

# David Shumway

## Software Engineer & Data Scientist

Email: [davidshumway@gmail.com](mailto:davidshumway@gmail.com)  
Github: [@davidshumway](https://github.com/davidshumway)  
LinkedIn: [@david-shumway-0b4661159](https://www.linkedin.com/in/david-shumway-0b4661159)

Phone: (773) 759-5970  
Address: 160 N. Elizabeth St.,  
Unit 1909, Chicago, IL 60607

### Professional Experience

**Graduate Research & Teaching Assistant**, University of Illinois at Chicago 2018-Present

- Innovated with solutions in knowledge graphs, AI, GIS, ETL, data science, visual analytics, machine learning, and domain adaptation toward solving real-world problems.
- Taught and led discussions for 11 semesters, engaging with groups of 5-250 students in courses such as Languages & Automata, Software Development for Mobile, Data & Web Semantics, Big Data Mining, Software Engineering, and Software Design.
- Mentored students in 4 semesters with NSF-REU, fostering research skills and academic growth.

**Co-Owner & Lead Developer**, Finger Lakes Business Systems, Plattsburgh, NY 2010-2018

- Collaborated with 750+ clients through Amazon MTurk, completing 1000+ unique projects and 3+ million microtasks while maintaining a 99.9% task approval rating, and continued off-site with key clients.
- Engineered in-house web scraping, data management, workflow optimization, and distributed systems solutions, writing 375K+ lines of in-house code in JavaScript, PHP, HTML/CSS, SQL, and Python, along with extensive use of the Google Chrome API and browser extension programming, including browser tabs, background scripts, event scripts, pop-up scripts, and storage.
- Created a simple open-source browser extension to help streamline the MTurk worker interface, published to the Chrome Store and Firefox Add-Ons, reaching up to 10K weekly users.

**Summer Intern**, École Polytechnique Fédérale de Lausanne, DSIL, Lausanne, Switzerland 2017

- Enhanced Kamusi.org's Neo4j natural language database by analyzing and integrating new datasets.

**Digital Archivist & Technical Specialist**, Paul Brunton Philosophic Foundation, Burdett, NY 2010-2013

- Curated 100K+ scanned documents and designed LAMP, JavaScript, Bash, and VBA solutions.
- Automated scanning process, resulting in 4x efficiency and saving two years of manual work.

**Tech Freelancer, Delivery Driver, Retail Associate, Library Clerk** 2000-2010

- Completed freelance work (Rent-A-Coder, Freelancer.com, MTurk) while working full-time in other roles.

### Education

**Ph.D. in Computer Science**, University of Illinois at Chicago (expected Fall 2025) 2018-Present

- Advisors: Isabel Cruz (in memoriam), Cornelia Caragea.
- Completed coursework and PhD qualifying exam in *Scientific Workflow Mgmt. Systems Provenance*.

**B.S. in Computer Science**, State University of New York at Plattsburgh 2015-2017

- Dean's List honors in all five semesters; Annual CS Department Academic Excellence Award recipient.

### Example Projects

*Microbial Risks to Divers at Coastal Locations*, 2021-Present

- Developed waterborne illness ontology using Protégé, and used BERT and LLMs for ontology learning.
- Built spatio-temporal time series datasets using Python (Pandas, GeoPandas), and built and evaluated supervised, unsupervised, and transfer learning models with scikit-learn, PyTorch, and TensorFlow.

*FDIC Quarterly Banking App*, 2023

- Created a single-page app using Flask, JavaScript, and DataTables to analyze FDIC API reporting data from 5,000+ US financial institutions, and implemented automated retrieval and caching mechanisms.

*MALDI-DB*, 2020-2021

- Developed a Django web portal hosted on Mass Open Cloud with OpenShift for analysis of bacteria and small molecules from MALDI-TOF spectra, and integrated R processing functionality via RPlumber.

### Technical Skills

**Languages:** Python, Java, PHP, JavaScript, HTML, CSS, C++, C#, R

**Databases & Knowledge Graphs:** PostgreSQL, MySQL, MariaDB, SQLite, Neo4j, SPARQL, Virtuoso, Jena, Protégé, RDF, OWL

**Other:** Android, Bash, Linux, Git, Pandas, Numpy, GeoPandas, Jupyter, D3.js, Google Colab, matplotlib, scikit-learn, PyTorch, TensorFlow, Docker, Mass Open Cloud, REST, AJAX, cURL, Apache web server, Nginx, MySQL Workbench, node.js, browser extensions, Chrome, Django, Flask, Apache Cordova, Scrapy, LaTeX

**Built and maintained PCs and laptops** including Linux installation and maintenance.

### Publications

*Building a Knowledge Graph for Water Quality & Health Risks*, Gautam et al. (2023, co-author)