Oleksiy Ts.
Senior Data Scientist / NLP Engineer

SUMMARY

- Data Scientist with a focus on Natural Language Processing and 16 years of experience in the IT industry; - Experienced in designing cloud-based architectures using BigQuery, Vertex AI, Cloud Functions, Datastore, and Cloud Storage; - Applies machine learning and deep learning techniques to solve problems in NLP, computer vision, and time series analysis; - Solid foundation in unsupervised learning methods, including clustering, anomaly detection, and recommender systems; - Skilled in full-cycle solution design, from requirements gathering to development and deployment; - Specialized in Generative AI, building advanced algorithms and models to address complex tasks using generative techniques; - Experience in ML model development for prediction, classification, and optimization use cases; - Proficient in Agentic AI, creating autonomous systems powered by LLMs for decision-making and workflow automation; - Experienced in text-to-speech technologies and implementation; - Experience with cloud platforms, including AWS (S3, EC2, Lex, Lambda, EMR), Azure OpenAI, and Google Cloud; - Developed applications across macOS, Linux, and Windows environments; - Creative and results-driven, with strong analytical, communication, and presentation skills.

TECHNICAL SKILLS

Main Technical Skills	Python, Data Science, NLP, AWS
AI Tools & Assistants	ChatGPT
Programming Languages	C#, Fortran, JavaScript, Python, R, VBA
AI & Machine Learning	AWS Lex, AWS ML (Amazon Machine learning services), AWS SageMaker (Amazon SageMaker), Azure OpenAI, BERT, CLIP, FastAi, Gen AI, GPT, Keras, LangChain, LLaMA, LlamaIndex, NLP, NumPy, OpenAI, RAG, Scikit-learn, STT (Speech-to-Text), T5, TensorFlow, Vertex AI, VITS, Whisper
Python Libraries and Tools	Beautiful Soup, Gensim, Keras, Matplotlib, NetworkX, NLTK, NumPy, Pandas, PySpark, Scikit-learn, SciPy, TensorFlow, Tkinter
Python Frameworks	Django, Django REST framework, FastAPI
.NET Platform	VBA
Data Analysis and Visualization Technologies	Apache Airflow, Jupyter Notebook, Pandas, Tableau
Databases & Management Systems / ORM	HDFS, HeidiSQL, Microsoft SQL Server, MongoDB, MySQL, MySQL Workbench, PostgreSQL, SQL
	CSS, HTML

UI Frameworks, Libraries, and Browsers	
Cloud Platforms, Services & Computing	AWS, Azure, GCP
Amazon Web Services	AWS EC2, AWS EMR, AWS Lambda, AWS Lex, AWS ML (Amazon Machine learning services), AWS S3, AWS SageMaker (Amazon SageMaker)
Azure Cloud Services	Azure Cloud Functions, Azure OpenAI
Google Cloud Platform	GCP Compute Instance, Google BigQuery
Platforms	Apache Solr
iOS Libraries and Tools	App Clips
Collaboration, Task & Issue Tracking	Atlassian Trello, Jira, Redmine
Virtualization, Containers and Orchestration	Docker, Kubernetes
Third Party Tools / IDEs / SDK / Services	Eclipse, MatLab, PyCharm, Sublime Text, Visual Studio
SDK / API and Integrations	FastAPI, Google API, GraphQL, MOZ API
Version Control	Git, TortoiseSVN
Methodologies, Paradigms and Patterns	Kanban, Scrum
Operating Systems	Linux, macOS, Windows
Mail / Network Protocols / Data transfer	PGM
QA, Test Automation, Security	Selenium
Other Technical Skills	BITS, Heap Analytics, MS CRM, Octave, Rest Framework, TTS

WORK EXPERIENCE

NLP Engineer, Document extraction model

Duration: Jan 2024 – Present

Summary: The project aims to develop an automated document extraction system to extract necessary documents from tender descriptions. The primary objective is to enhance document extraction accuracy to minimize errors during the document collection process. Additionally, the system will focus on increasing efficiency by automating the document identification process and enabling seamless integration with other applications through the API.

Responsibilities:

• Data Collection to train and evaluate an effective NLP model Implement the paraphrasing of questions;

• Document extraction;

• Integration and developed NLP model. NLP model integrated into existing procurement management systems;

• Testing and validation to ensure the accuracy and reliability extraction model;

• Web Scraping.

