Samuel Brotherton

CONTACT INFORMATION

Phone: (801) 927-0219
Email: sbrother@gmail.com
LinkedIn: https://www.linkedin.com/in/sbrother
Address: 7771 Buckboard Drive, Park City, UT 84098

EDUCATION

Harvard University, Cambridge, MA

B.A., Mathematics and East Asian Studies **Sep 2008 – May 2012**

Received highest honors for senior thesis analyzing over 200,000 Chinese blog posts, algorithmically detecting mutations in the Chinese language in response to censorship. Completed coursework in abstract algebra, Galois theory, topology, real and complex analysis, probability theory, and linguistics.

PROFESSIONAL EXPERIENCE

Google, Los Angeles, CA / Park City, UT (Remote)

Software Engineer **Feb 2021 – Present**

Software engineer on an infrastructure team within Ads Privacy and Security, using natural language processing and machine learning to prevent policy-violating ad impressions. Our systems handle 2M+ queries per second and combine modern deep learning techniques with multi-tier human evaluation pipelines. Additionally, I have a 20% role building remote sensing infrastructure and machine learning models for a stealth climate "startup" within Google.

Pachama, San Francisco, CA / Park City, UT (Remote)

Software Engineer **Jul 2020 – Feb 2021**

Worked on a small team to productionize Pachama's remote sensing models using Python, Tensorflow, Kubernetes, Cloud Dataflow, and other GCP technologies. Designed and built a data ingestion service to retrieve, transform and cache satellite imagery and LiDAR point clouds for easy use in both model training and production. Left to rejoin Google under their new remote work policy.

Cairn Labs, Park City, UT

President and Founder **Jan 2016 – Mar 2020**

Lead a cross-functional team of 4-6 to design and build software for clients, with a focus on applications that integrate state-of-the-art NLP and machine learning. Notable projects/clients include:

- Lead a remote (US-based) engineering team to build AI-powered audiovisual experiences, including a cloud deep learning server for song generation, a C++ constraint/transformation library for imposing additional musical style on deep learning output, and an automated Max/Ableton session for audio generation. Coordinated with creative and executive teams to balance engineering and IP development with artistic and business goals. See `warpsound.ai` for examples.
- A deep learning based conversational UI framework to power will.i.am's wireless earphones and other applications. Backed by Python, Tensorflow, Prolog, and other technologies. Supports multiple languages, extensible dialogue flows, and includes a type system that integrates with a custom knowledge base. Our software had a significant role in the success of the client's $117M fundraising round (`https://goo.gl/iMvVTt`).
- An offline-first, mobile first Electronic Health Records system designed for the unique challenges of mobile refugee health clinics. Currently deployed in Lebanon and Nicaragua (see Brotherton et al in *Selected Publications*).
- An intelligent agent for a healthcare client that performs realtime conversational analysis from raw audio, running a series of clinically validated health classifiers on audio and text from realtime transcription (see Demiris et al in *Selected Publications*).
- A risk scoring server for a client in the car insurance industry, collecting realtime driving data and assigning machine-learning based driver risk scores at a rate of 3000+ qps (see `motioninsurance.com`).
- An advertising server to display appropriate event ticket offers to users, based on user-level and page-level contextual targeting. Backed by Elixir/Phoenix, Python, PostGIS, and React.js.

Google, Venice, CA

Software Engineer **Jun 2014 – Mar 2016**

Worked on a small team using natural language processing and other machine learning techniques to improve advertisement quality. Led a 20% project related to mining semantic information from web data, which was adopted by several teams across different product areas. Built a named entity recognition system in C++ and a link detection algorithm that runs on very large graphs; contributed to a topic model for clustering semantic entities.

Whisper, Venice, CA

Software Engineer and Data Scientist **Mar 2013 – Apr 2014**

Sole data scientist at a rapidly expanding social media startup seeing upwards of three billion monthly pageviews. Designed and built an NLP service to extract topics and tags from posts, predict image searchterms from unstructured text, and target content to users. Implemented a new geographic search system using PostGIS that decreased search time by 90%. Worked closely with the front and backend development teams, writing production code in Erlang and Python.

Selected Publications

Brotherton, T., Brotherton, S., Ashworth, H., Kadambi, A., Ebrahim, H. and Ebrahim, S., 2022. Development of an Offline, Open-Source, Electronic Health Record System for Refugee Care. *Frontiers in Digital Health*, 4.

Demiris, G., Oliver, D.P., Washington, K.T., Chadwick, C., Voigt, J.D., Brotherton, S. and Naylor, M.D., 2022. Examining spoken words and acoustic features of therapy sessions to understand family caregivers' anxiety and quality of life. *International Journal of Medical Informatics*, 160, p.104716.

Demiris, G., Corey Magan, K.L., Parker Oliver, D., Washington, K.T., Chadwick, C., Voigt, J.D., Brotherton, S. and Naylor, M.D., 2020. Spoken words as biomarkers: using machine learning to gain insight into communication as a predictor of anxiety. *Journal of the American Medical Informatics Association*, 27(6), pp.929-933.

Stephens, M., Bensink, M., Brotherton, S., Chandler, D., Garcia, J. and Hollenbeak, C., 2016. Geographic Access to Oncology Services in the United States (US): Travel Disparities May Affect Granulocyte-Colony Stimulating Factor (G-CSF) Administration. *Blood*, 128(22), p.5905.

Stephens, M., Brotherton, S., Dunning, S., Emerson, L., Gilbertson, D., Harrison, D.J., Kochevar, J., McClellan, A., McClellan, W., Wan, S. and Gitlin, M., 2011. The End-stage Renal Disease (ESRD) Prospective Payment System (PPS) and Access to Care: Incremental Distance Traveled by Displaced Patients.

Programming Experience

Languages: Python, C++, Erlang, Elixir, Mathematica, C#, F#, Bash, C, LaTeX
Server Technology: Kubernetes, Docker, GCP, Cassandra, Redis, PostgreSQL, ElasticSearch/Lucene