

Elena Sofia Ruzzetti

elena.sofia.ruzzetti@uniroma2.it | Website | Google Scholar

 elena-sofia-ruzzetti |  elenasofia98

EDUCATION

• Ph.D in Data Science

2022 - Today

University of Rome Tor Vergata

- As part of the Human Centric ART group, my research currently revolves around Trustworthy AI principles, applied to language models. I have been working especially on robustness, security, interpretability, and fairness. Recently, I have been focusing on studying and mitigating privacy risks in LLMs.

• Master's Degree in Computer Science

2020 - 2022

University of Rome Tor Vergata

- Thesis Title: "Lacking the embedding of a word? Look it up into a traditional dictionary."
Thesis Supervisor: Fabio Massimo Zanzotto
- 110/110 cum laude

• Bachelor's Degree in Computer Science

2017 - 2020

University of Rome Tor Vergata

Rome, Italy

- Thesis Title: "Composing or not composing: This is the problem. . . in the Era of Neural Network"
Thesis Supervisor: Fabio Massimo Zanzotto
- 110/110 cum laude

PUBLICATIONS

C=CONFERENCE, J=JOURNAL, S=IN SUBMISSION

- [S.1] Venditti, D., **Ruzzetti, E. S.**, Xompero, G. A., Giannone, C., Favalli, A., Romagnoli, R., & Zanzotto, F. M. (2024). **Enhancing Data Privacy in Large Language Models through Private Association Editing**. In *arXiv preprint arXiv:2406.18221*.
- [C.1] **Ruzzetti, E. S.**, Xompero, G. A., Venditti, D., & Zanzotto, F. M. (2025). **Private Memorization Editing: Turning Memorization into a Defense to Strengthen Data Privacy in Large Language Models**. In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 16572–16592). Association for Computational Linguistics; <https://aclanthology.org/2025.acl-long.810/>
- [C.2] Zanzotto, F. M., **Ruzzetti, E. S.**, Xompero, G. A., Ranaldi, L., Venditti, D., Ranaldi, F., Giannone, C., Favalli, A., & Romagnoli, R. (2025). **Position Paper: MeMo: Towards Language Models with Associative Memory Mechanisms**. In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Eds.), *Findings of the Association for Computational Linguistics: ACL 2025* (pp. 15169–15180). Association for Computational Linguistics; <https://aclanthology.org/2025.findings-acl.785/>
- [C.3] Ranaldi, L., **Ruzzetti, E. S.**, Venditti, D., Onorati, D., & Zanzotto, F. M. (2024). **A Trip Towards Fairness: Bias and De-Biasing in Large Language Models**. In D. Bollegala & V. Schwartz (Eds.), *Proceedings of the 13th Joint Conference on Lexical and Computational Semantics (*SEM 2024)* (pp. 372–384). Association for Computational Linguistics; . <https://aclanthology.org/2024.starsem-1.30/>
- [C.4] Ranaldi, L., Pucci, G., Ranaldi, F., **Ruzzetti, E. S.**, & Zanzotto, F. M. (2024). **A Tree-of-Thoughts to Broaden Multi-step Reasoning across Languages**. In K. Duh, H. Gomez, & S. Bethard (Eds.), *Findings of the Association for Computational Linguistics: NAACL 2024* (pp. 1229–1241). Association for Computational Linguistics; . <https://aclanthology.org/2024.findings-naacl.78/>
- [C.5] Ranaldi, F., **Ruzzetti, E. S.**, Onorati, D., Ranaldi, L., Giannone, C., Favalli, A., Romagnoli, R., & Zanzotto, F. M. (2024). **Investigating the Impact of Data Contamination of Large Language Models in Text-to-SQL translation**. *Findings of the Association for Computational Linguistics: ACL 2024*, 13909–13920. <https://aclanthology.org/2024.findings-acl.827/>
- [C.6] Ranaldi, F., **Ruzzetti, E. S.**, Onorati, D., Zanzotto, F. M., & Ranaldi, L. (2024). **Termite Italian Text-to-SQL: A CALAMITA Challenge**. *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-It 2024)*, 1176–1183.
- [C.7] Onorati, D., Venditti, D., **Ruzzetti, E. S.**, Ranaldi, F., Ranaldi, L., & Zanzotto, F. M. (2024). **Measuring bias in Instruction-Following models with ItaP-AT for the Italian Language**. *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-It 2024)*, 679–706.
- [C.8] **Ruzzetti, E. S.**, Ranaldi, F., Onorati, D., Venditti, D., Ranaldi, L., Caselli, T., & Zanzotto, F. M. (2024). **Assessing the Asymmetric Behaviour of Italian Large Language Models across Different Syntactic Structures**. *CLiC-It 2024: Tenth Italian Conference on Computational Linguistics*

