Andrea Karlová

Webpage: andreakarlova.github.io
Github: https://github.com/AndreaKarlova
Email: andrea.karlova@gmail.com
Phone: +44 (0)7500 860642

University College London, London, United Kingdom
Department of Computer Science, 2015 - present
Research Student

- participation in drug discovery challenges focused on identifying promising hits for unexplored targets: MurD Ligase for antibiotics candidates and binding site of SARS-CoV-2 nsp13, binding site of SARS-CoV-2nsp3 with track-record of successfully identified potential hits
  used methods: 3D interaction fingerprints representations for structure similarity searching, scoring methods for molecular dynamics (GROMACS) computations, force-fields computations, molecular fragments recombination
- collaboration with the University of Chemistry and Technology, Prague, Czech Republic (Prof. Daniel Svozil) on virtual screening for the drug discovery
  used methods: VAEs applied to molecular fingerprints (ECFP and FCFP) producing chemically semantic latents, RL for molecular design generation, docking (Vina), E(n)-equivariant GNNs (exploration of EquiDock and EquiBind models)
- collaboration with the University of Surrey (Prof. Ravi Silva) and UCL AI Centre (Prof. Brooks Paige) on the high-throughput system for various perovskites photovoltaics discovery
  used methods: evaluation of the atomic structures using various quantum mechanical computations to determine the bandgap (GPAW and ASE packages, RDKit, pymatgen, psi4, pySCF, Material Project), various geometric deep learning methods (Graph Neural Networks with Message Passing, DNNs for Surfaces) , SMILEs compressing VAEs, material science pre-trained word embeddings
- robustness of latent representations
  used methods: Optimal Transport Theory, VAEs, GECO for fine-tuning beta-VAEs
- peer-reviews for Nature Scientific Reports, NeurIPS, ICML and ICLR conferences
- lectures on Big Data processing, Docker and ML pipelines for BioMed MSc students
- teaching assistant for Bayesian Deep Learning Methods
- MSc thesis co-supervised at UCL CS:
  Adriana Staudova, *Imitation Learning for De-Novo Drug Design* September 2023
  Alan Shen, *Active Learning Strategies for Binding Affinity Prediction* September 2023
  Cen, Huisan, *On the Topic of Density Functional Theory with Deep Learning*, September 2021
  Ortega Gonzales, Alvaro, *Enhancing Variational Monte Carlo with SVGD and Kernel Herding*, September 2021

Education

Pursuing PhD in Computer Science, University College London, United Kingdom present
Pursued PhD in Mathematics, Charles University, Prague, Czech Republic 2014
(PhD *title not gained* in 2013 due to death of supervisor)
Mgr. (BS + MSc) in Mathematics, Charles University, Prague, Czech Republic 2005

Employment

ML, Tech & Startups: 2017 - present
London North Eastern Railway, October 2022 - December 2023
ML Technical Lead

- people management: designed and hired a senior technical team for production-ready deployment of ML models, mentoring data scientists, designed and implemented methodologies for each step of the ML development life cycle, scrum of ML projects
- designed the architecture for the ML production pipeline on AWS, managed and contributed to the implementation the ISO27007 compliant CI/CD ML platform
- mentoring on various industry specific projects: anomaly detection for fraud detection, propensity modelling for customer in-market prediction, topic labelling of customer queries, prediction of in-station train delays

April19 Discovery Ltd., June 2021 - September 2022
ML R&D Consultancy

- Molecular Drug Design for drugs penetrating the central nervous system
  used methods: similarity searching, QSAR models and multi-task DNNs for target-assay pair prediction, bioactivity prediction, generative models in drug discovery, 3D ligand-protein fingerprints
  UCL CS MSc student projects supervision:
  Bedford, Rory, *Virtual Screening with 3D Molecular Fingerprints,* September 2022
  Penciu, Andrei, *Generative Modelling for Drug Design,* September 2021

Fractal Labs Ltd., July 2018 – August 2020
Senior Data Scientist, Data Science Lead

- Time-series forecasting for Small and Medium Enterprises' cash flow
  used methods: VAEs applied to time series, time warping loss function, Neural ODEs, various CNNs and RNNs based architectures, Random Forests, Kalman Filters
- Domain Specific Conversational chatbot for Small and Medium Enterprises
  used methods: various CNNs and RNNs-based architectures for Language Understanding Unit (NER model, Intent Classifier), Q-learning with memory reply (Dialog Manager component), VAEs (Language Generation)
- full production pipeline for ML models: Tensorflow Extended, docker
- worked on R&D grant applications: awarded 300k GBP InnovateUK grant (Project Lead)
- active collaboration with the University of Bristol and UCL CS on MSc. student projects supervision
  MSc thesis co-supervised:
  Stein, Adam, *Neural Variational Inference for Text Generation*, May 2019
  Pabrinkis, Aurimas, *Measuring the Spatial Distortions Caused by Embeddings*, September 2019

StatusToday, November 2017 – April 2018
Data Science and R&D

- created and developed employee productivity analytics based on user file activity recognition
  used methods: randomised intensities of Compound Poisson processes and change points detection

Academic and Research Positions: 2010 - 2016
University College London, London, United Kingdom
Department of Computer Science, 2015
Visiting Research Student

- preliminary research into constructing financial instruments based on intellectual property and related trading platform
  used methods: hidden Markov models, market-making strategies using RL
- calibration of tempered Levy Flights volatility surfaces to the FX market data
- supervision of MSc student final theses:
  Munir, Necati, *Creating a Trading Platform for Intellectual Property Trading*, April 2016.
  Zhao, Yaolin, *Trading Simulation in the Intellectual Property Market*, September 2015.

