

# DANIEL HOMOLA

GitHub · LinkedIn

---

Experienced senior data scientist / machine learning engineer with a proven track record of innovation and software development. Well-versed in applied deep learning and natural language processing. Delivered numerous successful complex modelling projects using proprietary big data. As co-founder and CTO, raised pre-seed investment for healthcare AI company.

## EXPERIENCE

---

### Founder

Nov 2019 -

*Wide Ridge*

*London, UK*

- Founded Wide Ridge, a boutique data science consultancy, specialising in applying the latest deep learning and natural language processing research to real world business problems.
- Leveraging BERT and other NLP technologies, created a news summarisation engine for journalists to pick out key points of lengthy articles quickly and efficiently.
- Built a modelling platform and data ingestion pipeline for automated discovery of profitable trading strategies, using deep reinforcement learning and advanced time-series modelling techniques.

### Co-founder and CTO

Apr 2019 - Oct 2019

*NeoGlia*

*London, UK*

- Created NeoGlia, a company aiming to build the world's first medical data market place, unlocking the potential of this extremely valuable but severely siloed data type, leveraging federated deep learning and a novel business model.
- Secured pre-seed investment (valued at \$1M) as part of the Entrepreneur First programme.
- Using PyTorch and PySyft, built our MVP: a federated deep learning system for healthcare data scientist. Demoed it to numerous pharmaceutical companies, using real world EMR data from the ICU units of US hospitals.

### Senior Data Scientist

Nov 2016 - Mar 2019

*IQVIA*

*London, UK*

- As a data scientist in the Predictive Analytics team of IQVIA, I leveraged the world's largest proprietary medical claims database, using advanced machine learning (Python, Tensorflow) and custom built ETL pipelines (PySpark, Hadoop) to build disease detection systems that could automate medical diagnosis making at the population level.
- Designed and implemented our team's deep learning methodology as a highly modular and extendible Python/TensorFlow package. Trained deep bidirectional LSTM and Temporal Convolutional Network models for disease detection, relying on embeddings that were learnt on the medical history of millions of patients.

- Developed custom ETL and data QC pipelines that scale to tens of millions of patients (Python/PySpark) and form the backbone of the team's data-science offerings.
- Built highly performant models using both clinical knowledge and data driven features and traditional ML methods such as clustering, SMOTE, imbalanced bagging, ensemble of XGBoosts.
- Carried out numerous research projects to enhance team's modelling pipeline, e.g. automated ICD9-ICD10 mapping through SNOMED-CT and Word2Vec based medical context matching.

### Science to Data Science

Jul 2016 - Sep 2016

*Marks and Spencer*

*London, UK*

- Using advanced NLP and ML techniques, our team of four built a prototype of a chatbot which could interact with customers and hence automate a portion of the customer service of M&S.

### Computational and Systems Medicine

Oct 2013 - Sep 2016

*Imperial College London*

*London, UK*

- Created **CorrMapper**: an online research tool for the integration, exploration and interactive visualisation of complex biomedical datasets. CorrMapper uses advanced feature selection and covariance estimation algorithms with state of the art visualisation modules to reveal the relevant correlation networks of biomedical datasets. The paper of this project is in preparation.
- Other notable projects from my PhD: [Science Flask](#), [BorutaPy](#), [MIFS](#).

## EDUCATION

---

### Imperial College London

Oct 2013 - Sep 2016

*PhD in Biomedical Machine Learning*

*London, UK*

- *Modules studied*: machine learning, multivariate statistics, advanced data visualisation & integration.
- *Thesis*: Integration and visualization of clinical-metabolic datasets for medical-decision making.

### Imperial College London

Oct 2011 - Sep 2012

*MSc in Bioinformatics and Theoretical System Biology – Distinction*

*London, UK*

- *Modules studied*: bioinformatics, statistics, programming (Python, R, MySQL), data analysis.
- *Thesis*: Tool development for 16S rRNA gene copy number normalisation and visualisation.

## TECHNICAL STRENGTHS

---

<b>Coding</b>	Python (advanced), TensorFlow/PyTorch (intermediate), JavaScript (intermediate)
<b>Database</b>	Hadoop, PySpark (intermediate), MySQL, postgresSQL, SQLite
<b>Web</b>	AWS, HTML5, CSS, Flask, Django
<b>Tools</b>	Certified Scrum Master, Git, CI/CD, Docker, Vagrant, Linux