Keith Wiley, PhD

Shoreline (Seattle), WA   505-615-4572   kwiley@keithwiley.com
https://keithwiley.com   https://www.kaggle.com/keithwiley   https://github.com/kebwi

SKILLS

Programming:	Python the last 10 years, Java 6 years, C/C++ 14 years
Programming envs:	Jupyter Notebook, many IDEs
MapReduce:	Hadoop/HDFS/Hive
Cloud:	Azure, Hadoop-on-Azure, AWS/EMR/EC2/S3
Analytics:	PyTorch, scikit-learn, numpy, machine learning (decision trees, random forests), matplotlib, Pandas, OpenCV
Science:	Image and Acoustic signal processing (FFT, wavelet, coaddition, spectrogram, octave-band)
Mobile development:	Android: 300,000 installs, 8500 sales (see below)
GUIs/plots/graphics/viz:	Mac, Android, Jupyter widgets, Tkinter, PySide (aka QT), Plotly, Pillow

EXPERIENCE

S3 Global & Seattle Sport Sciences Inc.   *Dec 2017 – Oct 2023*
Data Scientist
Data analytics for sports performance and strategic planning. Involvement in all stages of pipeline development (data capture, analysis, visualization, reporting). Most of the following tasks were critical and central to project success and business advancement. Many were developed in direct tandem with the CEO and other company leaders.
- Project owner: multi-camera deployment/configuration in professional soccer stadiums.
- Critical: Created custom tools for multi-camera GUI control and recording (PySide, Tkinter, Jupyter widgets, PTP-IP APIs).
- Created game-recording web-scheduler and driver (using *at*, similar to *cron*) to enable cross-timezone hands-free recordings in Europe driven from Redmond office. Other team members relied heavily on this tool.
- Critical: Created user-free (i.e., automated) on-the-fly video quality verification (OpenCV) of ~25 simultaneous 4K/30fps video streams, with immediate problem alerting, visualization, web-reporting.
- Created user-free, on-the-fly video transfer system to move videos from remote sites to in-office servers during recording for real-time (and post-recording) processing/analysis/reporting.
- Image/video-processing toward high-level analytics: Ball movement (optical flow analysis), player pose/motion/ visual-attention/skill, and team coordination (OpenCV, Pandas, lots of raw geometry).
- Support of computer-vision/machine-learning system (in-house GPU servers). Automated data shepherding between Pandas and SQL to build hands-free, heterogeneous Python data-processing pipelines and reporting.
- Project owner: Created business critical web-based reporting/publication system of player skill/tactics/strategy.
  - Analysis and visualization of player metrics.
  - Production and publication via web UI as Python-generated PDFs by a hands-free/turn-key system (my design) for consumption by high ranking soccer club owners/managers.

Integris Software Inc.   *Oct 2016 – Dec 2017*
Senior Software Engineer
Distributed data pipeline development for a nascent startup's GDPR compliance product. Processing in Python. Tools included PyCharm, Jupyter, and Apache Storm. Beyond technical implementation, work involved overall product design of data manipulations and analytics toward insight and business value.

Atigeo   *Aug 2013 – Sep 2016*
Senior Software Engineer & Data Scientist/Engineer
Hadoop/Hive/Hue/Spark Python for machine learning and predictive analytics on AWS. Implemented a Python distributed random forest in Hive/Python streaming. ETL for feature extraction and classification.
- Created a scikit-learn random forest and ensemble ML pipeline for cross-fold-validated predictive analytics and insight by feature importance exposure. ROC curves and AUC analysis/visualization of ML results.
- Created a pipeline to ingest/catalog/store/analyze/visualize datasets. This project implemented an SOW whose completion was the keystone of a seven-figure contract.

