Giuseppe **Attanasio**

Postdoctoral Researcher · Instituto de Telecomunicações

Instituto Superior Técnico, Av. Rovisco Pais, 1, 1049-001, Lisboa

⌂ gattanasio.cc | ✉ giuseppeattanasio6@gmail.com
🔳 g8a9 | 🔗 giuseppe-attanasio | 🎓 Scholar | bluesky

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

Bio

I am a Postdoctoral Researcher in the SARDINE group at Instituto de Telecomunicações, Lisbon, headed by scientific director André Martins. My research interests lie at the intersection of interpretability, fairness, and safety. I previously joined the MilaNLP group, working with Debora Nozza and Dirk Hovy.

I care about the impact of societal biases infused in AI technologies, about which I gave talks and published modest science. I believe education and associationism play essential roles. I co-founded a student team and an online community to connect passionate people. I believe deep understanding is rooted in hands-on exercise: I like coding for my research projects, developing automation tools, and writing practical tutorials.

I update my website more frequently than this document, with recent professional news, projects, and publications.

Education

University of Lisbon   *Lisbon, Portugal*
Postdoc at Instituto de Telecomunicações   *Feb. 2024 - ongoing*
- PI: André Martins

Bocconi University   *Milan, Italy*
Postdoc at Department of Computing Sciences   *Oct. 2022 - Feb. 2024*
- PI: Debora Nozza
- Advisor: Dirk Hovy

Politecnico di Torino   *Turin, Italy*
Ph.D. at Department of Control and Computer Engineering   *Nov. 2018 - Oct. 2022*
- Advisor: Elena Baralis

Politecnico di Torino   *Turin, Italy*
MD in Computer Engineering, Data Science track   *Oct. 2016 - Oct. 2018*
- Grade: 110/110 *cum Laude*

Politecnico di Torino   *Turin, Italy*
BD in Computer Engineering   *Oct. 2013 - Oct. 2016*
- Grade: 110/110

Publications

I firmly believe in open-access research. Most of my papers are free to read, and code and data are open-source under permissive license, whenever possible. All the pointers are collected on my website.

Selected

Attanasio, Giuseppe, Savoldi, Beatrice, Fucci, Dennis, and Hovy, Dirk. "Twists, Humps, and Pebbles: Multilingual Speech Recognition Models Exhibit Gender Performance Gaps". In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Ed. by Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 21318–21340. DOI: `10.18653/v1/2024.emnlp-main.1188`. URL: `https://aclanthology.org/2024.emnlp-main.1188/`

Attanasio, **Giuseppe**, Plaza del Arco, Flor Miriam, Nozza, Debora, and Lauscher, Anne. "A Tale of Pronouns: Interpretability Informs Gender Bias Mitigation for Fairer Instruction-Tuned Machine Translation". In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Ed. by Houda Bouamor, Juan Pino, and Kalika Bali. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 3996–4014. DOI: `10.18653/v1/2023.emnlp-main.243`. URL: `https://aclanthology.org/2023.emnlp-main.243`

Attanasio, **Giuseppe**, Nozza, Debora, Hovy, Dirk, and Baralis, Elena. "Entropy-based Attention Regularization Frees Unintended Bias Mitigation from Lists". In: *Findings of the Association for Computational Linguistics: ACL 2022*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 1105–1119. DOI: `10.18653/v1/2022.findings-acl.88`. URL: `https://aclanthology.org/2022.findings-acl.88`

Attanasio, **Giuseppe**, Pastor, Eliana, Di Bonaventura, Chiara, and Nozza, Debora. "ferret: a Framework for Benchmarking Explainers on Transformers". In: *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. Dubrovnik, Croatia: Association for Computational Linguistics, May 2023, pp. 256–266. URL: `https://aclanthology.org/2023.eacl-demo.29`

Delobelle, Pieter, **Attanasio**, **Giuseppe**, Nozza, Debora, Blodgett, Su Lin, and Talat, Zeerak. "Metrics for What, Metrics for Whom: Assessing Actionability of Bias Evaluation Metrics in NLP". in: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Ed. by Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 21669–21691. DOI: `10.18653/v1/2024.emnlp-main.1207`. URL: `https://aclanthology.org/2024.emnlp-main.1207/`

