Stanislav Jirak

jirakst@gmail.com Prague, Czechia +420-777090822

Profile

Taking a part in several high-profile projects and being with some of the largest clients and financial institutions in the region, Stanislav is highly regarded as a skilled and seasoned data professional with deep-rooted expertise across the entire data industry, especially in Generative AI (genAI) and Large Language Models (LLM). His experience and expertise can both drive the clients on strategical level to help them navigate in the every-changing landscape of AI, as well as implement complex projects hands-on to boost business performance and technical innovation. In addition to his professional endeavors, Stanislav is a highly regarded data science instructor, consistently earning high praise from his students. His dual mastery in practical and educational aspects of data science ensures that he shares his deep knowledge effectively. His achievements encompass winning multiple competitions, such as those on Blockchain or hosted by OpenAI, and publishing research in collaboration with Charles University. Stanislav is also an active member of the CityDAO blockchain community and believes in decentralized and fair future. He enjoys taking a part in competitions, meetups and conferences.

Employment History

Head of AI, SudoLabs, Remote, San Francisco - US

September 2024

I contribute to long-term AI strategy of the company and its positioning on the market. I participate in pre-sales activities and discussion with the client, to help open new opportunities and convert them into actual projects. I manage the implementation process and oversee all AI projects, typically spanning across several AI teams. I develop and shape long-term competencies within my AI department, and I support invidual team members in achieving their goals that align with the company strategy. I open new strategical partnerships with large vendors to support new sales opportunities and build the competency.

PROJECT EXPERIENCES:

- **IQor:** For a major customer experience (CX) client with over 40,000 employess in more than 10 country, I oversee and drive a project where we analyse a large amount of call transcripts on Amazon AWS Bedrock. To accomodate for cca 500,000 call transcripts per month per customer, we developed a custom solution that targets only a relevant sample of transcripts and preciesly analase them leveraging graph manner in a parallel way. Our solution is tightly integrated with Bedrock, and leverages its components such as OpenSearch, Pinecone, Llama 3.2, SQL agents and much more. For better rate limits, we decided to go multi-regional.

- **Kuwait Financial House (KFH):** I designed the solution architecture for KFH in three variations based on their needs, build on open-source technologies such as Apache Kafka, Flink, NiFi, HBase, MinIO, Apache Iceberg, Trino (Startburs), Apache Spark, Apache Airflow, JupyterLab Kubeflow, MLFlow, Apache Ambari, YARN, Zookeeper, Ranger, Kerberos etc. I included also technologies further enhancing the solution and supporting GenAI and LLM use-cases such as Pinecone, AirByte, dbt, Dataiku, Seldon etc. Two version were leveraing Cloudera Data Platform (CDP) such as Cloudera Data Flow, Data Engineering, Data Lake, Warehousing, Machine Learning, SDX. I calculated the hardware requirements, incl. GPU rack for LLM inference and training consisting of several NVIDIA A100 80GB graphic cards, and the costs and human resources associated with it. The client was impressed by our technical expertise, and I flew to Kuwait to meet the client in-person! My role was instrumental in increasing our chances to secure a multi-million contract with a long-term client.
- **Manufacturing Vertical:** I established this vertical based on the current competencies in the team upon my arrival, and the maturity curve of the company. Our clients are Orlen, US Steel, TRZ (major industry company in the Czech Republic).

LLM EXPERT, PriceWaterhouseCoopers (PwC), Prague

September 2023 — September 2024

As a part of Application Development (App Dev) within PwC, I specialize on designing and implementing enterprice-grade solutions based on Large Language Models (LLMs). I take a part in pre-sales activities such as RFPs and other events hosted or participated by PwC such as hackathons, conferences etc. I serve as a senior person and expert on Gen AI, AI Architecture, AI Governance and System Design.

Since PwC is a partner to Microsoft, most of our solutions are build around MS Azure such as DevOps, Azure OpenAI, AI Studio, Cognitive Services etc.

