Last updated in March 2026

Sean Hardesty Lewis

New York, NY 10044 | shl225@cornell.edu | +1 (346) 252-9443

linkedin.com/in/seanhardestylewis | github.com/seanhlewis | seanhlewis.com

Education

Cornell Tech (Cornell University), MS in Information Systems — **New York City, New York**
Technion — Israel Institute of Technology, MS in Information Science — August 2024 – May 2026

- GPA: 4.0
- Jacobs Certificate of Accomplishment in Computer Science
- Relevant coursework: Deep Learning, Machine Learning, Computer Vision, Trustworthy AI

The University of Texas at Austin, BS in Mathematics — **Austin, Texas**

- GPA: 3.7 — August 2022 – May 2024
- Relevant coursework: Advanced Calculus, Real Analysis, Scientific Computation, Software Design

Texas A&M University, BS in Computer Science* — **College Station, Texas**

- GPA: 4.0 — August 2021 – May 2022
- *Incomplete, transferred to UT Austin

Positions

Graduate Research Assistant, Urban Tech Hub, Cornell Tech — New York City, New York

- Developed data-driven climate resilience AI applications for NYC nonprofits. — August 2024 – Present
- Built NLP applications for the Urban Tech Hub; advised by Dr. Anthony Townsend.
- Developing embodied perception framework for robust scene understanding; advised by Dr. Wendy Ju.

Smart City Research Advisor, Urban Information Lab, UT Austin — Austin, Texas

- Took on an advisory role for directing research in the lab and leading development. — August 2024 – Present
- Spearheaded development of reproducible civic AI benchmarks for multilingual information retrieval.
- Authored papers on trustworthy, multilingual, and retrieval-augmented civic AI with Dr. Junfeng Jiao.

Software Engineer Intern, IBM — San Jose, California

- Built MoE chatbot integrated with IBM DB2, improving processing time by 20%. — May 2024 – August 2024
- Collaborated with NLP research team on prototype integration with Watsonx.ai APIs.
- Presented to Senior VPs of Software; secured executive support for continued development.

Undergraduate Research Assistant, Urban Information Lab, UT Austin — Austin, Texas

- Secured $30 k sponsorship from Dell Technologies (two ADA-6000 workstations). — August 2022 – May 2024
- Built harmonized dataset from 200+ U.S. cities for LLM fine-tuning; advised by Dr. Junfeng Jiao.
- Lead developer on $100 k grant for multilingual, grounded RAG chatbot for emergency communication.

Teaching

Teaching Assistant, Data Science in the Wild, Cornell University — January 2026 – Present

- Supported Dr. Allison Koenecke in delivering applied data science course for 110+ graduate students.
- Graded programming homeworks, exams, and projects spanning data, EDA, regression/inference, and ML.
- Provided project mentorship and technical feedback on end-to-end, supported course operations and Q&A.

Teaching Assistant, Deep Learning, Cornell University — August 2025 – December 2025

- Supported instruction and course operations with Dr. Hadar Elor for 80+ graduate students.
- Designed weekly quizzes and assignments on training/evaluating neural networks and computer-vision tasks.
- Graded coding assignments, quizzes, and projects covering CNNs, optimization, and representation learning.

Teaching Assistant, Machine Learning, Cornell University May 2025 – August 2025

- Aided Dr. Mohammad Al-Saad in teaching underrepresented Break Through Tech AI Fellows.
- Delivered weekly hands-on labs and provided 1-on-1 mentoring for 60+ undergraduate students.
- Built workshops and assignments to help launch the initiative's largest cohort (∼1000 Fellows).

Summer Academy Program Assistant, Department of Computer Science, University of Texas at Austin June 2022 – August 2022

- Co-taught three tracks: iOS/Swift (Xcode), Arduino C++ with PID, and HTML5/JS/Phaser for ∼60 students.
- Led Arduino robotics labs (C++/PID) with Dr. Justin Hart, guided teams to program autonomous navigation.
- Taught HTML5/JS game development with Phaser with Dr. Sarah Abraham, mentored projects to deployment.

Publications

StreamTwin: A Decentralized Digital Twin via Crowdsourced Sensing and Browser-Based Edge Computation January 2026

Sean Hardesty Lewis, Junfeng Jiao, Yiming Xu, Jihyung Park, Connor Phillips

Presented at AAAI 2026 Deployable AI Workshop
Association for the Advancement of Artificial Intelligence Conference (AAAI 2026)
Browser-based edge compute fuses detections from public webcams into a city-scale digital twin.

