

KELVIN M. LIU

Pittsburgh, PA ♦ (714) 684-6694 ♦ kmliu@cmu.edu

EDUCATION

Carnegie Mellon University (CMU), Pittsburgh, PA **2018**

M.S. Computational Biology

University of California, San Diego (UCSD), CA **2013**

B.S. Computer Science: Bioinformatics

EXPERIENCE

Data Scientist, PNC Financial Services **2019–Now**

- Product owner of Python pipeline for ThoughtSpot (TS), which visualizes loan data, projections, and reserves and generates insights for 600 business users
- Architected and developed four Python analytics modules to perform ETL for hundred-million-record 30-year datasets to TS database and automated controls against TS API query
- As secondary administrator of TS database, prototyped and implemented schemas, complexity analysis, regular expressions, and troubleshoot for 600 business users
- Scripted standardized reporting pipelines for Matplotlib automation of 100 CECL & CCAR finance report tables & plots visualization
- Devised a YAML notation for data filtering & manipulation and implemented automated translation into Python code
- Mentored eight junior software developers and wrote seven Confluence articles on topics including: Git, SQL, Python, Bash, Linux, regex, coding standards & paradigms, optimization, and framework architecture
- Led over a dozen tutorials & workshops (on Git, debugging, unit testing & UAT testing, ThoughtSpot ETL and frontend, and frameworks) for developers
- Conducted end-to-end agile methodology: requirements, sprint planning, design, writing & assigning Jira stories, development, unit / performance / UAT testing, and code deployment

Independent Scientist **2017-2019**

- Developed an a priori, declarative, ideographic spoken & written language which attempts to ground morphology and syntax in first principles using tools provided by propositional logic, set theory, type theory, number theory, object-oriented paradigm, metaphysics, linguistics, and classical field theory
- K. M. Liu-Huang. “Complexlang: a COMPact Logical EXperimental LANGuage.” *A Record of the Proceedings of SIGBOVIK 2019*. Paper presented at SIGBOVIK 2019, Carnegie Mellon University, Pittsburgh, 1 April 2019 (pp. 144-149). <kelvinliu.org/complexlang.pdf>.

Graduate Researcher, Murphy Lab, Dept. of Computational Biology, CMU **2014–2018**

- Formulated image-based generative model of endoplasmic reticulum (ER) morphology as a million-node, quadrillion-edge embedded random graph
- Trained that model through inverse modeling, nonpolynomial curve fitting, multivariate conditional kernel density estimation (KDE), et plus in Matlab & Python from 100 two-channel three-dimensional (3D) endoplasmic reticulum (ER) and microtubule (MT) images
- Constructed panel of 20 discriminative machine learning (ML) and deep learning classifiers

(neural networks (NN), support vector classifiers (SVC/SVM), boosters, random forests (RF), ridge regressor, linear discriminant (LDA)) in Matlab

- Used panel of classifiers to predict gastric cancer drug effectiveness with 80% accuracy from three-dimensional (3D) MT images in single-cell biopsies
- Quantified the relevance, redundance, and dependence (i.e. conditional mutual information) of 100 features through LASSO, MRMR, and cross-validation metrics
- Formulated R-tree-based data structure for connected & moving object insertion, query, and update with an order improvement in runtime over generic R-tree (10-fold faster in practice)
- Constructed agent-based physics dynamics model of ER comprising millions of objects of different types interacting through dynamics rules, using above data structure in Matlab

Research Intern, Stoica Group, NASA Jet Propulsion Laboratory **2017**

- Constructed HOM, a Mealy-machine-based machine-readable pictorial writing system for expressing humanoid movement, purposed for compact storage, query, and intercommunication of human-robot movement
- Devised & conducted experiment to translate 5000-frame video of human movement into HOM through pose estimation, causal inference, and sparsity compression in Python

Teaching Assistant (Programming for Scientists (in Golang)), CMU **2017**

Teaching Assistant (Introduction to Computational Biology), CMU **2015**

Private Tutor (ML, Statistical Methods, Swimming x8 students, Java, SPARC Architecture) **Various**

ADVANCED COURSEWORK

Active Learning, Online Learning, Machine Learning (incl. semi-supervised learning, reinforcement learning, graphical models), Intro to A.I. (incl. natural language processing (NLP)), Computer Vision (& Image Processing), Computer Architecture, Advanced Data Structures, Design & Analysis of Algorithms, Discrete Mathematics, Advanced Graph Theory, Intermediate Statistics, Probability & Statistics in Bioinformatics, Biological Databases, Computational Genetics, Computational Structural Biology (incl. simulated annealing), Cell & Systems Modeling (incl. Monte Carlo methods, evolutionary algorithms), Morphology, Phonetics

SKILLS

<i>Programming/Scripting</i>	Fluent in Python, Matlab, Go (Golang), Bash, HTML5; familiar with C, C++, Java, JavaScript, PHP, SQL, CSS 3, XML, SPARC Assembly, Django, Scikit-Learn, Keras, TensorFlow, Pandas, Matplotlib, Numpy
<i>Software</i>	Git, Jupyter, Matlab, MySQL, PostgreSQL, ThoughtSpot, PyCharm, Atom, Visual Studio, Eclipse, SSH, Confluence, SharePoint, Jira, VersionOne, MS Office, PowerPoint, Word, Excel VBA, LaTeX, Windows, Linux, Unix
<i>Design</i>	Websites, mobile applications, constructed languages, video & board games
<i>Teaching</i>	ML, Git, statistics, biology, Python, Golang, Java, C, HTML, swimming
<i>Spoken Langs.</i>	English (native), Mandarin Chinese (functional), Spanish (functional)