Siddhartha Jain

CONTACT INFORMATION

Email: tmfs10@gmail.com   Homepage: https://tmfs10.github.io

WORK EXPERIENCE

Amazon AWS AI, Pasadena, CA
Applied Scientist II   Oct 2021 – Present
Working on training Large Language Models (LLM) for code generation; prompting strategies for unit test generation using LLMs; reranking outputs of LLMs; prompting for LLMs using feedback (Product is AWS CodeWhisperer which is part of Amazon's recent Generative AI push)

Amazon Alexa, Cambridge, MA
Applied Scientist II   July 2020 – Oct 2021
Worked on applying BERT models to NLU problems – in particular domain classification and entity recognition

Massachusetts Institute of Technology, Cambridge, MA
Postdoctoral Associate   December 2017 – July 2020
Worked on statistical independence analysis for protein sequences, novel active learning techniques for deep learning, auxiliary objectives for deep network ensembles, interpretation of deep learning models for language, vision, and biology

EDUCATION

Carnegie Mellon University, Pittsburgh, PA
Ph.D. Computer Science   September 2011 – October 2017
advised by Prof. Ziv Bar-Joseph (Feb 2013 – October 2017)

Brown University, Providence, RI
B.Sc. Mathematics-Computer Science   September 2007 – May 2011

PUBLICATIONS UNDER SUBMISSION

Information Condensing Active Learning.
Siddhartha Jain, Ge Liu, David Gifford.

REFEREED PUBLICATIONS

Multi-lingual Evaluation of Code Generation Models.
Ben Athiwaratkun, Sanjay Krishna Gouda, Zijian Wang, Xiaopeng Li, Yuchen Tian, Ming Tan, Wasi Uddin Ahmad, Shiqi Wang, Qing Sun, Mingyue Shang, Sujan Kumar Gonugondla, Hantian Ding, Varun Kumar, Nathan Fulton, Arash Farahani, Siddhartha Jain, Robert Giaquinto, Haifeng Qian, Murali Krishna Ramanathan, Ramesh Nallapati, Baishakhi Ray, Parminder Bhatia, Sudipta Sengupta, Dan Roth, Bing Xiang *ICLR 2022*, 2022

Overinterpretation reveals image classification model pathologies.
Brandon Carter*, Siddhartha Jain*, Jonas Mueller*, David Gifford *Neurips 2021*, 2021

Machine learning optimization of peptides for presentation by class II MHCs.
Zheng Dai, Brooke Huisman, Haoyang Zeng, Brandon Carter, Siddhartha Jain, Michael Birnbaum, David

Gifford
Bioinformatics, 2020

Robust computational design and evaluation of peptide vaccines for cellular immunity with application to SARS-CoV-2.
Ge Liu*, Brandon Carter*, Trenton Bricken, Siddhartha Jain, Mathias Viard, Mary Carrington, David Gifford
Cell Systems, 2020

Maximizing Overall Diversity for Improved Uncertainty Estimates in Deep Ensembles.
Siddhartha Jain*, Ge Liu*, Jonas Mueller, and David Gifford.
AAAI 2020, 2020

Machine Learning Optimization of MHC class II Presented Peptides.
Haoyang Zeng, Brandon Carter, Siddhartha Jain, Brooke Huisman, Michael Birnbaum and David Gifford.
Machine Learning in Computational Biology, 2019

Transcriptional regulatory model of fibrosis progression in the human lung.
John McDonough, Farida Ahangari, Qin Li, Siddhartha Jain, Stijn E. Stijn E. Verleden, Jose Herazo-Maya, Milica Vukmirovic, Giuseppe DeIuliis, Argyrios Tzouvelekis, Naoya Tanabe, Fanny Chu, Xiting Yan, Johny Verschakelen, Robert Homer, Dimitris V Manatakis, Junke Zhang, Jun Ding, Karen Maes, Laurens De Sadeleer, Robin Vos, Arne Neyrinck, Panayiotis Benos, Ziv Bar-Joseph, Dean Tantin, James Hogg, Bart V Vanaudenaerde, Wim Wuyts, Naftali Kaminski.
JCI Insights, 2019

What made you do this? Understanding black-box decisions with sufficient input subsets .
Brandon Carter*, Jonas Mueller*, Siddhartha Jain, David Gifford.
AISTATS, 2019

Approximate Mutual Information-based Acquisition for General Models in Bayesian Optimization.
Siddhartha Jain, Nathan Hunt, David Gifford.
NeurIPS Workshop on Bayesian Deep Learning, 2018

Maximizing Overall Diversity to Control Out-of-Distribution Behavior of Deep Ensembles .
Siddhartha Jain, Ge Liu, Jonas Mueller.
NeurIPS Workshop on Bayesian Deep Learning, 2018

Using neural networks for reducing the dimensions of single-cell RNA-Seq data.
Chieh Lin, Siddhartha Jain, Hannah Kim, Ziv Bar-Joseph.
Nucleic Acids Research, 2017.

