Sophie Shih

New York, NY | sophie59595@gmail.com | LinkedIn | sophieshih.com

Education

Duke University | *Computer Science and Neuroscience* | 4.0/4.0 — Expected: May 2030
Relevant Coursework: Data Structures & Algorithms; *Upcoming:* Intro to Computer Systems, Proof-Based Linear Algebra, Applied Product Management | Involvements: Catalyst Tech Society, DTech, The Chronicle, StreetMed Dance Team, Club Tennis

Stuyvesant High School | GPA: 4.0/4.0 || 97 — Jun. 2026
Classes: AP Computer Science, AP Physics C Mech, AP Physics C E&M, AP Calculus BC, AP Chemistry, AP Mandarin, AP US Government, AP Macroeconomics, AP European History

Experience

AI Research Intern — Jun. 2025 – Sep. 2025
Mass General Brigham, PI: Dr. Synho Do — *Remote / Boston, MA*
- Built Neurobiome Navigator, an app that teaches patients and students how the gut and oral microbiome relate to Parkinson's disease (demo)
- Engineered an AI agent (Pydantic AI, LangChain, OpenAI API) that answers natural-language questions using the MINERVA knowledge graph (66,000+ microbe–disease associations) and semantic search over selected research papers
- Designed interactive visualizations that pull microbe and food data from the knowledge graph and let users choose how to sort and view it
- Wrote a thesis on using AI to turn complex biomedical findings into accessible educational tools

Code Coach — Nov. 2024 – Jun. 2025
TheCoderSchool — *New York, NY*
- Coached 5 students (ages 7–10) 1-on-1 in Scratch, adapting lessons to each student's skill level and interests
- Designed 15+ custom coding projects and challenges teaching core concepts (loops, conditionals, variables)
- Created interactive Scratch games introducing physics concepts (gravity, acceleration, collisions)

Radiology AI Research Intern — Jul. 2024 – Jun. 2025
Weill Cornell Medicine, Department of Radiology (PI: Dr. Martin Prince) — *New York, NY*
- Annotated 500+ abdominal and pelvic MRIs of kidney disease (ADPKD) patients to train and validate an AI model that segments kidneys, liver, and cysts
- Contributed to TraceOrg, a public web tool whose 3D U-Net/transformer model measures organ and cyst volumes (0.97 Dice)
- Used few-shot LLM prompting to generate patient-friendly explanations of incidental MRI findings (RSNA abstract)
- Experimented with imaging modalities (T1, T2, SSFP) and segmentation algorithms for 3D volumetric MRI data

Healthcare AI Intern — Jun. 2023 – Jul. 2023
MD.ai — *New York, NY*
- Presented weekly reports on LLMs for healthcare, curating high-quality medical data for question-answering
- Tested features in MD.ai's reporting products (e.g. LLM-based automated radiology report generation)
- Engineered an LLM-agent framework to generate patient-friendly letters and co-authored an arXiv paper, winning 2nd place at the SIIM Hackathon (slides / video)

Volunteer — Summers 2024, 2025
Believers — *Tokyo, Japan*
- Led [crafts, geography, cooking, and music] activities for 20 elementary school students at an alternative learning space for students who struggle in traditional schools
- Supported students' social skills and confidence and Japanese-language instruction through 1-on-1 and small-group work

Projects

The New York Census Line · Live site | *Python, JavaScript, Census API, NYC Open Data* — Oct. 2026
- Designed and built an interactive "subway map" website exploring housing, immigration, and income across NYC's 55 neighborhoods
- Sourced and combined 2020–2024 Census, NYC Open Data, and BLS data, using inflation-adjusted trends and correlations to surface patterns (e.g. rent burden vs. income, $r = -0.79$)

Online Purchase Intent Prediction · GitHub · Slides | *Python, scikit-learn, pandas · Catalyst Tech Society, team of 4* — Fall 2026
- Built a Random Forest classifier predicting whether online shopping sessions end in a purchase, trained on 12,000+ sessions (0.92 ROC-AUC, 90% accuracy)
- Designed a leakage-safe preprocessing pipeline (deduplication, one-hot encoding, stratified 80/20 split) and analyzed precision–recall trade-offs under class imbalance (15.6% buyers)

College App Atlas · Live site · GitHub | *Next.js, TypeScript, PostgreSQL, Claude & OpenAI APIs* — Apr. 2026
- Launched a full-stack web app where students brainstorm, draft, and track college applications and essays
- Built a 5,700+ college database (IPEDS, College Scorecard) with prompts verified against official sites
- Designed an "idea garden" where students plant ideas, grow them into unique essay threads, and get AI-assisted college research tailored to their interests

Awards & Certifications

SIIM (Society for Imaging Informatics in Medicine) Hackathon — 2nd Place, 2024 (slides / video) and 2025 (slides / video)
CITI Program Training: Biomedical Research Investigators and Key Personnel, IRB Members, Good Clinical Practice

Skills & Tools

Languages: Python, Java, JavaScript, TypeScript, SQL, HTML/CSS, NetLogo, Racket
Frameworks & Libraries: Next.js, React, Streamlit, scikit-learn, pandas, LangChain, Pydantic AI, FAISS
Tools: Git/GitHub, Docker, PostgreSQL, Neo4j, Vercel, OpenAI & Claude APIs, Notion, OnShape/CAD
Skills: Machine Learning, LLM Agents & Prompt Engineering, Medical Image Annotation, Biomedical Research, Scientific Writing