

Maria Valentini

Boulder, CO | maria.valentini@colorado.edu | (218)966-9905 | linkedin.com/in/maria-valentini

Education

University of Colorado Boulder, PhD in Computer Science & Cognitive Science Aug 2022 – May 2027 (planned)

- **Advisors:** Katharina von der Wense & Eliana Colunga
- **Dissertation:** *Leveraging Computational and Statistical Models to Understand and Improve Childhood Language Learning*
- GPA: 3.96/4.0

University of Colorado Boulder, MS in Computer Science Dec 2025

College of St. Scholastica, BS in Mathematics & Computer Information Systems Aug 2018 – May 2022

- GPA: 3.85/4.0; four-year varsity student-athlete (Academic All-Conference, ITA Scholar Athlete, Chi Alpha Sigma)

Research Interests

NLP for education; computational psycholinguistics (child language acquisition); LLM generation and evaluation; multilingual NLP; human-centered evaluation of AI

Publications

[†] denotes a research mentee; * denotes equal contribution

Main Conference Papers

An Investigation of Translationese in the Generations of Multilingual Large Language Models

Valentini, M., Wright, T., Granados, J., Colunga, E., & von der Wense, K. (2026). Accepted to *Proceedings of the Third Annual Conference on Language Modeling*. [Link]

Does Contextual Informativeness Predict Preschoolers' Word Learning from Stories?

Valentini, M., Granados, J., von der Wense, K., & Colunga, E. (2026). *Proceedings of the 48th Annual Meeting of the Cognitive Science Society*. [Link]

Massively Multilingual Joint Segmentation and Glossing

Outstanding Paper Award

Ginn, M., Tjuaatja, L., Rice, E., Marashian, A., Valentini, M., Xu, J., Palmer, A., & Neubig, G. (2026). *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*. [Link]

Measuring Contextual Informativeness in Child-Directed Text

Valentini, M., Wright, T.*[†], Marashian, A., Weber, J., Colunga, E., & von der Wense, K. (2025). *Proceedings of the 31st International Conference on Computational Linguistics*. [Link]

DeliGrasp: Inferring Object Properties with LLMs for Adaptive Grasp Policies

Xie, W., Valentini, M., Lavering, J., & Correll, N. (2024). *8th Annual Conference on Robot Learning*. [Link]

Evaluating LLMs as Tools to Support Early Vocabulary Learning

Weber, J., Valentini, M., Wright, T.[†], von der Wense, K., & Colunga, E. (2024). *Proceedings of the 46th Annual Meeting of the Cognitive Science Society*. [Link]

On the Automatic Generation and Simplification of Children's Stories

Valentini, M., Weber, J., Salcido, J.[†], Wright, T.[†], Colunga, E., & Kann, K. (2023). *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. [Link]

Workshop Papers

Generate Me a Bedtime Story: Leveraging Natural Language Processing for Early Vocabulary Enhancement

Hall, T., Valentini, M., Colunga, E., & Kann, K. (2022). *Proceedings of the Second Workshop on NLP for Positive Impact*. [Link]

Presentations

Invited Talks

- *Computational Models for the Understanding and Improvement of Childhood Language Learning*. Linguistics Colloquium, Johannes Gutenberg University Mainz, Germany. January 2027 (upcoming)

Guest Lectures

- *Introduction to NLP Research*. CSCI 3832: Natural Language Processing (undergraduate), CU Boulder. Fall 2024, Spring 2025 (2 lectures)
- *Transformers*. CSCI 5832: Natural Language Processing (graduate), CU Boulder. Fall 2025

Conference Posters

- Presented posters for all first-author papers: EMNLP 2023, COLING 2025, CogSci 2026, and COLM 2026

Research Experience

Graduate Research Assistant, University of Colorado Boulder Aug 2022 – present
(Profs. Katharina von der Wense & Eliana Colunga)

- Developed and evaluated LLM-based methods for generating vocabulary-rich shared-reading materials for young children, including metrics for contextual informativeness validated against human judgments
- Helped design a word-learning study with 49 preschooler–caregiver pairs and analyzed results with mixed-effects models
- Trained a classifier to identify ‘translationese’ in LLM generations and conducted ANOVAs to analyze its features
- Ongoing work models grammatical acquisition by training transformer language models on developmentally designed child-directed input
- Helped secure and maintain NSF EAGER grant funding

Data Science Intern, Digi-Key – Thief River Falls, MN May 2021 – Aug 2021

- Built machine learning models from scratch to produce a 52-week sales forecast used by marketing and sales executives
- Improved the website’s product recommender system by refining existing models

Undergraduate Research Assistant, College of St. Scholastica Jan 2019 – May 2022
(Prof. Tom Gibbons)

- Conducted research on applications of data analytics; presented at the 2019 STARS Computing Corps Celebration

Honors and Awards

- | | |
|--|---|
| • Top Reviewer, NeurIPS (2026; top 8% of reviewers) | • NSF GRFP Honorable Mention (2024) |
| • Outstanding Paper Award, ACL (2026) | • CEAS Dean’s Future Leadership Fellow (2022) |
| • Outstanding Teaching Assistant, CU Boulder CS (2026) | • Graduate School Diversity Fellow (2022) |
| • Outstanding Service Recognition (2026) | • Grace Hopper Celebration Scholar (2020) |
| • ARCS Scholar (2025) | • Tapia Conference Scholar (2020) |
| | • Clare Boothe Luce Research Scholar (2019) |

Research Mentorship

- **Téa Wright** (undergraduate, 2023–2024): co-author on 4 papers (EMNLP 2023, CogSci 2024, COLING 2025, COLM 2026); now a PhD student at UC Berkeley
- **Jesus Salcido** (community college student, 2023): co-author on EMNLP 2023 paper on generating and simplifying children’s stories
- **Samiksha Patil** (MS student, 2025–2026): co-author on paper in progress

Teaching Experience

Graduate Teaching Assistant, University of Colorado Boulder Aug 2022 – present

- CSCI 5832: Natural Language Processing (graduate, 3 semesters); CSCI 3202: Introduction to Artificial Intelligence; CSCI 3308: Software Development Methods and Tools; CSCI 1300: Introduction to Computer Science; CSCI 1200: Introduction to Computational Thinking
- Led recitation and lab sections of 40–50 students; delivered guest lectures (see Presentations)

Undergraduate Teaching Assistant, College of St. Scholastica May 2020 – Dec 2020

- CIS 3107: Database Modeling (2 semesters)

Service and Outreach

- Vice chair, CU Boulder CS Graduate Student Association (2026)
- PhD student representative, CU Boulder CS Graduate Committee (2026)
- Reviewer: NeurIPS (*Top Reviewer*); Queer in AI and CV4Edu workshops (2023, 2026)
- Volunteer committee: EMNLP, COLING, and NAACL (2023–2025)
- Mentored 4 prospective students through PhD admissions (2024–2026); peer mentor for incoming PhD students (2024, 2026)
- Co-founder and vice president, College of St. Scholastica Coding Club (2020–2022)
- Mentor for several FIRST LEGO League teams and a FIRST Robotics Competition team (2018–2022); volunteer judge (2022)

Technical Skills

Languages: Python, R, Java, C++, C#, SQL, HTML

Tools: PyTorch, TensorFlow, Hugging Face, NumPy, pandas, NLTK, spaCy, Matplotlib