Select experience examples:

- **AIVA2** - I spearheaded the development of AIVA2, an advanced iteration of our intelligent assistant system. Building upon the success of its predecessor, AIVA2 incorporates cutting-edge technologies to deliver enhanced performance and capabilities. At its core, AIVA2 utilizes the CrewAI framework for sophisticated agent orchestration and planning, while seamlessly integrating powerful tools from LangChain, LlamaIndex, and Composio. The system supports multiple inference methods, including PwC's Azure OpenAI instance, Ollama, and Groq, enabling the use of state-of-the-art language models such as GPT-4 and Llama 3.1. For comprehensive crew monitoring and performance analysis, I implemented AgentOps. The deployment leverages Knative on Kind, facilitating a serverless and event-driven architecture with Kubernetes.
- **LeanIX:** As the lead and expert engineer, I directed the integration of data from the SAP LeanIX application into a local Neo4J graph database. I implemented a user interface feature in Streamlit that includes quick access buttons for frequently asked questions. Additionally, I developed a sophisticated query-matching system using embeddings and cosine similarity calculations against a PostgreSQL database enhanced with the PGVector extension. This system evaluates user prompts, and if they don't meet the defined threshold, a language model agent dynamically composes the Cypher query for fallback scenarios. The responses, along with their respective Cypher queries, are presented to users for validation. Approved queries are then staged for review before inclusion in the blessed query database. I also integrated LangSmith for observability and LangServe as the backend, ensuring comprehensive system monitoring and robust backend support.

- **Smart Reporting:** I spearheaded the design and implementation of an advanced Retrieval-Augmented Generation (RAG) solution based on LangChain in the Legal & Tax domain. The solution retrieves and answers information from documents using both external inference via AzureOpenAI API and local inference with Llama3-8b-instruct-q4 on Ollama. I employed a self-querying technique for metadata enrichment, significantly improving recall. Depending on the type of inference chosen, the vector-search database used was either FAISS or Azure CosmosDB with PGVector. Evaluation was conducted over custom datasets using RAGAS and LangSmith.
- **Artificial Intelligence Virtual Assistant (AIVA):** I was the driving force behind designing and implementing an agentic virtual assistant leveraging LangChain and PwC OpenAI inference on Azure. AIVA was designed as a universal framework with modularity and flexibility in mind, orchestrating agents for several different tasks. Those use-cases included data retrival from SharePoint, Google Drive, a structured database, real-time information on the Internet, information from the official PwC website and much more. I implemented LangSmith for trace-catching and monitoring, and LangServer for the backend mode.
- **RFP - Customer Support Analytics for Raiffensen Bank:** I was pivotal in architecting and preparing a proposal for a comprehensive text and voice analytics platform on Amazon AWS, including AI Governance. The platform was designed for feature extraction to enhance downstream risk and marketing models. Key components included AWS Transcribe for voice call transcription and Bedrock with Anthropic's Claude LLM for feature extraction.
- **Knowledge Sharing**: I spearhead the knowledge sharing within our team on bi-weekly basis, showcasing sever key and interesting concepts in the LLM area. This includes fine-tunning of Large Language Models, destiling LLMs such as GPT4 into GPT2 for specific tasks and effecient inference, advanced prompting techniques and much more. This helps to faciliate the know-how in the company and key business stakeholders have the opportunity to gain valuable technical perspective.
- **Microsoft Envision:** I participated in the Microsoft Envision conference at Prague's O2 Universum on behalf of PwC. For the hands-on Build Day, I developed a Retrieval-Augmented Generation (RAG) system running locally on a Raspberry Pi 5. This system utilized Ollama and Microsoft's Phi-2 Large Language Model to showcase our capabilities and engage clients and conference visitors effectively.
- **ESET Hackathon**: As a mentor and co-organizer, I participated in a collaborative initiative between Microsoft and PwC for ESET, a cybersecurity company. The event focused on knowledge transfer through presentations, workshops, hacking sessions, and networking. Around 10 use-cases were tackled by developers, engineers, and scientists. The team I led aimed to prototype a solution that would suggest text corrections based on the company's comprehensive style guide, which encompassed many documents. To achieve this, we fine-tuned LLama3 using PEFT (Parameter Efficient Fine-Tuning) and QLoRA adapter, ensuring the model accurately adhered to the style guide's standards. We used Flesh-Kincaid readability score to evaluate the quality of the output.

