

Objective: To debunk ambiguities and provide helpful insights from data.

Employment

- Senior Software Engineer **eBay Inc.** 02/2021 -
Working with the Search Services team at Core Products. Focusing on automation and productivity.
 - Instituted code style standards, nightly builds, and automated static analysis team-wide. (Java)
 - Major maintainer of eBay's in-house end-to-end testing framework. (Python)

- Software Engineer **Google, Alphabet Inc.** 09/2019 - 11/2020
Worked on the team performing Anomaly Detection for data warehouse at the YouTube scale.
 - Full-stack developer: Used Golang, Python, TypeScript, and SQL almost daily.
 - Worked on defining and increasing data quality coverage.
 - Developed a distributed system with job queues. Managed databases in a data warehouse.
 - Collaborated across teams to integrate highly scalable query engines and to design APIs.

- Research Assistant **World Well-Being Project, Penn Positive Psychology Center** 03/2018 - 05/2019
Conducted social psychology research using a Big Data approach. Quantitatively compared cultures across countries using terabyte-level microblog corpora. Used Hadoop, Spark, and word embedding models. Published 3 research papers using data mining techniques. Answered questions include:
 - How much anger, surprise, fear, joy, etc. is behind every smiley emoji?
 - Which ideograms (emojis) pop into people's minds when they think about [topic]?
 - How do Chinese people and Americans express politeness differently?

- Research Associate **Wharton Research Data Services, Wharton School** 11/2017 - 05/2019
Extracted tabular data from plain-text financial reports. Used regular expressions, Jupyter Lab environment, Linux maintenance, and single-machine parallel computing. Coordinated two Research Assistants for teamwork on manually inspecting machine-extracted documents. Parsed 139,983 Form 10-K filings (99.92% of all filings since the 1990s) from plain text or HTML to JSON.

Publications

- Studying Politeness across Cultures Using English Twitter and Mandarin Weibo. Proceedings of the ACM on Human Computer Interaction. <https://dl.acm.org/doi/abs/10.1145/3415190>
- From A Glance to "Gotcha": Interactive Facial Image Retrieval with Progressive Relevance Feedback. Accepted at the SIGIR 2020 Workshop on Applied Interactive Information Systems. <https://arxiv.org/abs/2007.15683>
- Studying Cultural Differences in Emoji Usage across the East and the West. Proceedings of the International AAAI Conference on Web and Social Media. <http://arxiv.org/abs/1904.02671>

- Exploring (Dis-)Similarities in Emoji-Emotion Association on Twitter and Weibo. In Companion Proceedings of The 2019 World Wide Web Conference. <https://doi.org/10.1145/3308560.3316546>
- Quantum Fisher information of triphoton states. *Chinese Optics Letters*. <https://www.osapublishing.org/col/abstract.cfm?uri=col-14-3-032701>

Education

Research interests: data science, cross-cultural social psychology.

Degree	Field	Institution	Country	Class
- M.Sc.Eng.	<u>Data Science</u>	University of Pennsylvania (UPenn)	USA	2019
- B.Sc., Honors	<u>Materials and Nanosciences</u>	University of Waterloo (UW)	Canada	2017
- B.Eng.	<u>Nanomaterials and Technology</u>	Beijing Jiaotong University (BJTU)	China	2017

Skills

Recently used Python, Go, SQL, TypeScript, and more at Google. Experienced in natural language processing (esp. word embeddings), information extraction from textual documents, and parallel computing. Keen on data mining and data warehousing.

Computer Languages	Most proficient in Python. Obtained Google Readability in Go and TypeScript. Wrote $\geq 2,000$ lines of code in each of MATLAB, R, Mathematica, and Java.
Data Management	MongoDB, SQL (MySQL, PostgreSQL, Procella, Dremel), Hadoop.
Parallel Processing	Hadoop / Spark, Open Grid Scheduler, Python multiprocessing & threading.
Data Analysis	Jupyter, pandas, matplotlib, seaborn, scikit-learn, ggplot2.
Mathematics	Probability, Linear Algebra, Calculus, Mathematical Statistics.
Other Coursework	Algorithms, Machine Learning, Information Theory, Machine Perception.
Languages	Mandarin (native), English (fluent; professional writing), Japanese (reading).

Links

- LinkedIn: <https://www.linkedin.com/in/tslmy/>
- GitHub: <https://github.com/tslmy>
- Google Scholar: https://scholar.google.com/citations?user=rSJ_vnYAAAAJ