WILLIAM (BILL) XIA

ADDRESS
95 Winthrop St
Medford, MA 02155

LINKS
https://bill-from-ri.github.io/
https://www.linkedin.com/in/william-xia-ab40b2218/

CONTACT
401-834-5064
bill.xia.intern@3ds.com

EDUCATION

Bachelor of Science in Computer Science, Tufts University — September 2021 – May 2025
Minors in Mathematics & English
GPA: 3.81 / 4
Honors: *Summa cum laude* • Senior Honors Thesis • Dean's List (all semesters)
Clubs: Tufts Computer Science Exchange • Klezmer Ensemble (Violinist) • Tufts Greek Ensemble (Violinist) • Parnassus (Tufts Creative Writing Club) • Tufts Mountain Club

RESEARCH EXPERIENCE

Senior Honors Thesis: Grounding LLMs with Natural Language World Models September 2024 – May 2025
Advised by Professor Vasanth Sarathy and Professor Matthias Scheutz, Tufts University

- Analyzed the effect of natural language world models on LLMs' behavior when solving embodied reasoning puzzles, exploring the divide between natural language understanding and implicit knowledge.
- Implemented a modular MiniGrid environment for developing problem-solving-based puzzles and testing the spatial reasoning skills of LLM agents.
- Authored a 57-page thesis paper compiling my methods and results.

National Institutes of Health, Bethesda MD - Research Intern June 2023 – May 2025
Dr. Dina Demner-Fushman's Natural Language Processing Lab, National Library of Medicine

- Designed and released JEBS, a novel biomedical lexical simplification task and dataset comprising 21,595 annotated term-replacement pairs derived from 400 biomedical abstracts; developed end-to-end simplification system using BERT, Llama, and GPT-4o. Advised by Dr. Dina Demner-Fushman.
- First-authored a paper on JEBS accepted to ACL 2025; co-authored a companion paper submitted to Journal of Biomedical Informatics and presented its findings at the 2024 Text Retrieval Conference.
- Led development of JEBS baseline models, which achieved strong domain-expert human evaluation scores across simplicity, accuracy, and brevity.

Tufts University, Medford MA - Research Assistant May – December 2022
Professor Robert Jacob's Human-Computer Interaction Lab

- Conducted HCI research under Prof. Robert Jacob on differentiating mental workload states in users performing complex cognitive tasks (e.g. chess, n-back, mental rotation) using a MUSE EEG headset.
- Applied principal component analysis in Python and MATLAB to extract features from EEG and chess game data for random-forest-based classification of cognitive load.
- Co-authored 2 papers: one analyzing the efficacy of the MUSE headset for detecting mental workload states (submitted to IEEE TCDS) and another on correlations between chess move quality and brainwave signals (presented at Neuroadaptive Technologies 2025).

Tufts University, Medford MA - Research Volunteer June 2022
Professor Justin Hollander's Urban Attitudes Lab

- Performed data analysis on WebGazer.js eye tracking data using Python's MatPlotLib and Pandas Libraries.
- Communicated with a lab at Brown University to troubleshoot problems with WebGazer.js.

Brown University, Providence RI - High School Student Researcher July – August 2018
Professor Chun Guen Lee's Pathobiology Lab

- Read relevant literature and experimental protocols.
- Learned lab safety skills, how to take detailed research notes, and how to document experimental results.
- Accurately followed lab protocols while conducting experiments.