GENERATIVE AI ARCHITECT, Carvago (AEG), Prague

April 2024 — September 2024

At Carvago, my responsibilities encompass the identification and evaluation of generative AI and Large Language Models (LLM) use cases aimed at enhancing operational efficiency and driving sales growth. I meticulously assess the required time and resources for implementation, prioritizing these projects based on their business value. My role involves the development of Minimal-Valuable-Product (MvP) projects, as well as guiding the company towards a strategic, long-term framework for leveraging Large Language Models and generative AI technologies.

Additionally, I facilitate the dissemination of knowledge across the international group, promoting best practices to augment employee value to the organization. Operating in a pivotal position, I report directly to the CTO and CEO, maintaining close collaboration with other C-level executives and key stakeholders.

Selected project experience:

- **Car Audit Report:** I lead a high-visibility project and managed upto 4 people regarding development of application collecting technical information about a car from a technician, and generating an audit report for a customer with added business and sales value. The solution was based on LangChain, leveraging OpenAI API model GPT-4o and LangGraph for cyclic dialog design. Front-End was realized in Streamlit and Back-End in LangServe that builds on top of FastAPI and Uvicorn. I catched the traces via LangSmith and the users had the ability to submit feedback via a form that was stored in PostgreSQL database. Integration to internal systems leveraged RESTful API. The entire solution was containerized using Docker, enabling microservices architecture. Implementing the solution increased total sales revenue by 20%, and saved significant amount of time on human resources in logistics and operations.
- **Smart-Search:** I spearheaded the design and implementation of a hybrid-search solution, enabling customers to search a database of thousands of cars based on semantic meaning rather than specific keywords and parameters. The solution was built using LangChain and leveraged Snowflake Cortex Functions for embeddings. Snowflake's LLM model, Arctic, generated human-like descriptions of cars based on parameters scraped from sellers and the Internet. These descriptions were matched against encoded embeddings from the prompt using cosine similarity to return the most relevant results. Additionally, users could filter cars by price, making this a comprehensive hybrid-search solution.
- **Speech-2-Text Translation & Analytics**: I prototyped a solution using Amazon AWS to process call center recordings. The recordings were loaded from an S3 bucket into AWS Transcribe for transcription. The transcribed text was then processed by the OpenAI API for translation, summarization, and sentiment analysis.

Data Science & AI Engineering Mentor, Software Development Academy, Remote

January 2022 — September 2024

I specialize in instructing Deep Learning and advanced data science subjects, including CNN, RNN, NLP, LLM, Attention mechanisms, and Transformer architecture. I lead interactive sessions and design curriculum tailored to students with diverse levels of expertise. I mentor students on the practical application of data science techniques, ensuring they build a strong foundation for their professional development. Additionally, I conduct personalized Python interviews to prepare candidates for data science positions, and offer constructive feedback on coding, problem-solving, and communication skills. I consistently receive excellent feedback from my students and I'm a host to public events that aim to raise brand awareness.

Selected project experiences:

- **Study pre-dispositions**: I developed a comprehensive assessment to evaluate the logical and structured thinking abilities of students for their suitability in a Data Science and AI Engineering course. This initiative significantly reduced student dropout rates, enhanced learning efficiency, and improved overall student satisfaction.

- **Knowledge Tests**: I translated, corrected, and enhanced the knowledge assessments administered to students during the Data Science and AI Engineering course. My improvements ensured the tests accurately evaluated student comprehension and knowledge retention, leading to better academic outcomes and more precise identification of areas needing further instruction.
- **Syllabus Development**: I actively contributed to the creation and refinement of the course syllabus, providing valuable feedback and insights to ensure the curriculum met educational standards and effectively addressed the learning needs of students in the Data Science and AI Engineering course.
- **Brand Awareness**: I served as a keynote speaker at several company public appearances, including the IT Festival, to raise brand awareness and attract new talent. My presentations highlighted the company's innovative projects and work culture, effectively engaging the audience and enhancing the company's reputation within the tech community.
- **Personalized Training and Interviews**: I conduct tailored training sessions and test interviews in various technical domains, including Python, Machine Learning, Data Science, Convolutional Neural Networks (CNN), and Natural Language Processing (NLP). These sessions are designed to evaluate and enhance candidates' skills, ensuring they meet the high standards required for success in these fields. My approach focuses on identifying individual strengths and areas for improvement, providing targeted feedback and support to help candidates excel.

