

# Yurii

## Middle Data Scientist

### SUMMARY

- Data Scientist with 4+ years of experience in AI and machine learning; - Specialized in NLP, time series forecasting, and generative AI; - Built RAG systems using OpenAI, Langchain, and custom pipelines; - Developed multi-agent systems for sales enablement, education, and virtual assistants; - Proficient in Python, SQL, and ML libraries like Pandas, Sklearn, Keras, and PyTorch; - Created legal assistants with HuggingFace models and citation-based RAG responses; - Built voice-based chatbots using OpenAI Whisper and ElevenLabs for voice cloning and audio processing; - Designed pipelines for text-to-speech and image generation in mobile and cloud environments; - Extracted and analyzed financial data using AWS Textract and OpenAI VLMs; - Built production-ready support bots using Dialogflow CX, Twilio, and Google Firestore; - Experienced with AWS and GCP for scalable model deployment.

### TECHNICAL SKILLS

<b>Main Technical Skills</b>	Python, SQL, NLP, Gen AI, RAG
<b>AI Tools &amp; Assistants</b>	Claude, Gemini, Gemini 3, Grok
<b>Programming Languages</b>	Python
<b>AI &amp; Machine Learning</b>	AWS Textract, Claude, DialogFlow, Docling, ElevenLabs, Gen AI, Grok, Hugging Face, Keras, LangChain, LangGraph, NLP, NumPy, OpenAI, PyTorch, RAG, Scikit-learn, Spacy, Whisper
<b>Python Libraries and Tools</b>	Gensim, Keras, Matplotlib, NLTK, NumPy, Pandas, PyTorch, Scikit-learn, Seaborn
<b>Data Analysis and Visualization Technologies</b>	Pandas
<b>Databases &amp; Management Systems / ORM</b>	MySQL, Oracle Database, PostgreSQL, Sphinx, SQL, SQL queries
<b>Cloud Platforms, Services &amp; Computing</b>	GCP
<b>Amazon Web Services</b>	AWS Textract
<b>Google Cloud Platform</b>	Google Cloud Pub/Sub
<b>Deployment, CI/CD &amp; Administration</b>	Active Directory, Flux
<b>SDK / API and Integrations</b>	API, GraphQL, Twilio
<b>Operating Systems</b>	Unix

<b>Web/App Servers, Middleware</b>	WildFly
<b>Other Technical Skills</b>	Currency Cloud, Loco Translate

## WORK EXPERIENCE

### Data Scientist, Multi-Agent Assistant for Sales Enablement

**Duration:** 6 months

**Summary:** Developed a multi-agent assistant for sales enablement, capable of autonomously preparing lead summaries, company/industry reports, and meeting agendas.

**Responsibilities:**

- Designed and implemented a multi-agent system that used Tavily API for intelligent web search and synthesis;
- Automated generation of sales research reports (for person, company, industry) and meeting agendas based on lead data and found information;
- Built a knowledge base system allowing users to upload files or provide website links for ingestion and querying;
- Developed a RAG interface to let users interact with their documents and websites in natural language.

**Technologies:** Python, OpenAI, Tavily, Docling, pgvector, SQL, Langchain.

### Data Scientist, Legal Assistant using RAG System

**Duration:** 3 months

**Summary:** Legal assistant powered by a RAG system for U.S. constitutional and federal law. Enabled users to ask complex legal questions and receive accurate, citation-based responses grounded in the U.S. Constitution, federal law, and New York State regulations.

**Responsibilities:**

- Prepare dataset by parsing and embedding legal documents using text embeddings model;
- Researched and experimented with different text embedding models from HuggingFace to optimize performance and quality;
- Implemented a process for intelligent query reformulation step using RAG to improve retrieval accuracy based on the user's intent.

**Technologies:** Python, HuggingFace, Langchain, OpenAI, AWS.

### Data Scientist, Virtual Agent for FAQs and Ticket Logging

**Duration:** 1 year

**Summary:** Create a virtual agent for FAQs and logging tickets on TechSupport system. Extract names, mails, phone numbers, issues and details during the conversation fill the ticket on system or send self-guide.

**Responsibilities:**

- Design conversational flow and develop a chatbot for realistic conversation;



- Create a system for managing the connection to the TechSupport system for creating tickets;
- Extract detailed information from conversation and post-process conversational records and text;
- Implement issues classification system for smarter assignment of tickets on the TechSupport side;
- Integrate OpenAI Whisper to improve sound processing.

**Technologies:** Python, Dialogflow CX, Spacy, NLTK, Twilio, Google Firestore, OpenAI API, GCP.

### **Data Scientist, DriveED**

**Duration:** 6 months

**Summary:** Developed a multi-agent generative AI system to automate the creation of lesson plans aligned with U.S. educational standards.

#### **Responsibilities:**

- Designed a modular pipeline for generating lesson plans, integrating curriculum standards, success criteria, and textbook-based task generation;
- Implemented multi-agent orchestration for concept-based task design, visual asset generation, and final lesson assembly;
- Tuned agents to adapt lessons by grade level and complexity;
- Integrated RAG to access relevant textbook content and success criteria dynamically;
- Contributed to task formatting and refinement to meet pedagogical goals and improve classroom usability.

**Technologies:** Python, OpenAI, Langchain.

### **Data Scientist, Interpret AI**

**Duration:** 6 months

**Summary:** A mobile app designed for recording and interpreting users' dreams through a psychoanalytic lens. The bot engages users in interactive discussions to gather insights into their dreams and take into account their life context, aiming to provide meaningful interpretations.

#### **Responsibilities:**

- Crafting a natural dialogue flow to mimic friendly and insightful real-time conversations;
- Building a system for collecting and analyzing detailed information on dreams, and providing interpretations rooted in Jungian theory;
- Set Up ElevenLabs with cloning and tuning;
- Upgrade pipeline for real-time Text to Speech;
- Integrate Image Generation pipeline with Flux for visualization of Dreams;
- Image generation with Flux.

**Technologies:** OpenAI, Flux, Python, ElevenLabs, Grok, Claude, langchain, AWS.

### **Data Scientist, Tech Cargo System**

**Duration:** 1 year



**Summary:** The project focused on analyzing and extracting financial information from documents to assess the financial health of companies. Supported multilingual and multi-format document processing.

**Responsibilities:**

- Leveraged AWS Textract for automated data extraction from financial documents, including balance sheets, income statements, and cash flow statements;
- Designed workflows for financial health analysis using custom formulas and algorithms;
- Integrated OpenAI Visual Language Models for testing advanced data extraction capabilities;
- Built multilingual pipelines using AWS APIs;
- Enhanced support for diverse document types, ensuring scalability and robustness.

**Technologies:** Python, AWS Textract, AWS Textract Queries, OpenAI API (Vision and Text), AWS Translate, AWS Currency Converter.

**EDUCATION**

- **Master of Computer Science**
- **Bachelor of Computer Science**

