Robert Vacareanu

rvacareanu@arizona.edu | 520-704-8735 | robertvacareanu | Google Scholar Profile

Objective: A soon-to-be PhD graduate specializing in Natural Language Processing with relevant publications, internships, and work experience in the field and as a software developer, prior to PhD. Seeking opportunities as a Research Scientist.

Education

PhD in Computer Science | University of Arizona Aug. 2020 - July 2024 (Expected)
Thesis: Explainable Methods for Information Extraction GPA: 4.00
Advisor: Mihai Surdeanu
My thesis focuses on **explainability** and **interpretability** in information extraction. This work has resulted in 5 publications [11, 10, 9, 8, 5] and one patent [12].
In addition to my thesis work, I have also published on topics such as dependency parsing [14], representational learning for multi-word expressions [13], active learning [4], and reasoning with large language models and in-context learning [2].

MSc in Computer Science | Technical University of Cluj-Napoca Oct. 2018 - July 2020
AI-focused curriculum. I continued my undergraduate research in neuroscience and began working on NLP.

BSc in Computer Science | Technical University of Cluj-Napoca Oct. 2014 - July 2018
Completed an engineering-focused curriculum with comprehensive coverage of core areas in computer science. My thesis involved applying machine learning techniques to neuroscience, exploring the extent to which the shapes of neuronal spikes carry information about visual stimuli.

Work Experience

Applied Scientist Intern | AWS May 2023 - Aug 2023
- I interned in the Amazon Bedrock team, focusing on large language models and reasoning. My manager was Yassine Benajiba. I worked closely with Anurag Pratik, Evangelia Spiliopoulou, and Zheng Qi.
- This internship resulted in one article [1]. The approach has been to employ LLMs as specialized verifiers to improve their reasoning capabilities. The final approach improves upon vanilla chain-of-thought prompting in all cases and upon best-of-N sampling in over 65% of the cases.

Applied Scientist Intern | AWS May 2022 - Aug 2022
- I interned in the Amazon Comprehend team, working on sentiment analysis with a focus on aspect-based sentiment analysis and weak supervision. My manager was Miguel Ballesteros. I worked closely with Kishaloy Halder and Siddharth Varia.
- This internship resulted in two publications [7, 6]. The approach has been to handle all tasks through an unified architecture [7] and to construct a noisy dataset for aspect-based sentiment analysis through heuristics, without any supervision [6]. The final model obtains improvements as large as 15.84 absolute F1 in few-shot scenarios on the harder tasks [6].

Graduate Research Assistant | University of Arizona Aug. 2020 - Present
- I have been a research assistant throughout my PhD, working on NSF and DARPA grants.

Software Engineer | Catalysts Software Romania May 2017 - Feb. 2019
- As a Software Engineer, I contributed to satellite data processing projects, focusing on snow monitoring using data from the S3 and Landsat satellites. This role primarily involved developing software in C++ and taught me essential practices for production-ready software, including using Docker for containerization, participating in Agile development cycles, and collaborating effectively in team environments.

Service

- Reviewed for: PaN-DL Workshop @ COLING 2022, PaN-DL Workshop @ EMNLP 2023, NAACL 2024 (nominated for Great Reviewer), ACL 2024 (nominated for Great Reviewer)

Awards

- Outstanding Graduate Scholarship 2023
- Graduate College Fellowships 2020

References

Can provide additional contact information upon request.
- **Mihai Surdeanu**; My PhD Advisor; Professor at University of Arizona
- **Miguel Ballesteros**; My Manager at AWS in 2022; Principal Applied Scientist at AWS
- **Rebecca Sharp**; Collaborator during my first years of PhD; Principal Data Scientist at Lex Machina

- **Anurag Pratik**; My Mentor at AWS in 2023; Senior Research Engineer at AWS

Publications, Preprints, And Patents

[1] Robert Vacareanu, Anurag Pratik, Evangelia Spiliopoulou, Zheng Qi, Giovanni Paolini, Neha Anna John, Jie Ma, Yassine Benajiba, and Miguel Ballesteros. "General Purpose Verification for Chain of Thought Prompting". In: *ArXiv* abs/2405.00204 (2024). preprint. URL: https://arxiv.org/pdf/2405.00204.pdf.

[2] Robert Vacareanu, Vlad-Andrei Negru, Vasile Suciu, and Mihai Surdeanu. "From Words to Numbers: Your Large Language Model Is Secretly A Capable Regressor When Given In-Context Examples". In: preprint. 2024.

[3] Fahmida Alam, Md Asiful Islam, Robert Vacareanu, and Mihai Surdeanu. "Towards Realistic Few-Shot Relation Extraction: A New Meta Dataset and Evaluation". In: *Proceedings of the Fourteenth Language Resources and Evaluation Conference*. Torino, Italy: European Language Resources Association, May 2024. URL: http://arxiv.org/abs/2404.04445.

[4] Robert Vacareanu, Enrique Noriega-Atala, Gus Hahn-Powell, Marco A. Valenzuela-Escarcega, and Mihai Surdeanu. "Active Learning Design Choices for NER with Transformers". In: *Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation*. LREC-COLING 2024 (Poster). Torino, Italy: European Language Resources Association, May 2024.