SENIOR DATA SCIENTIST & ENGINEER, Ernst&Young, Prague

March 2023 — June 2024

At the big four company, my role predominantly extended into Solution Architecture and System Design, where I played pivotal roles in several projects. I was responsible for designing and calculating solutions for clients across various industries and actively participating in pre-sales engagements. Additionally, I mentored junior team members, providing guidance and assigning tasks to foster their professional growth and enhance their value to the company. This mentorship included technical coaching, best practices in system design, and hands-on support to ensure their successful integration into project teams. My contributions were instrumental in delivering robust and scalable solutions.

Selected project experiences:

- **Data Governance**: I created the proposal (incl. pricing) for Data Governance in CEPS, and energy sector provider, based on which we succeeded in securing the opportunity. I continued to play a pivotal role throughput the whole project which extended to selecting a Data Catalog for the operator. I analyzed the current architecture and company processes in Sparx Enterprice Architect (EA) and in other systems, in supplementation to frequent meetings. Based on the inputs, I designed the standards, policies, roles and data domains as a part of the governance, which were consistently communicated with the client. Additionally, I concluded a throughout analysis of all available data catalogs on the market and evaluated roughly 25 features for each. I performed a dimensional reduction on top of the comprehensive multidimensional matrix with evaluation criterias, to create a readable representation I called "Analytical Cube". I was systematically in touch with the key stakeholders such as enterprice architecture, cybersecurity, data science and analytics etc.

- **A Large Analytical Cluster CRANE**: For Austrian ERSTE, a major banking group in the region, I played a leading role in resolving performance and stability issues on a 10 TB Cloudera Distributed Hadoop (CDH) cluster. This involved a deep dive into the cluster's architecture and operations, identifying bottlenecks, and implementing optimizations. Key technologies used included Hadoop, HDFS, Hive, Impala, and PySpark. I optimized PySpark applications, adjusted resource allocation via YARN, and configured HDFS for improved data throughput. Additionally, we implemented performance tuning on Hive and Impala queries, and created fine-tunned Spark session configuration that resulted in performance gains of two orders of magnitude. The project also required extensive use of Cloudera Manager for monitoring and managing the cluster's performance and stability.
- **Solution Redesign for Eurostat Data Integration:** I reverse-engineered a malfunctioning production solution that failed due to poor design and completely rewrote it using Python and PySpark to ensure it adhered to best practices. The objective was to download information from Eurostat via API to enrich downstream risk models. Key technologies and practices included Python for scripting the API interactions and data processing, and PySpark for efficient distributed data processing and transformation. I integrated the Eurostat API for automated data retrieval and implemented robust error handling and comprehensive logging to ensure reliability and maintainability. Additionally, I incorporated data validation and cleaning steps to maintain data quality.
- **Advanced Analytical Cluster Optimization**: As part of Project CRANE, I conducted a series of workshops providing architectural consultations to the client on the solution design of their 10 TB Hadoop and Cloudera distributed big-data analytical cluster. The primary challenges addressed included optimizing large data transfers between HDFS systems, orchestrating ETL and other data processing jobs, and implementing robust monitoring and logging solutions. From an architectural perspective, I provided strategies to enhance data replication and network configurations to optimize large-scale data transfers. I proposed the adoption of Apache Airflow for orchestrating ETL processes and other data jobs, facilitating efficient scheduling, execution, and monitoring of workflows. For comprehensive monitoring and logging, I recommended the ELK stack (Elasticsearch, Logstash, Kibana) in combination with Grafana and Prometheus. This architectural setup provided real-time performance metrics, alerting, and visualizations, ensuring system stability and quick issue resolution. This architectural approache resulted in significant improvements in data transfer effeciency, streamlined the processing workflows, and enhanced system monitoring capabilities, ensuring the client analytical cluster operated at peak efficiency.
- **AI Conference of Tomorrow:** I actively participated in co-organizing a conference hosted by EY, which focused on Artificial Intelligence, Blockchain, and Virtual and Augmented Reality for our current and prospective clients. This conference aimed to showcase cutting-edge technologies and their applications, fostering engagement and collaboration with industry leaders and innovators. My involvement included planning sessions, coordinating with speakers, and ensuring the event provided valuable insights and networking opportunities for all attendees.