OpenCityCorpus: A Large-Scale, Harmonized, and LLM-Ready Corpus of Urban Data for Scientific Research December 2025

Junfeng Jiao, **Sean Hardesty Lewis**, Yiming Xu, Jihyung Park, Connor Phillips

Presented at NeurIPS 2025 AI4Science Workshop
Conference on Neural Information Processing Systems (NeurIPS 2025)
∼200 GB harmonized corpus spanning 200+ cities, with a schema-harmonization pipeline.

SafeMate: Providing Reliable, Step-by-Step Emergency Assistance with a Guideline-Grounded Agent December 2025

Junfeng Jiao, Jihyung Park, Yiming Xu, **Sean Hardesty Lewis**, Lucy Atkinson, Kristen Sussman

Under Review
Retrieval-grounded assistant that turns official emergency guidance into actionable, step-by-step instructions.

Minecraft to 3D: A Pipeline for High-Fidelity Reconstruction of Minecraft Worlds August 2025

Sean Hardesty Lewis

10.1145/3721250.3743044
ACM Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH 2025)
End-to-end pipeline for reconstructing Minecraft worlds as high-fidelity, interactive 3D scenes.

Working Papers

Why Re-infer the Known? Utilizing Place, Event, and Infrastructure Data in Scene Understanding Workflows 2026
Sean Hardesty Lewis, Matt Franchi, Wendy Ju

Manuscript in preparation.
Improving embodied scene understanding by utilizing spatiotemporal priors.

Workshop Publications

From Walled Gardens to Open Streets: A Pipeline for Cross-City Data Harmonization August 2025

Sean Hardesty Lewis, Junfeng Jiao

NeurIPS 2025 UrbanAI Workshop
Conference on Neural Information Processing Systems (NeurIPS 2025)
A novel workflow that harmonizes urban data from Socrata, ArcGIS, and CKAN.

Posters & Invited Talks

The Broken Promise of Data-Driven Urbanism - Designing a Design Process Workshop at Cornell Tech — Feb 2026

Slides

Specialization Research Projects - Invited Panelist at Cornell Tech (panel chaired by Dr. Deborah Estrin) — September 2025

NLP Workshop - Break Through Tech AI Workshop — July 2025

Slides

Agentic Exploration in Games - Cornell Tech Innovation Meets Impact — Dec 2024

Poster

AI Emergency Communication: Deterministic Models for Critical Information - 2nd Annual Smart Cities and Generative AI Symposium (Good Systems) — May 2024

Slides

Finding the Optimal Way Out: A Study on Bellman's Lost in a Forest Problem - Math for All Conference — April 2024

Slides

Using Generative AI for Digital Planning - Good Systems Symposium — March 2024

Poster

Multilingual AI-Assisted Emergency Preparedness - Good Systems Symposium — March 2024

Poster

Digital Planning for Sustainable Urban Future - Utrecht University, Netherlands (audience included Dr. Michael Batty) — January 2024

Slides

AI-Powered Insights: Extracting Value from Complex Data Ecosystems - 1st Annual Smart Cities and Generative AI Symposium (Texas Advanced Computing Center) — August 2023

Slides

Open Source Tools & Infrastructure

Searchable.City — Dec 2025 – Present

- Created the **first open-vocabulary semantic map** of New York City by running a vision-language model over millions of street-view images, converting streetscape visuals into structured, searchable urban signals.
- Deployed an interactive map with more than 1,000+ daily visits that lets users "Ctrl+F" the city (ex. *scaffolding*, *gothic*), surfacing neighborhood-scale patterns without curated labels or zip codes.

Human Operator (MIT Media Lab Hackathon Winner) — March 2026

- Won MIT Media Lab's HARDMODE (Hardware x AI) Hackathon by developing a human augmentation tool that allows AI to briefly take control of your body to help you learn or do things you cannot do.
- Led backend development of system, integrating camera stream with VLM and decision support system for voice commands to specific muscle actuation for AI controlling human body (wave, clench, finger actuation, etc.).

StreamTwin (Decentralized Digital Twin) — May 2025 – January 2026

- Developed browser-based edge compute system which fuses crowdsourced webcam detections into a live world model via an Aggregate Spatiotemporal Cache (ASC), no raw video leaves clients.
- Reconstructs scenes with 0.73 IoU on 10 live cameras while cutting per-stream bandwidth from 5 Mbps to 20 kbps ($>200\times$) with interactive traffic visualization.

