Greg Lever | Ph.D. (Cantab.)

32 Ashridge Way, Morden, London SM4 4EF

+44 (0) 7814 676 385 • greglever@cantab.net • greglever.github.io
levergreg • greglever

I am a Software and Data professional with extensive Python, Java, SQL, NoSQL, and Machine Learning experience. I gained my PhD in Computational Enzymology at the Cavendish Laboratory, University of Cambridge. My PhD thesis won the Springer Thesis prize and was published by Springer (now with over 2400 combined chapter downloads) and my publications in peer-reviewed journals have received over 70 citations so far. I had four years of experience in Computational Biophysics labs, including work as an independent Postdoctoral Research Associate at MIT. I have since had three years commercial software experience and I now work in the Bioinformatics and Genome Analysis team at Genomics England. I have been invited to review manuscripts at the Journal of Biological Systems and Physica Scripta. I am passionate about continual and lifelong learning.

Current Position

Software Engineer, Genomics England **Farringdon, London**

Delivering the 100,000 Genomes Project (100KGP) and starting the NHS Genomic Medicine Service (GMS) April 2017 – present

- Lead engineer for the Interpretation Platform API (Python/Django/PostgreSQL/RabbitMQ/Celery), a scalable webservice acting as a central data hub with external clinicians relying on the genomic interpretation to provide clinical diagnosis and treatment. The service is also used significantly by internal bioinformaticians and requires excellent communication and collaboration.
- Main contributor to the Data Models (Avro/Java) used company wide and migrations required for reverse and forward compatibility (Python). Contributor to the Genomic Interpretation Portal (Javacsript/React).
- Contributing to the creation of the company's Genotype/Phenotype database and advanced query engine (Java/MongoDB/Jetty) relied on by internal researchers and external clinicians.
- Docker and continuous integration (Jenkins/CircleCI/Travis) and Salt for deployments to nginx or tomcat.
- Sprints are managed with Jira and Confluence, we have daily scrums and fortnightly retrospectives and sprint planning sessions.

Computation and Modelling Skills

Languages: *Strong in* Python (Django/Celery/Scikit-Learn/Pandas/NumPy/SciPy/SpaCy/Flask/SQLAlchemy/Cython), SQL (PostgreSQL/MySQL/BigQuery), Linux (BASH scripting along with awk, sed and grep), AWS, Git (I encourage rebasing and gladly help others with merge conflicts), NoSQL (MongoDB) and LaTeX typesetting. *Confident with* Java (Jetty/Spring Boot), R, Javascript (React), Neo4j/Cypher Query, CSS/HTML, Mathematica and the Google App Engine. *Experience of* Perl, Fortran, C++ and Matlab. *Learning* Cirq for quantum computing.
Operating Systems: Strong capabilities in UNIX/Linux and High Performance Computing clusters, Mac OS X, and Windows-based environments.
Certifications: Medicinal Chemistry (EdX), R Programming (Coursera) and Data Scientist's Toolbox (Coursera).

Experience

Data Scientist (Data Science Developer), Arkera **Westminster, London**

A fintech startup at the intersection of Data Science and Wealth Management (established March 2015) Dec 2015 – April 2017

- I was Arkera's first Data hire and was instrumental in building the Data team, adding an additional three members.
- Python (Flask/SQLAlchemy) and SQL (complex queries for PostgreSQL, migrated from MySQL) for our backend and web scraping, along with the design of our database schema and dimensional data modelling.
- I am a strong advocate of Test-Driven Development and implemented the company's unit and functional testing framework.
- I played a key role in moving the Data team to an agile development process (stand-ups, sprint planning, code reviews, retrospectives, continuous integration) initially through Trello and then to JIRA.
- Organised the code base into an ecosystem of microservices providing HTTP endpoints for our iOS and web applications.

Senior Data Scientist, Stylect **WeWork, Liverpool St**

An e-commerce startup with an app and website designed to enable women to find their perfect shoes April 2015 – Dec 2015

- I headed the Data Science efforts at Stylect, utilising my skills in Python, SQL and Machine Learning.
- I led the design, engineering and deployment of production-worthy Machine Learning architectures for Recommender Systems to improve the user experience and ultimately increase sales.
- Involved extensive Python development (web scraping, pytest) with Pandas (including NumPy, SciPy and Matplotlib) and Scikit-Learn along with maintaining and adding new endpoints serving web, iOS and Android applications.
- Required SQL to use services from the Google Cloud Platform including Big Query and the App Engine.
- Utilised Natural Language Processing (Natural Language Toolkit) to perform Feature Engineering for Machine Learning (Scikit-Learn) approaches for classifying retail products.
- Devised, implemented and interpreted A/B tests to assess the accuracy of the predictive models.

