# Dhruv Srikanth

dhruvsrikanth5@gmail.com | +44 7858599685 | GitHub | LinkedIn | Website        London, UK

## EDUCATION

**The University of Chicago**                                                                *Grad. April 2023*
*Master of Science, Computer Science (Machine Learning, High Performance Computing)*, GPA: 3.7
Relevant Coursework: Algorithms, Data Analytics, High Performance Computing, Parallel Programming, Compilers

**Toyota Technological Institute Chicago**                                                   *Grad. April 2023*
*Joint with UChicago Master of Science, Computer Science*
Relevant Coursework: Machine Learning, Computational Learning Theory [paper], Deep Learning [paper] [code]

**PES University**                                                                           *Grad. Jun 2021*
*Bachelor of Technology, Electronics & Communication Engineering (Signals & Systems)*, GPA: 3.9
Relevant Coursework: Computer Vision, Computer Architecture, Machine Learning, Image Processing, Signals Analysis

**Programming Languages:** Python, C/C++, GoLang, Java, MATLAB, SQL, JavaScript, Mojo, Bash, Lua

**Frameworks/Libraries:** PyTorch, TensorFlow, Hugging Face, Jax, OpenAI API, Ray, CUDA, Triton, MPI, OpenMP, Kubernetes, Pandas, Scikit-learn, Docker, Node, vLLM

## WORK EXPERIENCE

**Weco AI**                                                                                  London , *UK*
*Founding Engineer*                                                                          *May 2024 – Present*
- **Developed best performing ML coding agent** (OpenAI evaluation) **& R&D coding agent** (METR evaluation) outperforming top Human scientists from DeepMind, OpenAI, Anthropic, Stanford, Berkeley, CMU and MIT [package] [paper]
- Leading research on improving LLM code generation through interaction scaling and evolution-based tree search policies for kernel engineering, model development and prompt optimization
- Trained LLMs (Mistral, Mixtral, Llama 3) for code generation using SFT, qlora & quantization with HuggingFace and setup inference pipeline to locally host models using vLLM
- Leading core backend - architected backend system for AIDE using GCP and Vercel for autoscaling & load balancing, Postgres for database management and FastAPI to develop backend API
- Co-leading consumer tools - developed Python CLI for interacting with AIDE backend API enabling customers to iteratively optimize their code using AIDE LLM agent

**Auton Lab: Carnegie Mellon University, Robotics Institute**                                Pittsburgh , *USA*
*Research Programmer | Dr. Artur Dubrawski*                                                   *May 2023 – Apr 2024*
- Developed AI models & graph-based algorithms for semantic segmentation and identification of vessel bifurcations deployed on autonomous robotic arm **for blood vessel surgery** in animal subjects [paper]
- Trained and fine-tuned **multimodal foundation models** over distributed clusters (transformers, VAEs, autoencoders, bayes nets) for detecting patient vitals through drones deployed during disaster relief
- Published AutonFeat [code] [docs]; distributed automatic featurization library for time series analysis (forecast + classification)
- Developed RNNs, LSTMs, transformers with 0.81 AUC with UPMC doctors, **improving renal failure outcomes for ICU patients**

**UChicago Booth Center for Applied Artificial Intelligence**                                Chicago , *USA*
*Machine Learning Researcher | Dr. Sendhil Mullainathan*                                      *Apr 2022 – Mar 2023*
- Trained 50+ models for identifying, measuring and mitigating algorithmic and architectural bias in ML models via transfer learning
- Empirically proved presence of inductive biases (induced correlations between covariates – e.g. race, gender) in pretrained (ImageNet) weights and ubiquitous large vision model architectures (AlexNet, VGG, ResNet, DenseNet, ViT, Diffusion models)
- Developed recommendation engine, expert system and API that utilizes user context to modify dynamic knowledge graphs

**Myelin Foundry**                                                                           *Bangalore, India*
*Machine Learning Engineer (Intern)*                                                         *Jan 2021 – July 2021*
- **Created revenue stream of $1m+ USD** developing real-time competitor analysis tool with language models & generative AI
- Developed and deployed full stack web applications with Python, MySQL, HTML, CSS, JavaScript, Flask, Azure and AWS VMs

**General Electric**                                                                         *Bangalore, India*
*Machine Learning Engineer (Intern)*                                                         *June 2019 – Aug 2019*
- Reported  4+ leading causes of defects across X-ray insert manufacturing process through ML data analysis with Pandas and sklearn
- Developed-deployed X-ray defect identification model (84% accuracy) with Random Forest & Naïve Bayes classifier (QA pipeline)

## PUBLICATIONS

Z. Jiang, D. Schmidt, **D. Srikanth**, D. Xu, I. Kaplan, D. Jacenko, and Y. Wu. AIDE: AI-driven Exploration in the Space of Code. *arXiv preprint arXiv:2502.13138*, 2025. [paper]

C. G. Morales, **D. Srikanth**, J. H. Good, K. A. Dufendach and A. Dubrawski, "Bifurcation Identification for Ultrasound-driven Robotic Cannulation," 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [paper]

P. Shukla, **D. Srikanth**, L. Cohen, and M. Turk. Utilizing Adversarial Examples for Bias Mitigation and Accuracy Enhancement. *arXiv preprint arXiv:2404.11819*, 2024. [paper]

S. Bidnur, **D. Srikanth**, and S. Gurugopinath. Resource-Conscious High-Performance Models for 2D-to-3D Single-View Reconstruction. *IEEE Region 10 Conference*. [paper] [code]

## OPEN SOURCE

**Weco AI: aideml** (930+ stars)
- Contributing to internal R&D at Weco AI exploring agentic tree search algorithms, LLM agent system design and core backend development. Improvements are rolled out periodically to the package by our team. [package]

**HuggingFace: Text Generation Inference** (10.2k+ stars)
- Enabled 4 bit quantization in HuggingFace's LLM inference engine by fixing bugs in Multi-Layer Perceptron block [commit]

**TSFresh** (8.8k+ stars)
- Implemented fine-grained control over multiprocessing of parallelized functions enabling developers to configure number of processes/jobs in model training and feature extraction [commit]

## PROJECTS

GoLLUM - A Compiler between C++ and Go [code], PyNN (Deep Learning Framework) [code], GAN Experimentation Packages [code], Neural style transfer CLI tool [code], RTConcierge; Road trip recommendations via LLMs, Speech Transcription & Translation [code]

## AWARDS

**Myelin Intern Award**
- Awarded to highest performing intern at the company
- Second person in company history to receive award (2021)

**Prof. M R Doreswamy Merit Scholarship Award (x2)**
- Awarded to top performing students in the Department
- Awarded in Semesters 5 and 6 (2019-2020)

**Prof. C N R Rao Merit Scholarship (x2)**
- Awarded to top performing students in the Department
- Awarded in Semesters 3 and 4 (2018-2019)

**Distinction Award (x6)**
- Awarded to students securing a grade of First Class Distinction
- Awarded in Semesters 1, 2, 3, 4, 5 and 6 (2017-2020)