NATARAJAN KRISHNASWAMI

Email nkrish@acm.org
LinkedIn nkrishnaswami
Cell +1 (845) 546 0732
City New York NY

Skills & Tools

Data stewardship & ethics
- PETs, primarily DP, some familiarity with SMPC
- PII, anonymization, de-identification, re-identification
- CKAN/CCMS metadata, provenance, veracity, archival
- Bias/fairness measurement and mitigation
- Big data management and analytic infrastructures

Data Science, machine learning & domains
- Decision trees, logistic & other regression methods
- NN methods (CNN, LSTM, Attention, VAE, GAN)
- NLP (embeddings, grammar induction, topic modeling)
- GIS/geographic analytics
- Visualization (D3, matplotlib, ggplot2, Vega-Lite)

Mathematics
- Analysis (real, functional, harmonic)
- Graph theory (spectral, algebraic, geometric)
- Statistics, learning theory, inverse problem theory
- Stochastic processes, dynamical systems

Software development
- Software Dev Life Cycle, agile methodologies, DVCS
- Robust, maintainable APIs, software, and architectures
- Portable software development (mostly C/C++/Python)
- Distributed and lock-free algorithms
- Performance, optimization, & scalability
- Typescript, JavaScript, node.js, Web Components

Education

Master of Information and Data Science ....University of California, Berkeley, School of Information

Bachelor of Science Cum Laude, Mathematics .... Case Western Reserve University

Distinctions

Affiliations
- Association for Computing Machinery
- International Association of Privacy Professionals
- New York Academy of Sciences

Patents
- US 8739171, High-Throughput Computing in a Hybrid Computing Environment
- US 8914805, Rescheduling Workload in a Hybrid Computing Environment

Public Service
- NYC Open Data Ambassador, 2022-present
- D4BL volunteer co-tech lead for COVID Data, 2020-2021
- NYC Open Data Queens Borough Ambassador, 2019-2021
- Lesbians Who Tech LGBTQ Tech & Innovation Fellow, 2017
- White House LGBTQ Tech & Innovation Fellow, 2015

Awards
- IBM Outstanding Technical Achievement (OTA) Award ISV porting consultation driving $300+m revenue

Professional Experience

Twitter ....Sep 2021- Jun 2022

Staff ML Engineer, Machine Learning Ethics, Transparency, and Accountability (META)
- Identified and prioritized internal client needs satisfiable using privacy-enhancing technologies (PETs) [differential privacy, federated ML, secure multiparty computation, requirements analysis]
- Investigated and documented current state of policy research on PETs for US and EU regulatory compliance
- Eng. point of contact on partnership for privacy-preserving data analytics for research replication [PySyft, Docker, Kubernetes]
- Designed/developed service to integrate privacy-preserving model bias assessment into key points in ML model dev/deployment lifecycle and product A/B testing platform [algorithmic bias metrics; Python, PipelineDP, Spark, TensorFlow, TFMA, Swagger; GCP: Compute Engine, Endpoints, IAM, Dataflow; API design, UX, agile development]
- Worked with fairness researchers to identify engineering opportunities to improve research workflow/iteration velocity [UX]

Google 2017-2021

Senior Software Engineer, Public Statistics and Query Understanding in Search
- Gathered requirements, collaboratively designed, and implemented support for complex, nested, fact-seeking queries in Search and Assistant: "debbie reynolds grandchildren fathers ages" [NLP/NLU/NER, C++17, Knowledge Graph, ML, ranking]
- Co-authored & designed/implemented tooling for open statistical dataset metadata spec: DSPL 2 [JSON-LD, schema.org, Python, open standards, open source, agile development]
- Greatly expanded population query coverage by identifying most up-to-date, authoritative sources, and automating ingestion of fine-grained population data for European and Canadian administrative regions & municipalities [open/public data]
- Collaboratively developed new schema to support answering faceted statistical queries: "hispanic population california"
- Prototyped Dataset Search integration into main Search Results Page [Java, UX, requirements analysis, agile development]
- Prototyped augmenting stats-seeking news queries (e.g. unemployment rates) with relevant context [petascale, MapReduce]

