Olga Seleznova

Generative AI Engineer / Machine Learning Engineer

Ottawa, ON | 343-322-1181 | olgaselesnyova@gmail.com | linkedin.com/in/olga-seleznova

Enthusiastic ML Engineer with extensive experience in deep learning modeling, natural language processing (NLP), and speech technologies. Proficient in machine learning (ML), model debugging and optimization, and competency in translating complex data into actionable insights. Successfully improved research and decision-making processes by managing large-scale data environments and implementing strategic solutions. Experienced in fostering projects within cross-functional teams, with a proactive approach to problem-solving and a proven ability to leverage advanced statistical, data-driven, and modeling techniques to achieve strategic goals and deliver impactful results.

CORE COMPETENCIES

- 3+ years of expertise in progressive data science roles in large-scale organizations, with hands-on expertise in delivering AI solutions to address complex business challenges.
- Proficient in Python programming and handling unstructured and structured data for machine learning, NLP, and speech recognition.
- Highly competent in interpreting and communicating complex data findings to business stakeholders and delivering meaningful presentations
- Experienced in managing the entire machine learning lifecycle, from model development to deployment, using Docker in AWS cloud environments.
- Strong interpersonal and team-building skills and a proven ability to mentor and support teams in various projects and foster an atmosphere of trust, collaboration, and high performance in a dynamic, fast-paced environment
- Highly organized with keen attention to detail and proficiency in managing projects, risks, and opportunities and a proven track record of driving continuous improvement initiatives and end-to-end machine learning solutions

TECHNICAL EXPERTISE

Programming languages: Python, Bash
Machine learning tools: NumPy, Pandas, Scikit-learn, Matplotlib, and others
Deep learning modeling: PyTorch, Transformers, Hugging Face, Weights & Biases
Natural language processing Tools: Word2Vec, SpaCy, BERT, vector similarity (FAISS)
Speech Recognition packages: Whisper, Torch audio, Wave2vec, HuBERT, Librosa
Large Language Models (LLM), AI tools and methodologies: Prompt engineering, Ollama, Llama, Mistral, OpenAI, LangChain, LlamaIndex, Retrieval Augmented Generation (RAG), GitHub copilot
Version control: Git, GitHub
Cloud technologies: AWS, Docker
Databases: SQL, MongoDB
IDEs: VSCode, PyCharm, Jupyter Notebooks, Google Colab
UI frameworks for Machine learning: Gradio, Streamlit

PROFESSIONAL EXPERIENCE

Research Engineer March 2023 – May 2024
Technion - Israel Institute of Technology, Israel

Operated within the 'Speech, Language, and Deep Learning' Lab, driving the modeling and research of deep learning projects in the speech and language domains, utilizing Python, PyTorch, Whisper, Torchaudio, NumPy, Matplotlib, and other advanced technologies.

- Developed over 4 clean and reusable packages for academic purposes by creating templates and streamlining submission processes
- Enhanced research efficiency by contributing to a Master's degree publication through the implementation of advanced multithreading and multiprocessing models, significantly improving computational performance

- Guided students by resolving complex technical challenges related to Python, PyTorch, environment setup, and large-scale data management (terabytes), facilitating smooth project execution and enhancing research outcomes
- Managed virtual machines and created server and Docker environments to support deep learning projects, ensuring a functional and scalable infrastructure
- Conducted research in automatic speech recognition using Whisper, HuBERT, and wav2vec libraries and performed speech alignment with Montreal Forced Aligner and restoration with UMAP model
- Created and maintained reusable packages (tools) tailored to the laboratory's needs, streamlining research processes and improving resource efficiency
- Collaborated on the creation of Hebrew speech recognition datasets, leading to their publication on Hugging Face by ivrit.ai and supporting the broader AI community
- Interacted closely with Master's and PhD researchers on innovative research projects in deep learning and speech processing

Data Scientist September 2021 – September 2022
Skai (Software Development), Israel

- Provided valuable insights from eCommerce reviews to management, facilitating data-driven decision-making and enhancing strategic product development based on customer feedback
- Led the research, development, and implementation of advanced NLP systems, including the fine-tuning and error analysis of BERT for text classification and T5 for question-answering tasks
- Designed and deployed semantic similarity and keyword extraction modules to advance insights mining, improving the accuracy and relevance of extracted information
- Analyzed and processed large datasets of unstructured reviews from specific product verticals to extract actionable insights, guiding product development and enhancing customer understanding
- Managed the dataset labeling process and model fine-tuning to optimize model performance in close collaboration with project managers and analysts
- Coordinated a multidisciplinary team of Python and R developers, data scientists, and statisticians under a product manager's guidance to achieve project deliverables and ensure compliance with technical standards

Data Science Intern February 2021 – July 2021
TRG Solutions (IT Services and IT Consulting), Israel

- Performed Named Entity Recognition and developed sentiment prediction models for social media posts (Twitter and Facebook)
- Enhanced fine-tuning of the DistilBERT model from HuggingFace, increasing sentiment classification accuracy from 47% to 86.20% across three categories (positive, negative, and neutral)
- Implemented advanced text classification models, including Random Forest, CatBoost, XGBoost, RoBERTa, and DistilBERT, utilizing Scikit-learn and Transformers packages to enhance predictive accuracy and performance
- Delivered a comprehensive presentation of findings to students and lecturers, effectively communicating technical details and insights
- Collaborated closely with team members and operated under the guidance of a senior data scientist, efficiently sharing responsibilities and contributing to team success

Researcher August 2020 – December 2020
University of Haifa, Israel

- Achieved approximately 85% accuracy in classifying Hebrew posts from medical forums, addressing low-scale, unbalanced data challenges
- Conducted comprehensive EDA, text preprocessing, and classification using Python libraries, including Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, and PyTorch transformers
- Collaborated with the Integrative Pain (iPain) laboratory on a project focused on understanding chronic pain conditions, contributing to the development of new measurement, prevention, and treatment methods
- Contributed to a published paper on informativeness versus emotionality in medical forums, identifying key features to differentiate between these concepts

EDUCATION AND PROFESSIONAL DEVELOPMENT

Master's Degree in Data Science 2022
University of Haifa, Israel
Advancement Data Science Program 2021
Ydata, Intensive one-year career advancement program in Data Science, Israel

CERTIFICATIONS

Generative AI with Large Language / DeepLearning AI - Coursera 2023
Introduction to Machine Learning in Production / DeepLearning AI - Coursera 2023
Mathematics for Machine Learning: Multivariate Calculus and Linear Algebra / Coursera 2020
Data Structures and Algorithms in Python / Udacity 2020
Using Python to Access Web Data / University of Michigan – Coursera 2018
Python Data Structures / University of Michigan – Coursera 2018

EXTERNAL PROJECT

VoyageVocab Project: Language Learning Assistant for Travellers
Developed an application utilizing Llama and LangChain to generate travel-related phrases for tasks such as ordering food or buying tickets, with a user-friendly interface implemented through the Gradio package.
Link: github.com/OlgaSeleznova/VoyageVocab

Job Posting Relevance
an AI-powered tool designed to determine how relevant a job posting is to a candidate's resume. The system leverages a large language model to provide an analysis of how well a resume aligns with the specific requirements of a job posting.
Link: github.com/OlgaSeleznova/JobPostingRelevancy