Cercas Curry, Amanda, **Attanasio**, **Giuseppe**, Talat, Zeerak, and Hovy, Dirk. "Classist Tools: Social Class Correlates with Performance in NLP". in: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Lun-Wei Ku, Andre Martins, and Vivek Srikumar. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 12643–12655. DOI: `10.18653/v1/2024.acl-long.682`. URL: `https://aclanthology.org/2024.acl-long.682/`

Bianchi, Federico, Suzgun, Mirac, **Attanasio**, **Giuseppe**, Röttger, Paul, Jurafsky, Dan, Hashimoto, Tatsunori, and Zou, James. *Safety-Tuned LLaMAs: Lessons From Improving the Safety of Large Language Models that Follow Instructions*. arXiv:2309.07875 [cs]. Sept. 2023. DOI: `10.48550/arXiv.2309.07875`. URL: `http://arxiv.org/abs/2309.07875` (visited on 09/15/2023)

Röttger, Paul, Kirk, Hannah, Vidgen, Bertie, **Attanasio**, **Giuseppe**, Bianchi, Federico, and Hovy, Dirk. "XSTest: A Test Suite for Identifying Exaggerated Safety Behaviours in Large Language Models". In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Ed. by Kevin Duh, Helena Gomez, and Steven Bethard. Mexico City, Mexico: Association for Computational Linguistics, June 2024, pp. 5377–5400. DOI: `10.18653/v1/2024.naacl-long.301`. URL: `https://aclanthology.org/2024.naacl-long.301/`

Koudounas, Alkis, Pastor, Eliana, **Attanasio**, **Giuseppe**, Mazzia, Vittorio, Giollo, Manuel, Gueudre, Thomas, Cagliero, Luca, Alfaro, Luca de, Baralis, Elena, and Amberti, Daniele. "Exploring subgroup performance in End-to-End speech models". In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023

Chronologically

This list has been automatically extracted and parsed from my Google Scholar page on March 30, 2025.

Röttger, Paul, **Attanasio**, **Giuseppe**, Friedrich, Felix, Goldzycher, Janis, Parrish, Alicia, Bhardwaj, Rishabh, Di Bonaventura, Chiara, Eng, Roman, Geagea, Gaia El Khoury, Goswami, Sujata, et al. "MSTS: A Multimodal Safety Test Suite for Vision-Language Models". In: *arXiv preprint arXiv:2501.10057* (2025)

Attanasio, **Giuseppe**, Basile, Pierpaolo, Borazio, Federico, Croce, Danilo, Francis, Maria, Gili, Jacopo, Musacchio, Elio, Nissim, Malvina, Patti, Viviana, Rinaldi, Matteo, et al. "CALAMITA: Challenge the Abilities of LAnguage Models in ITAlian". In: *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024), Pisa, Italy*. 2024

Attanasio, **Giuseppe**, Delobelle, Pieter, La Quatra, Moreno, Santilli, Andrea, and Savoldi, Beatrice. "Itaeval and tweetyita: A new extensive benchmark and efficiency-first language model for italian". In: *CLiC-it 2024: Tenth Italian Conference on Computational Linguistics, Date: 2024/12/04-2024/12/06, Location: Pisa, Italy*. 2024

Pernisi, Fabio, **Attanasio**, **Giuseppe**, and Nozza, Debora. "MONICA: Monitoring Coverage and Attitudes of Italian Measures in Response to COVID-19". In: (2024)

Attanasio, **Giuseppe**, Savoldi, Beatrice, Fucci, Dennis, Hovy, Dirk, et al. "Twists, humps, and pebbles: multilingual speech recognition models exhibit gender performance gaps". In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. 2024

Delobelle, Pieter, **Attanasio**, **Giuseppe**, Nozza, Debora, Blodgett, Su Lin, Talat, Zeerak, et al. "Metrics for what, metrics for whom: assessing actionability of bias evaluation metrics in NLP". in: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. 2024