University College London, London, United Kingdom
Department of Civil, Environmental & Geomatic Engineering, 2015

Developer

- testing scalability of a high-volume Fusion and Analysis platform for geospatial clouds, which was co-funded by European Commission (FP7 2002 – 2013, IQmulus project)

Gorilla Science, 2011 - 2016
PhD and research funding

- long-running collaboration with Dr. Patrick S. Hagan on the term structure volatility surface models using tempered Levy Flights to produce robust hedges for forward-starting products
  results first presented at: RISK Event Quant Summit 2016 Europe, April 2016
  used methods: partial integral differential equations, stochastic calculus for pure jump processes, perturbation theory, complex variables

Imperial College London, Department of Mathematics, 2015-2016
Visiting Research Student

- fixed income models driven by tempered Levy Flights

University of Oxford, The Oxford-Man Institute of Quantitative Finance, 2010
Visiting Research Student

- computational methods for option pricing models driven by stable laws
  results presented at: QMF Quantitative Methods in Finance, Sydney, 2010

Columbia University, Department of Statistics, 2010
Visiting Research Student

- developing numerical solutions for option pricing models driven by stable laws
  results presented at: seminar talk at Rutgers University

Financial Industry: 2005 - 2010
PricewaterhouseCoopers Czech Republic, 2009 - 2011
Consultant

- pricing exotic financial derivatives and illiquid products

KBC Global Services, Group Value and Risk Management, 2006 – 2008
Model Validation, Associate

- validation of Interest Rate Derivatives and Life Insurance Derivatives models

CSOB (member of KBC Group), Risk Management Department, 2005 - 2009
Market Risk, Associate

- market risk methodology, independent price verification, reporting to the Board of Executives
- quantitative researcher for commodity trading models

Selected Publications

1. (2024) Karlová A., Kabra R., Augusto de Souza D., Paige B.: *Variational Inference with Censored Gaussian Process Regressors*, July 2024, ICML 2024 Workshop on Structured Probabilistic Inference & Generative Modeling
2. (2024) Karlová A., Paige B. : *SWUS: Active Learning with Structure Weighted Uncertainty Score*, July 2024, ICML 2024 Workshop AccMLBio
3. (2022) Karlová A., Dehaen W., Penciu A., Dallaway R., Goonatilake S.; *PEPSI: Post-docking Evaluation with Protein-Small Molecules Interaction*, July 2022, Conference Workshop, WCB ICML 2022
4. (2021) Karlová A., Dehaen W., Penciu A.; *How to Reward Your Drug Agent.* December 2021, Conference Workshop: DGMs and Applications NeurIPS 2021

5. (2021) Karlová A., Underwood C., Silva R.: *Virtual Screening For Perovskites Discovery,* July 2021, Workshop: Tackling Climate Change with ML, ICML 2021
6. (2021) Karlová A., Dehaen W., Svozil D.; *Molecular Fingerprint VAE,* July 2021, Conference Workshop: WCB ICML 2021
7. (2020) Karlová A.; *Robustness of Latent Representations,* July 2020, Conference Workshop Proceedings UDL ICML 2020
8. (2018) Klebanov L., Karlová A.; *Distributional Tail Estimation through its Characteristic Function,* Mar 2018, Journal of Mathematical Sciences, Vol. 229, Issue 6, pp 714–718
9. (2016) Karlová A., Hagan P.S.; *On Beyond Black*; Sept. 2016, Wilmott, p.42-69
10. (2013) Karlová A.; *On distribution of product of stable laws*; SIAM Theory of Probability and its Applications (submitted September 2013, accepted July 2015)
11. (2008) A. Karlová; *Levy Processes and Balayage Spaces*; J. Antoch., G. Dohnal, (Eds.) ROBUST 2008, Proceedings of the 13th Summer School of the Union of Czech Mathematicians and Physicists, 171—176.

Selected Talks

1.	AI UK, London	2022
2.	QuantInsights, London	2019
3.	Women in Data Science, 2019, Tel Aviv,	2019
4.	49th London PyData Meetup	2018
5.	RISK Event Quant Summit 2016 Europe	2016
6.	Quantitative Methods in Finance, QMF 2010, Sydney, Australia	2010

Programming Skills and Languages:

- programming languages: Python, C/C++, Bash
- libraries: Python – Tensorflow, SciPy, Numpy, Pandas, ScikitLearn, PyTorch
- other tools: SQL, MongoDB, git, docker, AWS
- operating system: Linux (Debian)

Grants, Honours, Awards:

Best Poster Award WCB ICML 2021, InnovateUK (2019), Bio-Math Summer School and Workshop. Middelfart, Denmark (2008), 'Stochastic Differential Equation Models with Applications to the Insulin-Glucose System and Neuronal Modelling' at Satellite Summer School on Levy Processes, funded under Marie Curie EU grant contract. Sandbjerg, Denmark (2007), 'Statistics for stochastic differential equations models' at EMS Summer School, Séminaire Européen de Statistique, funded under Marie Curie EU grant contract. Cartagena, Spain (2007), Global Association of Risk Professionals Seminar under a scholarship from Agence de Transfert de Technologie Financiere. Luxembourg (2005)