

Parker Seegmiller

Hanover, NH 03755 · (385) 219-8093 · pkseeg.gr@dartmouth.edu · pkseeg.com

Selected Publications

1. **Seegmiller, P.** & Preum, S. (2026). “Measuring Distribution Shift in User Prompts and Its Effects on LLM Performance” *ACL 2026*. (distribution shift, LLM robustness, deployment evaluation)
2. **Seegmiller, P.**, Gatto, J., Greer, S., Isingizwe, G., Ray, R., Burdick, T., & Preum, S. (2026). “How Much Would a Clinician Edit This Draft? Evaluating LLM Alignment for Patient Message Response Drafting” *ACL 2026*. (clinical NLP, LLM alignment, evaluation methodology)
3. **Seegmiller, P.**, Mehta, K., Saha, S., Tao, C., Oraby, S., Gupta, A., Chung, T., Bansal, M., & Peng, N. (2025). “FLAMES: Improving LLM Math Reasoning via a Fine-Grained Analysis of the Data Synthesis Pipeline” *Findings of EMNLP 2025*. (LLM reasoning, data synthesis)
4. **Seegmiller, P.** & Preum, S. (2023). “Statistical Depth for Ranking and Characterizing Transformer-Based Text Embeddings” *EMNLP 2023*. (text embeddings, representation learning)
5. **Seegmiller, P.**, Gatto, J., Johnston, G., Basak, M., & Preum, S. (2023). “HealthE: Recognizing Health Advice & Entities in Online Health Communities” *ICWSM 2023*. (health NLP, named entity recognition)

Additional Publications

6. Gatto, J., **Seegmiller, P.**, Burdick, T. E., Khayal, I. S., DeLozier, S., & Preum, S. M. (2025). “Follow-up Question Generation For Enhanced Patient-Provider Conversations” *ACL 2025*. (clinical NLP, dialogue systems)
7. Gatto, J., Sharif, O., **Seegmiller, P.**, & Preum, S. M. (2025). “Document-Level Event-Argument Data Augmentation for Challenging Role Types” *ACL 2025*. (data augmentation, information extraction)
8. **Seegmiller, P.**, Gatto, J., Sharif, O., Basak, M., & Preum, S. M. (2024). “Do LLMs Find Human Answers To Fact-Driven Questions Perplexing? A Case Study on Reddit” *Workshop at ICWSM 2024*. (LLM evaluation, human-AI comparison)
9. Gatto, J., **Seegmiller, P.**, Burdick, T. E., & Preum, S. M. (2024). “In-Context Learning for Preserving Patient Privacy: A Framework for Synthesizing Realistic Patient Portal Messages” *Machine Learning for Health (ML4H 2024)*. (privacy-preserving NLP, synthetic data generation)
10. Amazon AGI Foundations. (2024). “The Amazon Nova Family of Models: Technical Report and Model Card” . (contributor; commercial foundation models)
11. Gatto, J., Sharif, O., **Seegmiller, P.**, Bohlman, P., & Preum, S. M. (2023). “Text Encoders Lack Knowledge: Leveraging Generative LLMs for Domain-Specific Semantic Textual Similarity” *EMNLP 2023*. (semantic similarity, domain adaptation)
12. **Seegmiller, P.**, Gatto, J., Basak, M., Cook, D., Ghasemzadeh, H., Stankovic, J., & Preum, S. (2023). “The Scope of In-Context Learning for the Extraction of Medical Temporal Constraints” *HealthNLP Workshop, ICHI 2023*. (in-context learning, clinical information extraction)
13. Gatto, J., **Seegmiller, P.**, Johnston, G., & Preum, S. (2022). “Identifying the Perceived Severity of Patient-Generated Telemedicine Queries Regarding COVID” *Journal of Medical Internet Research (2022)*. (transfer learning, clinical text classification)
14. Masters, T., Nayakankuppam, D., Yu-Buck, G., & **Seegmiller, P.** (2023). “Using the Behavioral Learning Theory to Examine COVID-19 Pandemic Stress” *EJHPE 2023*.
15. Masters, T., Nayakankuppam, D., Yu-Buck, G., & **Seegmiller, P.** (2022). “Consumption as Therapy: Individual and Country Factor Effects on Stress and Optimism During a Sustained Stressor” *AMA Winter Academic Conference 2022*. (paper presentation)

Experience

PhD Researcher

September 2021 – Present

Dartmouth College / Persist Lab, Hanover, NH

- Research focus: robustness of LLMs in high-stakes, low-data deployment settings
- Advisor: Prof. Sarah Masud Preum

Research Scientist Intern

June 2026 – August 2026

Collinear AI, Sunnyvale, CA

- Data synthesis for agentic large language models

Applied Science Intern

August 2024 – December 2024

Amazon — Foundations of AGI Team, Sunnyvale, CA

- Developed data synthesis pipeline for LLM mathematical reasoning; contributor to Amazon Nova, a deployed family of commercial foundation models (see: FLAMES, EMNLP 2025)

Teaching Assistant

September 2021 – Present

Dartmouth College, Department of Computer Science, Hanover, NH

- Machine Learning, Discrete Mathematics, Android Programming

Data Science Intern

June 2020 – August 2020

Aetna, a CVS Health Company (Payment Integrity Team), Remote

- Built ML model predicting healthcare provider abusive upcoding

Research Assistant

September 2018 – December 2020

Brigham Young University, Provo, UT

- ML and statistical research across CS, Marketing, and Statistics departments (fake news classification, sports analytics, behavioral science modeling)

CS Instructor

July 2019 – August 2021

Juni Learning, Inc., Remote

- Mentored 20+ students in Python and Java; prepared students for USA Computing Olympiad

Teaching Assistant

January 2019 – April 2021

Brigham Young University, Departments of CS & Statistics, Provo, UT

- Introductory data science, statistics, and programming courses

Education

PhD, Computer Science

September 2021 – November 2026

Dartmouth College, Hanover, NH

- Awards: Guarini PhD Innovation Fellow; Guarini School Graduate Student Poster Session 2023 Winner

BS, Statistics: Data Science

August 2017 – April 2021

Brigham Young University, Provo, UT

- Minors: Computer Science, Mathematics

Skills

Programming: Python, Java, C++, SQL, R

Frameworks & Libraries: PyTorch, HuggingFace Transformers, vLLM, sentence-transformers, scikit-learn, pandas, NumPy, SciPy

LLM & NLP: Fine-tuning (SFT), preference learning (RLHF/DPO), retrieval-augmented generation (RAG), agentic frameworks, LLM reasoning, chain-of-thought prompting, instruction tuning, prompt

engineering, text embeddings, synthetic data generation, data mixing & optimization, distributional robustness, evaluation methodology, clinical NLP, benchmark development

Infrastructure: Git, Linux/Bash, SLURM, Weights & Biases (W&B), MLflow, Jupyter

Human Languages: English (Native), Lithuanian (Advanced)

Service

Reviewer: NeurIPS (2025–2026), ACL (2026), EMNLP (2023–2025)