Attanasio, **Giuseppe**, Delobelle, Pieter, La Quatra, Moreno, Santilli, Andrea, and Savoldi, Beatrice. "Itaeval and tweetyita: A new extensive benchmark and efficiency-first language model for italian". In: *CLiC-it 2024: Tenth Italian Conference on Computational Linguistics, Date: 2024/12/04-2024/12/06, Location: Pisa, Italy*. 2024

Zerva, Chrysoula, Blain, Frédéric, De Souza, José GC, Kanojia, Diptesh, Deoghare, Sourabh, Guerreiro, Nuno M, **Attanasio**, **Giuseppe**, Rei, Ricardo, Orasan, Constantin, Negri, Matteo, et al. "Findings of the Quality Estimation Shared Task at WMT 2024 Are LLMs Closing the Gap in QE?". In: *Proceedings of the Ninth Conference on Machine Translation,* Association for Computational Linguistics. 2024, pp. 82–109

Zaranis, Emmanouil, **Attanasio**, **Giuseppe**, Agrawal, Sweta, and Martins, André FT. "Watching the watchers: Exposing gender disparities in machine translation quality estimation". In: *arXiv preprint arXiv:2410.10995* (2024)

Lardelli, Manuel, Dill, Timm, **Attanasio**, **Giuseppe**, and Lauscher, Anne. "Sparks of fairness: Preliminary evidence of commercial machine translation as English-to-German gender-fair dictionaries". In: *Proceedings of the 2nd International Workshop on Gender-Inclusive Translation Technologies*. 2024, pp. 12–21

Lardelli, Manuel, Lauscher, Anne, and **Attanasio**, **Giuseppe**. "GeFMT: Gender-Fair Language in German Machine Translation". In: *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 2)*. 2024, pp. 37–38

Lardelli, Manuel, **Attanasio**, **Giuseppe**, and Lauscher, Anne. "Building Bridges: A Dataset for Evaluating Gender-Fair Machine Translation into German". In: *ACL 2024 (Findings)* (2024)

Dinkar, Tanvi, **Attanasio**, **Giuseppe**, Curry, Amanda Cercas, Konstas, Ioannis, Hovy, Dirk, and Rieser, Verena. "Proceedings of Safety4ConvAI: The Third Workshop on Safety for Conversational AI@ LREC-COLING 2024". In: *Proceedings of Safety4ConvAI: The Third Workshop on Safety for Conversational AI@ LREC-COLING 2024*. 2024

Raus, Rachele, Tonti, Michela, Cerquitelli, Tania, Cagliero, Luca, **Attanasio**, **Giuseppe**, La Quatra, Moreno, and Greco, Salvatore. "L'analyse du discours et l'intelligence artificielle pour réaliser une écriture inclusive: le projet EMIMIC". in: *SHS Web of Conferences*. Vol. 138. EDP Sciences. 2022, p. 01007

Cercas Curry, Amanda, **Attanasio**, **Giuseppe**, Talat, Zeerak, and Hovy, Dirk. "Classist Tools: Social Class Correlates with Performance in NLP". in: *ACL 2024*. 2024

Attanasio, **Giuseppe**, Nozza, Debora, Bianchi, Federico, and Hovy, Dirk. "Is It Worth the (Environmental) Cost? Limited Evidence for the Benefits of Diachronic Continuous Training". In: *arXiv preprint arXiv:2210.07365* (2022)

Koudounas, Alkis, Pastor, Eliana, **Attanasio**, **Giuseppe**, Alfaro, Luca de, and Baralis, Elena. "Prioritizing Data Acquisition For End-to-End Speech Model Improvement". In: *ICASSP 2024. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2024

Koudounas, Alkis, Pastor, Eliana, Mazzia, Vittorio, Giollo, Manuel, Gueudre, Thomas, Reale, Elisa, **Attanasio**, **Giuseppe**, Cagliero, Luca, Cumani, Sandro, De Alfaro, Luca, et al. "Leveraging confidence models for identifying challenging data subgroups in speech models". In: *2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*. IEEE. 2024, pp. 134–138