Expedia via Slalom Consulting placement   *Dec 2012 – Jul 2013*
Big Data Engineer, Consultant
Member of a team building a new user feature on the web app.
- Brief MongoDB project when I first started.
- Hadoop/Hive on AWS, using EMR and nonEMR-Hadoop in EC2. Tasks: EC2-to-S3 data synch., Hive stand-up, AWS profiling.
- Accomplishments: Hadoop 2.0/YARN EC2 deployment. Amazon's own engineers were curious about my progress. https://keithwiley.com/writing/HowToDeployHadoopYarnOnEC2.shtml

Slalom Consulting Feb 2012 – Jul 2013
Big Data Engineer, Consultant
National Mobility team (mgr. Jeff Rubingh), National BI team (mgr. Kevin Gregory), developing big data processing techniques.
- Focus: Hadoop MapReduce, Hive, Cloudera, Tier 3, Hadoop-on-Azure.
- Topics: CRM, NY MTA, Linked-In/Twitter APIs, some OpenLayers visualization.

University of Washington, Dept. of Astronomy Feb 2010 – Jan 2012
Research Scientist IV
LSST group (mgr. Andrew Connolly). Massively parallel image processing (coaddition) in Hadoop. Dataset: SDSSDB (30TB, 4-mil images), future applications to LSST (60PBs). Cluster (NSF CluE): 892 machines, 700TB storage, 3568 concurrent processes. This work culminated in an as-first-author PASP (journal) paper, and a first-author ADASS (conference) presentation & paper.

University of Washington, Applied Physics Lab May 2007 – Feb 2010
Software Engineer IV
Two projects:
- *Sonar Simulation Toolkit* (eigenray model of underwater acoustics, mgr. Robert Goddard). Incorporated external libraries, OO design, feature development, optimization/performance-redesign, refactorization, unit-testing.
- Real-time data-acquisition/FFT-proc. with low data-loss tolerances, rapid throughput, and amenability to parallelism.

The Institute for Genomic Research Sep 1997 – Aug 1999
Software Developer
C++ bioinformatics software development for DNA shotgun sequencing and closure analysis.

EDUCATION

PhD Computer Science	University of New Mexico, Albuquerque	Jul 2006
M.S. Computer Science	University of New Mexico, Albuquerque	Dec 2003
B.A. Psychology	University of Maryland, College Park	Dec 1997

PERSONAL PROJECTS
Sample only. Please see my website for a comprehensive listing and github for a few public disseminations.

Image/Acoustic Signal Processing
Keith's Image Stacker: Multi-threaded (parallel) image stacking, Laplacian sharpening, wavelet denoising. Used by amateur astrophotographers, reviewed online and in Astronomy and Sky & Telescope.
WildSpectra (collaboration: Dr. R. Haven Wiley, Biology dept, UNC-CH): Mac real-time spectrogram analyzer, used in Dr. Wiley's research lab and by researchers throughout the acoustic-biology community.

Data Analytics and Neuromorphic Computing
Neuromorphic CM1K Emulator: A Python emulator of General Vision's CM1K neuromorphic chip, including slides presenting modeling experiments. See personal website or github for more info.
Ships in Satellite Imagery by neuromorphic chip: Neuromorphic computer vision and image classification
 https://www.kaggle.com/code/keithwiley/ships-in-satellite-imagery-by-neuromorphic-chip/notebook
Leeds Butterfly Dataset by neuromorphic chip Neuromorphic computer vision and image classification
 https://www.kaggle.com/code/keithwiley/leeds-butterfly-dataset-by-neuromorphic-chip/notebook

Predictive Modeling and Dynamic Websites
Movie Hurl (http://moviehurl.com): A Perl-driven website of "shaky-cam" movie ratings, offering weighted averages and personalized predictions from correlated user-pair ratings.
Houseplant soil moisture monitoring system: MicroPython on ESP microcontrollers with web server, status, notifications, & analysis
 https://keithwiley.com/houseplantSoilMoistureMonitor.shtml

Android
WildSpectra Mobile: Real-time scrolling spectrograms (FFT and octave-band) on Android devices. Also: real-time waveform/FFT/ octave spectrum, and post-recording editing/playback and file I/O.
Shead Spreet: Spread sheet for Android devices with 300,000 installs, 8500 sales, and a 4.3/5 rating.

HCI
Druid (PhD thesis): Vector drawing program which permits interwoven surfaces (Celtic knots, Olympic rings, etc.) and which provides an isomorphic efficient user interface.