- [C.9] Ranaldi, L., Pucci, G., Ranaldi, F., **Ruzzetti, E. S.**, & Zanzotto, F. M. (2024). The limits of Italian in Reasoning Tasks. *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-It 2024)*
- [C.10] **Ruzzetti, E. S.**, Ranaldi, F., Logozzo, F., Mastromattei, M., Ranaldi, L., & Zanzotto, F. M. (2023). **Exploring Linguistic Properties of Monolingual BERTs with Typological Classification among Languages.** In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 14447–14461). Association for Computational Linguistics; <https://aclanthology.org/2023.findings-emnlp.963/>
- [C.11] Onorati, D., **Ruzzetti, E. S.**, Venditti, D., Ranaldi, L., & Zanzotto, F. M. (2023). **Measuring bias in Instruction-Following models with P-AT.** In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 8006–8034). Association for Computational Linguistics; . <https://aclanthology.org/2023.findings-emnlp.539/>
- [C.12] Ranaldi, L., **Ruzzetti, E. S.**, & Zanzotto, F. M. (2023). **PreCog: Exploring the Relation between Memorization and Performance in Pre-trained Language Models.** In R. Mitkov & G. Angelova (Eds.), *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing* (pp. 961–967). INCOMA Ltd., Shoumen, Bulgaria; . <https://aclanthology.org/2023.ranlp-1.103/>
- [C.13] Ranaldi, L., Nourbakhsh, A., **Ruzzetti, E. S.**, Patrizi, A., Onorati, D., Mastromattei, M., Fallucchi, F., & Zanzotto, F. M. (2023). The Dark Side of the Language: Pre-trained Transformers in the DarkNet. In R. Mitkov & G. Angelova (Eds.), *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing* (pp. 949–960). INCOMA Ltd., Shoumen, Bulgaria; <https://aclanthology.org/2023.ranlp-1.102/>
- [C.14] Onorati, D., Ranaldi, L., Nourbakhsh, A., Patrizi, A., **Ruzzetti, E. S.**, Mastromattei, M., Fallucchi, F., & Zanzotto, F. M. (2023). **The Dark Side of the Language: Syntax-Based Neural Networks Rivaling Transformers in Definitely Unseen Sentences.** 2023 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 111–118.
- [C.15] **Ruzzetti, E. S.**, Onorati, D., Ranaldi, L., Venditti, D., & Zanzotto, F. M. (2023). **Investigating Gender Bias in Large Language Models for the Italian Language.** *CLiC-It 2023: 9th Italian Conference on Computational Linguistics*
- [C.16] Ranaldi, F., **Ruzzetti, E. S.**, Ranaldi, L., Venditti, D., Giannone, C., Favalli, A., Romagnoli, R., & Zanzotto, F. M. (2023). **Prompting LLMs in Italian language for Text-to-SQL translation.** *CLiC-It 2023: 9th Italian Conference on Computational Linguistics*
- [C.17] Ranaldi, L., Pucci, G., **Ruzzetti, E. S.**, Zanzotto, F. M., & Freitas, A. (2023). **Teasing LLMs adapted to Italian.** *CLiC-It 2023: 9th Italian Conference on Computational Linguistics*
- [C.18] **Ruzzetti, E. S.**, Ranaldi, L., Mastromattei, M., Fallucchi, F., Scarpato, N., & Zanzotto, F. M. (2022). **Lacking the Embedding of a Word? Look it up into a Traditional Dictionary.** In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 2651–2662). Association for Computational Linguistics; . <https://aclanthology.org/2022.findings-acl.208/>
- [J.1] **Ruzzetti, E. S.**, Venditti, D., Zanzotto, F. M., & Fallucchi, F. (2024). **Using distributional models for studying the influence of school textbooks in children bias.** *IEEE Access*, 12, 158207–158214.
- [J.2] Miranda, M., **Ruzzetti, E. S.**, Santilli, A., Zanzotto, F. M., Bratières, S., & Rodolà, E. (2025). **Preserving Privacy in Large Language Models: A Survey on Current Threats and Solutions.** *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=Ss9MTTN7OL>

ACADEMIC EXPERIENCE

- **Understanding Memorization and Generalization in Transformer-based LLMs and beyond** June 2025
Tutorial at the 2025 International Joint Conference on Neural Networks (IJCNN 2025), Zanzotto, F. M., Ruzzetti, E. S.
 - An analysis of Transformer-based Large Language Models and how a part of the capabilities of LLMs depends on their ability to memorize: the memorization mechanisms inside the Transformer architecture are discussed and the possibility for different architectures is discussed.
- **Practical NLP 2023-2024** 2023 - 2024
Hands-on lessons offered to Master students as a complement to the NLP course [🔗]
- **Reviewer in the ACL Rolling Review (ARR)** 2021 - 2025