

SANJA SIMONOVIKJ

sanja_sss@yahoo.com ♦ sanjass.github.io ♦ github.com/sanjass

Full-stack Data Scientist with 3+ years of industry experience leveraging AI/NLP to drive business value at scale. MEng and BSci alumna of MIT, award winner of IMO (International Math Olympiad), multi-hackathon winner. Looking for the right opportunities in Toronto, ON, including remote roles. Fully authorized to work in Canada.

EDUCATION

Massachusetts Institute of Technology

Master of Engineering, Concentration in Artificial Intelligence

Jun 2020 - Jun 2021

Bachelor of Science, Computer Science and Engineering

Sep 2016 - May 2020

INDUSTRY EXPERIENCE

Indeed

Senior Data Scientist, Metadata and Taxonomy

Toronto (Remote), Austin, TX (Remote), Jan 2023 - present

Data Scientist, Metadata and Taxonomy

Austin, TX (Remote), Jul 2021 - Jan 2023

- Crafting and guiding the implementation of tailored, production-grade AI/NLP solutions for extracting and representing metadata from jobs and resumes powering matching systems across domestic and global markets
- Led a pivotal initiative to transition 8 core markets to a fully AI-powered approach to job occupation metadata extraction with an outsized impact, providing a 10-20% increase in key coverage and quality metrics
- Leveraged LLMs (fine-tuning, prompting, embeddings) to accelerate the generation and validation of new types of metadata, improving job and jobseeker data collection, matching, and measurement
- Foundational member of a cross-team effort to create in-house HR domain-specific LLM; curated over 20 HR training and evaluation datasets, guided and tested the implementation of a rigorous evaluation framework
- DSci lead for a company-wide initiative to transform metadata collection and usage in a core market (Japan), owning workstreams requiring creative solutions, effective change management, and inter-team collaboration
- Conceptualized and improved upon numerous human-in-the-loop AI-based workflows, facilitating ongoing model improvement and accelerating the expansion of the occupation Taxonomy in over 12 new international markets
- Advising fellow Scientists, Taxonomists, and Engineers on best practices, automation, and trade-off management

Microsoft - Data Science Intern, Docs ML team, Microsoft Fluid Office

Cambridge, MA (Remote), Jan 2021

- Created and optimized lightweight classification models that perform 20% better than a heuristic baseline

Qualcomm - Software Engineering Intern

San Diego, CA, Jun-Aug 2019

- Built an AI-powered triaging framework, reducing the team's troubleshooting sessions from hours to minutes

RESEARCH AND TEACHING EXPERIENCE

MIT EECS Department Graduate Teaching Assistant for a popular ML course (6.036)

Feb - May 2021

Improbable AI Group, MIT CSAIL Graduate Research Assistant - Prof. Pulkit Agrawal Feb 2020 - Jun 2021

AWARDS AND PROJECTS

- **Indeed Hackathon:** Indeed Tweakify (AI Resume Writer) - 1st place track winner Jun 2023
- **Indeed Hackathon:** Indeed Rhizome (Social Graph) - 1st place overall, Global winner Aug 2021
- **HackMIT:** LabelLearn (Smart Data Labeling), 1st place overall, Grand Prize winner Sept 2019

SKILLS

Programming Languages: Python, Java, Javascript, Bash, R, SQL, Julia

Technologies: Git, Docker, AWS (S3, EC2, Lambda, CloudWatch, EBS), Elasticsearch, Mongo, Presto, Neo4j

ML Tools: OpenAI, TensorFlow, TFX, Pytorch, Kubeflow, Streamlit, standard ML packages (pandas, dask, pyspark, numpy, scikit-learn, keras, matplotlib), Jupyter Notebooks, Apache Beam, Apache Spark, Optuna, Langchain

Human Languages: English, Macedonian (native), Serbian (proficient), Spanish (intermediate), French (limited)