Joshua Fan

Email: jyf6@cornell.edu
Phone: (425)-628-7677
205 Dryden Road (Apt 401) -- Ithaca, NY 14850
Website: joshuafan.github.io
GitHub: github.com/joshuafan

EDUCATION

Ph.D. in Computer Science, Cornell University, Ithaca, NY (in progress)   **Aug. 2019 – present**
  ➢ Advisor: Prof. Carla Gomes   ➢ Research focus: Machine Learning, Computational Sustainability

M.S. in Computer Science, University of Washington, Seattle, WA   **Mar. 2017 – Jun. 2019**
  ➢ **GPA: 3.84/4.0**   ➢ Advisor: Prof. Sreeram Kannan
  ➢ **Selected courses:** Deep Reinforcement Learning, Online and Adaptive Machine Learning, Machine Learning for Big Data, Probabilistic Graphical Models, Theory of Optimiszation and Continuous Algorithms, Databases

B.S. in Computer Science, University of Washington, Seattle, WA   **Sep. 2013 – Mar. 2017**
  ➢ **GPA: 3.97/4.0** *(summa cum laude)*

RESEARCH EXPERIENCE

Research Assistant, Cornell Institute for Computational Sustainability *(with Prof. Carla Gomes)*   **Aug. 2019 – present (part-time)**
  ➢ Exploring techniques for multi-entity modelling using deep learning, with applications to predicting agricultural crop yield

Research Assistant, UW Institute for Protein Design *(with Prof. David Baker)*   **Jan. – Jun. 2019 (part-time)**
  ➢ Implemented 3-D CNN, GAN, and 3-D Variational Autoencoder, and used them to assess and refine protein structure quality
          Jun. – Sep. 2017 (full-time)
Research Assistant, UW Information Theory Lab *(with Prof. Sreeram Kannan)*   *Mar. – Jun. 2017; Jan. – Jun. 2018 (part-time)*
  ➢ Developed scalable algorithms inspired by Latent Dirichlet Allocation and matrix factorization to discover cell types and find structure in large single-cell RNA-seq datasets (over 1 million cells) ([Poster](#), [Paper](#), [Code](#))
      o **Publication**: Sumit Mukherjee, Yue Zhang, **Joshua Fan**, Georg Seelig, and Sreeram Kannan. "Scalable preprocessing for sparse scRNA-seq data exploiting prior knowledge." *Bioinformatics*, 34, 2018, i124–i132.
  ➢ Used Siamese Seq2Seq neural networks and deep reinforcement learning to learn an "edit embedding" ([Poster](#), [Report](#))

Research Assistant at **UW Computing for Development Lab** *(with Prof. Richard Anderson)*   **Mar. 2015 – Jun. 2016 (part-time)**
  ➢ Helped redesign a survey app which helps public health workers collect data according to medical protocol

INDUSTRY EXPERIENCE

NLP Research Intern, Docugami   **Jun. – Sep. 2018, Jun. – Aug. 2019 (full-time);** *Mar. – Jun. 2018, Jan. – Jun. 2019 (part-time)*
  ➢ Researched, implemented, and designed state-of-the-art NLP algorithms to address open problems in document analysis

Software Engineer Intern, Facebook *(Integrity Computer Vision Team)*   **Sep. – Dec. 2018**
  ➢ Trained a clip-based convolutional neural network model to detect violent and dangerous content in videos

Software Engineer Intern, Facebook *(Search, Whole Page Ranking Team)*   **Sep. - Dec. 2017**
  ➢ Trained a sequence neural network to improve ranking of search result modules, given user's recent query history

Software Engineer Intern, Facebook *(Search Indexing Team)*   **Jun. – Sep. 2016**
  ➢ Built an indexing console tool which enables engineers to debug and test changes to the search indexing pipeline

Software Design Engineer Intern, BitTitan   **Jun. – Sep. 2015**
  ➢ Implemented infrastructure in C# to allow mailbox migrations to be simulated and tested in memory

TEACHING EXPERIENCE

Teaching Assistant, **University of Washington** and **Cornell University**   **Sep. 2015 – present**
  ➢ Went above and beyond to create additional resources, mentor other TAs, and host extra review sessions to clarify concepts
  ➢ Won UW's **Bob Bandes Memorial Excellence in Teaching Award** in 2019 (one of 3 winners)

ADDITIONAL PROJECTS (see http://joshuafan.github.io/Projects for more details)

Storage and Retrieval of Robotic Laser Range Data in Database Systems ([Poster](#), [Report](#))   *(Course project: Graduate Databases)*
  ➢ Implemented a database for laser-range scans to allow for efficient content-based querying and retrieval of images

Contextual Bandits Notes ([Notes](#))   *(Course project: Online and Adaptive Machine Learning)*
  ➢ Surveyed recent research on contextual bandits and created a report synthesizing important results/algorithms

LANGUAGES AND TECHNOLOGIES

  ➢ **Significant experience:** Python, Java, C#, SQL, C++, PHP/Hack
  ➢ **Some familiarity:** R, Matlab, Julia, HTML/CSS, JavaScript, JQuery
  ➢ **Libraries/tools:** Tensorflow, Pytorch, Eclipse, Git, Visual Studio, Linux, Nuclide