Deniz Bölöni-Turgut

db823@cornell.edu (407)-480-6687 denizbt.github.io linkedin.com/in/deniz-bt github.com/denizbt

EDUCATION

B.S. Computer Science (GPA: 3.94) *Aug 2022 – May 2026*
Cornell University, College of Engineering Ithaca, NY

Selected Coursework: OOP & Data Structures, Functional Programming, Honors Discrete Math, Analysis of Algorithms, Natural Language Processing, Deep Learning, Large-Scale ML, Probability & Statistics, Linear Algebra, Multivariable Calculus

RESEARCH EXPERIENCE

Undergraduate Research Assistant (NLP) *Jan 2024 – present*
Cornell University, Computer Science Department Ithaca, NY

- **Multi-Author Style Analysis (CLEF 2025 Shared Task)**: Designed and trained an ensemble model from multiple fine-tuned transformer models for multi-author writing style analysis: given pairs of sentences, identify whether the sentences were written by the same author or different authors (binary classification). Submitted system and published paper to PAN Lab at CLEF 2025, achieving average macro F1 score of 0.8. Advised by Prof. Claire Cardie.
- **BioLaySumm (ACL 2024 Shared Task)**: Fine-tuned T5 and BART for 91.7% increase in ROUGE score for lay summarization of biomedical articles with HuggingFace transformers. Improved readability scores by 31% with prompt engineering and creative use of the capabilities of OpenAI GPT API. Advised by Prof. Claire Cardie.
- **Open-Ended Question Evaluation**: Evaluated length bias and self-bias in open-ended essay responses by GPT-3.5 Turbo, GPT-4o and Llama3 LLMs. Utilized prompt engineering techniques to generate concise and verbose responses, while controlling and measuring semantic overlap. Developed a novel rubric for grading open-ended question responses. Computed inter-annotator agreement (Krippendorff's alpha) among human and LLM essay graders. Advised by Prof. Tanya Goyal and Prof. Claire Cardie.

Undergraduate Research Assistant (Parameter Efficient ML) *Aug 2025 – present*
University of Central Florida, Mathematics Department Hybrid (Orlando, FL)

- Developing sparse low-rank adaptation (LoRA) variants for language and vision models. Advised by Prof. Aritra Dutta.

Undergraduate Research Assistant (ML for Software) *May 2023 – May 2025*
Carnegie Mellon University, Software and Societal Systems Department Pittsburgh, PA

- Compiled an 800GB+ dataset of obfuscated binary code, starting from open-source C-language repositories. Designed a data pipeline to automate GitHub scraping, obfuscation, and compilation with g++ and clang. Refactored a large-scale Python codebase. Published paper in proceedings of DIMVA 2025. Advised by Prof. Claire Le Goues.

SOFTWARE EXPERIENCE

Software Development Engineer Intern *Jun 2025 – Aug 2025*
Amazon (Live Creator Tools team) New York, NY

- Designed and implemented an automated content moderation system for Amazon Live streams using computer vision models, reducing operational costs by about 88%. Built scalable solution leveraging AWS Step Functions, Lambda, and DynamoDB. Presented design document and product demo to stakeholders.

Data Scientist & Technical Project Lead *Feb 2023 – present*
Cornell Data Science (Engineering Project Team) Ithaca, NY

- Collaborated with peers to design, test, and deploy machine learning software projects with real-world applications. Completed projects include an ML system to predict winners for Formula One races, a scalable search engine for mathematic equations, and a natural language system to filter papers for open-access online research repository arXiv.
- Mentored new members at annual 24-hour Datathon. Presented at final showcase and to advisor Prof. Kilian Weinberger.

AWARDS

Honorable Mention, CRA Outstanding Undergraduate Researcher Award *Jan 2025*
Annual award given to top ≈ 200 undergraduate student researchers in North America with outstanding potential in computing research. Nominated by Computer Science Department of Cornell University.

National Merit Scholar *May 2022*

Dean's List, College of Engineering *Fall 2022 – present*

PUBLICATIONS

[1] **D. Bölöni-Turgut**, D. Verma, and C. Cardie, "Team cornell-1 at PAN: Ensembling Fine-Tuned Transformer Models for Writing Style Analysis," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Sixteenth International Conference of the CLEF Association (CLEF 2025)*, 2025.