Technologies: GPT, Langchain, LlamaIndex, Azure.

NLP Engineer, AI Customer Support

Duration: Sep 2023 – Dec 2023

Summary: The project idea involves developing a chatbot for the client's website. The bot is capable of answering customer inquiries regarding experience and capabilities. Additionally, it can integrate into the company's internal ERP system, aiding in extracting essential information from a vast array of data.

Responsibilities:

• Data Collection and Pre-processing;

• Model Selection and Development;

• Integration and Deployment;

• Integrating various APIs and web services;

• Utilizing cloud infrastructure for deployment and management.

Technologies: LangChain, Sentence Transformers, AWS Sagemaker, OpenAI ChatGPT API, LLMs, RAG.

NLP Engineer, Question-answering ML solution

Duration: Jun 2023 – Aug 2023

Summary: A question-answering ML solution for a fantasy game to serve as a chatbot with world knowledge at hand. The player can interact with a persona to retrieve knowledge about the universe. Except for retriever-reader components, the solution also includes grammar correction and paraphrasing to improve the quality of answers as well as integration with ChatGPT to ensure the ability to keep the context of the previous discussion.

Responsibilities:

• Implement the question-answering machine learning solution for the chatbot;

• Implement Extractive Q&A, including grammar correction;

• Compare between pre-trained models Bert family, Electra, Roberta, Sentence transformer;

• Implement the paraphrasing of questions.

Technologies: Python, Electra, TFIDF, Hugging Face T5, OpenAI ChatGPT.

NLP Engineer, Under NDA

Duration: Mar 2023 – Jun 2023

Summary: The solution for generating the text summary and generating the most relevant keywords for unpublished web pages to help run the SEO process.

Responsibilities:

• Implement keywords extraction solution for unpublished web pages;

• Implement text summarization for web content;

• Train Bert models for keyword extraction;

• Train GPT 3 models for keyword extraction;

• Improve prompt engineering for GPT 4 for keyword extraction;

• Integrating various APIs and web services;

• Utilizing cloud infrastructure for deployment and management.

Technologies: Python, Bert, OpenAI GPT, USE.

NLP Engineer, Simulation of Communication

Duration: Nov 2022 – Feb 2023

Summary: The goal of the project is to create a simulation of communication as similar as possible to certain human, considering topics that they like or dislike, their writing style, and their voice.

Responsibilities:

• Leveraged CLIP model to acquire the topics from images;

• Applied generative model for composing descriptions of customer sentiments to topics;

• Fine-tuned GPT model to consider generic style and attitude of certain persons to simulate human communication;

• Train the VITS model for text-to-speech conversion.

Technologies: Python, CLIP, OpenAI GPT 3.0, ChatGPT, VITS.

NLP Engineer, Review analysis assistant

Duration: Aug 2022 - Dec 2022

Summary: The tool to assist with analyzing product reviews. It is an ML solution based on fine-tuning GPT-3 for evaluation of evaluation of completeness, grammar, and relevancy of product reviews.

Responsibilities:

• Implement solution for evaluation of completeness, grammar, and relevancy of product reviews;

• Develop API and a simple web app to work with the solution.

Technologies: Python, OpenAI GPT 3.0.

Data Scientist / ML Engineer, Segmentation of retail customers

Duration: May 2022 – Aug 2022

Summary: The project aims to segment retail customers into homogeneous groups based on purchases history including recency, frequency, and monetary metrics. The discovered patterns of customers' behavior allow us to recommend engagement campaigns to prevent customer churn.

Responsibilities:

• Implement RFM model for customer segmentation;

• Prepare presentations for stakeholders.

Technologies: Python, GCP, k8s

Data Scientist, Estimation of outcomes odds of sports games

Duration: Jan 2022 – Apr 2022

Summary: The system for predicting probabilities of outcomes of sports games in the betting domain to assist bookmakers with setting rate coefficients

Responsibilities:

• Executed EDA;

• Defined metrics;

• Evaluated available models;

• Designed the solution;

• Prepared reports with insights towards improvements;

• Managed A/B testing.

Technologies: Python, Predictive Modelling.

