

# gavin leech

data scientist, software engineer

## about

London based.

[Email](#)

[Blog](#)

[Github](#)

[LinkedIn](#)

[Reviews](#)

## research

[Empirical AI safety](#)

[Speaker similarity](#)

[Automation & welfare](#)

[Aid impact bounds](#)

[On Sturgeon's law](#)

[On data science](#)

[Maximum wage](#)

[Controversy vs impact](#)

[Formal analogies](#)

## programming (in production)

Python/SciPy/Spark,  
Keras, \*nix, Azure, C#,  
JS (ES6, node, TS),  
PHP, Pig, LaTeX  
...Powershell.

## metadata

[vs publication bias](#)

[Career JSON](#)

[Microblog](#)

## experience

Sep'16 – Now **Data Scientist** AXA Insurance, London  
Creating ML models (GLMs, GBMs, deep ANNs) and infrastructure for pricing, medical risk, & image recognition. Also web scraping at scale, also managing a TB/day Azure + Hadoop pipeline. Tech lead for 4 'scientists.

*Buzzwords:* Python, Spark, Keras, xgboost, Azure, Oozie, Avro, Pytest.

May – Aug'16 **Frontend Developer** Furo, London  
Developed an Angular SPA for a procurement startup. Sole responsibility for the design & implementation of tendering. TS controllers from C# APIs.

*Buzzwords:* C# 6, Typescript, Angular, Visual Studio, MS SQL Server.

Oct'15 – Jun'16 **Software Developer** Freetobook, Glasgow  
Improved a large legacy accommodation engine: PHP back-end, JS front-end, REST APIs. Sole design of a PCI compliant credit-card storage system.

*Buzzwords:* PHP (5&7), LAMP, SQL, Laravel, cURL, Gitlab, New Relic.

Jun – Sep'15 **Graduate Software Developer** Stugo, Glasgow  
Jumped into web tech from scratch; writing production code within 2 weeks, deployed to pharmacies worldwide. Worked on a medicine pump manager (C#) & a secure pharmacy record system (nodeJS). 4th employee.

*Buzzwords:* C#, WPF, node.JS, MEAN, Ractive, Mongo, Vagrant, VMWare.

Dec'14 – May'15 **Database Developer** STAND International, Edinburgh  
Built a small ops database for an international NGO. Greenfield design & deployment; delivered at zero infrastructure and licence tcost.

*Buzzwords:* MS Access, Office365, SQL, UML, data protection.

Jan – Apr'14 **Assistant Statistician** The Scottish Government, Edinburgh  
Derived & explained official stats for Education Analytical Services. Worked on the 'Insight' reporting platform, giving teachers hard data & viz on their classes & themselves. Prepped for the national Social Care Survey 2015.

*Buzzwords:* SAS, SQL, OLAP, PCA, data protection, official statistics.

(...)

## software

### **Side effects in Gridworlds.**

Original work on the side-effects problem in reinforcement learning, with Tom McGrath, Jessica Cooper and Karol Kubicki. Implemented DQN and MaxEnt IRL in Tensorflow and evaluated their safety properties. Written up [here](#), code [here](#)

### **sesame: SEArch for StAtistical MEthods.**

Decision tree / search engine for which stats you can/should use. Scala.js

### **The First Computers**

Deployed [here](#). Tiny app (and conceptual analysis) to solve a historical question. Uses the [Transcrypt](#) Python transpiler (unfortunately).

### **Py2HTK**

On github [here](#). Wrapper for the venerable (creaky) ML library, the Hidden Markov model ToolKit. Used in [this research](#).

## education

### **BSc Mathematics & Statistics**, the Open University.

For rigour. Probability theory up to diffusion processes, regression up to GLMs, numerics up to BFGS and simplex, Bayes up to HMC, linear algebra up to multilinear, Analysis up to Laplacians. Also medical statistics, experiments, RCTs. Part-time during full-time work. 4.33 GPA expected.

### **MSc Software Development**, University of Glasgow.

For power. Java, SQL, HTML/CSS/PHP, MC6800 assembly plus Patterns, data structures & algorithmics. Psychology of security, law of safety-critical systems, design of UX. Thesis in computational linguistics: '[Hidden Markov Models for Accommodation](#)'. 4.33 GPA

### **MA (Hons) Economics with Philosophy**, University of Aberdeen.

For direction. Formal logic, micro- & macro- economics, social statistics, health policy, development policy. Alumni Fund scholarship on merit. Dissertation: '[Evaluating aid abolitionism](#)'. [3.86 GPA](#).

**Apache Spark Developer Certificate**, Databricks. For PySpark 2.x. RDDs, DFs, ML, Streaming, advanced configuration. Cert [here](#).

## interests

effective altruism, open science, [self-calibration](#), failures of human reasoning, failures of human reasoning about human reasoning, reinforcement learning, lists.