Data Science Fellow, The ASI **London**

The ASI Fellowship enables academic scientists to become data scientists and engineers *February 2015 – April 2015*

- Built a prototype recommender system for an e-commerce startup (Stylect) to improve their user experience and increase sales.
- Developed strategies for growth hacking and user retention for a community-building startup (Quiet Riots).
- Expert-led training and practical workshops in Hadoop, Spark and MapReduce as well as Statistics and Business-specific concepts.

Postdoctoral Associate, MIT and University of Cambridge **Cambridge, MA (USA)**

MIT Department of Chemical Engineering and Cavendish Laboratory *March 2014 – February 2015*

- Further developed and galvanised existing skills in Python gained from my PhD.
- Built upon my knowledge of R from my MSci, learning new concepts and tools for analysing results of simulations.
- Mentoring of graduate students, taught undergraduates to use linux and High Performance Computing for the very first time.
- Independent exploratory analysis using Python/Pandas (including NumPy, SciPy and Matplotlib) of an existing body of data.
- Collaborated closely with teams based in the UK (Cambridge, Bristol and Southampton) and the USA (Yale).

Education

PhD, University of Cambridge **Cambridge**

Computational Enzymology, Cavendish Laboratory *September 2010 – March 2014*

- Long-term independent research and analysis from self-driven projects that delivered scientific results and expanded my field.
- Applied mathematical theories to complex real-world problems and generated new methodologies to improve existing approaches.
- Extensive Python (Pandas, NumPy, SciPy) development for data analysis of computational chemistry simulations.
- Immersed in a High Performance Computing Environment utilising linux, BASH, perl, awk, sed and grep.

MSci (First Class Honours), University College London **London**

Theoretical Physics *September 2006 – July 2010*

- Through summer projects in 2009 and '10, extended self-taught Linux and Python programming knowledge through real research.
- Contributed to Bayesian Inference software using R and Java and developed extensions in C++.
- Course Highlights: Java,C++, Mathematica, Advanced Quantum Theory, Stochastic Processes, Group Project.

Extracurricular Activities

Supervisor for MIT Undergraduate Research Opportunities Program (UROP) **Cambridge, MA (USA)**

MIT Department of Chemical Engineering *July 2014 to Oct. 2014*

Provided modelling & simulation training for a candidate who had never previously run any simulations or used the linux terminal.

Co-supervised Cambridge MPhil Student in Scientific Computing **Cambridge**

Cavendish Laboratory, University of Cambridge *October 2012 to Oct. 2013*

Provided electronic structure expertise and biological expertise, to complement the MPhil candidate's additional co-supervisor.

Physics At Work - Outreach Event for the Cavendish Laboratory **Cambridge**

Cavendish Laboratory, University of Cambridge *September 2011/'12/'13*

Organised exhibitors for and participated in a three day outreach exhibition in Cambridge aimed at inspiring 14-17 year olds to become the next generation of scientists, engineers and technology specialists by showcasing the many and varied ways computational physics is used in the real world.

Procter & Gamble R&D European PhD Seminar 2013 **Brussels**

Procter & Gamble (P&G) Brussels Innovation Centre *April 2013*

Applied technical knowledge to solve real problems found in the typical day of a Scientist or Engineer in R&D at P&G.

i-Teams: Commercialising Creativity **Cambridge**

Finding real-world applications for a synthetic reversible molecular lego system *October 2011 – December 2011*

Led a team of 6 PhD students and 1 MBA student asked to investigate an invention that came out of a Cambridge laboratory.

Business Challenge hosted by Cambridge University Technology and Enterprise Club **Cambridge**

Building a Business Case for more in silico *modelling in the Pharmaceutical Industry* *March 2011*

Researcher, London Centre for Nanotechnology **London**

Successfully applied for funding from the Engineering and Physical Sciences Research Council *June 2010 – September 2010*

Researcher, University College London **London**

Successful in funding competition for all department undergraduates *June 2009 – September 2009*

References available on request