Jacob Daniel Moss

jm2311@cam.ac.uk
+44 7817200229

Academic

University of Cambridge, Computer Lab 2020 - present

PhD Computer Science

Researching the impact of varying strengths of mechanistic inductive biases on dynamical models in biology. Improved a popular framework, RNA velocity, by removing unrealistic assumptions and providing a data-driven approach to integrate single cell epigenetic markers and continuous transcription rates. This was achieved by designing a transformer-based diffusion model to produce trajectories of cell fate.

Advisers: Prof. Pietro Liò & Prof. Jeremy England.

Research areas: machine learning in single cell genetics, dynamical modelling.

University of Cambridge, Computer Lab 2019 - 2020

MPhil Advanced Computer Science: **Distinction**

Developing machine learning models for the stochastic modelling of dynamics. Extended the Neural ODE framework to a distribution of initial conditions. Investigated methods of solving PDE-based latent force models.

Adviser: Prof. Pietro Liò.

Research areas: machine learning, drug discovery, genomics.

Thesis: "Gene Regulatory Network Inference with Latent Force Models"

Specialism: Probabilistic Machine Learning, Digital Signal Processing, Natural Language Processing

University College London 2016 - 2019

BSc Computer Science: **First class honours**, top 10% of cohort

Final project on large-scale data aggregation of unstructured auction house data, with a grammar-based NLP processing pipeline.

Adviser: Prof. Philip Treleaven

Thesis: "Cross-Auction Arbitrage with Machine Learning"

Specialism: Statistics, Machine Learning, Linear Algebra, Stochastic Calculus, Uncertainty Analysis

Experience

Freelance AI Contracting Mar 2023 - present

- Creation of a graph database of biosynthetic gene clusters with end-to-end analysis and sequence model training for antibiotic discovery for a biotechnology client. Integrated a novel masked objective in the protein domain.
- Implemented for a Poker platform client a Pluribus-like Poker bot in Python and C++ capable of superhuman performance on 2-player no-limit Texas hold 'em. The algorithm uses counter-factual regret minimisation, Monte Carlo tree search and Bayesian hand probabilities.
- Developed a full-stack algorithmic trading bot for initiating orders on several cryptocurrency exchanges operating in different countries. Orders were optimised for speed and traded based on AI-tagged news events. Trained, quantised, and deployed large GPT model on AWS Neuron hardware.
- Deployed machine learning pipeline for processing large planning database for a property technology startup.
- Building a data analysis framework for clinical trials data and NICE guidance.
- Led a workshop on understanding GPT models and their uses within the risk, investments, and insurance industry to senior executives at Aviva.

Researcher, GlaxoSmithKline Jun 2022 - Sep 2022

Worked alongside Ed Curry in functional genomics on the analysis of a high-throughput CRISPR assay, focussing on determining genetic interactions and drug targets.

Machine Learning Engineer, Synteny Jan 2022 - Jun 2022

Developed machine learning methods for protein embedding and disease-sequence association prediction. Implemented several models from literature, extending to larger datasets and evaluating performance on a large range of experimental datasets.

Machine Learning Engineer - Cassandre Investments Nov 2018 - May 2019

Built an asset price prediction model using autoregressive Gaussian processes for forecasting. Feature engineering including grammar-based named entity extraction and sale frequency calculations. NLP for clustering items based on description.

Data Scientist - Ocado Technology Jul 2018 - Sep 2019

Developed an ensemble classifier to determine the sentiment of Ocado-related tweets at 80% accuracy. Wrote an algorithm to incorporate a new sentiment class in pre-trained models.

Freelance R&D 2016-2019

Visijax: Developed a smart cyclist jacket with fabric-integrated LEDs and inbuilt MCU controller providing navigational prompting and geofencing.

Movebubble: Developed processing pipeline of vast property information datasets, serving two hybrid mobile apps via a C# Web API backend.

Skills

Machine learning: TensorFlow, PyTorch, numpy, pandas

Programming languages: Python, C++, Java, C#, JavaScript

Frameworks: AWS (incl. Neuron), GCP, React, RxJS, .NET, Django, Kubernetes, git, SQL

Publications

University of Cambridge, AI Research Group Dec 2019 - present

- **Moss, J. D.**, England, J. L., and Liò, P., 2023. Pseudotime Diffusion. *NeurIPS 2023 Machine Learning and the Physical Sciences workshop*.
  Analysis of whole-genome sequencing data has been outpaced by the experimental techniques that generate those datasets. Conventional methods like dimensionality reduction limit the information obtainable from a dataset. We expand on the biophysical model of RNA velocity to yield meaningful insights into the functional trajectories of individual cells, leading to downstream applications such as the identification of genes driving a disease pathway. We achieve this by learning a diffusion model to compute pseudotime and velocity distributions.
- **Moss, J. D.**, Opolka, F. L., England, J. L., and Liò, P., 2023. Deep Kernel Learning of Nonlinear Dynamical Systems. *ICML 2023 SynS & ML workshop* and in review at *TMLR*.
  Scientific processes are often modelled by sets of differential equations. We drastically improving the scalability of latent force models, balancing data-driven and mechanistic inference in dynamical systems, whilst improving inference accuracy. This overcomes the computational challenges associated with quantifying uncertainty in large datasets. We learn the solution operator by employing a deep kernel along with a function embedding.
- **Moss, J. D.**, Opolka, F. L., Dumitrascu, D., and Liò, P., 2021. Gene Regulatory Network Inference with Latent Force Models. *NeurIPS 2021 Machine Learning and the Physical Sciences workshop & AAAI 2021 Science-Guided AI symposium*.
  The existing inference techniques associated with physically inspired latent force models rely on the exact computation of posterior kernel terms which are seldom available in analytical form, rendering them intractable to most applications. We overcome these problems by proposing a variational solution to a general class of non-linear and parabolic partial differential equation latent force models. Further, we show that a neural operator approach can scale our model to thousands of instances, enabling fast, distributed computation. We achieve competitive performance on several tasks where the kernels are of varying degrees of tractability.
- Norcliffe, A., Bodnar, C., Day, B., **Moss, J. D.**, & Liò, P., 2021. Neural ODE Processes. In *ICLR 2021 main conference* and *NeurIPS 2020 Machine Learning and the Physical Sciences workshop*
  We introduce Neural ODE Processes (NDPs), a new class of stochastic processes determined by a distribution over Neural Ordinary Differential Equations (NODEs). This resolves two issues with NODEs: the inability to adapt to incoming data-points and the inability to explain a sparse set of measurements with many possible underlying dynamics. We show that our model can successfully capture the dynamics of low-dimensional systems from just a few data-points. At the same time, we demonstrate that NDPs scale up to challenging high-dimensional time-series with unknown latent dynamics such as rotating MNIST digits.
- Norcliffe, A., Day, B., **Moss, J. D.**, & Lió, P., 2021. Meta-learning using privileged information for dynamics. In *ICLR 2021 Learning to Learn and SimDL workshops*.

Achievements

Stanford Machine Learning Course

Stanford Coursera courses on Sequence Models and Deep learning. Certs: GUF2CZ62FQHD, 8US78QT64Q3K

Grants & Exhibitions

Accelerate C2D3 Grant 2023: Awarded the Accelerate-C2D3 grant for AI research and innovation.
Wearable Technology Show 2016: Exhibited the Visijax jacket.