

Gabriel Franco

LinkedIn : www.linkedin.com/in/gaabrielfranco E-mail: gvfranco@bu.edu
Github : <https://www.github.com/gaabrielfranco> Mobile: +1 (617) 685-9112
Personal Webpage: <https://cs-people.bu.edu/gvfranco/> Address: 665 Commonwealth Ave, Boston, MA

- EDUCATION**
- Boston University**, Boston, MA
PhD, Computer Science
Sep 21 - Present
Advisor: Mark Crovella
- Universidade Federal de Viçosa**, Viçosa - MG, Brazil
MSc, Computer Science
Aug 2018 - Jun 2021
Advisor: Giovanni Comarela
- Universidade Federal de Viçosa**, Viçosa - MG, Brazil
B.S., Computer Science
July 2018
- RESEARCH INTERESTS** Mechanistic Interpretability, LLMs, Machine Learning, Weakly Supervised Learning
- SCHOLARSHIP**
- Boston University Research Scholarship, PhD Student** Sep 21 - Present
Worked on the Learning from Label Proportions problem. Currently, working on mechanistic interpretability and evaluation of Large Language Models (LLMs).
- CAPES Research Scholarship, Master Student** Aug 18 - Jul 20
Worked on better cross-validation strategies for the Learning with Label Proportions problem.
- CNPq Research Scholarship, Scientific Initiation (PIBIC/CNPq)** Aug 17 - Jul 18
Worked with improving the malaria model on Autosimmune, a multiagent human immune system simulator made in JAVA.
- CNPq Research Scholarship, Scientific Initiation (PIBIC/CNPq)** Aug 15 - Jul 16
Developed the Bio-Oracle software. Bio-ORACLE is a software written in JAVA which uses data mining techniques to help decision-making in bioethics. Bio-ORACLE was developed at the Laboratory of Epidemiological and Computational Methods in Health of Department of Medicine and Nursing in Universidade Federal de Viçosa.
- FAPEMIG Scientific Initiation Fellow (PIBIC/FAPEMIG)** Mar 15 - Jul 15
Worked with modeling the plasmodium on Autosimmune, a multiagent human immune system simulator made in JAVA.
- INDUSTRY EXPERIENCE**
- Data Scientist Intern at Microsoft** May 24 - Aug 24
- Worked on fine-tuning Small Language Models (SLMs) for Windows products using different modalities (text and image).
- Data Scientist at SEEK** Sep 20 - Jul 21
- Designed, implemented, and maintained recommender systems to provide personalized job ad recommendations for customers.

- Improved average response time of a recommender system by more than 50%.
- Increased business metrics for a recommender system with statistical significance after an A/B test.

Data Scientist at Localiza

Jul 20 - Sep 20

- Developed machine learning models to identify possible reliable customers.
- Performed analysis of clients' behaviors in the platform.
- Presented results to stakeholders.

**TEACHING
EXPERIENCE**

Boston University

Sep 23 - Dec 23

CDS DS 701: Tools for Data Science. Teaching Assistant.

Activities:

- Prepare and teach discussions to the students.
- Design the practical homeworks and the midterm exam (Kaggle style competition).

Departamento de Informática (UFV)

Nov 19

L^AT_EX short course. Instructor.

Universidade Federal de Viçosa

Mar 19 - Jul 19

INF 100 - Introduction to Programming I. Teaching Assistant.

**VOLUNTEER
EXPERIENCE**

NoBugs Empresa Júnior de Informática

Jan 17 - Jan 18

NoBugs is a junior enterprise of the UFV Computer Science course. We made low-cost web systems to regional enterprises.

PUBLICATIONS

- Mechanistic Interpretability:
 1. Franco, Gabriel, and Mark Crovella. "Sparse Attention Decomposition Applied to Circuit Tracing." <https://arxiv.org/abs/2410.00340> **Under submission.**
- LLM evaluation:
 1. Calais and Franco et al. "Beyond accuracy: understanding the performance of LLMs on exams designed for humans." **Under submission.**
- Learning from Label Proportions (LLP):
 1. Franco, Gabriel, Giovanni Comarela, and Mark Crovella. "The Complex Landscape of LLP Evaluation." **Under submission.**
 2. Franco, Gabriel, Mark Crovella, and Giovanni Comarela. "Dependence and Model Selection in LLP: The Problem of Variants." Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023.
- Others:
 1. Franco, G., Ribeiro, M. H. F., and Comarela, G. "Towards an Interpretable Metric for DOTA 2 Players: An Unsupervised Learning Approach." 8th Brazilian Conference on Intelligent Systems (BRACIS). 2019

2. Gomes, Andréia Patricia, et al. "Plasmodium Falciparum Infection: In Silico Preliminary Studies." Abakós 5.1 (2016): 63-83.

**OTHER
ACADEMIC
PRODUCTIONS**

1. Siqueira-Batista, Rodrigo, et al. Parasitologia: Fundamentos e Prática Clínica. Guanabara, 2020. ISBN 9788527735735. (I co-wrote Chapter 4 about the computational approach in the study of parasitic diseases)
2. Comarela, G.; Franco, G. ; Trois, C. ; Liberato, A. ; Martinello, M. ; Corrêa, J. H. ; Villaça, R. Introdução à Ciência de Dados: Uma Visão Pragmática utilizando Python, Aplicações e Oportunidades em Redes de Computadores SBRC 2019 (Short Course)

SKILLS

Python; Git; TransformerLens; Numpy; Scipy; Pandas; Matplotlib; Seaborn; Scikit-learn; Statsmodel; Pytorch; Transformers; PEFT; LoRA; Fine-tuning LLM; TransformerLens; Problem Solving; Research; Machine Learning; Data Mining; Probability; Mechanistic Interpretability; C/C++

SERVICE

- Reviewer ICLR 2025

AWARDS

- KDD'22 Student Travel Award

**ADDITIONAL
ACTIVITIES**

- ACM International Collegiate Programming Contest - Regionals - 2014, 2015, 2017
- Minas Gerais State Programming Contest - 2014, 2015, 2017

**RELEVANT
COURSES**

Boston University:

- CS542 - Machine Learning
- CS565 - Algorithmic Data Mining
- CS537 - Randomness in Computing
- CS655 - Graduate Computer Networks
- CS511 - Formal Methods 1
- DS563 - Algorithmic Techniques for Taming Big Data
- LX690 - Metrics and Evaluation in Natural Language Processing (Audited)

Universidade Federal de Viçosa:

- INF623 - Artificial Intelligence
- INF610 - Data Structures and Algorithms
- INF723 - Data Visualization

LANGUAGES

- Portuguese (Native or bilingual proficiency)
- English (Professional working proficiency)
- Spanish (Elementary proficiency)

REFEREES

Prof. Mark Crovella
Boston University
crovella@bu.edu

Prof. Giovanni Ventrone Comarela
Universidade Federal do Espírito Santo
gc@inf.ufes.br

Emanuel Vale
SEEK AI
evale@seek.com.au