

Daniel C. Burfoot, PhD.

CONTACT INFORMATION 1935 Addison St #510
Berkeley, CA 94704
Cell: (617) 216-4249
E-mail: daniel.burfoot@gmail.com
Web: <http://danburfoot.net>
GitHub: <http://github.com/comperical>

SUMMARY Full-stack data architect who can handle big data systems, data analysis, and ML algorithm design. Ability to develop complex software systems from scratch. Research emphasis on natural language processing, machine learning, statistics, information theory, and philosophy of science. Strong academic background in mathematics and computer science. Excellent communication skills in both written and spoken English.

WORK EXPERIENCE

- Senior ML Engineer, LinkedIn. (5/2019-present). Built data engineering pipelines and ML algorithms for member language inference and position date inference. Migrated several systems to new LinkedIn ML framework.
- Founder, Ozora Research. (7/2013 - 5/2019). Primary developer for Ozora's advanced English text analysis and data compression system. This work involved algorithm design, NLP and linguistics research, statistical modeling, AWS architecture, performance tuning, web crawling, database design, and Java software engineering. Successfully developed a working, good-quality sentence parser without the use of labelled training data; this achievement appears to be unprecedented in the field. Parser can be viewed at <http://ozoraresearch.com/crm/public/parseview/UserParseView.jsp>
- NLP Contractor, Cargo Chief. (part-time, 1/2018 - 12/2018). Developed algorithms in Python to extract truck information (location, truck type, etc) from email text; key challenge is building robustness to widely varying email text layout. Built a suite of evaluation, management, and analysis tools for the system using MySQL, EC2, and CI tool. Created an admin web app console in Flask to help developers control, analyze, and debug the core NLP components.
- NLP/ML Contractor, UserTesting Inc. (part-time, 8/2017 - 12/2017). Research engineer for sentiment analysis tools using deep neural networks with Python and Keras. The essential challenge for this project was to obtain good results in spite of having limited labelled training data. Successful result due to creative solution to problem of "finding" more useful training data in client's system.
- Lead Scientist, Digilant Inc. (full-time 7/2011 - 7/2013, part-time 7/2013 - 7/2014). Primary developer for Audience Index, one of the Digilant's main data science offerings. This system involved work in machine learning, big data crunching on Hadoop and AWS, SQL database design, and visualization. Wrote the first version of the big data ETL system for Digilant's reporting solution that allowed customers to visualize their advertising usage statistics.
- Lead Developer, SmartCoachPlus training schedule software (full-time: 6/2010 - 12/2010) Improved version of original SmartCoach, implemented in server-side Java. Built all components of application from scratch, including SQL back-end, application logic, and presentation layer.

EDUCATION

University of Tokyo, Hongo, Bunkyo-Ku, Japan

Ph.D., Machine Learning (graduation date: April 2010)

- Thesis Title: Statistical Modeling as a Search for Randomness Deficiencies
- Areas of Study: Statistical Learning, Computer Vision, Robot Motion Control

McGill University, Montreal, Quebec, Canada

M.Sc., Computer Science (graduation date: September 2006)

- Thesis Title: Limitations of and Extensions to Heuristic Search Planning

University of Connecticut, Storrs, Connecticut, USA

M.Sc., Physics. (graduation date: June 2004)

Harvard University, Cambridge, Massachusetts, USA

B.A., Applied Math, specialization in Comp. Sci. (Graduation Date: June 1999)

AWARDS

- Japanese Government Monbusho Scholarship (2006-2010).
- Best Philosophy Paper Award, 10th International Conference on Simulation of Adaptive Behavior (2008).
- Dean's Honor List for M.Sc. Thesis, McGill University (2006).
- Andre J. Courtemanche Graduate Fellowship, McGill University (2005-2006).
- Sigma Pi Sigma (Physics Honor Society) Invitee, University of Connecticut (2004).
- Graduated *Cum Laude*, Harvard University (1999).
- Runner-up in "Best Web-Only Tool" category of Digital Awards competition of Magazine Publishers of America for SmartCoach (2007) (see above).

KEY

PUBLICATIONS

- *Notes on a New Philosophy of Empirical Science*. Book describing a philosophy of science based on large scale lossless data compression. The philosophy was developed to address chronic, serious limitations in the evaluation methodologies used in fields such as NLP and computer vision. Online at <https://arxiv.org/abs/1104.5466>
- *Statistical Modeling as a Search for Randomness Deficiencies*. PhD Thesis. Formulate an approach to statistical modeling based on process of 1) transform data set into encoded form (analogous to compressed bit stream) and 2) search encoded form for non-randomness.

SOFTWARE SKILLS

Highly Proficient: Java, Python, SQL, Web programming.

Significant Experience: Hadoop, Spark, Scala, JavaScript, HTML, AWS, Unix.

TECHNICAL

SCREENINGS

TopTal Core, TopTal AI, TripleByte

ACTIVITIES

President, Berkeley Entrepreneurial Toastmasters. Help to organize a club dedicated to the practice of public speaking.