

July 2023

## SUMMARY

### **Software Engineer, Trainer, Data Scientist, Technologist, Speaker, and Writer**

I am a data scientist, software developer, trainer, and a writer & researcher about software topics. I have been very active in Open Source communities, especially around Python; I am a member of the Python Software Foundation, was a Director of the PSF for 6 years, and co-chair its Trademarks Committee for 15 years, and of its Scientific Python Working Group. In a related capacity, I frequently speak at conferences, often as keynotes, including OSCON, PyCon US, and AnacondaCon, and have given technical keynote talks at PyCon India, PyCon UK, PyCon South Africa, PyCon Belarus, Conferencia Internacional de Software Libre (Havana), PyData SF, PiterPy (Leningrad), and PyCon MEA (Middle East & Africa), PyCon Estonia.

I bring to both development-oriented projects and in training a broad background in research and coding of projects around scientific and numeric computing, machine learning, natural language processing, data analysis, cryptography/security, information visualization, and databases. In general I am familiar with many programming technologies and tools, a quick learner, and comfortable choosing and using the most appropriate technology for a given project.

As a writer and trainer, I bring experience as a hands-on developer. My ongoing *Charming Python* column was, for 8 years, the most widely read discussion of the Python programming language worldwide. Articles and columns I write on other topics are also well known and well received.

My 2003 Addison-Wesley book *Text Processing in Python* was a standard reference used in both industry and academia. My 2021 *Cleaning Data for Effective Data Science: Doing the Other 80% of the Work* filled an important gap in existing data science literature. My 2022 *Regular Expression Puzzles and AI Coding Assistants* was the first technical book about Copilot and ChatGPT. Over the last few years, I have presented webinars and made video recordings about scikit-learn, PyTorch, broader data science, and general programming topics.

As well as being an experienced coder, I can explain technologies and techniques to team members and collaborators. I bring comparative and analytic perspective to development or documentation goals, and often assume a project lead role and facilitate work flow. I am well familiar with the tools and techniques around version control systems, project management, collaborative tools, unit and integrity testing tools, performance monitoring, and can evaluate a range of software and infrastructure options.

## EXPERIENCE

### **Data Scientist and Trainer, KDM Training • 2020-Present**

Consultant, developer and trainer. Developed cloud-native software. Created, recorded, and/or delivered a wide variety of material covering general Python programming, data science, machine learning, deep neural networks, cloud native architecture, and other topics, for a variety of clients.

Consultant, developer, and technical writer for (partial):

- Service Employers International Union, Data Systems and Analytics Department.

- Expert witness in a case involving a prominent technology company.

- Trainer and content developer for INE.

- Zimagi open source distributed declarative data processing platform.

- Union.ai/Flyte open source orchestrator for data and ML workflows.

- Consultant for Anaconda, Inc. on legal and technical issues in generative AI.

### **Principal Software Engineer, Atlantis Technology • 2021-2022**

Consultant, database expert, and web architect for Atlantis clients. Roles have included the following. Orchestration of micro-services within large financial firm, and application of data analysis and machine learning techniques. Provision of web APIs and database schemata for a medical devices company, to enable pharmacology research; this role also brought to the client far better testing, code-review, and documentation practices. Languages for development at these placements included Python, JavaScript, Golang, SQL, bash, YAML, to script automation and CI/CD pipelines, and smaller amounts of a half dozen other programming languages.

### **Chief Technology Officer, Bold Metrics Inc., San Francisco SF • 2019**

Supervised programmers and consultants at a small, distributed startup that provides industry-specific machine learning models for large clothing designers and retailers. Developed improved ML models, data cleaning pipelines, deployment automation, data analysis, and custom metrics specific to client needs.

The system infrastructure was microservice and cloud based, using Kubernetes for container resource management; I worked with and helped improve documentation, code, and procedures for deployment and system testing, and am familiar with cloud-based deployments and tools. Code I developed and maintained was primarily in Python, with bash used for automation, and JavaScript used from front-end code.

### **Senior Trainer and Senior Software Engineer, Anaconda Inc • 2015-2018**

Created and oversaw the training curriculum used by Continuum Analytics/Anaconda Inc.

and set a rubric for content development and evaluation of training sessions. I trained extensively in Scientific Python and Data Science for a broad array of clients in scientific research, security agencies, financial analysts, and other groups needing high performance numeric computing. I created a special focus in these materials on machine learning tools, such as TensorFlow, Keras, and scikit-learn. Assisted with in-house and FLOSS development of tools and systems published and created by Continuum/Anaconda, including on the DARPA XDATA project, the Navigator GUI incorporated into recent Anaconda versions, and IOPro, a tool that assists with working with large data sets.

Enhanced and developed customer-facing documentation and in-house project planning documents. Contributing editor to a book project, the 2<sup>nd</sup> edition of *Guide to NumPy*. Developed presentation material for conference talks and keynotes about Continuum/Anaconda products and processes.

Most of my work was in Python, and especially in Jupyter Notebooks for pedagogical and data-exploration materials, but also worked in Cython, Numba, C, C++, Fortran, bash, and other tools used in the Continuum tool chain.