Transcriptome analyses identify key cellular factors associated with HIV-1 associated neuropathogenesis in infected men.
Narasimhan J. Venkatachari, Siddhartha Jain, Leah Walker, Shalmali Bilavaker-Mehla, Ansuman Chattopadhyay, Ziv Bar-Joseph, Charles Rinaldo, Ann Ragin, Eric Seaberg, Andrew Levine, James Becker, Eileen Martin, Ned Sacktor, Velpandi Ayyavoo.
AIDS Journal, 2017.

Reconstructing the temporal progression of HIV-1 immune response pathways.
Siddhartha Jain, Joel Arrais, Narasimhan J. Venkatachari, Velpandi Ayyavoo, Ziv Bar-Joseph.
Intelligent Systems for Molecular Biology, (ISMB), 2016

Temporal transcriptional response to latency reversing agents identifies specific factors regulating HIV-1 viral transcriptional switch.
Narasimhan J Venkatachari, Jennifer M Zerbato, Siddhartha Jain, Allison E Mancini, Ansuman Chattopadhyay, Nicolas Sluis-Cremer, Ziv Bar-Joseph, Velpandi Ayyavoo.
Retrovirology, 2015.

Multitask Learning of Signaling and Regulatory Networks with Application to Studying Human Response to Flu.
Siddhartha Jain, Anthony Gitter, and Ziv Bar-Joseph.
PLOS Computational Biology. 10:12, 2014 and
Society for Laboratory Automation & Screening, (SLAS), 2015

Large Neighborhood Search for the Dial-a-Ride Problem.
Siddhartha Jain and Pascal Van Hentenryck.
17th International Conference on Principles and Practices of Constraint Programming, (CP), 2011.

A General Nogood-Learning Framework for Pseudo-Boolean Multi-Valued SAT.
Siddhartha Jain, Ashish Sabharwal, and Meinolf Sellmann.
25th Conference on Artificial Intelligence, (AAAI), 2011.

A Complete Multi-Valued SAT Solver.
Siddhartha Jain, Eoin O'Mahony, and Meinolf Sellmann
16th International Conference on Principles and Practice of Constraint Programming, (CP), 2010.

Upper Bounds on the Number of Solutions of Binary Integer Programs.
Siddhartha Jain, Serdar Kadioglu, and Meinolf Sellmann.
7th International Conference on Integration of AI and OR Techniques in Constraint Programming, (CP), 2010.

POSTERS

Reconstructing the temporal progression of HIV-1 immune response pathways.
Siddhartha Jain, Joel Arrais, Narasimhan J. Venkatachari, Velpandi Ayyavoo, Ziv Bar-Joseph.
Probabilistic Modeling in Genomics, 2015.

Transfer learning for reconstructing dynamic signaling and regulatory networks.
Siddhartha Jain, Anthony Gitter, and Ziv Bar-Joseph.
18th Annual International Conference on Research in Computational Molecular Biology, (RECOMB), 2014.

TECHNICAL REPORTS

Parallel Heuristics for TSP on MapReduce.
Siddhartha Jain and Matthew Mallozzi.
Brown University Tech Report, 2011

HONOURS AND AWARDS

Wally George Fellowship offered at Georgia Tech for Ph.D. studies
Ontario Trillium Scholarship offered at U. of Toronto for Ph.D. studies
Undergraduate Teaching and Research Assistantship Award Grant for doing Research for Summer 2010
Perry and Dr. Hilary Hoffmeister Brown Annual Fund Scholarship for years 2007-8, 2008-09, 2010-11

TEACHING EXPERIENCE

Carnegie Mellon University

- Graduate teaching assistant for Introduction to Machine Learning in Fall 2014
- Graduate teaching assistant for Principles of Imperative Computation in Spring 2014

Brown University
- Head Teaching Assistant for *Introduction to Computer Systems* in Fall 2009 and Fall 2010
- Teaching Assistant for *Design & Analysis of Algorithms* in Spring 2010
- Teaching Assistant for *Introduction to Computer Systems* in Fall 2008

SERVICE

Reviewer for ICLR 2022, ICLR 2021; ICML 2020; UAI 2019-20; AISTATS 2020-21; PLOS One 2014, 2019; Bioinformatics 2015-16,2019; NeurIPS 2019-20; TCBB 2019; RECOMB 2016-19; ICML WCB 2017; ISMB 2015-16; ACM-BCB 2015; CP 2013; CPAIOR 2013

Member of the Program Committee for UAI 2019, ICML Workshop on Computational Biology 2017

REFERENCES

On request.