AI ARCHITECT CONSULANT, FLSmidth, Copenhagen

January 2024 — March 2024

At FLSmidth, a leading industrial enterprise based in Denmark, my principal responsibilities within my role were centered on the conceptualization and design of system architectures specifically tailored for generative AI applications. Additionally, I played a pivotal role in the formulation and development of AI Governance guidelines, ensuring robust and ethical management of AI

operations. This engagement was structured as a predefined three-month contract, during which I contributed to enhancing the company's strategic initiatives in AI implementation and governance.

Big Data Scientist, Adastra Digital, Prague

January 2021 — March 2023

As a Senior Data Scientist at Adastra Digital, I took a part in numerous projects as below, both with clients within and outside the European Union. I took the lead in establishing and strengthening the internal expertise in Databricks, delivering seminars and presentations on Spark and Databricks in Data Science at the University of Economics. Additionally, I successfully implemented an automated screening system for evaluating potential candidates and developed comprehensive test questions in the areas of data science, Spark, and Python.

Our tech. stack was mostly based on Apache Hadoop ecosystem and Cloudera.

Projects:

1. **TRENCH (KB Bank - Societe Generale Group)*:*** For a major banking institution, I was responsible for developing an advanced analytic solution using Hadoop to classify payment transactions in real-time and predict clients' end-of-month account balances. The solution was built on the Cloudera Data Platform (CDP) for core operations, utilizing Apache Kafka for real-time messaging and HDFS for data storage. I implemented the ELK stack (Elasticsearch, Logstash, Kibana) along with Prometheus and Grafana for comprehensive monitoring and logging. Task scheduling was managed with Apache Airflow, and automation processes were facilitated by Apache Jenkins. Model development was carried out using PySpark, with production models deployed in Scala, ensuring robust and scalable performance.
2. **Valeo (Autonomous Vehicle Data Infrastructure Design)**: As a Business Analyst and Data Infrastructure Designer, I collaborated on designing scalable infrastructure for processing autonomous vehicle data. I translated business requirements into technical specifications, ensuring seamless communication with the client. I implemented data processing solutions using Apache Hadoop and Apache Spark, and integrated real-time data ingestion with Apache Kafka. My responsibilities included developing ETL processes, maintaining data schemas, and optimizing workflows for high-throughput data streams. Additionally, I ensured data security compliance and provided comprehensive documentation and training for client teams.
3. **TBC - Major Bank in Georgia**: As a Data Owner and Data Science Consultant, I developed a default scoring engine using LightGBM and an anti-fraud scoring engine with IsolationForest. I maintained close communication with the business to ensure alignment with their needs and objectives. My role involved data preprocessing, feature engineering, model training, and validation. I also implemented the models into production, monitored their performance, and adjusted parameters to optimize accuracy and efficiency.
4. **GRC - Customs Administration of the Czech Republic**: Contributed to a pilot project for migrating data to the cloud, enhancing data accessibility and scalability. I performed BI analytics to optimize customs processes and support data-driven decision-making. My responsibilities included developing ETL pipelines, ensuring data integrity during migration, and creating insightful reports and dashboards to provide actionable intelligence for operational improvements.
5. **Coderbyte**: I spearheaded the development of test questions for assessing new candidates for positions such as Data Scientist, Data Engineer, Data Analyst, Big Data Analyst, and Solution Architect. My role involved designing comprehensive assessments to evaluate candidates' technical skills and problem-solving abilities, ensuring the selection of highly qualified professionals for these roles.