NATARAJAN KRISHNASWAMI

Email nkrish@acm.org
LinkedIn nkrishnaswami
Cell +1 (845) 546 0732
City New York NY

Professional Experience, continued

Bloomberg LP... 2010-2017

Senior Software Engineer, Portfolio & Risk Analytics

- Designed, implemented, and owned minimally intrusive tracing & diagnostics API for SoA platform (observability/telemetry); crucial for Portfolio Analytics service and support, leading to wider adoption [C++, SOA, UNIX, agile dev]
- Co-designed and implemented industry-leading rules-based/conditional/hierarchical classification engine for assets, grounded in concepts and best practices from information retrieval & CS/automata theory; critical for retaining customers from a recent acquisition and reducing technical debt by unifying 3 earlier systems [C++, systems design, UX, agile dev]

IBM Corporation.. 1998-2010

Advisory Software Engineer, z/OS and Linux

- Contributed to vibrant z/OS UNIX app ecosystem by providing POSIX, C/C++, Java application performance & portability expertise to many internal & independent software vendors, and porting much open-source software to z/OS [open stds, open source]
- Developed I/O performance monitoring (CMF) driver for Linux for zSeries [C, z/Architecture, Linux kernel, open source]
- Designed and implemented many standards conformant, high performance, reliable C/C++/POSIX runtime & filesystem features
- Planned and led team developing shared-disk, heterogeneous OS, distributed file system. [C, kernel dev, networking, filesystems]
- Increased speed 100x of very heavily used POSIX thread-specific data APIs with a novel, constant-time, lock-free algorithm
- Increased speed 2x of approximate name list search (onomastic analytics) with bit-parallel similarity scoring algorithm [C++, IR]

Publications

Persistent Bias Among Local Election Officials.. Journal of Experimental Political Science

- Audit study to determine if local elections officials respond at disparate rates to emails containing different racial/ethnic cues
- Local-election-authority level randomized design blocking on political and demographic covariates
- Tools: R/blocktools, IPython, pandas, custom python wrapper for Census API data, mongodb, selenium webdriver to scrape email addresses, pydkim and marrow.mailer for contacting subjects

Selected Public Data-Related Projects

Data for Black Lives COVID-19 Dataset...Dataset overview, GitHub repo

- 2020 D4BL volunteer team to build automated open-source scrapers for US state and city COVID-19 web dashboards, co-lead
- Wrote scraping framework and several per-state scrapers [Python, web scraping, ETL]
- Onboarded new volunteers [technical leadership, documentation]
- Instituted and led code review and, static checking, and other quality processes [technical leadership, documentation]
- Tools: custom Python 3.8 scraper framework, requests.py and selenium for retrieval, census API for demographic data, many more

Eviction&..Twitter, Outcomes slides

- 2017 LWT LGBTQ Tech & Innovation Summit. Salvation Army-sponsored volunteer project to support full utilization of unused grant funds for keep those at-risk in their homes in the Houston area
- Analysis and visualization of temporal and geospatial trends in Harris County, TX, eviction filing patterns
- Responsible for scraping, geocoding, data analysis and exploratory visualization
- Tools: requests.py for scraping, census API for geocoding, Carto for storage/visualization

#TransNeeds ....Twitter, Outcomes slides

- 2015 White House LGBTQ Tech & Innovation Summit. Project for civic/economic/political inclusion volunteer working group
- National Transgender Listening Campaign: hashtag and SMS campaign to gain insight into needs & priorities of trans* community, culminating in recommendations for trans-inclusive federal policies
- Responsible for data analysis: identifying salient topics and harassing tweets
- Tools: IPython for EDA, tweepy for data acquisition, sklearn for text analytics, clustering, topic modeling

AidSight.. aidsight.org

- Master's degree capstone project
- Platform for international development/aid workers to search, explore collaboration structure in, and assess data quality for International Aid Transparency Initiative (IATI) standard data sets.
- Tools: AWS S3, EC2; MongoDB for cleaned/normalized data; Neo4J for graph data; IPython/matplotlib/sklearn for EDA; Flask for API/backend; JavaScript/D3 front end.