Antonio Laverghetta Jr.

580-583-9743 • aml7990@psu.edu
[LinkedIn] • [GitHub] • [Google Scholar] • [Website]

BRIEF BIOGRAPHICAL SKETCH

I am an Assistant Professor of the Practice in Quantitative Methods and Wesleyan University. My research lies at the intersection of artificial intelligence, natural language processing, creativity, and psychometrics, with a particular focus on applications in STEM education. I am broadly interested in how AI can be used to enhance learning and creativity, especially in STEM. My work has appeared in highly respected conferences and journals across disciplines, including ACL, EMNLP, IMPS, PACA, and SfNC. At Wesleyan, I am the director of the Machine Psychometrics and Computational Creativity Lab. My work has been supported by the Penn State Center for Language Science and Google Cloud.

EDUCATION

Ph.D., Computer Science and Engineering December 2023
University of South Florida Tampa, FL
Dissertation Title:
A Psychometric Analysis of Natural Language Inference Using Transformer Language Models

M.S., Computer Science May 2021
University of South Florida Tampa, FL

B.S., Computer Science December 2018
University of South Florida Tampa FL

RESEARCH EXPERIENCE

Assistant Professor of the Practice July 2025 – Present
Wesleyan University Middletown, CT
Director: Machine Psychometrics and Computational Creativity Lab`

Postdoctoral Scholar January 2024 – June 2025
Pennsylvania State University State College, PA
Cognitive Neuroscience of Creativity Lab Mentor: Roger Beaty

- *Automated Item Generation for Creativity Assessments*
  Created a framework for automated generation of free response items using automatic prompt engineering, multi-agent systems, large language models. The framework can autonomously create a diverse pool of valid and reliable test items when provided a validity metric to optimize. [C3]
- *Detection of Invalid Responses to Survey Tasks using Deep Learning*
  Developed machine learning classifiers to detect careless responding for free-response assessments. The best performing model achieves up to 0.7 precision score on a benchmark dataset while being lightweight enough to run without GPUs. [J1, J2]
- *Automated Scoring of Creativity Assessments using AI*
  Contributed to automated scorers for creativity assessments by developing LLM few-shot scoring models, which predicted human originality and quality/effectiveness judgements at a correlation strength of 0.6 or greater. [J4]
  Investigated biases in LLM creativity ratings for STEM assessments, using linguistic analysis of LLM

and human justifications of originality scores. Results demonstrated statistically significant differences between LLM and human explanation structures and Likert ratings of fine-grained scoring rubrics. [C2]

- *Preference Modeling for Creativity Scoring*
  Curated a dataset consisting of more than 200,000 human ratings to creativity tasks spanning dozens of languages, hundreds of prompts, and multiple modalities. Developed a preference model for scoring response creativity, using a Llamma 3.1 model trained using direct preference optimization. The resulting model generated responses that were on average semantically more diverse than strong baselines, including GPT-4o, and represents the first successful attempt to train scoring models using reinforcement learning. [C1]

Graduate Research Assistant January 2019 – December 2023
University of South Florida Tampa, FL
Advancing Machine and Human Reasoning Lab Mentor: John Licato

- *Modeling Psychometric Properties of Test Items using Language Models*
  Demonstrated that transformer language models (BERT, RoBERTa, etc.) strongly predict latent characteristics of test items, including their difficulty and reliability. [C7, C8]
  Building on these results, a system was developed that used GPT-3 and prompting to generate new items with equivalent psychometric properties to human written ones. [C4]

- *Developing a New Benchmark of Negation Reasoning for Language Models*
  Created a new taxonomy of negation based on natural language inference training examples. The taxonomy is inspired by research on negation use in developmental psychology and provides a first step towards a more fine-grained analysis of negation competencies in language models. [C6]

- *Predicting Word Acquisition Norms Using Transformer Embeddings*
  Explored the feasibility of using contextual embeddings taken from transformers to predict the age at which children will acquire unfamiliar words in their vocabulary, using a probing methodology. The majority of the probes explored were found to more strongly predict acquisition norms than baseline feature engineering approaches. [C9]

- *Generating and Solving Word Analogies Using Word Embeddings*
  Created a model of analogical reasoning, using FastText word embeddings and the ConceptNet knowledge graph. The model achieved state-of-the-art performance on multiple analogy datasets and can generate novel word analogies given a seed term. [C12]

- Designed prompts and meta-classifiers for instruction-tuned LLMs to perform legal entailment reasoning as part of the COLIEE 2023 competition, achieving the best performance of all submitted systems using a GPT-4 legal reasoner. [J4]

INDUSTRY EXPERIENCE

Applied Scientist Intern May 2023 – August 2023
Amazon Sunnyvale, CA

- Studied how to optimize annotator workflows for intent detection using language models.
- Developed monte-carlo and chain-of-thought reasoning prompting strategies for prompt optimization on intent classification. Submitted findings from this work to the Amazon ML Conference.
- Achieved 5% accuracy improvement on intent detection benchmarks, compared to strong baselines.

