Giuseppe **Attanasio**

PH.D. STUDENT · POLITECNICO DI TORINO
Corso Duca degli Abruzzi, 24, 10129, Torino, Italy

giuseppe.attanasio@polito.it
gattanasio.cc | g8a9 | giuseppe-attanasio | @peppeatta | Scholar

"Computers aren't the thing. They are the thing that gets us to the thing." - Halt And Catch Fire

Bio

I'm a third-year Ph.D. student at the Department of Control and Computer Engineering of Polytechnic of Turin. Presently, I work on the understanding and regularizion of large Language Models in the context of bias and fairness application. In the past, I worked on modeling and forecasting financial time series.

I currently live in Turin. I love reading Sci-Fi and playing basketball, while circumstances led me to discover that I'm not so bad at cooking. I also like DIY and automating boring stuff. Besides that, I am a passionate learner. I spend countless hours on lectures and tutorials about languages, frameworks, and technologies which I deem interesting.

I update much more frequently my personal website with projects, recent publications, and other fun stuff.

Research interests. Sequence Modeling and Forecasting Algorithms, Neural Language Models, Explainable AI, ML-enabled Quantitative Trading

Education

Politecnico di Torino *Torino, Italy*
PH.D. AT DEPARTMENT OF CONTROL AND COMPUTER ENGINEERING *2018 - 2021 (exp.)*
- Advisor: Elena Baralis

Politecnico di Torino *Torino, Italy*
MD IN COMPUTER ENGINEERING, DATA SCIENCE TRACK *2016 - 2018*
- Grade: 110/110 *cum Laude*

Politecnico di Torino *Torino, Italy*
BD IN COMPUTER ENGINEERING *2013 - 2016*
- Grade: 110/110

Publications

Bianchi, F., Attanasio, G., Pisoni, R., Terragni, S., Sarti, G., Lakshmi, S. (2021). Contrastive Language-Image Pre-training for the Italian Language. ArXiv:2108.08688 [Cs]. http://arxiv.org/abs/2108.08688

Attanasio, G., Cagliero, L., Baralis, E. (2020). Leveraging the explainability of associative classifiers to support quantitative stock trading. Proceedings of the Sixth International Workshop on Data Science for Macro-Modeling, 1–6. https://doi.org/10.1145/3401832.3402679

Attanasio, G., Pastor, E. (2020). PoliTeam @ AMI: Improving Sentence Embedding Similaritywith Misogyny Lexicons for Automatic Misogyny Identificationin Italian Tweets. In V. Basile, D. Croce, M. Maro, L. C. Passaro (Eds.), EVALITA Evaluation of NLP and Speech Tools for Italian—December 17th, 2020 (pp. 48–54). Accademia University Press. https://doi.org/10.4000/books.aaccademia.6807

Cagliero, L., Garza, P., Attanasio, G., Baralis, E. (2020). Training ensembles of faceted classification models for quantitative stock trading. Computing, 102(5), 1213–1225. https://doi.org/10.1007/s00607-019-00776-7

Attanasio, G., Giobergia, F., Pasini, A., Ventura, F., Baralis, E., Cagliero, L., Garza, P., Apiletti, D., Cerquitelli, T., Chiusano, S. (2020). DSLE: A Smart Platform for Designing Data Science Competitions. 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), 133–142. https://doi.org/10.1109/COMPSAC48688.2020.00026

Attanasio, G., Cagliero, L., Garza, P., Baralis, E. (2019a). Combining News Sentiment and Technical Analysis to Predict Stock Trend Reversal. 2019 International Conference on Data Mining Workshops (ICDMW), 514–521. https://doi.org/10.1109/ICDMW.2019.00079

Attanasio, G., Cagliero, L., Garza, P., Baralis, E. (2019b). Quantitative cryptocurrency trading: Exploring the use of machine learning techniques. Proceedings of the 5th Workshop on Data Science for Macro-Modeling with Financial and Economic Datasets, 1–6. https://doi.org/10.1145/3336499.3338003

Attanasio, G., Cannavò, A., Cibrario, F., Lamberti, F., Montuschi, P., Paravati, G. (2017). HOT: Hold your own tools for AR-based constructive art. 2017 IEEE Symposium on 3D User Interfaces (3DUI), 256–257. https://doi.org/10.1109/3DUI.2017.7893369

I believe in open-access research. Most of my papers are free to read, and the experimental code is open-sourced whenever possible. All the pointers are collected on my website.