[2] L. Dramko, **D. Bölöni-Turgut**, C. Le Goues, and E. Schwartz, "Quantifying and Mitigating the Impact of Obfuscations on Machine-Learning-Based Decompilation Improvement," in *Proceedings of the IEEE Conference on Detection of Intrusions, Malware, and Vulnerability Assessment*, 2025.

TEACHING & MENTORSHIP

Teaching Assistant, Mathematical Foundations of Computing (CS 2800) *Jan 2025 – May 2025*

- Led weekly 30 student recitation section to review discrete mathematics and probability concepts (i.e. random variables, finite automata, binary relations, induction) and teach proof writing. Hosted office hours to help students solve problems. Graded and provided feedback on proof-based problem sets.

Education Chair & Course Lecturer, Cornell Data Science *Jan 2024 – Dec 2024*

- Developed course materials and programming assignments for introductory machine learning class of 50-80 students. Delivered 50 minute weekly lectures for two university semesters. Mentored underclassmen and hosted office hours. Managed all administrative tasks and ensured effective communication with faculty advisor Prof. Rene Kizilcec.

Guest Lecturer, Introduction to ML Workshop, Cornell University *May 2025*

- Designed and led interactive 1.5 hour workshop in collaboration with Women in Computing at Cornell to teach machine learning fundamentals to the greater Cornell community.

PROJECTS

LoRA Code Reproduction | *Python, PyTorch, HuggingFace Transformers* *Apr 2025 – May 2025*

- Implemented LoRA for BERT and RoBERTa models from scratch, using methodology in seminal paper. Used PyTorch to inject custom written LoRA modules to self-attention layers of transformers. Fine-tuned BERT and RoBERTa with hyperparameters from paper to reproduce results on natural language understanding benchmark GLUE.

MathSearch | *Python, AWS (Lambda, Sagemaker, SQS), Docker, YOLOv8* *Sep 2023 – May 2024*

- Developed a full-stack ML web application, MathSearch, to solve problem of identifying math equations in PDFs. Led an Agile team of 15 developers through design and implementation of product with an end of semester deadline. Implemented a CI/CD ML pipeline with YOLOv8, MathPix API, and Levenshtein String similarity. Designed scalable and low-latency backend using queuing service, AWS SQS. Deployed with Docker and cloud services AWS Lambda and Sagemaker.

OScrabble Command Line Game | *OCaml, Scrum* *Sep 2023 – Dec 2023*

- Created 1200 line game OScrabble in functional OCaml, inspired by the titular board game. Implemented single and multi-player, easy/hard mode, automatic scoring, and board UI. Collaborated with two developers using Scrum methodology and test-driven development. Achieved 95% line coverage with 100 unit tests for bug-free gameplay.

Formula One Winner Predictor | *Python, PyTorch, NumPy, BS4, Pandas* *Feb 2023 – May 2023*

- Trained a model to predict the winners of the 2023 Formula One races, achieving a 61% accuracy. Used multilayer perceptron, Monte-Carlo simulation and custom loss functions implemented in PyTorch. Collected training data by scraping historical betting odds from multiple sources using BeautifulSoup4 and Selenium. Applied data mining techniques to enhance models.

TECHNICAL SKILLS

Programming Languages: Java, Python, OCaml, C/C++, HTML/CSS/JavaScript, MySQL, MATLAB, LaTeX

Machine Learning: PyTorch, NumPy, Pandas, HuggingFace Transformers, Sentence Transformers, Scikit-learn, SciPy, Matplotlib, OpenCV, Pillow, OpenAI/GPT API, Beautiful Soup, Selenium, Wandb, Streamlit

Other: AWS (Lambda, Sagemaker, DynamoDB, Step Functions, SQS, S3, EC2), Docker, Git, Slurm, Figma, Linux, VS Code

PERSONAL

(Natural) Languages: English, Turkish, Hungarian, French

Hobbies & Interests: Reading (sci-fi, fantasy, Jane Austen), stationery, long walks, cognitive philosophy, linguistics