Yuliia K.

**AI/ML Engineer /
Back-End Developer**

Summary of Qualifications	AI/ML Engineer and Back-End Developer with 3 years of experience designing scalable data-driven systems, modular multi-agent platforms, and LLM-powered conversational agents and chatbots. Proficient in building RESTful APIs with FastAPI and Pydantic, integrating SQL/NoSQL databases, and implementing asynchronous, production-grade back-end architectures. Skilled in machine learning, prompt engineering, and deploying solutions on AWS/GCP using Docker.

Skills

Programming Languages/Technologies

- Asynchronous Programming
- XML/JSON/YAML
- HTML/CSS
- SQL
- Python

RDBMS

- MySQL
- PostgreSQL
- Alembic
- SQLite

NoSQL

- Redis

Virtualization Tools

- Docker / Docker Compose

Methodologies

- Agile, Scrum, Kanban
- Pair Programming

Operating Systems

- Microsoft Windows
- Debian/Ubuntu/ Arch/Linux

Versions Control

- Git
- GitHub
- GitLab

Frameworks/Libraries

- FastAPI
- SQLAlchemy
- OpenAI (Whisper, GPT)
- Numpy/Pandas
- Requests
- Aiohttp
- Matplotlib
- Seaborn
- Pytorch
- Pillow
- Boto3
- Streamlit
- Tensorflow
- Scikit-Learn
- LangChain
- LlamaIndex
- Whisper
- Sentence Transformers

Cloud Providers

- AWS

Development Tools

- Visual Studio
- PyCharm

Testing Tools

- Pytest
- Unitest, Mock

Experience

AI Voice Verification Assistant

Project Description:	Developed an automated outbound-calling system to confirm the person's employment status. The solution detects voicemail, IVR systems, and other bots to ensure the call flow interacts only with real humans. It uses intelligent dialogue understanding to identify IVR options and guide the call through the correct path. A retrieval-based approach supports dynamic conversation flows by referencing similar past dialogues stored in a vector

database. The system integrates language models and embeddings to enhance call accuracy, decision-making, and overall reliability.

Domain:	AI/ML | STT | TTS | LLM | RAG
Involvement Duration:	0.5 years
Project Role:	AI/ML Engineer
Responsibilities:	- Implemented retrieval-augmented generation (RAG) pipelines using LLMs and vector search to drive dynamic and context-aware conversation flows for the voice bot; - Developed and integrated embedding-based similarity search using AWS Bedrock and pgvector to match ongoing calls with the most relevant historical dialogues; - Built LLM-powered components for IVR detection, voicemail/bot classification, and generation of structured call summaries; - Implemented Python best practices by adding type hints, defining request/response schemas, asynchronous request handling, logging.
Project Team Size:	6-7 team members
Tools & Technologies:	Python, FastAPI, Pydantic, Asyncio, Groq API (LLM), AWS Bedrock & Sentence-Transformers embeddings, pgvector, RAG pipeline.

Hotel Security System

Project Description:	Developed a voice bot receptionist for a law firm that handles incoming calls, verifies service requests using RAG, and assists in booking appointments. The bot collects and confirms contact information, including names and email addresses, and can process uploaded documents to generate real-time summaries. It interacts naturally with callers, answers follow-up questions, and concludes conversations efficiently.
Domains:	AI/ML | Computer Vision
Involvement Duration:	1 year
Project Role:	AI/ML Engineer
Responsibilities:	- Designed and developed AI models for human detection, tracking, and re-identification across multiple camera feeds; - Implemented face recognition algorithms for identifying and verifying individuals in elevator camera systems; - Built and optimized deep learning pipelines using frameworks such as TensorFlow / PyTorch for efficient video analytics; - Conducted data preprocessing, augmentation, and annotation workflows to improve model robustness and accuracy; - Performed model evaluation, tuning, and continuous improvement using real-world data from hotel surveillance cameras.
Project Team Size:	3-4 team members
Tools & Technologies:	Python, Pydantic, PyTorch, TensorFlow, YOLO family models, DeepSORT, CNNs (ResNet, OsNet), feature embedding, FaceNet, Vector databases / Similarity Search.

Voice Receptionist

Project Description:	Developed a voice bot receptionist for a law firm that handles incoming calls, verifies service requests using RAG, and assists in booking appointments. The bot collects and confirms contact information, including names and email addresses, and can process uploaded documents to generate real-time summaries. It interacts naturally with callers, answers follow-up questions, and concludes conversations efficiently.
Domain:	AI&ML |STT | TTS | LLM | RAG
Involvement Duration:	1.5 years
Project Role:	Back-End Developer

Responsibilities:	• Designed and implemented a CRM service for managing contacts and scheduling appointments; • Developed LLM-integrated tools to enable automated interaction with the CRM system; • Built a scalable document upload and processing service using S3, enabling real-time document summarization; • Implemented conversation flow functionality to handle calls, verify requests, and guide interactions.
Project Team Size:	5-7 team members
Tools & Technologies:	Python, FastAPI, Pydantic, Asyncio, Whisper, OpenAI Realtime GPT-4 Mini, S3, GoHighLevel, RAG, Embeddings.

Education	**Bachelor's degree in Applied Computer Science** University of Bamberg

Languages	**English –** Fluent **German –** Fluent **Ukrainian –** Native **Russian–** Native