Research Data Scientist Intern May 2022 – August 2022
Meta Seattle, WA

- Created a machine learning model using PyTorch to diagnose lagging alerts in streaming pipelines, increasing detection accuracy by 10%.
- Performed time series analysis using Kats to detect anomalies in streaming jobs.
- Worked with partner teams to perform feature engineering and selection for auto-diagnosis models.

R&I Intern May 2021 – August 2021
Interdigital Los Altos, CA

- Developed Natural Language Understanding models for patent search and document ranking.
- Used topic models to analyze internal datasets, applying findings to optimize training pipelines for contrastive learning.
- Achieved a 3% average increase in precision over baselines on a ranking benchmark, significantly reducing the number of irrelevant documents returned.

Software Developer Intern in Test May 2018 – September 2018
ConnectWise Tampa, FL

- Wrote Python scripts to automate nightly UI testing for ConnectWise Manage web application.

Full Stack Developer Intern August 2016 – July 2017
Tenex Software Solutions Tampa, FL

- Used git, SQL, and web technologies to resolve ~100 bugs and implement improvements to Tenex's voter registration platform.

FUNDING ACQUISITION

Google Gemma Academic Program GCP Credit Award Spring 2025
Aligning Large Language Models with Human Creativity Preferences $10,000
PI: Roger Beaty, co-PI: **Antonio Laverghetta Jr.**

Google Gemma Academic Program GCP Credit Award Summer 2024
Generating Psychometrically Valid Creativity Ratings using Large Language Models
PI: Roger Beaty, co-PI: **Antonio Laverghetta Jr.** $5,000

PSU Center For Language Science Seed Grant Spring 2024
Unbiased Detection of Invalid Responses to Open-Ended Tasks using Large Language Models
PI: Roger Beaty, Janet van Hell; co-PI: **Antonio Laverghetta Jr.** $2,100

AWARDS

OPA Postdoc Travel Award March 2025

- Received a competitive $500 travel award to present research at an international conference.

IMPS 2021 Duolingo Student Award July 2021

- One of three to receive an award for best student presentation, out of 130 participants.

TEACHING EXPERIENCE

Instructor – Deep Learning and Large Language Models: An Introduction Spring 2026

- Built a new undergraduate course on generative AI.

Instructor – Exploratory Data Analysis and Pattern Discovery Spring 2026
Instructor – Machine Learning Methods for Audio and Video Analysis Fall 2025, Spring 2026

Instructor – Applications of Machine Learning in Data Analysis Fall 2025
Teaching Assistant – EdX Data Analytics Bootcamp 2020 – 2025

- Average instructional effectiveness score of ~4/5.

Teaching Assistant – Natural Language Processing Fall 2019
Teaching Assistant – Intro to Linux Summer 2019
Teaching Assistant – Programming Fundamentals Spring 2019

SERVICE

Invited Reviewer – NSF EDU CORE 2024, ERC 2025 COG, NSF EDU STEM K-12 2025
Journal Reviewer – ACM Transactions on Intelligent Systems and Technology, Thinking Skills and Creativity, Nature Communications, Creativity Research Journal, Journal of Creative Behavior
Conference Reviewer – EMNLP 2025, ECAI 2025, BEA 2025, EMNLP 2024, ECAI 2024, BEA 2024, CogSci 2024, NAACL 2024, BEA 2023
Member - Penn State Center for Language Science, Penn State Center for Socially Responsible AI, Penn State Neuroscience Institute
Coffee Hour Coordinator – Penn State Postdoctoral Society

- Facilitated professional networking events for postdocs at the University Park campus.

PRESENTATIONS

[1] Laverghetta Jr., A., Pronchick, J., Bhawsar, K. & Beaty, R. (2025) How do Humans and Language Models Reason About Creativity? A Comparative Analysis. Talk presented at the Society for the Neuroscience of Creativity. Paris, France.

[2] Laverghetta Jr., A., Luchini, S., Pronchick, J., & Beaty, R. (2025) Automated Detection of Invalid Responses to Creativity Assessments. Talk presented at the Society for the Psychology of Aesthetics, Creativity, & the Arts Conference. New Haven, CT.

[3] Laverghetta Jr., A. (2024) Psychometric Assessment Using Large Language Models. The first International Conference on Cognitive and Psychological Assessment and Enhancement. Beijing Normal University, Beijing, China. **Invited Keynote Speaker**.

[4] Laverghetta Jr., A., & Licato, J. (2023) Automatic Generation of Cognitive Test Items Using Large Language Models. Talk presented at the International Meeting of the Psychometric Society. University of Maryland, College Park.

[5] Laverghetta Jr, A., Nighojkar, A., Mirzakhalov, J., & Licato, J. (2021) Predicting Human Psychometric Properties Using Transformer Language Models. Talk presented at the International Meeting of the Psychometric Society. University of Maryland, College Park.