Koudounas, Alkis, Pastor, Eliana, **Attanasio**, **Giuseppe**, Mazzia, Vittorio, Giollo, Manuel, Gueudre, Thomas, Reale, Elisa, Cagliero, Luca, Cumani, Sandro, Alfaro, Luca de, et al. "Towards comprehensive subgroup performance analysis in speech models". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 32 (2024), pp. 1468–1480

Nozza, Debora and **Attanasio**, **Giuseppe**. "Is It Really That Simple? Prompting Language Models for Automatic Text Simplification in Italian". In: *CLiC-it 2023*. CEUR Workshop Proceedings, 2023

Attanasio, **Giuseppe**, Plaza-del-Arco, Flor Miriam, Nozza, Debora, and Lauscher, Anne. "A Tale of Pronouns: Interpretability Informs Gender Bias Mitigation for Fairer Instruction-Tuned Machine Translation". In: *EMNLP 2023*. 2023

Cassani, Giovanni, Guenther, Fritz, **Attanasio**, **Giuseppe**, Bianchi, Federico, and Marelli, Marco. "Meaning Modulations and Stability in Large Language Models: An Analysis of BERT Embeddings for Psycholinguistic Research". In: (2023)

Attanasio, **Giuseppe**, Pastor, Eliana, Di Bonaventura, Chiara, and Nozza, Debora. "ferret: a Framework for Benchmarking Explainers on Transformers". In: *EACL 2023, System Demonstrations*. 2023, pp. 256–266

Bianchi, Federico, Suzgun, Mirac, **Attanasio**, **Giuseppe**, Röttger, Paul, Jurafsky, Dan, Hashimoto, Tatsunori, and Zou, James. "Safety-tuned llamas: Lessons from improving the safety of large language models that follow instructions". In: *ICLR 2024* (2023)

Pastor, Eliana, Koudounas, Alkis, **Attanasio**, **Giuseppe**, Hovy, Dirk, and Baralis, Elena. "Explaining Speech Classification Models via Word-Level Audio Segments and Paralinguistic Features". In: *EACL 2024* (2023)

Bonaldi, Helena, **Attanasio**, **Giuseppe**, Nozza, Debora, and Guerini, Marco. "Weigh Your Own Words: Improving Hate Speech Counter Narrative Generation via Attention Regularization". In: *arXiv preprint arXiv:2309.02311* (2023)

Röttger, Paul, Kirk, Hannah Rose, Vidgen, Bertie, **Attanasio**, **Giuseppe**, Bianchi, Federico, and Hovy, Dirk. "XSTest: A test suite for identifying exaggerated safety behaviours in large language models". In: *NAACL 2024*. 2023

Curry, Amanda Cercas, **Attanasio**, **Giuseppe**, Nozza, Debora, Hovy, Dirk, et al. "Milanlp at semeval-2023 task 10: ensembling domain-adapted and regularized pretrained language models for robust sexism detection". In: *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. Association for Computational Linguistics. 2023

Attanasio, **Giuseppe**. "Dissecting Deep Language Models: The Explainability and Bias Perspective". In: (2022)

Koudounas, Alkis, La Quatra, Moreno, Vaiani, Lorenzo, Colomba, Luca, **Attanasio**, **Giuseppe**, Pastor, Eliana, Cagliero, Luca, and Baralis, Elena. "ITALIC: An ITALian Intent Classification Dataset". In: *INTERSPEECH 2023*. IEEE. 2023

Cassani, Giovanni, **Attanasio**, **Giuseppe**, and Bianchi, Federico. "Sound symbolic associations via contrastive learning of text and images". In: *Architectures and Mechanisms for Language Processing (2022)*. 2022

Chia, Patrick John, **Attanasio**, **Giuseppe**, Tagliabue, Jacopo, Bianchi, Federico, Greco, Ciro, Moreira, Gabriel de Souza P, Eynard, Davide, and Husain, Fahd. "E Pluribus Unum: Guidelines on Multi-Objective Evaluation of Recommender Systems". In: *arXiv preprint arXiv:2304.10621* (2023)