Data Scientist / ML Engineer, Retail Discovery

Duration: Sep 2021 – Dec 2021

Summary: The project aims to discover the items frequently bought together by association rules mining.

Responsibilities:

• Implement market basket analysis algorithms;

• Research with the most relevant association rules;

• Present results to business stakeholders.

Technologies: Python, GCP, k8s

Data Scientist / ML Engineer, MistakePredict

Duration: May 2021 – Sep 2021

Summary: Predicting the probability of potential mistakes during customer self-scanning in retail stores.

Responsibilities: RFM model implementation for segmentation, Feature engineering, Cloud deployment, Presenting to stakeholders

Technologies: Python, Sklearn, GCP Vertex AI

Data Scientist, Estimation of churn prediction

Duration: Jan 2021 – May 2021

Summary: Solution for churn prediction in the domain of gambling and casinos to assist with corresponding advertisement campaigns.

Responsibilities:

• Designed the ML solution;

• Worked on feature engineering;

• Trained a variety of ML models for estimating the date of the next credit request.

Technologies: Python, Core ML algorithms

Data Scientist / ML Engineer, Fruit labels classification

Duration: Mar 2020 – Dec 2020

Summary: The AI-based solution for the classification of fruits and vegetables using image recognition technologies.

Responsibilities:

• Develop image recognition service for fruit label classification;

• Deployed it to GCP VM/k8s;

• Deploy web app;

• Integrating various APIs and web services;

• Utilizing cloud infrastructure for deployment and management;

• Implement market basket analysis algorithms;

• Implemented models for retail fraud detection;

• Create solution design;

• Prepare presentations for stakeholders;

• Estimate data science tasks.

Technologies: Python, Tensorflow, Keras, Cloud Vision AutoML, GCP, k8s, FastAPI, Vertex AI.

Data Scientist, ML anti-fraud solution

Duration: Oct 2019 – Mar 2020

Summary: An innovative ML anti-fraud solution that detects fake mobile app installations with the help of supervised and unsupervised machine learning algorithms including deep learning helps mobile application advertisers stay ahead of fraudsters who charge for fake installations.

Responsibilities:

• Develop sequential models for fraud detection;

• Develop unsupervised models for fraud detection;

• Develop rules set models for fraud detection;

• Work on features engineering.

Technologies: Python, Tensorflow, Keras, HDFS.

Software Architect, ML Engineer, ML-Powered Online Shopping Recommendations

Duration: Mar 2019 – Oct 2019

Summary: ML solution for session-based recommendations for users of an online shop with the help of LSTM / STAMP models that are trained on historical events sequences such as view, add to bags, and purchase.

Responsibilities:

• Building a pipeline from data engineering to model deployment and creating a microservice for its usage;

• Designing architecture incorporating all specified components;

• Performing cost estimation for the developed solution;

• Defining KPIs;

• Setting up monitoring (observability) and visualization (visibility) for the system;

• Analyzing tools for each stage of development and deployment, determining reasons for their usage, and justifying these choices;

• Developing database table architecture for storing and processing project data;

• Implement improvements to a model, in particular, added additional features;

• Develop model offline testing solution;

• Develop and tested model API using the Gatling load tool;

• Develop algorithms for calculating product popularity.

Technologies: Python, Tensorflow, GCP, KubeFlow, BigQuery, Vertex AI, Scheduled Queries, Cloud Functions, Datastore, Cloud Storage.

NLP Engineer, Neural Machine Translation

Duration: Dec 2018 – Mar 2019

Summary: The project aim is to improve the neural machine translation with the help of state-of-the-art recurrent neural networks with attention.

Responsibilities:

• Work on data preprocessing for NMT;

• Predict placeholders POS using LSTM.

Technologies: Python, Pandas, NumPy, Sklearn, Tensorflow, NLTK, SciPy, GCP Translation AutoML.

NLP Engineer, Automated analysis of infrastructure logs

Duration: Apr 2018 –Dec 2018

Summary: System for automated analysis of infrastructure logs which are row text data to discover the associated groups of resources that generate the logs based on extracted tags, IDs, names, and other recognized entities.

Responsibilities:

• Develop the ML pipeline of text processing;

• Implement various algorithms to capture relevant quality clusters and groups of log resources.

