Michal Malyska

647 780 4095 • malyskamichal@gmail.com • michalmalyska.com • Github

Work Experience

Machine Learning Lead **Toronto, Canada**
Semantic Health (acquired) *January 2021 - December 2022*

- Spearheaded the research and development of the company's clinical ML offerings (variety of clinical NLP + OCR + Tabular), successfully delivering multiple client deployments, leading to a US expansion, and driving a $60M valuation.
- Hired and led a team of 7 ML Scientist / Engineers: Managed their professional growth, setting objectives, conducting standups, and nurturing their leadership and technical capabilities.
- Created and deployed multiple new ML product lines based on customer calls and product insights from internal QA efforts leading to increases of 100k ARR per contract
- Introduced a novel non-intrusive data labelling and feedback scheme, increasing the granularity of labels without impacting the end user's workflow, which led to creation of new models with >50% performance gain on downstream KPIs
- Roadmapped and oversaw a new MLOps initiative to allow for better data and model monitoring

Machine Learning Scientist **Toronto, Canada**
Semantic Health (acquired) *June 2019 - December 2020*

- Joined as the first employee - architected and built the ML stack setting the foundation for the company's NLP capabilities
- Developed a custom multi-armed bandit model deployment scheme quadrupling ML deployment velocity to production
- Pioneered a novel architecture for generating supporting evidence for ML model predictions, implemented and deployed it leading to a 300% increase in business KPIs for the product
- Authored a series of Architecture Decision Records that still serve as the model for decision documents for the company
- Implemented a testing suite for the ML stack, including model minification allowing for pre-commit end-to-end ML testing

Sessional Lecturer **Toronto, Canada**
University of Toronto *2020-Present*

- Developed educational materials and delivered instruction for undergraduate and graduate courses in Statistics and ML
- Communicated sophisticated technical topics to a non-technical audience
- Modernized the CSC412 - Probablistic ML course curriculum by adding Attention, Transformers, and Diffusion models

Education

University of Toronto **2022, 2019**
MSc. Statistics, BSc. Statistics

- Research Visitor at St Michael's Hospital, President of Statistical Sciences Union, Dean's List, Teaching Assistant

Publications

- BioNLP - ICDBigBird: A Contextual Embedding Model for ICD Code Classification
- ACM CHIL - Active learning for medical code assignment
- EMNLP - Multiple Sclerosis Severity Classification From Clinical Text

Skills

- **Programming Languages**: Python, R, SQL, JavaScript
- **Libraries** : Pytorch, Tensorflow, SpaCy, Git, MLFlow, Kubernetes, Docker, Dagster, Airflow, Feast, Seldon, ArgoCD
- **Machine Learning** : Natural Language Processing, LLM, Neural Networks, Deep Learning, Statistical Learning
- **Open Source Contributions**: MedSpaCy, AllenNLP, SciSpaCy, MS-BERT