Data Scientist, TIP IT Solutions, Prague

July 2019 — January 2021

Projects:

1. **Virtual Banking Assistant KATE: (*CSOB Bank - KBC Group):*** As a NLP Engineer with the core team, I significantly contributed in the development of the major banking virtual assistant KATE. Working within the core team for a large banking group, I developed a solution based on the RASA framework. My responsibilities included leveraging a transformer-based dual entity extraction and intent classification model, creating Czech word embeddings, and developing custom NLP components. The core application was developed in Python, incorporating unit and smoke testing to ensure reliability. The project utilized a range of technologies including RASA, Transformer models, Scikit, SpaCy, NLTK, Python, Docker, and AWS.
2. **Virtual Banking Assistant GEORGE - Ceska Sporitelna Bank (ERSTE Group)**: Analyzed data from GEORGE, the virtual banking assistant, to classify user feedback and identify areas for improvement. Leveraged data-driven insights to optimize the assistant's performance, ensuring it met customer needs and expectations. Regularly communicated with stakeholders to report on GEORGE's progress, key performance indicators, and proposed future enhancements.
3. **Keboola & Databricks - Ceska Sporitelna Bank (ERSTE Group)**: Collaborated on a strategic initiative to integrate JIRA system data with an existing data warehouse solution, designing and supervising the seamless implementation of the technical solution. Coordinated and managed cross-functional teams, fostering effective collaboration. Ingested data into Databricks and performed analysis within the environment. Ultimately, implemented Keboola as the central platform for business analysis and data science.

Data Engineer, ProfiCredit, Pardubice

June 2017 — June 2019

- Primarily focused on developing and maintaining a global data warehouse solution, ensuring its performance and scalability.
- Conducted data analytics to derive insights and improve decision-making processes within the organization.
- Developed a Default Score Engine for a financial institution, which automatically assigned a financial risk score to clients based on their predicted ability to repay a loan.
- Leveraged data-driven methods and advanced analytics techniques to enhance the accuracy and effectiveness of the Default Score Engine.

BI Developer, Mikroelektronika, Vysoke Myto

May 2016 — May 2017

- Developed efficient SQL Stored Procedures to handle complex data processing tasks and optimize database performance.
- Created a wide range of reports to provide actionable insights for clients and stakeholders, enabling informed decision-making.
- Utilized MS SQL Server and SSRS (SQL Server Reporting Services) to develop, deploy, and maintain reporting solutions tailored to specific needs.

PCB Designer, AWOS, Pardubice

October 2015 — May 2016

- Designed Printed Circuit Boards (PCBs) for a diverse clientele, meeting unique specifications and requirements.

- Had the privilege of designing the main boards for the first European electric motorbike, based in Austria, showcasing expertise in innovative and sustainable solutions.
- Collaborated with clients and engineering teams to ensure that the PCB designs met performance, reliability, and manufacturability standards.

Education

Bachelor, Brno Technical University, Brno

September 2012 — August 2015

- Studied microprocessor design, gaining a solid foundation in the principles and techniques related to the design and development of microprocessors.
- Acquired a comprehensive understanding of electrical engineering concepts, preparing for a successful career in the industry.
- Engaged in practical, hands-on learning experiences to further develop technical skills and real-world problem-solving capabilities.

Certifications

Generative AI Fundamentals, Databricks

May 2024

Azure AI Engineer, Microsoft

April 2023 — April 2024

Azure Data Scientist Associate, Microsoft

September 2022 — September 2023

Developer for Apache Spark 3.0, DataBricks

February 2022 — February 2024

Data Science Workshop, Adastra Digital

November 2018

Implementing DWH Solution, GOPAS

December 2017

Research Publications

Classifying Medical Diagnosis using LLMs, MEDOSFT, 1.LF UK (Charles University)

December 2023 — January 2024

The main problem of digitalization of health records is the unstructured nature of many medical records, which are often recorded as free text, complicating their machine processing. During the CEE Healthcare Hackathon 2023 event at Prague's IKEM (Institute of Clinical and Experimental Medicine), the possibility of converting digital unstructured records into a structured form using artificial intelligence was explored. A prototype was developed that uses AI methods for categorical evaluation of messages and allows querying in natural language. The key method for text analysis was the use of pre-trained GPT (Generative Pre-trained Transformer) models from OpenAI, which were instructed to identify specific diagnoses, such as diabetes mellitus.