Koudounas, Alkis, Pastor, Eliana, **Attanasio**, **Giuseppe**, Mazzia, Vittorio, Giollo, Manuel, Gueudre, Thomas, Cagliero, Luca, Alfaro, Luca de, Baralis, Elena, and Amberti, Daniele. "Exploring subgroup performance in End-to-End speech models". In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023

Pipoli, Vittorio, **Attanasio**, **Giuseppe**, Lovino, Marta, and Ficarra, Elisa. "Squeeze and Learn: Compressing Long Sequences with Fourier Transformers for Gene Expression Prediction". In: (2023)

Tagliabue, Jacopo, Bianchi, Federico, Schnabel, Tobias, **Attanasio**, **Giuseppe**, Greco, Ciro, Souza Moreira, Gabriel de, and Chia, Patrick John. "A challenge for rounded evaluation of recommender systems". In: *Nature Machine Intelligence* 5.2 (2023), pp. 181–182

Chia, Patrick John, **Attanasio**, **Giuseppe**, Bianchi, Federico, Terragni, Silvia, Magalhães, Ana Rita, Goncalves, Diogo, Greco, Ciro, and Tagliabue, Jacopo. "Contrastive language and vision learning of general fashion concepts". In: *Nature Scientific Reports* 12.1 (2022), p. 18958

Attanasio, **Giuseppe**, Nozza, Debora, Bianchi, Federico, and Hovy, Dirk. "Is It Worth the (Environmental) Cost? Limited Evidence for the Benefits of Diachronic Continuous Training". In: *arXiv preprint arXiv:2210.07365* (2022)

Bellocca, Gian Pietro, **Attanasio**, **Giuseppe**, Cagliero, Luca, and Fior, Jacopo. "Machine Learning with Applications". In: ()

Tagliabue, Jacopo, Bianchi, Federico, Schnabel, Tobias, **Attanasio**, **Giuseppe**, Greco, Ciro, Moreira, Gabriel de Souza P, and Chia, Patrick John. "EvalRS: a Rounded Evaluation of Recommender Systems". In: *arXiv preprint arXiv:2207.05772* (2022)

Nozza, Debora, Bianchi, Federico, and **Attanasio**, **Giuseppe**. "HATE-ITA: Hate speech detection in Italian social media text". In: *Proceedings of the Sixth Workshop on Online Abuse and Harms (WOAH)*. 2022, pp. 252–260

Attanasio, **Giuseppe**, Nozza, Debora, Bianchi, Federico, et al. "MilaNLP at semeval-2022 task 5: Using perceiver IO for detecting misogynous memes with text and image modalities". In: *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Association for Computational Linguistics. 2022

Raus, Rachele, Tonti, Michela, Cerquitelli, Tania, Cagliero, Luca, **Attanasio**, **Giuseppe**, La Quatra, Moreno, and Greco, Salvatore. "L'analyse du discours et l'intelligence artificielle pour réaliser une écriture inclusive: le projet EMIMIC". in: *SHS Web of Conferences*. Vol. 138. EDP Sciences. 2022, p. 01007

Bellocca, Gian Pietro, **Attanasio**, **Giuseppe**, Cagliero, Luca, and Fior, Jacopo. "Leveraging the momentum effect in machine learning-based cryptocurrency trading". In: *Machine Learning with Applications* 8 (2022), p. 100310

Attanasio, **Giuseppe**, Nozza, Debora, Pastor, Eliana, Hovy, Dirk, et al. "Benchmarking post-hoc interpretability approaches for transformer-based misogyny detection". In: *Proceedings of NLP Power! The First Workshop on Efficient Benchmarking in NLP*. Association for Computational Linguistics. 2022

Chia, Patrick John, **Attanasio**, **Giuseppe**, Bianchi, Federico, Terragni, Silvia, Magalhães, Ana Rita, Goncalves, Diogo, Greco, Ciro, and Tagliabue, Jacopo. "Fashionclip: Connecting language and images for product representations". In: *arXiv preprint arXiv:2204.03972* (2022)

Attanasio, **Giuseppe**, Nozza, Debora, Hovy, Dirk, and Baralis, Elena. "Entropy-based attention regularization frees unintended bias mitigation from lists". In: *ACL 2022 (Findings)*. 2022

