

EDUCATION

- New York University, Courant Institute:** *Ph.D. in Computer Science* **2026 – Present**
- Working with Professors Eunsol Choi and Pavel Izmailov.
 - Supported by the *National Science Foundation Graduate Research Fellowship* (\$159,000) and the *NYU MacCracken Fellowship*.
- Yale University:** *BS in Computer Science, BS in Applied Mathematics, GPA 3.93, Cum Laude* **Class of 2026**
- Distinction in Computer Science major
 - Distinction in Applied Mathematics major
- Middle Tennessee State University:** *Non-Degree Student, GPA 4.00* **2018-2021**
- **Relevant Coursework:** Cloud Computing for Data Analysis, Advanced Python Programming, Computer Science 1 & 2
- University School of Nashville:** *High School Diploma, GPA 3.99 (unweighted)* **Class of 2022**

TECHNICAL EXPERIENCE

- Undergraduate Researcher in Computation and Language Lab (CAL), EECS/CSAIL, MIT** **May '24 – December '25**
- Selected for **2024 and 2025 MIT Summer Research Program (MSRP)**; chosen for **MSRP 2024 and 2025 Extensions**
 - Worked under Professor Yoon Kim
- Undergraduate Researcher in YaleNLP Lab, Computer Science, Yale** **August '23 – May '26**
- Worked under Professor Arman Cohan
- Software Engineering Intern: Space and Airborne Systems, L3Harris Technologies** **May '23 – August '23**
- Software engineer on a team developing automatic collision avoidance software for satellites, where I designed unit tests, fixed bugs, and implemented hierarchical state machines for mission applications
- Student Researcher in Robotics and Autonomous Systems Lab (RASL), Mech E, Vanderbilt** **July '18 – June '22**
- High school researcher focusing on software and mechatronic design for healthcare applications

PROJECTS

- Enhancing Data Efficiency for Transformer Models Using Probabilistic Context-Free Grammars** **CAL @ MIT: 2024 – 2025**
- Trained a Probabilistic Context-Free Grammar (PCFG) on a small text corpus to generate synthetic data which was used to pretrain language models with varying architectures
 - Synthetic data consisted of both text and string representations of the tree structures generated by the PCFG
 - The pretrained language models were then fine tuned on the original small corpus and evaluated
 - Observed that PCFG pretraining improved syntactic generalization and perplexity at small scales, though gains were modest and sometimes inconsistent compared to unstructured baselines
 - Showed that explicit hierarchical structure in pretraining data alone is insufficient to induce stronger inductive biases, motivating further study of which properties of data most effectively support syntactic generalization
 - [Project writeup](#) and [code](#)
- SciDQA: A Deep Reading Comprehension Dataset over Scientific Papers** **YaleNLP @ Yale: 2023 – 2024**
- Curated a question-answer dataset based on scientific reviews to generate deep-level questions for scientific literature
 - Extracted and refined questions and answers from reviews
 - Ran experiments with various language models and parameters
 - **Publication:** Shruti Singh, **Nandan Sarkar**, Arman Cohan, *SciDQA: A Deep Reading Comprehension Dataset over Scientific Papers*, EMNLP Main '24: <https://aclanthology.org/2024.emnlp-main.1163/>
- Design of a Communication and Data Collection App for Autism Intervention** **RASL @ Vanderbilt: 2021 – 2022**
- Coauthor publication in **JMIR Human Factors**: <https://humanfactors.jmir.org/2023/1/e45957>
- Design of an Augmented Reality System for Autism Intervention** **RASL @ Vanderbilt: 2019 – 2021**
- **Joint first author** in publication in **ACM Transactions on Accessible Computing**: <https://doi.org/10.1145/3481642>
- Design of an Intelligent System for Tremor Stabilization for Parkinson's Disease** **RASL @ Vanderbilt: 2018 – 2019**
- **Joint first author** in publication in **Lecture Notes in Computer Science**: https://doi.org/10.1007/978-3-030-22219-2_21

TECHNICAL SKILLS

- **Selected Programming Skills:** Deep learning, PyTorch, large language model prompting, pretraining and finetuning language models, object-oriented programming, virtual and augmented reality programming, microcontroller programming, embedded programming, CAD, Unity, Git, Agile/Scrum software development
- **Languages:** Python, C, C++, C#, Rust, Java, JavaScript, Racket