

# Karolis Jankauskas

A Biochemical Engineer turned Data and Machine Learning Engineer.

**Languages** Python • Cython • C++ • R • Julia • CUDA  
**Libraries** Keras • TensorFlow • Xgboost • Sklearn • SciPy ecosystem • rpy2 • OpenCV  
**Backend** Flask • Django • Celery  
**Databases** PostgreSQL • MySQL • RabbitMQ • Redis • Apache  
**Tools** Git • CircleCI • Docker • AWS • Jupyter  
**Other** Deep Learning • CNN • LSTM • Sequence Models • Time Series  
**PhD research** Scheduling • Evolutionary Programming • Mathematical Programming

London, UK

+44 7858 271487  
info@kjankauskas.com

[www.kjankauskas.com](http://www.kjankauskas.com)  
[www.github.com/karolisjan](https://www.github.com/karolisjan)  
[www.linkedin.com/in/karolis-jan](https://www.linkedin.com/in/karolis-jan)

---

## EXPERIENCE

---



**Data Engineer (internship)**  
**WorldQuant**  
London, UK, 2018 September – Present

WorldQuant is a quantitative asset management firm with more than 700 employees across 26 offices in 15 countries.

- Developing an encoder-decoder LSTM-based system for early detection of anomalies in high-frequency time series data.



**Machine Learning Engineer (worked part-time whilst doing a PhD)**  
**Aiden.ai**  
London, UK, 2018 February – 2018 September

Aiden is a Natural Language Processing (NLP) powered virtual assistant who helps marketers make better decisions.

- Developed and deployed a REST API to detect anomalies in advertising time-series data in real-time.
- Built a multi-task deep neural network-based model with entity embeddings of categorical variables to answer questions such as “What will the total number of impressions, link clicks, and app installs be next week if I spend \$X on ads ABC targeting Y audience in Z location?”.
- Created a neural network and genetic algorithm-based system for recommending an optimal way to distribute weekly advertising budget to maximise ad sets performance, e.g. impressions, link clicks, and app installs.



**Data Scientist (worked part-time whilst doing a PhD)**  
**Picasso Labs**  
London, UK, 2017 Feb – 2018 February

Picasso Labs is recognised as one of Unilever Foundry’s most ambitious and innovative start-ups of the past 5 years.

- Using a pre-trained Inception v3 CNN, HSV color histograms, and entity embeddings of date, time, and image tags developed a system for recommending images to improve marketing performance on social media.
- Re-used the system above to create a content-based image retrieval engine.
- Built a custom facial expressions recognition model and applied it on over 5000 web-scraped images from US online media to investigate “visual bias”. See [www.newsweek.com/liberal-media-not-biased-trump-thinks-703291](http://www.newsweek.com/liberal-media-not-biased-trump-thinks-703291)
- Applied regression and non-parametric statistical tests to determine best performing image categories and segments.



**Teaching Assistant**  
**UCL Biochemical Engineering**  
London, UK, 2016 February – 2018 February

Supervised research and taught MSc and MEng students discrete-event simulation, mathematical programming, evolutionary programming, and multi-objective optimisation.



**Consultant (internship)**  
**Sphere Fluidics Ltd**  
Cambridge, UK, 2014 June – 2014 August

Created fluid-flow models of microfluidic chips for a novel single-cell screening and analysis system.



**Research Associate**  
**UCL Advanced Centre for Biochemical Engineering**  
**London, UK, 2013 June – 2013 August**

Performed multi-variate data analysis on mass spectrometry data to improve the expression of virus-like particles from *Pichia pastoris* cells for a universal influenza vaccine project.

---

## EDUCATION



**Udacity**  
**2017 – 2018**  
**Nanodegree, Artificial Intelligence, Certificate of Completion**

- Implemented Depth-First Search, Constraint-Propagation, Minimax, Alpha-Beta Search, and Iterative Deepening algorithms to solve Sudoku and beat human players in the game of Isolation.
- Using Planning Domain Definition Language (PDDL), A\*, and propositional logic developed a solution to find the most efficient route to route air cargo to their respective destinations.
- Using a preprocessed dataset of tracked hand and nose positions extracted from video, trained a set of Hidden Markov Models (HMM) to identify individual sign language words.
- Implemented and applied feed-forward NN, CNN, and RNN on a variety of problems such as classification of over 100 different dog breeds, facial keypoints detection, time-series prediction, and language models.



**UCL**  
**London, UK, 2014 – 2018 (thesis submitted in 2018 September, awaiting Viva)**  
**PhD, Biochemical Engineering (Operational Research, Combinatorial Optimisation)**

Thesis title: “Biopharmaceutical Capacity Planning using a Flexible Genetic Algorithm Approach”

Using Python, C++, CUDA, and Docker, developed a cross-platform genetic algorithm-based tool for continuous-time multi-objective planning and scheduling of biopharmaceutical facilities under deterministic and stochastic demand.

- Accomplishments*
- Presented a keynote lecture at the 27<sup>th</sup> European Symposium on Computer Aided Process Engineering (ESCAPE 27), Barcelona, Spain, 2017.
  - Awarded a Year 1 Research Project Prize for Best PhD Project and Poster.



**UCL**  
**London, UK, 2010 – 2014**  
**Master of Engineering (MEng), Biochemical Engineering, First-Class Honors**

- Accomplishments*
- Received Jacobs Engineering Design Project Prize.
  - Received Head of Department Commendation Award.

- Activities*
- Fitness instructor at UCLU Muay Thai Club.

---

## PUBLICATIONS & TALKS

Jankauskas K, Papageorgiou, L.G. and Farid, S.S., Fast Genetic Algorithm Approaches to Solving Discrete-Time Mixed Integer Linear Programming Problems of Capacity Planning and Scheduling of Biopharmaceutical Manufacture, Computers and Chemical Engineering (2018), DOI: <https://doi.org/10.1016/j.compchemeng.2018.09.019>

Jankauskas, K., Papageorgiou, L.G. and Farid, S.S., 2017. Continuous-Time Heuristic Model for Medium-Term Capacity Planning of a Multi-Suite, Multi-Product Biopharmaceutical Facility. In *Computer Aided Chemical Engineering* (Vol. 40, pp. 1303-1308). Elsevier, DOI: <https://doi.org/10.1016/B978-0-444-63965-3.50219-1>

Jankauskas, K., McCartney, GR., Osborne, MD., Papageorgiou, LG., Farid, SS. 2017. Multi-Objective Capacity Planning for Multi-Product Biopharmaceutical Facilities Under Uncertainty, 253rd ACS National Meeting, San Francisco, USA, April 2-6.

Jankauskas, K., Papageorgiou, LG., Farid, SS. 2016. Production Scheduling of a Multi-Product Biopharmaceutical Facility Using a Genetic Algorithm, 28th European Conference on Operational Research (EURO), Poznan, Poland, July 4-8.

---

## PROJECTS

- [www.github.com/karolisjan/BiopharmaScheduling/tree/development](https://github.com/karolisjan/BiopharmaScheduling/tree/development) a scheduling library for biopharma
- [www.github.com/karolisjan/DeepLearning](https://github.com/karolisjan/DeepLearning) a collection of deep learning projects
- [www.github.com/karolisjan/Genetic-Programming](https://github.com/karolisjan/Genetic-Programming) AI controllers created using genetic programming