SORODOC IONUT-TEODOR

CONTACT INFORMATION

MOBILE: +40-747-897-834
E-MAIL: ionut.sorodoc@gmail.com

RESEARCH INTERESTS

Retrieval augmented generation, Computational Linguistics, Machine Learning, LLMs, Artificial Intelligence.

WORK EXPERIENCE

Principal LLM Researcher
Huawei Ireland Research Center, Cluj-Napoca.   *August 2025-*

Applied Scientist
Amazon AGI, Barcelona.   *July 2022 - July 2025*

Applied Science Intern
Amazon Alexa, Berlin.   *November 2021 - March 2022*

Research Intern
Facebook AI Research, New York.   *June 2018 - October 2018*

Research Assistant
University of Trento, Italy.   *February-September 2017*

Research Intern
University of Melbourne, Australia.   *August 2016 - November 2016*

EDUCATION

Ph.D. in Translation and Language Sciences   *2017-2022*
Pompeu Fabra University, Barcelona, Spain.
Supervisor: Gemma Boleda

M.Sc. in Cognitive Science, Language and Multimodal Interaction   *2014-2016*
CIMeC, University of Trento, Rovereto, Italy.
Supervisor: Raffaella Bernardi

M.Sc. in Human Language Science and Technology   *2014-2016*
University of Malta, Malta.
Supervisor: Michael Rosner

B.Sc. in Computer Science   *2010-2013*
University of Bucharest, Romania.

PUBLICATIONS

The Amazon Nova Family of Models: Technical Report and Model Card (*paper*)
Contributor   *2024*

Retrieving contextual information for long-form question answering using weak supervision (*paper*)
Co-authors: Philipp Christmann, Svitlana Vakulenko, Bill Byrne, Adria de Gispert   *November, 2024*
The 2024 Conference on Empirical Methods in Natural Language Processing - Findings

Challenges in including extra-linguistic context in pre-trained language models (*paper*)
Co-authors: Laura Aina, Gemma Boleda   *May, 2022*
Conference: Third Workshop on Insights from Negative Results in NLP

Probing for referential information in language models (*paper*)
Co-authors: Kristina Gulordava, Gemma Boleda   *July, 2020*
Conference:The 58th annual meeting of the Association for Computational Linguistics (ACL)

What do entity-centric models learn? Insights from entity linking in multi-party dialogue (*paper*)
Co-authors: Laura Aina, Carina Silberer, Matthijs Westera, Gemma Boleda   *June, 2019*
Conference: The 7th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), Minneapolis, United States of America.

Comparatives, Quantifiers, Proportions: A Multi-Task Model for the Learning of Quantities from Vision (*paper*)
Co-authors: Sandro Pezzelle, Raffaella Bernardi   *June, 2018*
Conference: The 6th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), New Orleans, United States of America.

"Pay Attention to Those Sets!" Learning Quantification from Images. (*paper*)
Co-authors: Raffaella Bernardi, Aurelie Herbelot, Sandro Pezzelle,
Mariella Dimiccoli   *Accepted*
Journal: Natural Language Engineering 24.3, 363-392.   *2018*

Multimodal topic labelling (*paper*)
Co-authors: Tim Baldwin, Jey Han Lau, Nikolaos Aletras   *April, 2017*
Conference: The 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Valencia, Spain.

PROJECTS
- **Member** - Student board of the European Chapter of the ACL (EACL) 2020 - 2022
- **Researcher** - EU, Horizon 2020 European Research Council Starting Grant 715154 ("2017 - 2022 AMORE: A distributional MOdel of Reference to Entities"), Universitat Pompeu Fabra
- **Member** - The European Network on Integrating Vision and Language (iV&L Net) ICT COST Action IC1307

TEACHING
Teaching Assistant: Corpus Linguistics and Analysis
BA Applied languages, Pompeu Fabra University
Teaching Assistant: Empirical methods for studying language
BA Applied languages, Pompeu Fabra University

RESEARCH ACTIVITIES
- Organizer of EACL Student Research Workshop, 2021.
- Organizer of GeCKO Symposium, Barcelona, 2020.
- Member of the Program Commitee of LANTERN Workshop 2020, 2019.
- Reviewer for EMNLP 2019, 2020, ACL 2019, COLM 2024 and ARR.

FURTHER TRAINING
Eastern European Machine Learning (EEML) summer school
July, 2019, Bucharest, Romania.

2nd Summer School on Integrating Vision & Language: Deep Learning.
March, 2016, Malta.

PROGRAMMING
- Python, PyTorch (Advanced).
- C/C++, Keras (Intermediate).
- Tensorflow, Torch, Java, Scripting (Basic).
- Strong knowledge of algorithms and data structures.

LANGUAGES
- Romanian: native.
- English: proficient.
- Spanish and German: basic knowledge.