### **Software Engineer and Technical Writer, D. E. Shaw Research • 2008-2015**

Documented programming interfaces to the Anton supercomputer. Worked with developers to consolidate, expand, clarify and organize documentation of the world's fastest supercomputer for research simulations in computational biochemistry. Work involved understanding a range of hardware issues and programming tools: custom assemblies, M4, C, C++, Python, RTL, simulation environments. As well, worked with research chemists to understand motivations in physical sciences for DESRES hardware and software systems. Created an internal document-production tool chain using a variety of scripting and text processing tools. Assisted in writing the internal coding standards used at this research lab.

Created and maintained a software tool, written in Python, to validate chemical simulation configurations. Worked with other users developers to create the queuing and job logging system used internally. Maintained and enhanced the external website which requires export validation of downloaders. Created tools to auto-generate embedded source code for the Anton computer. Created an in-house accounting front-end interface using Django which communicated with back-end data sources. Created a parsing tool to transform a custom specification language into C++, Verilog, and Python source code, and to generate documentation.

### **Senior Programmer/Consultant, Madison Tyler LLC, Beverly Hills CA • 2006-2007**

Worked at high-volume financial trading firm specializing in “black box” and “gray box” automated and computer-assisted trading. Primary development language was Python, with additional use of shell scripting, SQL and RDBMS optimization, Java, C/C++, and numerous other specialized tools. Responsible for analysis of latency and runtime efficiency issues. Developed in-house code related to computational linguistics, statistical modeling, pattern extraction, field-specific ontology, and real-time data processing. Created AJAX interfaces for

user interaction with server-generated results.

### **Writer/Columnist, IBM; Intel; O'Reilly, etc • 2000-2011**

Wrote articles well known in the Python and XML communities, and in other technical areas (cryptography, functional programming, Linux administration, networking, etc). Columns include *Charming Python* and *XML Matters*. Author of Addison-Wesley book *Text Processing in Python*. Publication history includes several hundred programming/technical articles.

### **Consultant, Gnosis Software & KDM Training • 1998-2018**

Freelance consultant on various projects, typically ones where an expertise in algorithms, data design, or general technologies is needed at a more sophisticated level than in-house staff can provide. Sample clients/projects have included:

- *Service Employees International Union*, Washington DC. (2009-2014). Generation of data, largely via web-scraping, for use in organizing campaigns.
- *Ripe Media*, Beverly Hills CA (2007). Developed JavaScript library for enhanced user interaction on a portion of Grammy Award's web site. Debugged and tested against various platforms/browsers.
- *Snowtide Informatics*, Hadley MA (2006). Improved algorithmic structure of flagship product, PDFTextStream, leading to both greater accuracy and speed of PDF parsing and analysis. Programming performed in Python and Java.
- *Polimetrix*, Palo Alto CA (2005). Text processing and natural language processing problems for normalization of names and postal addresses. Assured strict conformance of processed results with USPS Pub28, even on approximate, incorrect, or poorly formatted input data.

### **Chief Technology Officer, Open Voting Consortium • 2004-2008**

Board member and technical advisor to non-profit organization devoted to providing fair and transparent voting technologies (open source code and voter-verifiable paper ballots). Duties include code inspection, security analysis, development of licensing language, coding standards, press relations, promotion of frameworks to academic, press, and professional colleagues, and interaction with government bodies. Also served on the IEEE P1622 committee, *Voting Systems Electronic Data Interchange*.

### **Senior Programmer, Human Technology Partnership, Greenfield MA • 1995-1998**

Maintained and upgraded existing software products used by major financial services companies, while simultaneously developing new product lines. Worked with legacy code in several xBase dialects (Clipper, Foxpro) and Pascal to provide client customization; performed much revamping of user interface and created better modularization, transportability, and especially internal documentation of existing code to streamline future customization. Also created several new information-centric products using various

hypertext tools to develop computer-based training and reference works for commercial and custom markets.

**Adjunct Professor, University of Hartford; University of Massachusetts;  
Massachusetts College of Liberal Arts • 1990-1994**

Developed curriculum and taught courses in medical ethics, general ethics, feminism, social and political philosophy, logic, epistemology, and general introduction to philosophy. Extensive academic publication in philosophy and related fields.

**EDUCATION**

University of Massachusetts at Amherst; Ph.D., Philosophy, 1999

University of Massachusetts at Amherst; M.A., Philosophy, 1992

University of Colorado at Boulder; B.A. Philosophy/Mathematics, 1987

**SPECIALITIES**

Technologies: Python; machine learning; SQL/RDBMS; schema design; textual analysis; statistical natural language processing; XML and web technologies (JavaScript, Typescript, CSS); data structures; shell scripting. Prior familiarity with large range of programming languages and tools, including Go, Rust, Java; PHP; C; C++; Ruby; Perl; R; Julia; Scala; xBase dialects (Clipper, Foxpro, Fivewin, XBase++); Unix text tools (sed, awk, grep, tr, etc); Pascal/Delphi; Fortran; Paradox PAL; Alpha Four; Basics (VB/VBA, RealBasic, etc); REXX; TCL/TK; Bash; Lisp/Scheme; Smalltalk; Haskell; etc.

Industries: Scientific computing; finance and trading; regulatory issues in banking and insurance; education and training systems; technical certification; business workflow analysis; standards bodies and standards compliance.

**REFERENCES**

Available upon request.