Alexander S. Tregub

GitHub: github.com/AlexTregub alexs.tregub@gmail.com Cell: +1 (330) 671-5352
Scholar: scholar.google.com/citations?user=yvmxP5MAAAAJ

Objective

Data scientist with dual background in Applied Mathematics and Computer Science, seeking opportunity to apply skills in mathematical modeling, statistical analysis, and algorithm development. Experienced in computational analysis, machine learning, and modeling applications through hands-on research projects and coursework. Other interests include developing simulations, optimization problems, and challenging puzzles.

Education

Kent State University Graduation: May 2026
Bachelor's of Science in Applied Mathematics, Computer Science GPA: 3.9/4.0

- NASA/Ohio Space Grant Consortium 2025-2026 Fellowship
- KSU Summer Undergraduate Research Experience Fellowship Summer 2023-2025
- President's List 2023-2024
- Dean's List 2021-2025

Skills

- Programming Languages and Tools: C++, Python, C, R, Bash, HTML, LaTeX, MatLab, Mathematica, PyTorch, TensorFlow
- Experience working with multidisciplinary teams, Scrum teamwork, Seminar presentations, Data visualization, Joint software development

Research and Work Experience

NASA/Ohio Space Grant Consortium Fellowship, "Developing Dimensionality Reduction Metric for RNA Editing Data from Microgravity Experiments" **2025 - Present**

- Implemented additional functionality to extend existing C tool for additional genomic data processing.
- Developed C++ tool for selecting and merging BAM (binary alignment map) read count data on multiple attributes.
- Developed utilities for processing VCF (variant calling) data in C++ and Python

Dr. Ruoming Jin's Machine Learning and Computational Science Research Lab (KSU) **2022 - Present**

- Presented topics in machine learning, computational modeling, and mathematical approaches.
- Developed Markov-Chain Monte-Carlo based simulation of protein-protein interactions in C++
- Derived proof of uniformity for placement approaches for modeling protein-protein interactions

Volunteer in Dr. Maimuna Majumder's Data Science and Digital Epidemiology Lab (Harvard Medical School) **2022 - 2025**

- Implemented Python toolset to collect Twitter data based on user-selected metadata and keywords for annotation.
- Assisted with the parallelized automation of Twitter data collection using shell scripting
- Developed Python tool to perform random resampling and linear regression analyses
- Assisted with data visualization tools to produce interpretable interim results

Summer Undergraduate Research Experience Fellowship, Kent State University **2023 - 2025**

- Developed simulations and proofs for "Modeling and simulating protein-protein interactions with an extension of random geometric configurations of varying-size proteins in three dimensions" with Dr. Ruoming Jin (KSU), in collaboration with Dr. Jack Su (CWRU)
- Explored "Analysis of Combinatorial Optimization on Matroids, and its Applications" with Dr. Ruoming Jin (KSU) (2023)
- Served as student leader for the SURE program. Represented the SURE program during Kent State's summer preview days for prospective students (2025)
- Presented posters at KSU Undergraduate Research Symposium (2024,2025)

Publications and Preprints

1. Lubwama, B., Ontiveros, J., Correll Carlyle, R., Kumar, S., Berkane, T., Puri, A., Tregub, A., Nitirahardjo, C., Morgan, E., Lawler, B.C., Aimone, E., H. Piontkivska, and M.S. Majumder. 2025. **Practical Considerations for Fine-Tuning BERT-Based Language Models in Health Research: Lessons from Classifying Anti-Vaccine Posts on Social Media.** Available at SSRN https://ssrn.com/abstract=5276034. http://dx.doi.org/10.2139/ssrn.5276034

Keywords: natural language processing, text sentiment categorization, fine-tuning, optimizing BERT model performance, benchmarking LLM models

2. Ontiveros, J., Correll Carlyle, R., Puri, A., Kumar, S., Tregub, A., Nitirahardjo, C., Morgan, E., Lawler, B.C., Aimone, E., Piontkivska, H., and Majumder, M.S., 2023. **Classification Performance Thresholds for BERT-Based Models on COVID-19 Twitter Misinformation.** Available at SSRN 4489865. http://dx.doi.org/10.2139/ssrn.4489865

Keywords: training data selection, logistic linear classifier, NLP, distilling, fine tuning, web scraping, data labeling, social media sentiment analysis

Personal Contributions: These two studies examined parameterization of BERT language models for sentiment classification. I developed Python tools utilizing Twitter's APIs to (i) extract tweet text based on a query consisting of relevant metadata and keywords, and (ii) to extract tweet IDs based on text queries, automating the collection of IDs to be used with the Hydrator tool to extract full text. Using parallelized Bash scripts, I automated the data collection. I also created a script for parsing and storing user message data encoded such that both labelers and language models could interpret emojis for additional context for sentiment analysis.

3. Nitirahardjo, C., Morgan, E., Lawler, B.C., Aimone, E., Tregub, A., Puri, A., Ontiveros, J., Correll Carlyle, R., Majumder, M.S., and Piontkivska, H., 2024. **Comparing the Usage of Russian-and Ukrainian-Derived Search Terms to Evaluate the Impact of Misinformation, Disinformation, and Propaganda in the US.** *Disinformation, and Propaganda in the US (June 20, 2024)*. Available at SSRN: https://ssrn.com/abstract=4871612. http://dx.doi.org/10.2139/ssrn.4871612

Keywords: orthography, search trend analysis, user sentiment analysis, linear regression analysis, bootstrapping

Personal Contributions: This study explored the use of Google Search Trend data for user sentiment analyses. I developed Python tool to perform automated resampling and regression analysis of search trend data, and created visualizations for interim results.