PUBLICATIONS / MANUSCRIPTS

- Brian Ondov, **William Xia**, Dina Demner-Fushman. "Lessons from the TREC Plain Language Adaptation of Biomedical Abstracts (PLABA) track". Submitted to *Journal of Biomedical Informatics*, 2025.
- **William Xia**. "Grounding Large Language Models with Natural Language World Models". Senior Honors Thesis. Submitted to *Tufts Digital Library*, 2025.
- **William Xia**, Ishita Unde, Brian Ondov, Dina Demner-Fushman. "JEBS: A Fine-grained Biomedical Lexical Simplification Task". Presented at the *63rd Annual Meeting of the Association for Computational Linguistics*, 2025.
- Matthew Russel, **William Xia**, Sam Youkeles, and Robert J.K. Jacob. "Neural Correlates of Move Quality During Chess Games: A Low-Cost EEG Study". Presented at *Neuroadaptive Technologies*, 2025.
- Matthew Russell, Samuel Youkeles, **William Xia**, Kenny Zheng, Aman Shah, Robert J.K. Jacob. "Decoding chess puzzle play and standard cognitive tasks for BCI: A low-cost EEG study". Submitted to *IEEE Transactions on Cognitive and Developmental Systems*, 2025.

PRESENTATIONS

- **William Xia**. "JEBS: A Fine-grained Biomedical Lexical Simplification Task". 63rd Annual Meeting of the Association for Computational Linguistics. July 28, 2025.
- **William Xia**. "Grounding Large Language Models with Natural Language World Models". Senior Honors Thesis Defense. April 25, 2025.
- Brian Ondov (Co-presenter), **William Xia** (Co-presenter), Ishita Unde, Hoa Dang, Dina Demner-Fushman. "PLABA @ TREC 2024". Text Retrieval Conference (TREC) 2024. November 20, 2024.
- **William Xia** (Presenter), Brian Ondov, Dina Demner-Fushman. "Identifying and Simplifying Non-consumer Terminology in Biomedical Abstracts". National Library of Medicine Summer Internship Program Poster Day. August 2, 2024.
- **William Xia** (Presenter), Brian Ondov, Dina Demner-Fushman. "Identifying and Simplifying Non-consumer Terminology in Biomedical Abstracts". National Institutes of Health Summer Internship Program Poster Day. August 1, 2024.
- **William Xia** (Co-presenter), Amelia Hawks (Co-presenter), Sarah Rose Odutola (Co-presenter). "Rewilding of Laboratory Mice Enhances Granulopoiesis". National Institutes of Health Summer Intern Journal Club: *The Last of Us: The Threat of Emerging Fungal Pathogens*. July 11, 2024.
- **William Xia** (Presenter), Brian Ondov, Dina Demner-Fushman. "Identifying and Defining Non-consumer Medical Terminology". National Institutes of Health Summer Internship Program Poster Day. August 9, 2023.
- **William Xia** (Presenter), Brian Ondov, Dina Demner-Fushman. "Identifying and Defining Non-consumer Medical Terminology". National Library of Medicine Summer Internship Program Poster Day. August 8, 2023.
- **William Xia** (Co-presenter), Ike Keku (Co-presenter), Jerry He (Co-presenter). "Improving Language Understanding by Generative Pre-Training". National Institutes of Health Summer Intern Journal Club: *Exploding Gradients: The Rise of Large Language Models in AI*. July 19 2023.

PROJECTS

Senior Capstone Project: Recommender System for ASL Learning Platform September 2024 – May 2025
Role: Machine Learning Engineer, Programmer

- Collaborated with a team to create an AI-powered lesson generation system for ASLdeafined, an online American Sign Language education program.
- Integrated the AI lesson generation system with ASLdeafined's database and front-end learning tools.

Language Model: RAG Pinterest Customer Support Chatbot April 2025
Role: Programmer

- Developed a retrieval-augmented generation (RAG) LLM chatbot, which sourced information from a knowledge graph representing hypothetical Pinterest user data.
- Developed a full stack webapp to host the chatbot, using Flask for the backend, React.js for the frontend, and Neo4j for the graph database.

Neural Network: Custom Model May 2024
Role: Programmer

- Programmed a neural network from scratch using Python and the NumPy library.
- Trained and evaluated the performance of the neural network using the *Iris* flower data set.
    - Optimized the neural network's performance by tuning hyperparameters with a validation set.

Language Model: Statistical Complex Term Identifier March – May 2024
Role: Programmer

- Developed a Categorical-Dirichlet statistical language model to identify complex biomedical terms in academic texts.
- Deepened my understanding of statistical machine learning and natural language processing.

