

# 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

Python/SciPy/Spark,

Keras, \*nix, C#, Scala,

JS (ES6, node, TS),

PHP, Pig, LaTeX

...Powershell.

## metadata

[vs publication bias](#)

[Career JSON](#)

[Microblog](#)

## interests

effective altruism, reinforcement learning, failures of human reasoning, failures of human reasoning about human reasoning, statistics as applied philosophy, lists.

## projects

2018

### Side effects in Gridworlds

[writeup here](#)

Original work on the side-effects problem in reinforcement learning, with Tom McGrath and Karol Kubicki. Implemented DQN and MaxEnt IRL in Tensorflow and evaluated their safety properties. For the 1st RAISE [AI Safety Camp](#).

2017

### sesame: SEarch for StAtistical MEthods

[work in progress](#)

Decision trees / search engine for which stats you can/should use. [Scala.js](#)

2017

### The First Computers

[deployed here](#)

Tiny app (and a conceptual analysis) to solve a historical question. Uses the Transcript Python transpiler (unfortunately).

2015

### Py2HTK

[github here](#)

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

## education

2012 – Now

### BSc Mathematics & Statistics

[The Open University](#)

For rigour. First expected. 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.

2014 – 2015

### MSc Software Development

[University of Glasgow](#)

For power. Conversion Masters: Java, SQL, HTML/CSS/PHP, MC6800 assembly plus 'patterns, data structures & algorithmics. The psychology of security, law of safety-critical systems, design of UX. Thesis in computational linguistics: '[Hidden Markov Models in the study of accommodation](#)'. [4.33 GPA](#)

2018

### Apache Spark Developer Certificate

[Databricks](#)

PySpark versions 2.x. RDDs, DFs, ML, Streaming, config, arch. Cert [here](#).

(...)

## 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 Hadoop pipeline. Tech lead for 4 'scientists'.  
  
*Buzzwords:* Python, Spark, Keras, xgboost, Jupyter, 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, VS2015, SQLServer, Bitbucket.
- 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.
- 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.

(...)