Minecraft to 3D May 2025 – August 2025

- Trained 3D CNN and created end-to-end pipeline to upscale voxel worlds to interactive high-resolution environments, exports to Blender, Godot, Unity, and Unreal Engine.
- Network achieves 97.8% mean IoU on isolated structures with processing a 1 km² map (about 65 million blocks) taking 147s on RTX 4090 and never exceeding 3.2 GB of system memory due to sparse-voxel octree.

OpenCityCorpus August 2024 – July 2025

- Harmonized ~200 GB, 200+ U.S. city corpus from Socrata/ArcGIS/CKAN data sources into unified, semantically enriched schema for LLM training and RAG.
- Created queryable dataset with documentation and loaders.

SafeMate September 2024 – July 2025

- Led development on $100 k City of Austin grant for MCP-based, retrieval-augmented agent that routes to tools for policy retrieval, checklist generation, and structured summarization of trusted sources.
- Outperforms GPT-4o and GPT-3.5 on emergency preparedness queries (correctness, groundedness, completeness, relevance, fluency).

OpenCityAI March 2023 – August 2024

- Built ingestion + RAG pipeline over combined city portal data, retrieving citation-backed answers.
- Outperforms Google Bard and Microsoft Bing on city-data QA, with higher answer accuracy and groundedness.

SmartCityData August 2022 – February 2023

- Cross-city search and linking for municipal open-data portals across heterogeneous datasets.
- Normalizes schemas, ranks relevance, and visualizes coverage to reduce discovery friction.

Professional Service & Community Involvement

Member, EAAMO Urban Data Science Working Group August 2025 – Present

- Contributed to research discussion on data harmonization and digital twins.

President, Cornell Game Development Club August 2024 – Present

- Directed club and sponsors, fostering a collaborative environment for game development.
- Organized and facilitated industry-focused workshops, hackathons, and networking events.

SuperMaker, Cornell Tech Maker Lab August 2024 – Present

- Coordinated workshops and mentorship programs for aspiring makers and creatives.
- Organized monthly events to foster collaboration between students and industry professionals.

Researcher, Alan Turing Institute Data Study Group January 2026 – February 2026

- Embedded millions of job-posting descriptions in regression models to improve economic forecasts.
- Tested whether job-posting signals can support near-term unemployment prediction, vacancies, and skill-demand forecasts; contributed to forthcoming Turing Institute and Innovate UK BridgeAI report.

Mentor, Macaulay Honors Datathon (CUNY) September 2025

- Guided teams as they worked with MTA datasets to solve real-world urban challenges.
- Provided technical guidance, feedback on approaches, and helped participants create effective solutions.

Reviewer, NeurIPS 2025 Workshops: Reach & Limits of AI for Scientific Discovery, Structured Probabilistic Inference & Generative Modeling August 2025 – September 2025

- Reviewed submissions for NeurIPS 2025 Workshops (AI4Science, SPIGM, UrbanAI).

Technology Officer, Laurel Cooperative August 2023 – May 2024

- Managed technology operations, including maintenance, website creation, and event coordination.
- Organized collective house meetings in leadership role, overseeing decision-making processes, and resolving conflicts through consensus-based methods. Helped onboard new members into collaborative culture.

Honors & Awards

MIT Media Lab Hackathon Winner (HARD MODE) 2026

- Won MIT Media Lab hackathon with a custom-built EMS device that uses AI to control human body.

Hispanic Scholarship Fund (HSF) Scholar 2025

- Awarded to exceptional Hispanic students in higher education (also 2024, 2023).

Cornell Merit Scholarship 2024

- Awarded to exceptional graduate students upon entry to Cornell.

City of Austin Grant 2024

- Won $100,000 from CoA for developing AI-assisted multilingual emergency preparedness chatbot.

Dell Technologies Grant 2023

- Won $30,000 in form of two $15,000 RTX 6000 Ada Dell workstations to develop digital twins.

Fuller Endowed Scholarship 2023

- Recognized for academic excellence during undergraduate studies.

Merner Scholarship 2022

- Awarded for outstanding academic performance.

Eagle Scout, AP Scholar with Distinction, Spanish V Academic Award 2020

- Awards before enrollment in higher education.

Skills

Programming: Python, C++, C#, JavaScript, TypeScript, Java, HTML, CSS, SQL, XML, Bash, R, RISC-V

Tools: Linux, Visual Studio, Cloudflare, Unreal/Unity/Godot, PyTorch, Firebase, TensorFlow, Flask, Docker, Git, OpenCV, React, Blender, OSRM, AWS, Stata, ArcGIS, Selenium, Scrapy, WebRTC

Languages: English, Spanish