Language Model: Sentiment Analyzer June 2023
Role: Programmer

- Programmed two sentiment analysis models from scratch using Python and the NumPy library.
    - One model used logistic regression to classify text; the other was a Naïve Bayes classifier.

Web Application: BitCalc December 2022
Role: Programmer

- Designed and built a programming calculator web app using HTML, JavaScript, and the React framework.
    - Functionalities included basic arithmetic, bitwise operations, and base changing.
- Maintained version control using GitHub.

Digital Game: Catacombs February – May 2022
Roles: Project Manager, Programmer

- Led a team of three classmates to create a video game using Unity and C#.
- Coordinated weekly team meetings and maintained a project timeline to keep the team organized.
- Designed and programmed the enemy AI system.

OTHER WORK EXPERIENCE

Dassault Systèmes, Waltham MA – SolidWorks AI Product Management Intern June 2025 – December 2025

- Developed and deployed a full-stack, local LLM agent that interfaces with Dassault's internal web APIs; the agent improved access to analytics and documentation for 20+ SolidWorks project managers.
- Deployed the AI assistant within the Jenkins CI/CD pipeline, enabling secure, scalable access across internal cloud systems and reducing manual data retrieval time by 25%.

Tufts University, Medford MA - Programming Languages Teaching Assistant September 2023 – May 2025

- Led weekly recitations and office hours for 30+ students a semester.
- Regularly answered student questions on the course *Piazza* forum.
- Graded students' assignments at teaching assistant grading meetings.

Ruffalo Noel Levitz - Telefund Student Fundraiser January – May 2022

- Fundraised for Tuft's various colleges and programs through engaging phone calls with alumni and parents.
- Handled confidential and sensitive information in a professional, appropriate manner.

Cumberland Public Library, Cumberland RI - Teen Volunteer June 2019 – December 2020

- Reshelved books and designed posters for library public events.
- Assisted with library children's events.

Moses Brown RISE Summer Camp, Providence, RI - Counselor in Training June – August 2018

- Assisted full counselors with camp activities and supervised campers ages 3-13 years old.

AWARDS & HONORS

- Dean's List, Tufts University, Fall 2021 – Spring 2025
- Intramural Research Training Award, National Institutes of Health, Summer 2023 – Summer 2024

COMPLETED ONLINE COURSES & CERTIFICATES

- **Reinforcement Learning Course.** Taught by David Silver. Published by Google DeepMind on YouTube. Completed October 2024.
- **Natural Language Processing Specialization.** Taught by Łukasz Kaiser and Younes Bensouda Mourri. Published by DeepLearning.AI on Coursera. Completed June 2023.
- **Python Data Structures.** Taught by Charles Russell Severance. Published by University of Michigan. Completed and certified May 2022.
- **Programming for Everybody (Getting Started with Python).** Taught by Charles Russell Severance. Published by University of Michigan. Completed and certified May 2022.

SKILLS

Programming Languages: Python, Java, C/C++/C#, JavaScript, Linux, SQL, SML/OCaml, PHP, R, MATLAB
Other Technical: Hugging Face Transformers, LangChain, PyTorch, MySQL, AWS, Unity, React.js, HTML, Git
Spoken Languages: English, Mandarin, Greek

HOBBIES

Creative Writing / Storytelling

- Published a novel called *A Merchant's Feud* through Amazon Kindle Direct Publishing.
- Served as Dungeon Master for various *Dungeons & Dragons* groups between 2019 and 2025.

Violin

- 16 years of experience, with training in classical, jazz, and various Eastern European folk traditions, including Hungarian, klezmer, and Greek music.
- Played in Tufts University's Klezmer Ensemble and Greek Music Ensemble.

LINKS

Personal Website: https://bill-from-ri.github.io/
LinkedIn: https://www.linkedin.com/in/william-xia-ab40b2218/
GitHub: https://github.com/bill-from-ri
Google Scholar: https://scholar.google.com/citations?hl=en&user=KoY_mHQAAAAJ