Attanasio, **Giuseppe**, Greco, Salvatore, La Quatra, Moreno, Cagliero, Luca, Tonti, Michela, Cerquitelli, Tania, and Raus, Rachele. "E-mimic: Empowering multilingual inclusive communication". In: *2021 IEEE International Conference on Big Data (Big Data)*. IEEE. 2021, pp. 4227–4234

Bianchi, Federico, **Attanasio**, **Giuseppe**, Pisoni, Raphael, Terragni, Silvia, Sarti, Gabriele, and Balestri, Dario. "Contrastive language-image pre-training for the italian language". In: *CLiC-it 2023*. 2021

Attanasio, **Giuseppe** and Pastor, Eliana. "PoliTeam@ AMI: Improving Sentence Embedding Similarity with Misogyny Lexicons for Automatic Misogyny Identification in Italian Tweets." In: *EVALITA 2020*. 2020

Attanasio, **Giuseppe**, Giobergia, Flavio, Pasini, Andrea, Ventura, Francesco, Baralis, Elena, Cagliero, Luca, Garza, Paolo, Apiletti, Daniele, Cerquitelli, Tania, and Chiusano, Silvia. "DSLE: a smart platform for designing data science competitions". In: *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE. 2020, pp. 133–142

Attanasio, **Giuseppe**, Cagliero, Luca, and Baralis, Elena. "Leveraging the explainability of associative classifiers to support quantitative stock trading". In: *Proceedings of the sixth international workshop on data science for macro-modeling*. 2020, pp. 1–6

Attanasio, **Giuseppe**, Cagliero, Luca, Garza, Paolo, and Baralis, Elena. "Combining news sentiment and technical analysis to predict stock trend reversal". In: *2019 International Conference on Data Mining Workshops (ICDMW)*. IEEE. 2019, pp. 514–521

Cagliero, Luca, Garza, Paolo, **Attanasio**, **Giuseppe**, and Baralis, Elena. "Training ensembles of faceted classification models for quantitative stock trading". In: *Computing* 102.5 (2020), pp. 1213–1225

Attanasio, **Giuseppe**. "Comparing time series and associative classification approaches to quantitative stock trading". PhD thesis. Politecnico di Torino, 2018

Attanasio, **Giuseppe**, Cagliero, Luca, Garza, Paolo, and Baralis, Elena. "Quantitative cryptocurrency trading: exploring the use of machine learning techniques". In: *Proceedings of the 5th Workshop on Data Science for Macro-modeling with Financial and Economic Datasets*. 2019, pp. 1–6

Attanasio, **Giuseppe**, Cannavò, Alberto, Cibrario, Francesca, Lamberti, Fabrizio, Montuschi, Paolo, and Paravati, Gianluca. "HOT: Hold your own tools for AR-based constructive art". In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2017, pp. 256–257

Research projects

The following is a list of research projects I carried out or where I have contributed to the conceptualization of the work. Projects include collaborations with both academic peers and industry.

Fairness in End-to-End Speech Representation Models   In this ongoing collaboration, we aim to study how modern neural end-to-end speech models (e.g., Wav2Vec 2.0 or HuBERT) behave on specific subgroups of the dataset. The end goal is to detect bias toward under-represented social groups or recording conditions where models under-perform.

Entities involved: Amazon Alexa AI.

Gender Inclusive Communication in Italian.   E-MIMIC (Empowering Multilingual Inclusive comMu-nICation) is a joint effort of classical and computational linguistics communities in fighting against non-inclusive, prejudiced language forms. We use modern transformer-based language models to learn inclusive language in written text. Currently, we focus on Academic and Public Administration Italian documents with the goal of non-inclusive content identification and rephrasing.

Entities involved: University of Turin, University of Bologna, University of Bergamo.

Contrastive Language-Image Pre-training for the Italian Language.   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 extended 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 image retrieval and zero-shot classification tasks.
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.

Entities involved: Bocconi University, University of Milano-Bicocca, University of Groningen.