SKILLS

Programming Languages: Python, R, SQL, JavaScript, JSON, HTML, CSS, Java, REST, C, C++
Libraries: NumPy, SciPy, scikit-learn, PyTorch, Transformers, Matplotlib, Pandas, Jupyter, Plotly, LangChain, TensorFlow, Keras, OpenAI, Bootstrap, vLLM, PySpark
Software: git, Visual Studio, Anaconda, Postman, Microsoft Office, Qualtrics, Prolific, Docker
Cloud & Databases: Google Cloud, AWS, Postgres, SQLite, MongoDB

JOURNAL PUBLICATIONS

[J1] Laverghetta Jr., A., Luchini, S., Pronchick, J., & Beaty, R., Automated Detection of Invalid Responses to Creativity Assessments. *Thinking Skills and Creativity* (under review)

[J2] de Chantal, P. L., Beaty, R., Laverghetta, A., Pronchick, J., Patterson, J., Organisciak, P., ... & Karwowski, M. Artificial intelligence enhances human creativity through real-time evaluative feedback. *Journal of Experimental Psychology: General* (under review)

[J3] Luchini, S. A., Maliakkal, N., T., DiStefano, P. V., Laverghetta Jr., A., Patterson, J. D., Beaty, R. E., & Reiter-Palmon, R. Automated Scoring of Creative Problem-Solving with Large Language Models: A Comparison of Originality and Quality Ratings. *Psychology of the Aesthetics, Creativity, and the Arts.* (2025)

[J4] Bilgin, O., Fields, L., Laverghetta Jr, A., Marji, Z., Nighojkar, A., Steinle, S., & Licato, J. (2024). Exploring Prompting Approaches in Legal Textual Entailment. The Review of Socionetwork Strategies, 1-26.

[J5] Laverghetta Jr, A., Tran, M., Braynen, A., Steinle, S., Moydinboyev, B., Daas, H., & Licato, J. (2023). A survey of fatigue measures and models. The Journal of Defense Modeling and Simulation, 15485129231158580.

REFERRED CONFERENCE PUBLICATIONS

[C1] Ismayilzada, M., Laverghetta Jr, A., Luchini, S. A., Patel, R., Bosselut, A., van der Plas, L., & Beaty, R. Creative Preference Optimization. (2025) In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing* (**acceptance rate ~20%**)

[C2] Laverghetta Jr., A., Tuhin, C., Hope, T., Pronchick, J., Bhawsar, K., & Beaty, R. How do Humans and Language Models Reason About Creativity? A Comparative Analysis. (2025) *Proceedings of the Annual Meeting of the Cognitive Science Society* (**acceptance rate ~20%**)

[C3] Laverghetta Jr, A., Luchini, S., Linnell, A., Reiter-Palmonn, R., & Beaty, R. The creative psychometric item generator: a framework for automated item generation and validation using large language models. *CREAI 2024: Workshop on Artificial Intelligence and Creativity* (2024)

[C4] Laverghetta Jr., A., & Licato, J. (2023). Generating Better Items for Cognitive Assessments Using Large Language Models. In Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023) (pp. 414–428). Association for Computational Linguistics.

[C5] Nighojkar, A., Laverghetta Jr., A., & Licato, J. (2023). No Strong Feelings One Way or Another: Re-operationalizing Neutrality in Natural Language Inference. In Proceedings of the 17th Linguistic Annotation Workshop (LAW-XVII) (pp. 199–210). Association for Computational Linguistics.

[C6] Laverghetta A., Jr., & Licato, J. (2022) Developmental Negation Processing in Transformer Language Models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 545–551.* (**acceptance rate ~20%**)

[C7] Laverghetta Jr, A., Nighojkar, A., Mirzakhalov, J., & Licato, J. (2021). Predicting human psychometric properties using computational language models. In *The Annual Meeting of the Psychometric Society* (pp. 151-169). Cham: Springer International Publishing.

[C8] Laverghetta, A., Jr., Nighojkar, A., Mirzakhalov, J., & Licato, J. (2021). Can Transformer Language Models Predict Psychometric Properties? In *Proceedings of *SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics* (pp. 12-25). Online: Association for Computational Linguistics.

[C9] Laverghetta A., Jr., & Licato, J. (2021). Modeling Age of Acquisition Norms Using Transformer Networks. *The International FLAIRS Conference Proceedings, 34.*

[C10] Mirzakhalov, J., Babu, A., Ataman, D., Kariev, S., Tyers, F., Abduraufov, O., ... & Chellappan, S. (2021). A Large-Scale Study of Machine Translation in the Turkic Languages. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 5876 – 5890) (**acceptance rate ~ 23%**).

[C11] Laverghetta, A., Jr., Mirzakhalov, J., & Licato, J. (2020). Towards a Task-Agnostic Model of Difficulty Estimation for Supervised Learning Tasks. In *Proceedings of the 1st Conference of the Asia-Pacific*

Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: Student Research Workshop (pp. 16-23). Suzhou, China: Association for Computational Linguistics.

[C12] Boger, M., Laverghetta, A., Jr., Fetisov, N., & Licato, J. (2019). Generating Near and Far Analogies for Educational Applications: Progress and Challenges. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)* (pp. 1968-1975). Boca Raton, FL: IEEE.

PATENTS

[P1] Licato, J. & Laverghetta Jr, A. AUTOMATICALLY CREATING PSYCHOMETRICALLY VALID AND RELIABLE ITEMS USING GENERATIVE LANGUAGE MODELS. 20240339042 (2024)