Natural Language Processing Co-op — *July 2023 – Present*
  - Distilling capabilities from large DL models to smaller ones, **reducing long-term computation costs** by over 95% and reducing reliance on closed source models.
  - **Curated a synthetic dataset** of over 100k samples for **training transformer-based models** for specialized medical QA.
  - Contributing to experimental **medical NLP research**, designing experiments and reviewing new methods.
  - Developed an embedding-based semantic search engine to extract information from medical corpora of over 150k documents.

- **Professor Huaizu Jiang's Visual Intelligence Lab — Northeastern University** — Boston, MA
  Research Assistant — *Aug 2022 – Present*
  - Currently **leading a new project** on egocentric video understanding.
  - Working toward presenting our novel ideas for **action anticipation in a long-form egocentric setting**.
  - Previously worked on projects focused on visual commonsense reasoning abilities – the ability to answer commonsense questions regarding images.
  - Developed systems to easily apply new prompt templates to QA datasets, evaluated novel methods from papers, and sampled over 10 different models on relevant tasks.

- **Khoury College of Computer Sciences — Northeastern University** — Boston, MA
  Teaching Assistant — *Jan 2022 – July 2022*
  - Mentored over 80 students for two semesters as a TA for the Fundamentals of Computer Science course.
  - Led labs in which students practiced new material, provided guidance in office hours, and graded assignments.

Projects

- **MarkovPatch: Random Image Masks for Attention-based Explainable AI** — *Fall 2022*
  - Applied image masks to a pre-trained classification deep neural network during inference to identify spatial features of significance.
  - Developed a random mask generator by sampling a second-order Markov chain. The distribution parameters were adjusted to alter the size and spatial correlation of the masks' patches.
  - Applying this method to a cat classifier, **demonstrated that the model is paying more attention to contour features** and specific feline attributes like ears and whiskers.
- **Assigning TAs to Labs Using Evolutionary Computing** — *Fall 2022*
  - Formulated matching teaching assistants to lab sessions with constraints as a **cost optimization** problem.
  - Developed a program that applies evolutionary computing principles to search for the minimum-cost solution.
  - Wrote scoring functions to quantify progress and agents that "mutated" solutions to search for better ones.
  - Was able to **reliably and quickly find optimal solutions** using the developed method.

Certifications

- **Deep Learning Specialization**: Neural Networks and Deep Learning Fundamentals, Hyperparameter Tuning, Regularization and Optimization, Structuring Machine Learning Projects, Convolutional Neural Networks, Sequence Models. (online course taught by Professor Andrew Ng)

Interests

- Running — Road Cycling — Star Wars — Skiing — Learning French through books and podcasts