EvalRS: Rounded Evaluation of Recommender Systems   Standard evaluation metrics for Recommender Systems fail to capture important aspects, such as behavioral testing and fairness on subgroups of users. In this project, we proposed EvalRS, an open-source framework, a data challenge, and a workshop (CIKM 2022) to foster more rounded evaluations of RecSys.

Entities involved: Coveo, Coveo Labs, Stanford University, Microsoft, NVIDIA.

Vision and Language Multi-Modality for General Fashion Concepts Learning   The cost of training tailored supervised models can be prevented by learning general representations of concepts and reusing them. We show CLIP (Contrastive Language-Italian Pre-Training) can learn general fashion concepts (e.g., long/short sleeves, high/low heels) efficiently and effectively.

Entities involved: Coveo, Coveo Labs, Stanford University, Farfetch.

MALTO: Machine Learning at Polito.   I co-founded MAchine Learning At poliTO (MALTO), a university-funded student team at Politecnico di Torino.

The team's main goal is to participate in various data science competitions. These competitions typically involve achieving high performance on machine learning problems (e.g., classification or regression) and proposing novel, cutting-edge approaches.

Entities involved: Politecnico di Torino.

Unsupervised Learning for Smart Fleet Management.   I joined 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.

Entities involved: KMaster (Telepass Group).

Public Activities & Teaching

Research Community

Workshops.

- I co-organized the first edition of EvalRS: a Rounded Evaluation of Recommender Systems which has been a data challenge and workshop during CIKM 2022 (`https://reclist.io/cikm2022-cup/`). The event fosters a more rounded evaluation of recommender systems, encompassing aspects beyond standard evaluation, such as behavioral testing and group fairness.

- I co-organized the third edition of the Safety for Conversational AI (`https://sites.google.com/view/safety-conv-ai-workshop`).

University

M2L School.   I co-tutored the second edition of the Mediterranean Machine Learning (M2L) Summer School, with ownership of the Natural Language Processing part. We prepared a shared notebook to guide students through 1) implementing a Transformer from scratch, 2) training the model for language modeling and machine translation, and 3) testing on gender bias.

Teaching Assistant.   I have been a teaching assistant for several courses at Politecnico di Torino. In total, I have held 171 hours of complementary teaching.

- (2021) **Data science lab: processes and methods**, 60h, MD in Data Science and Engineering
- (2020) **Data science lab: processes and methods**, 39h, MD in Data Science and Engineering
- (2020) **Business Intelligence for Big Data**, 21h, MD in Industrial engineering and management
- (2019) *** Data science lab: processes and methods**, 39h, MD in Data Science and Engineering
- (2019) **Business Intelligence for Big Data**, 21h, MD in Industrial engineering and management
- (2018) **Basi di dati**, 21h, BD in Computer Engineering
- (2018) **The Fourth Industrial Revolution: Promises and Pitfalls in Blending New and Traditional Approaches in Manufacturing and Service Sectors**, 30h, Alta Scuola Politecnica School

* 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 have devised and contributed to the launch of Research Bites, a series of short research talks and seminars held by Ph.D. students for students of the course Data Science Lab: process and methods. The goal of RBs is to disseminate cutting-edge research topics, in short, high-level pills. The series is now in its second edition.

Master Thesis Supervisor.   I have supervised the work of 9 master students.

Industry

I have held consulting hours and courses twice.

- (2021) **Data Theory: Data Visualization with Python**
  Reply: Digital Services, Technology and Consulting, Turin (online)
- (2020) **Python technologies for Data Analytics**
  Applied Mechatronic Engineering & Technologies, Turin

Peer reviewing

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

- ACL Rolling Review. Editions: 2021, 2022
- 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.   *Turin, Italy*
Founder   *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.   *Turin, 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 rusty with *italicized* entries.

Programming	Python, *C, C++, Java, C#, JavaScript, PHP, Matlab, SQL*
Deep Learning	PyTorch, *Tensorflow, JAX, Keras*
Scripting	Bash, *Awk*
DevOps	*AWS, GCP, Docker*
Front-end	*Hugo, React, Dash*
Graphics	Inkscape, GIMP
Languages	Italian, English, *Spanish, Portuguese, French*