Technologies: Python, Sklearn, NLTK, regexp.

Data Scientist, System for automated process

Duration: Dec 2017 – Apr 2018

Summary: System for the automated process of validating and extracting data from the documents to hint the most probable values at entering corresponding types of fields such as email, date, organization name, etc. Solution based on tuning OCR model, and applying custom algorithms for improvement.

Responsibilities:

• Execute research of available solutions;

• Implement the pipeline of image processing and OCR;

• Evaluate and optimize the results.

Technologies: Python, Tesseract, OpenCV, Docker.

Data Scientist, Under NDA

Duration: Aug 2017 – Dec 2017

Summary: ML solution for clustering the keywords by semantic similarity and for estimating the missing values of monthly search volume.

Responsibilities:

• Implement entity recognition (location, trend, fabric, color, etc.);

• Develop web app (UI/API) for demo of data categorization;

• Develop automated reports;

• Develop custom algorithms to compute missed data.

Technologies: Python, pandas, NumPy, Sklearn, docker, AWS s3, NLTK, USE.

ML engineer, Data operation

Duration: Jul 2017 – Aug 2017

Summary: The project aim is to estimate the project state health as the probability of its completion based on current project characteristics. The solution has been implemented as an inference of a probabilistic graphical model with parameters provided by an expert with the possibility of training the model on factual observations.

Responsibilities:

• Develop functionality to gather data from the confluence, Excel, and My SQL to use in the model;

• Develop functionality to automatically validate the data gathered automatically as well as provided manually;

• Develop an algorithm for learning the PGM model from observed data and feedback;

• Develop an algorithm for generating simulated data to discover possible optimized model structures;

• Develop a library to convert the PGM model from SamIam format to pgm.py format and further export of pgm.py model to Python text code.

Technologies: Python, pgm, MySQL, SEMAIM.

EDUCATION

- **Ivan Franko National University of L'viv**
- Specialist's degree in Mathematics
- 2000
- **Institute of Mathematics NASU**
- Candidate of science in Theory of dynamical systems
- 2011

CERTIFICATION

- **Aerial Image Segmentation with PyTorch** (Oct 2022)
- **Launching into Machine Learning** (Oct 2022)
- **How Google does Machine Learning** (Oct 2022)
- **Facial Expression Recognition with PyTorch** (Sep 2022)
- **Practical Reinforcement Learning** (Feb 2021)
- **Architecting with Google Kubernetes Engine: Foundations** (Nov 2020)
- **Google Cloud Platform Fundamentals: Core Infrastructure** (Nov 2020)
- **Scalable Machine Learning on Big Data using Apache Spark** (Aug 2020)
- **Google Cloud Platform Big Data and ML Fundamentals** (Apr 2020)
- **TensorFlow in Practice Specialization** (4 courses) (Apr 2020)
- **Sequences, Time Series, and Prediction** (Apr 2020)
- **Convolutional Neural Networks in TensorFlow** (Jul 2019)
- **Natural Language Processing in TensorFlow** (Jul 2019)
- **Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning** (Jul 2019)
- **Deep Learning Specialization** (5 courses) (Aug 2018)
- **Sequence Models** (Aug 2018)
- **Convolutional Neural Networks** (Aug 2018)
- **Neural Networks and Deep Learning** (Jun 2018)
- **Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization** (Jul 2018)

- **Structuring Machine Learning Projects** (Jul 2018)
- **Applied Data Science with Python Specialization** (5 courses) (May 2018)
- **Applied Machine Learning in Python** (Feb 2018)
- **Applied Social Network Analysis in Python** (Feb 2018)
- **Applied Text Mining in Python** (Oct 2017)
- **SEO Tutorial for Beginners** (Feb 2018)
- **Introduction to Data Science in Python** (Oct 2017)
- **Applied Plotting, Charting & Data Representation in Python** (Sep 2017)
- **Probabilistic Graphical Models Specialization** (3 courses) (Jul 2017)
- **Probabilistic Graphical Models 2: Inference** (Jul 2017)
- **Machine Learning** (Jun 2017)
- **Probabilistic Graphical Models 1: Representation** (Jun 2017)
- **Probabilistic Graphical Models 3: Learning** (Jun 2017)