[5] Robert Vacareanu, Fahmida Alam, Md Asiful Islam, Haris Riaz, and Mihai Surdeanu. "Best of Both Worlds: A Pliable and Generalizable Neuro-Symbolic Approach for Relation Classification". In: *Findings of the Association for Computational Linguistics: NAACL 2024*. Findings of NAACL 2024 (Poster). Mexico City, Mexico: Association for Computational Linguistics, June 2024. URL: https://arxiv.org/pdf/2403.03305.pdf.

[6] Robert Vacareanu, Siddharth Varia, Kishaloy Halder, Shuai Wang, Giovanni Paolini, Neha Anna John, Miguel Ballesteros, and Smaranda Muresan. "A Weak Supervision Approach for Few-Shot Aspect Based Sentiment Analysis". In: *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Yvette Graham and Matthew Purver. EACL 2024 (Oral). St. Julian's, Malta: Association for Computational Linguistics, Mar. 2024, pp. 2734–2752. URL: https://aclanthology.org/2024.eacl-long.167.

[7] Siddharth Varia, Shuai Wang, Kishaloy Halder, Robert Vacareanu, Miguel Ballesteros, Yassine Benajiba, Neha Anna John, Rishita Anubhai, Smaranda Muresan, and Dan Roth. "Instruction Tuning for Few-Shot Aspect-Based Sentiment Analysis". In: *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*. Ed. by Jeremy Barnes, Orphée De Clercq, and Roman Klinger. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 19–27. DOI: 10.18653/v1/2023.wassa-1.3. URL: https://aclanthology.org/2023.wassa-1.3.

[8] Enrique Noriega-Atala, Robert Vacareanu, Gus Hahn-Powell, and Marco A. Valenzuela-Escárcega. "Neural-Guided Program Synthesis of Information Extraction Rules Using Self-Supervision". In: *Proceedings of the First Workshop on Pattern-based Approaches to NLP in the Age of Deep Learning*. Ed. by Laura Chiticariu, Yoav Goldberg, Gus Hahn-Powell, Clayton T. Morrison, Aakanksha Naik, Rebecca Sharp, Mihai Surdeanu, Marco Valenzuela-Escárcega, and Enrique Noriega-Atala. Gyeongju, Republic of Korea: International Conference on Computational Linguistics, Oct. 2022, pp. 85–93. URL: https://aclanthology.org/2022.pandl-1.10.

[9] Robert Vacareanu, Dane Bell, and Mihai Surdeanu. "PatternRank: Jointly Ranking Patterns and Extractions for Relation Extraction Using Graph-Based Algorithms". In: *PANDL*. PaN-DL Workshop At COLING. 2022. URL: https://aclanthology.org/2022.pandl-1.1.pdf.

[10] Robert Vacareanu, George Caique Gouveia Barbosa, Enrique Noriega-Atala, Gus Hahn-Powell, Rebecca Sharp, Marco Antonio Valenzuela-Escarcega, and Mihai Surdeanu. "A Human-machine Interface for Few-shot Rule Synthesis for Information Extraction". In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: System Demonstrations* (2022). URL: https://aclanthology.org/2022.naacl-demo.8.pdf.

[11] Robert Vacareanu, Marco A. Valenzuela-Escárcega, George Barbosa, Rebecca Sharp, and Mihai Surdeanu. "From Examples to Rules: Neural Guided Rule Synthesis for Information Extraction". In: *Proceedings of the 13th Language Resources and Evaluation Conference (LREC)*. 2022. URL: https://arxiv.org/abs/2202.00475.

[12] Mihai Surdeanu, Marco A. Valenzuela Escarcega, Gustave Hahn-Powell, Dane Bell, Thomas Hicks, Enrique Noriega, Clayton Morrison, Rebecca Sharp, Robert Ionut Vacareanu, and George Barbosa. "Methods For Extracting And Assessing Information From Literature Documents". 17/344,774. 2021.

[13] Robert Vacareanu, Marco A. Valenzuela-Escarcega, Rebecca Sharp, and Mihai Surdeanu. "An Unsupervised Method for Learning Representations of Multi-word Expressions for Semantic Classification". In: *The 28th International Conference on Computational Linguistics in Barcelona (COLING 2020)*. 2020. URL: http://clulab.org/papers/coling2020-mwe.pdf.

[14] Robert Vacareanu, George C. G. Barbosa, Marco A. Valenzuela-Escarcega, and Mihai Surdeanu. "Parsing as Tagging". In: *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC)*. 2020. URL: http://clulab.org/papers/pat.pdf.

Coding Skills

- **Experienced:** Pytorch, Huggingface Transformers, NumPy, Python, Scala, Java, C++
- **Familiar:** Jax, AWS (S3, EC2, SageMaker), Jenkins, Shell Scripting, SQL,
- **Other:** Git, Algorithms, Unix, Pandas, Matplotlib, Seaborn, Plotly, SciPy, Scikit-Learn, SpaCy, LaTeX