

Ian Yang

iyang30@gatech.edu | ianyng.com | linkedin.com/in/ianyang02 | U.S. Citizen
<https://scholar.google.com/citations?user=7A4ZCDoAAAAJ&hl=en>

Education

B.S. in Computer Science, Georgia Institute of Technology (GPA: 3.92)

Expected May 2024

Threads: Artificial Intelligence and Theoretical Computing

Thesis: Relationship Extraction via Language Models for Normativity Analysis

Advisor: Mark Riedl

Relevant Completed Coursework: Natural Language Processing; Machine Learning; Artificial Intelligence; Science, Technology, and Human Values; Computing and Society; Cognitive Science; Advanced Algorithms; Automata and Complexity Theory

Pre-Prints

Louis Castricato, Alexander Havrilla, Shahbuland Matiana, Michael Pieler, Anbang Ye, **Ian Yang**, Spencer Frazier, and Mark Riedl. 2022. "Robust Preference Learning for Storytelling via Contrastive Reinforcement Learning".

Full Paper: <https://arxiv.org/abs/2210.07792>

Poster Presentations

Ian Yang, Norman Sadeh. 2023. "You Shouldn't Know That! Natural Language Processing Towards Enhancing Android Privacy Label Disclosures". *REU final research symposium at Carnegie Mellon University*.

Research Experience

Research Assistant | Carnegie Mellon University

August 2023 - Present

Research assistant in the Usable Privacy Lab advised by Prof. Norman Sadeh.

Exploring applications of NLP with respect to privacy policies and regulations. Projects include:

- Building and training classifiers with machine learning models (logistic regression, multilayer perceptron, etc) for predicting potential privacy disclosure compliance issues.
- Evaluating capabilities of LLMs and Generative AI tools (BERT, RoBERTa, ChatGPT, Llama2) to perform information extraction and classification of data collection practices based on the text of privacy policies. Manuscript in progress.
- Processing data from human study involving annotating privacy policies for automatically answering questions about data collection practices using GenAI tools. Manuscript in progress.

Research Assistant | Georgia Tech

Jan 2022 – Present

Research assistant in the Human-Centered AI Lab advised by Prof. Mark Riedl and mentored by PhD candidate Spencer Frazier. Have worked on a variety of projects, including:

- Extension of the *Moral Stories* dataset with human authoring of datapoints given world states.
- Human evaluations via Qualtrics for storytelling with fine-tuned language model preference learning. Paper submitted to EMNLP and IJCAI.
- Relational triple generation using large language models for knowledge graph creation, to be used for character action evaluation with respect to cultural norms (senior thesis).
- Generation of TextWorlds for navigation of reinforcement learning agents engaging in imaginative play. Research funded by Amazon.

REU Student | Carnegie Mellon University

May 2023 – August 2023

Visiting researcher as part of the REUSE (Research Experience for Undergraduates in Software Engineering) program at CMU advised by Prof. Norman Sadeh. Poster presented at final research symposium.

- Used *PoliPy*, a Python library for dynamically scraping privacy policies, to build large datasets (>700,000 rows) including privacy policies and data types collected.
- Built and trained machine learning classification models with privacy data to predict data labels for Android app privacy policies.

Research Assistant | The Ohio State University

Apr 2021 – May 2022

Research assistant at The Ohio State University advised by Prof. Dong Xuan and mentored by PhD student Yunqi Zhang. Relevant Projects:

- Implemented software detecting voice biometrics using ASR (Automatic Speech Recognition) supporting 19+ languages and reaching over 80% accuracy.
- Used *Vosk*, a speech recognition API, for speaker ID via cosine distance of x-vectors.
- Literature reviews of 15 distinct facial recognition algorithms.

Teaching Experience

CSE 6040 | Tutor | Georgia Tech

Aug 2022 – Present

Tutor for 1000+ students enrolled in CSE6040, Computing for Data Analytics, in the Master of Analytics program. Tasked with holding hourly sessions daily for topics including but not limited to: linear algebra, multivariate calculus, NumPy, pandas, machine learning algorithms (k-means, DBSCAN, PCA, logistic regression, image compression).

CS 1100 | Teaching Assistant | Georgia Tech

Aug 2021 – Present

Teaching Assistant for CS1100, Freshman Leap Seminar. Graded assignments including career goals, 4/5-year schedules, and resumes. Worked directly with Director of Computing Engagement for planning activities and class assignments.

Peer Mentor | Georgia Tech

Aug 2021 – Present

Mentored 55 freshmen in their transition to the College of Computing at Georgia Tech. Helped with adjusting to workloads, planning schedules, and minimizing stress when entering a rigorous and fast-paced academic environment.

AP Computer Science A | Tutor

May 2021 – Feb 2022

Tutored 4 students in preparation for the AP Computer Science A exam. Focused on beginner to intermediate level Java concepts from basic object-oriented programming including but not limited to: polymorphism, simple data structures, recursion.

Projects

Explainable Moral Alignment via Natural Language | Georgia Tech

May 2023

Fine-tuned GPT-3 on text data scraped from the subreddit *r/AmITheAsshole* to build a generative language model that provides both a moral judgement given a scenario and natural language explanation for the judgement.

- Found that fine-tuned model outperforms base GPT-3 both in accuracy of judgement and reasonableness of explanation (using crowdsourced judgements as ground-truth in training).

Industry Experience

Software Development Intern | FlightBridge, Inc.

Aug 2021 – May 2022

FlightBridge is the technology platform connecting Private Aviation with their trusted network of trip service vendors and partners. One-stop online to book and manage FBO and handling requests, hotel reservations, rental car bookings with FBO delivery, catering, fuel, ground transportation, aircraft detailing, security, and much more.

- Crafted full-stack solutions for 300+ features/bugs combining C#, JavaScript, SQL, and front-end frameworks.
- Optimized and overhauled more than 100 different sites and APIs with ASP.NET MVC and Subversion QC.
- Designed UI/UX of webpages visited by over 200,000 clients annually.

Technical Skills

Programming:

- Proficient: Python (NumPy, SciKit, pandas), Java, C# (ASP.NET)
- Familiar: C, SQL, HTML, CSS, JavaScript

Software/Tools: Git, Subversion, TeX, Bootstrap (Dashkit), Qualtrics, Android, Unix