Research projects

CLIP-Italian. During the HuggingFace JAX Community Week, we had the chance to specialize OpenAI's CLIP for the Italian language. CLIP is one of the most recent multi-modal models that connect images and text. The original model is limited to English, so we decided to extend its capabilities. We hence presented the first CLIP model for the Italian Language, trained on more than 1.5 million high-quality image-text pairs. The released model outperforms the multilingual CLIP model on the tasks of image retrieval and zero-shot classification.

The project was selected among the 15 finalist projects - out of 100 - of the Flax/JAX Community Week organized by HuggingFace in partnership with Google and received a special nominee in the final evaluation round.

Telepass's KMaster Fleet Management. I took part in a joint research project between my research group and the company KMaster (Telepass SpA). We designed and implemented an end-to-end machine learning and clustering-based pipeline to characterize driving behaviors and fleet management.

Teaching activities

Academy

Teaching Assistant. I have been a teaching assistant to several courses at Politecnico di Torino.

2021 **Data science lab: processes and methods**, MD in Data Science and Engineering
2020 **Data science lab: processes and methods**, MD in Data Science and Engineering
2020 **Business Intelligence for Big Data**, MD in Industrial engineering and management
2019 *** Data science lab: processes and methods**, MD in Data Science and Engineering
2019 **Business Intelligence for Big Data**, MD in Industrial engineering and management
2018 **Databases**, BD in Computer Engineering

* While Introduction to Databases and Business Intelligence were pre-existent and consolidated courses, I have been a major contributor in shaping *Data Science Lab: process and methods*, launched in September 2019 and currently one of the central courses in the Data Science and Engineering master degree at Politecnico.

The course is the first introduction to the Python programming language and the basics of Data Science and Machine Learning libraries for MD students. We worked hard to provide students with comprehensive exercises and solutions (10 laboratories, for a total of 60+ pages of lab exercises and 250+ pages of solutions). All the material is freely available on the course website.

Research Bites. I contributed to the launch of Research Bites, a series of short research talks and seminaries held by PhD students for students of the course Data Science Lab: process and methods. The goal of RB is to disseminate cutting-edge research topics, in short, high-level pills. The series is now in its second edition.

Industry

I have been a consultant to several companies.

2021 **Reply: Digital Services, Technology and Consulting**, Data Visualization with Python — *Turin (online)*

2020 **Applied Mechatronic Engineering & Technologies**, Python technologies for Data Analytics — *Turin*

Peer reviewing

I reviewed at least one work submitted to the following venues:

- ACM KDD SIGKDD Conference On Knowledge Discovery And Data Mining. Editions: 2020, 2021
- ACM SIGMOD/PODS International Conference on Management of Data. Editions: 2021
- IEEE ICDE: IEEE International Conference on Data Engineering. Editions: 2020
- IEEE ICDM: IEEE International Conference on Data Mining. Editions: 2021
- ACM SAC: ACM/SIGAPP Symposium On Applied Computing. Editions: 2021
- EDBT: International Conference on Extending Database Technology. Editions: 2020, 2021
- DaWaK: International Conference on Big Data Analytics and Knowledge Discovery. Editions: 2019, 2021

I reviewed at least one work submitted to the following journals:

- Expert Systems With Applications, Elsevier
- Future Generation Computer Systems, Elsevier
- Machine Learning With Applications, Elsevier

Work Experience

Kupata S.r.l. — *Torino, Italy*

Founder, Shareholder — *Nov. 2016 - 2020*

- Kupata's main goal is to streamline the Lost and Found process. It brings innovation with a solution that helps people in returning items in a simple, secure, and undisclosed way. The business involves a physical object, the Kupa, and a social community that encourages members to act in the right way.

Consoft Sistemi S.p.A. — *Torino, Italy*

Curricular intern — *Mar. 2016 - Jul. 2016*

- I participated to the bootstrap phase of the Consoft's proprietary Knowledge Base platform. We built our solution upon Orange HRM, an open source PHP-based platform.

Skills

I am familiar with *italicized* entries.

Programming C, C++, Python, Java, *C#, JavaScript, PHP, Matlab, SQL*
Framework *Hadoop, Spark*
Scripting Bash, *Awk*
DevOps *AWS, GCP, Docker*
Front-end *Hugo, React, Dash*
Graphics Inkscape, GIMP, *Blender, Unity 3D*
Languages Italian, English, *Spanish, French*