Competitions

OpenAI GPT3 Hackathon, 1st place

June 2022 — June 2022

We desgined a solution that was scrapping web sources such WebMD, Reddit etc. and recommending the best course of action based on class classification from the image encoder. My role as a senior data scientist and AI engineer was pivotal, and more specifically focused on the synthesis part from OpenAI API GPT3 model. Our solution secured to top (1st) place out of numberous teams and hunderdens of participants.

PragueAI, special prize

May 2023

We recieved a special award by prg.ai for our project of virtual asistent for city planing. The solution was based on RAG architecture from public documents and was attempting to leverage a virtual asistant using D-ID as the interface. My role was rather in solution architecture and consulation than in development.

CommerzBank Hackathon, winner

May 2022

For this internal hackathon competition hosted by Czech branch of CommerzBank, I won the first place. I perfomed low-rank matrix factorization using Singular Value Decomposition (SVD) to compress word embeddings and hence cope around memory limilations for the purpose of effective NLP over large base of text.

HackToTheMoon, 1st place in category Security&Privacy

October 2019

In the category "security and privacy" we secured the first place at the on-site hackathon sponsored by Rockaway Blockchain, Binance and other major companies in the space. We created a Dapp (decentralized application) leveraging Oasis framework and Rust to create a robust, transparent and secure voting system employing the Janecek voting method. We estimated the immediate costs saving to 97.5%. I played a pivotal role in both developing the Dapp and designing the solution.

ProfiReal Hackathon, winner of prize money

May 2019

In the context of this international internal hackathon, I won the first place out of the entire group. My solution implemented default score engine into the data engineering jobs to automatically assign a financial calculator to a client based on the probability of loan default.

Technologies

Linux

Git

T-SQL

Python

PySpark

NLP (Nature Language Processing)

DataBricks

Microsoft Azure

Amazon AWS

GCP

Hadoop

MLFlow

Kubernetes

TensorFlow

ETL

Reinforcement Learning

Matplotlib / PyPlot / Seaborn

Spark

Data Governance

Nature Language Processing

Large Language Models

System Design

Azure Cognitive Services

Hadoop

Apache Kafka

ELK Stack (Elasticsearch, Logstash, Kibana)

Prometheus

Apache Jenkins

Scala

Apache HBase

RASA Framework

Word Embeddings

Docker

BERT (Bidirectional Encoder Representations from Transformers)

SpaCy

FastText

OpenAI API

IsolationForest

Multi-Agent Systems

Game Theory

Jupyter

Warehousing

Apache Airflow

Lanchain

Solution Architecture

MS Azure

Azure OpenAI

Cloudera Data Platform (CDP)

HDFS

Grafana

Apache Airflow

PySpark

Apache Hive

Apache Spark

Transformer-based Models

Amazon AWS

Kubernetes

GPT (Generative Pre-trained Transformer)

NLTK (Natural Language Toolkit)

Hugging Face Transformers

Gradient Boosting Machine (GBM)

Spark Cluster

MLFlow

Scikit-learn

Keras

Jupyter Notebooks

NumPy

Seaborn

EU AI Act

Fairness, Accountability, and Transparency (FAT) Principles

Model Risk Management (MRM)

Large Language Models (LLMs)

Attention Mechanism

Event-Driven Architecture

Data Governance Frameworks

Data Lineage

Master Data Management (MDM)

GDPR Compliance

AzureOpenAI API

Self-querying Technique

Azure CosmosDB with PGVector

LangSmith

Semantic Search

Knowledge Graphs

Parameter Efficient Fine-Tuning (PEFT)

Azure Functions

Python

TensorFlow

PyTorch

Pandas

Matplotlib

AIGA Framework

Ethical AI Guidelines

Model Risk Management (MRM)

Explainable AI (XAI)

Transformer Architectures

Model Optimization Techniques

RESTful API

Data Catalogs

Data Quality Management

Data Privacy and Security

LangChain

Ollama

FAISS (Facebook AI Similarity Search)

RAGAS

Vector Databases

Document Retrieval Systems

Inference Engines

QLoRA

Azure LogicApps

Azure Cognitive Services

Azure Databricks

Azure Synapse Analytics

Azure Kubernetes Services (AKS)

Azure Data Factory

Azure Stream Analytics

Azure HDInsight

Azure Data Lake

Power BI

Languages

Czech

English

German