SASHA <ALEXANDRE> DOUBOV

doubovs@cs.toronto.edu | cs.toronto.edu/~doubovs/ | sashadoubov |
sashaDoubov | Sasha Doubov | sashadoubov

Work Experience

MosaicML/Databricks Aug 2023 -
Research Scientist *San Francisco*

- Core member of pre-training team for DBRX, a 132 billion Mixture of Experts (MoE) open-source model: highest MMLU, HumanEval, GSM8k scores for an open model at the time of release
- Performed model architecture and scaling experiments to select final hyperparameters and architecture for DBRX model
- Contributed to MegaBlocks MoE training library, becoming the #2 contributor behind the creator of MegaBlocks
- Mentored a research scientist intern, offering technical support and steering project goals
- Working on improving MoE routing, fine-tuning and quantization-aware training

MosaicML April 2023 - Aug 2023
Research Scientist Intern *San Francisco*

- Explored techniques for hyperparameter search (ex. muP) for pre-training LLMs
- Added long-context and domain-specific evals for comparing various finetuning recipes of LLMs

Cohere Oct 2022 - Mar 2023
Machine Learning Intern *Toronto*

- Proposed and evaluated structured sparsity techniques for pre-training LLMs, including block-diagonal butterfly matrices and layer dropout
- Integrated structured pruning approaches into Cohere's training stack, including rewriting PyTorch implementations into JAX

Cerebras Apr 2022 – Aug 2022
Research Intern *Toronto*

- Investigated unstructured sparsity algorithms to improve CNN model performance
- Developed algorithms to accelerate sparse neural network training

Uber ATG Sep 2019 - Dec 2019 & Jan 2019 – Jul 2019
Research Intern *Toronto*

- Developed novel deep learning algorithms for large-scale retrieval-based localization using LiDAR

Conference Publications

Beyond Chinchilla-Optimal: Accounting for Inference in Language Model Scaling Laws
ICML 2024

- Nikhil Sardana, Jacob Portes, **Sasha Doubov**, Jonathan Frankle

Scalable Neural Data Server: A Data Recommender for Transfer Learning
NeurIPS 2021

- **Sasha Doubov***, Tianshi Cao*, David Acuna, Sanja Fidler

Pit30M: A Benchmark for Global Localization in the Age of Self-Driving Cars
IROS 2020 *Finalist Best Application Paper*

- Julieta Martinez, **Sasha Doubov**, Ioan Andrei Bârsan, Shenlong Wang, Gellért Máttyus, Raquel Urtasun

Workshop Publications

Sparse Upcycling: Inference Inefficient Finetuning

NeurIPS 2024 ENLSP Workshop

- **Sasha Doubov**, Nikhil Sardana, Vitaliy Chiley

How many trained neural networks are needed for influence estimation in modern deep learning?

NeurIPS 2022 I Can't Believe It's Not Better Workshop

- **Sasha Doubov**, Tianshi Cao, David Acuna, Sanja Fidler

Studying BatchNorm Learning Rate Decay on Meta-Learning Inner-Loop Adaptation

NeurIPS 2021 Meta-learning Workshop

- **Sasha Doubov***, Gary Leung*, Alexander Wang*

Education

University of Toronto — Sep 2020 – Apr 2022

MSc Computer Science — *cGPA: 3.93*

- Advisor: Prof. Sanja Fidler
- Courses: Information Theory, Neural Net Training Dynamics

University of Waterloo — Sep 2015 – Apr 2020

BASc Electrical Engineering — *cGPA: 94%*

- First in Class for graduating cohort

Awards

Sandford Fleming Award for Academic Excellence	2020
Gerry Heckman Scholarship	2020
First in Class Engineering Scholarship	2020, 2018
Waterloo North Hydro Electrical Engineering Scholarship	2019
President's Research Award	2018
Hatch Entrance Scholarship	2016
University of Waterloo President's Scholarship of Distinction	2016

Skills

Languages: Python, C/C++, Java, MATLAB
Frameworks & Tools: PyTorch, Jax, Tensorflow, Git