Peter Pollak

Full Stack Software Developer, Gen AI, Large Language Models, Machine Learning, NLP, Computer Vision

1220 Vienna, Austria
peter@exautomata.ai
pollak.peter89@gmail.com
Github: Syzygy2048

EXPERIENCE

Self Employed - *Founder & Developer* 2024 - Now

Build AI tools to automate processes, improve efficiency, and drive better decisions.

- Development of Automation and Personal Assistant tools involving closed and open source LLM systems as conversation, decision and reasoning engines involving function calling, MCP, RAG, Agents and other techniques. (See Project: PAIGE below).
- Freelancing

Artificial Researcher IT GmbH - *Sr. Software Developer (ML, Backend, Web)* 2019 to 2023

NLP based on rules and Machine Learning. Search Optimization.

- Managed the complete data lifecycle, including sourcing, extraction, analysis, model training, storage, and delivery.
- Train, deploy and use Machine Learning Systems, including gathering, selecting, and preparing training data.
- Creation, updating, and maintenance of cloud server infrastructure, encompassing VMs, Docker, Kubernetes, backend applications, APIs, databases, and websites, along with local hardware. Similar tasks for international business clients.
- Implemented comprehensive policies and systems covering security, workflows, deployment, version control, authentication and billing.
- Helped implement SCRUM-based agile workflows to the company.
- Supervise PhD. Students in their NLP related projects.

Parkbob GmbH - *Software Developer (Android)* 2016 to 2019

App development for Parking Regulations and Parking state detection.

- Update, extend, maintain, refactor and test the app and SDK along with the implementation of several secondary apps.
- Communicated requirements to external contractors for the iOS app.
- Delivering SDK to partners like A1 HandyParken and Gelbe Seiten.
- Optimized/Automated KPI Pipeline
- Completely rewrote the core functionality to meet changing Android platform requirements.

mySugr GmbH - *Software Developer (Android)* 2014 to 2016

App development focused on diabetes management, certified as a Class I medical device.

- Helped grow the company from 100.000 to 500.000 users in the limited market of Type 1 diabetics.
- Introduced SE/DevOps techniques like CI, Unit Tests, A/B Testing, Analytics, Feature Switches, Dependency Injection.

PROGRAMMING LANGUAGES

- Python - Proficient
- JavaScript - Prof.
- Java - Prof.
- C++ - Interm.

TOOLS & TECHNOLOGIES

- Haystack-AI
- ChatGPT, Gemini, Gemma, Llama, Ollama, OpenSource LLMs,
- MCP, RAG, Agents, Swarms
- Kubernetes, Docker, Microservices
- REST, JSON, Flask, FastAPI, Express
- spaCy, NLTK, React
- TensorFlow, PyTorch
- GIT, Jira, Confluence
- Google Cloud & APIs
- Android, Multiplat.
- SQL, NoSQL, Elastic & OpenSearch Stack, Vector DB
- Gradle, Maven, Ant
- Unit Tests, Feature Switches, A/B Testing CI Deployment
- Windows, WSL, Linux

SKILLS & SPECIALIZATIONS

- Full Stack, DevOps, DataOps
- NLP, AI, GenAI
- Visual Computing
- Machine Learning
- Pattern Recognition
- Data Analysis
- Cloud Computing
- Agile/SCRUM

Workflows

EDUCATION

Vienna University of Technology, Vienna — *BSc (Media & Visual Computing)*
All coursework for my Bachelor's degree has been completed. Work on my thesis, which is focused on the topic of patent classification, is currently paused.

HTL Donaustadt, Vienna — *High School Diploma*
Emphasis on Network Programming and Network Management

SPOKEN LANGUAGES

- German - Native
- English - C2/Prof.

PUBLIC PROJECTS

Query Formulator - Solo Development — *Convert patent text into search queries*
The web application transforms patent text into queries compatible with major patent databases. The project involved designing and implementing backend and frontend logic, covering data extraction, ranking algorithms, and query expansion features as well as deployment.

Onto-Graph - Solo Development — *Visualize Hyponymy Relationships between words*
The web application offers a visualization of hyponymy relationships across various domains. The project involved extracting the data from its sources as well as setting up distributed servers for the workload, processing and storing the data in a database, including selecting and configuring a database, displaying this data in a useful way as well as deploying the web app.

Parkbob - Mostly Solo Development — *Visualize Parking regulations and spot availability*
The app provides drivers with essential information on current and future parking rules, regulations, prices, and maximum allowed parking times. Over a span of two years, the project was developed and maintained as a solo endeavor, handling all aspects of its creation and ongoing updates.

StreetCrowd - Multiple Developer — *Fleet optimization, Crowdsourcing & Gig economy*
Developed for ReachNow (DriveNow in the EU), this app incentivizes users to relocate cars from low to high activity zones. My role was significant in both conceptualizing a solution to this unique challenge and in the actual implementation of the app.

mySugr - Multiple Developers — *Diabetes Diary/Companion*
This app is designed to help diabetics monitor their current blood sugar levels, carbohydrate intake, and other factors impacting future blood sugar. My involvement was in the development of all aspects of the app, with a focus on ensuring a streamlined and safe user experience, aligning with its classification as a Class I medical device.

mySugr Monster Companion - Solo Development — *Diabetes Awareness, Marketing*
In a collaborative marketing campaign between Colgate Total and mySugr, focusing on raising awareness about oral health issues related to diabetes, I was responsible for implementing the Android app and all its features.

Beyond The Black Rainbow - 2 Developers — *Basic Video Game Engine*
Developed as a university project, this video game engine was created from scratch using C++ and OpenGL/GLSL. My primary responsibilities included designing the software architecture, managing rendering and event systems, integrating physics, and implementing post-processing.

RadCloud - 2 Developers — *Text Rendering*
As part of a university project, this Android app was developed to create word clouds from multiple sources, with the position and size of each word reflecting its relative frequency and significance per source.

NON-PUBLIC PROJECTS

PAIGE - Proactive Automated general IntelliGence Engine

A highly customizable and extensible personal assistant written in Electron (Javascript) and Python. Serves as a desktop client for arbitrary LLMs (local, onsite and cloud based) including ChatGPT. Through plugins it can be seamlessly extended with a wide range of capabilities like creating time sheets automatically, managing local files, calendars, emails, JIRA tickets, etc. and chain these tools together to accomplish more complex tasks.

NutriTracker

An AI based calorie and nutrition tracker web-app.

AI Book Club

A project to give readers a companion to discuss their books, enabled by LLMs, advanced RAG and summarization algorithms. Discuss the deeper meanings, get a recap of the story after putting the book down for a while and be reminded of who that character that hasn't appeared since chapter 1 even is. All without spoiling the ending of the book.

Text Summary

Isolates the most relevant topics and sections of patents through statistical means, then reconstructs the core ideas out of selected token sequences.

Data & Code Sharing Detection

Analyzes scientific publications based on rule based NLP, machine learning and metadata information to determine if the authors claim to generate or share data or code, and if so, if they are actually doing so.

Ontology Extractor

Analyzes scientific and medical text to extract word relationships, which serve as the foundation for constructing knowledge graphs. For example, connecting concepts like 'organ -> liver -> fatty liver disease' and 'alcohol -> fatty liver disease.' This approach enables the establishment of connections between various concepts, such as diseases, their causes, or potential treatments.

Vector-Space Service

A backend service capable of converting text into vectors, which can then be aggregated to construct embedding models. This service utilizes these models to facilitate vector space operations like similarity measurements, clustering, and pruning.