Matteo Tolloso

25 years

☎ *+39 388 3749544*
✉ *matteotolloso@gmail.com*
in *matteotolloso*
matteotolloso
✉ *matteo.tolloso@phd.unipi.it*

Education

November 2024
Present

PhD in Artificial Intelligence, *University of Pisa*

I am a PhD student investigating the fundamental limitations of Graph Neural Networks for scientific applications. My research analyzes the challenges of long-range information propagation in large-scale graphs and, in parallel, focuses on new paradigms for uncertainty quantification. I am developing methods that allow models to diagnose the specific sources of their uncertainty, distinguishing between noisy data, model ignorance, and novel inputs to build more trustworthy AI for molecular simulation and computational health.

September 2022
October 2024

Master Degree in Artificial Intelligence, *University of Pisa*

Grade 110/110 with honors.
This degree path gave me the ability to understand and use the state of the art in all fields of Artificial Intelligence through courses such as “machine learning”, “deep learning” and “data mining”.

September 2019
July 2022

Bachelor Degree in Computer Science, *University of Pisa*

Grade 110/110 with honors.
This degree program has given me a good horizontal knowledge of all fields of computer science, as well as strong mathematical and methodological foundations for deeper studies.

Experience

April 2024
October 2024

Machine Leaning Engineer Intern, *H&M*, Stockholm

Geometric Deep Learning and Continual Learning for personalized product recommendation. I’ve developed here my master’s thesis.

February 2024
May 2024

Tutor for the course “Laboratory 1”, *University of Pisa*

I guided students in mastering programming tools and concepts, particularly in JavaScript and TypeScript. I provide hands-on assistance with debugging and versioning systems like Git. Additionally, I support students in implementing algorithms on various data structures, emphasizing effective problem-solving strategies.

October 2023
April 2024

Research Assistant: single cell RNA sequencing, *University of Pisa*

I worked with a multidisciplinary team of researchers whose objective is to improve the accuracy in the identification of differentially expressed genes tackling the problem from a mathematical, biological, and computational point of view.

August 2023

Research School, *Max Plank Institute*, Saarbrücken, Germany

I’ve been selected to attend the summer school organized by Cornell University, the University of Maryland, and the Max Plank Institute. I had the opportunity to get in touch with some of the best researchers in the world, together with the most talented students. This experience gave me a broad vision of the frontier of science in different fields of computer science.

January 2023
July 2023

Research Fellowship: protein embedding, *University of Pisa*

I focused on discovering similarities, differences, and special properties among various methods for representing genetic sequences. Specifically studying various deep learning models for protein embedding, looking for new properties that they can represent, and comparing them with traditional approaches such as multiple sequence alignment algorithms.

October 2022
December 2022

Tutor for the course “Laboratory 2”, *University of Pisa*

I had the opportunity to help other students with some of the main topics I learned during my bachelor degree such as advanced C programming, fundamentals of operating systems, techniques for coordinating and managing resources, synchronization mechanisms, multi-thread and multi-process applications, memory and file system management.

Publications and Theses

September 2025

“Real-time and personalized product recommendations for large e-commerce platforms”, **M. Tolloso** , *D. Bacciu, S. Mokarizadeh, M. Varesi* , Accepted for publication at the International Conference on Artificial Neural Networks (ICANN) 2025, Kaunas, Lithuania

We present a methodology to provide real-time and personalized product recommendations for large e-commerce platforms, specifically focusing on fashion retail. Our approach aims to achieve accurate and scalable recommendations with minimal response times, ensuring user satisfaction, leveraging Graph Neural Networks and parsimonious learning methodologies. Extensive experimentation with datasets from one of the largest e-commerce platforms demonstrates the effectiveness of our approach in forecasting purchase sequences and handling multi-interaction scenarios, achieving efficient personalized recommendations under real-world constraints.

Springer Link

June 2024

“How much do DNA and protein deep embeddings preserve biological information?”, **M. Tolloso** , *S.G. Galfrè, M. Podda, A. Sîrbu, C. Priami* , Accepted for publication at the 22nd Conference on Computational Methods in Systems Biology (CMSB24), Pisa, Italy

Models that map DNA and protein sequences into deep embeddings have been recently developed. While their ability to improve prediction in downstream tasks has been demonstrated, clear advantages and disadvantages of embedding types, and different means of applying them, are not yet available. In this paper we compare five different models (one for DNA, four for proteins) and different embedding aggregation methods with respect to their ability to preserve evolutionary and functional information, using a hierarchical tree approach. Specifically, we introduce a novel procedure that builds hierarchical clustering trees to assess the relative position of sequences in the embedding latent space, compared to the phylogenetic and functional similarities between sequences. The methods are benchmarked on five different datasets from various organisms. The ESM protein language model and DNABert emerge as best performers in different settings.

Springer Link

May 2024

“From Gene to Cluster: COTAN v2 for Improved scRNA-seq Analysis and Unsupervised Clustering”, *S. G. Galfrè, M. Fantozzi, I. Testa,* **M. Tolloso**, *A. Alberti, A. Sîrbu, F. Morandin, C. Priami*, Presented at Bioinformatics Italian Society (BITS) Conference 2024, Trento, Italy

The COTAN library provides a unique approach in analyzing scRNA-seq data. Its emphasis on gene co-expression at the single-cell level, including for lowly expressed genes, promises to offer significant advantages in extracting meaningful insights from sparse datasets. COTAN (v2) was rigorously evaluated across various stages of the scRNA-seq workflow: from its core gene-gene correlation analysis to cluster marker identification.

October 2024

Real-time and personalized product recommendations for large e-commerce platforms, *Master Thesis*, supervisors: Prof. Davide Bacciu, Dr. Mokarizadeh Shahab

July 2022

Automated biomedical literature mining to investigate ZTTK syndrome, *Bachelor Thesis*, supervisor: Prof. Alina Sîrbu

Using automated literature analysis, machine learning, and graph theory algorithms, we expanded our knowledge of Zttk by incorporating information on similar diseases and exploring the broader context of the SON gene.

GitHub link

Projects

Pandas
Scikit-learn
Deep learning
Tensorflow
Explainability

Bots on Twitter

This teamwork has focused on cleaning the data from Twitter API from unwanted noise that may hinder knowledge discovery; applying and comparing different clustering and classification techniques for deeper insight and learning; finally, investigate techniques to explain and interpret "black box" predictions to understand its results better.

GitHub link

Classical AI
Planning
Heuristics

Google hashcode: Router Placement

In this challenge we tackled a real-world problem with high complexity using classical artificial intelligence algorithms combined with ad-hoc heuristics such as genetic algorithms, multi-objective A* and simulated annealing.

GitHub link

C++
FastFlow

Huffman code

Fast and scalable implementation of Huffman Code using native C++ threads and FastFlow library.

GitHub link

NLP
Pytorch

Automatic question generation

During my involvement in a question generation project, I collaborated on an analysis of various models aimed at automatically creating questions from textual inputs, focused primarily on answer-aware and answer-agnostic question generation tasks.

GitHub link

Languages

English (B2), Italian (native)