Henry Wilde

Cardiff, Wales henrydavidwilde@gmail.com
github.com/daffidwilde References available upon request

Summary

I am a thoughtful, ethically minded scientist with a track record of pragmatism and efficient, impactful work. I have a breadth of projects under my belt from large-scale health data analysis with machine learning to productionising secure enclaves for record linkage. I find great joy in picking up new tools and techniques, and in putting those skills to use at pace.

Currently, I am leveraging LLMs to realise business efficiencies in the ONS, and I champion the increased use of privacy-enhancing technologies (PETs) across the Civil Service and Government.

Having successfully led numerous high-impact projects in academia and government, I am now looking to apply my expertise as a data scientist and software engineer in a new venture.

Employment

May 2022 – present

Data scientist *Data Science Campus, Office for National Statistics*

- Developing a LLM-based reader to summarise ONS activity in parliamentary debates, leading to significant cash savings for the Office
- Core developer of a privacy-preserving record linkage toolkit, including an accompanying secure computation architecture on GCP
- Mentored a team of apprentices in creating a Python interface to the England and Wales 2021 Census API
- Technical lead and project owner in creating high-fidelity synthetic census microdata using distributed computing and differential privacy

Python (data science stack, BeautifulSoup) | **Version control** (Git, GitLab, GitHub) | **Google Cloud Platform** | **Docker** | **Automated testing** (pytest, hypothesis, GitHub Actions) | **Publishing** (Quarto, Streamlit, GitHub Pages, Markdown, LaTeX) | **LLMs** (Gemini, OpenAI, LangChain) | **Distributed computation** (Dask, PySpark, Google BigQuery)

Feb 2021 – May 2022

Research associate *Water Research Institute, Cardiff University*

- Designed and implemented the software infrastructure for the Welsh Government wastewater surveillance programme
- Taught myself the principles of R for data science in the first month to establish reproducible ETL pipelines for biochemical data
- Developed two core models for monitoring COVID-19 prevalence across Wales: a hierarchical GAM for predicting case rates and a Bayesian model to account for dilution in the wastewater system
- My analysis and reporting had a direct impact on Welsh Government policy at the height of the pandemic

R (tidyverse, mgcv, Shiny, RStan, RMarkdown) | **Version control** (Git, GitHub) | **LIMS**

2019–2020

Volunteer consultant *School of Biosciences, Cardiff University*

- Commissioned by the largest school in the University to improve their dissertation allocation process
- Implemented a hands-off, programmatic framework using a Python research library I developed during my PhD
- Reduced the workload from a week across the team to a matter of seconds on one computer, and guaranteed mathematical fairness

Python | **Version control** (Git, GitHub) | **Jupyter** | **Microsoft Excel**

Dissertation supervisor *School of Mathematics, Cardiff University*

- Co-supervisor for a MMORS final-year project on Folk Theorems in game theory
- Mentored the student in how to produce a sustainable piece of research software to accompany their dissertation
- Assisted in editing the final report prior to submission

Python | **Version control** (Git, GitHub) | **SQL** | **LaTeX**

2017–2021

PhD studentship teaching *School of Mathematics, Cardiff University*

- Heavily involved in teaching modules and services, including courses on statistical inference and Python for mathematics, the university maths support service, and hackathons for Masters students
- Founded an Advanced Python Workshop for my fellow PhD students covering topics like distributed computing, automated testing, and version control
- Mentored a high school student during a Nuffield Research Placement

Python (data science stack, SymPy, Dask) | **Version control** (Git, GitHub) | **Testing** (pytest, hypothesis, Travis CI) | **Writing** (LaTeX, Markdown, reStructuredText, Sphinx)

Education

2017–2021 **PhD Applied Statistics, Operational Research and Data Analytics** *School of Mathematics, Cardiff University*

- My thesis focuses on the thorough and ethical utilisation of machine learning in healthcare settings
- Key results include new perspectives on algorithm evaluation through data synthesis, and fair clustering
- My research provided actionable insights for my co-funders into a critical healthcare population in their care using only administrative data
- Accompanied by a suite of sustainably developed research software packages

2014–2017 **BSc Mathematics (First Class Honours)** *School of Mathematics, Cardiff University*

- Maintained a breadth of interests, including operational research, computing, and pure mathematics
- Received perfect scores for two projects: a simulation and analysis of a hospital emergency department, and an empirical comparison of two strategies in an iterated Prisoner's Dilemma

Awards

2022–2024 **Reward and Recognition** *Office for National Statistics*

- Received a total of eight awards across all three bands, rewarding me for going above and beyond in my work
- Two awards for giving particularly accessible and engaging technical talks to colleagues in the Office
- Three awards for my involvement in high-priority surge work between governmental departments and with our international partners
- A sustained excellence award for my work on synthetic data and its impact on the ONS Data Strategy
- Two awards for fostering a culture in my teams that values software sustainability and effective project management practices

2022 **PETs Hackathon** *United Nations PET Lab*

- Finished third out of two hundred international teams
- The hackathon was centred around a real-world application of privacy-enhanced data analysis
- Accurately predicted three hidden characteristics of Kenyan refugee households using open-source tools for differential privacy inside a secure enclave

2018 **Support for NATCOR Bursary** *Association of European Operational Research Societies*

- Received financial support to attend postgraduate courses in operational research
- Courses covered approximation algorithms and heuristics, and predictive analysis and forecasting

Publications

A list is also available online.

Thesis

2021 Wilde, H. *New methods for algorithm evaluation and cluster initialisation with applications to healthcare*. Cardiff University. PDF. GitHub repository.

Journals

2022 Wilde, H., et al. *Accounting for dilution of SARS-CoV-2 in wastewater samples using physico-chemical markers*. Water, 14(18):2885. DOI:10.3390/w14182885

2020 Wilde, H., Knight, V. and Gillard, J. *Evolutionary dataset optimisation: learning algorithm quality through evolution*. Applied Intelligence, 50:1172-1191. DOI:10.1007/s10489-019-01592-4

Wilde, H., Knight, V. and Gillard, J. *Matching: a Python library for solving matching games*. Journal of Open Source Software, 5(48):2169. DOI:10.21105/joss.02169

Pre-prints

2024 Jones, O., et al. *Estimating wastewater dilution using chemical markers and incomplete flow measurements: application to normalisation of SARS-CoV-2 measurements.* DOI:10.20944/preprints202402.1109.v1

2022 Houssiau, F., et al. *A framework for auditable synthetic data generation*. arXiv:2211.11540

Interests

Cooking

I taught myself to cook as a child, and then worked as a chef while at sixth form, including at a former Michelin star restaurant. Cooking for friends and family is now one of my dearest pastimes.

Cycling

During the height of the COVID-19 pandemic, I desperately needed something to occupy myself outside of writing my thesis. So, I taught myself bike mechanics and renovated a vintage steel-frame touring bike.

D & D

I adore fantasy in all its forms. Now, after years of listening to Dungeons & Dragons podcasts, I serve as the game master in a homebrew campaign for my three brothers.