

UMBERTO SELVA

Updated: July 2025

AI Engineer, PhD, Founder @ 森 Selva AI Studio

@ u.o.selva@gmail.com

linkedin.com/in/umberto-selva

github.com/umbertoselva

calendly.com/selva-ai-studio/discovery-call

+39 3335218014

EU (Italy / Netherlands)

English, Italian



PROFILE

- EU-based freelance AI engineer with 5+ years of experience in Python software development • Former post-doctoral researcher and university lecturer with a PhD in linguistics from Leiden University, the Netherlands
- Wide-ranging background in computational linguistics, natural language processing (NLP), machine learning (ML) and data science • Broad experience developing NLP/AI apps across academia and the pharmaceutical industry.

AI Engineering Freelancing Consultancy Services @ 森 Selva AI Studio

• GenAI applications and APIs

End-to-end design, development, deployment and optimization of scalable, event-driven architecture-based AI applications, APIs, MCP servers and microservices, using modern cloud-native strategies and covering everything from proof of concept to MVP to production-ready systems.

• Chatbots

LLM-powered conversational assistants that know your company data, provide reliable answers, and leverage user profiling and custom memory strategies, available 24x7 for customer service, internal support, or lead qualification.

• AI agents and workflows

Intelligent agents that can reason, retrieve information, use tools, execute multi-step tasks, and make smart decisions without human input, built with type-safe structured output and function-calling techniques, following state-of-the-art chain of responsibility architecture guidelines for orchestration, routing, parallelization, etc. to handle complex workflows and save your team time and effort.

• Data and document processing

Automated and intelligent pipelines for extraction, parsing, classification, and labelling and reformatting of data from structured and unstructured sources (PDFs, invoices, legal documents, reports, tables, databases, etc.), cutting down manual repetitive work and improving accuracy.

• RAG (retrieval-augmented generation) pipelines

Scalable custom retrieval systems leveraging vector databases, semantic search (embeddings), keyword-based and hybrid search, BM25, reranking, and filtering, on top of custom chunking strategies, to reduce hallucination and improve factual grounding of LLM output.

• Evaluation and optimization pipelines

Design of automated testing pipelines for continuous performance optimization of your AI system, focusing on output relevance, faithfulness, context precision, context recall, or custom metrics, using state-of-the-art tools, LLM-as-judge, human-in-the-loop, and domain expert and user feedback strategies.

• Prompt engineering

Structured, modular, and model-specific prompt design using techniques such as zero-shot, few-shot, iterative, chain-of-thought, tree-of-thought, self-reflection, dynamic-template and meta-prompting, temperature adjustment, and tool-based augmentation to make sure your LLMs do exactly what you want them to do.

• Strategic guidance and training

Strategic guidance on how to integrate AI into your business and company workflows, transform your ideas into feasible products, discover new opportunities, evaluate cost-performance tradeoffs, and organize hands-on tutoring so that your team can use AI tools with confidence.

SELECTED EXPERIENCE

• Founder and AI Consultant

@ 森 Selva AI Studio

📅 2025–today

- Selva AI Studio brings together a group of freelancing specialists to offer consultancy services across artificial intelligence, natural language processing, machine learning, data science, and software development.
- I act as team leader and coordinator across projects and offer AI engineering services (see previous page).

Selected tech stack: Python, OpenAI, Anthropic, PydanticAI, LangChain, LangGraph, LlamaIndex, Langfuse, Unstructured, Docling, Ragas, FastAPI, PostgreSQL, PGVector, Supabase, Celery, Redis, Github Actions, Docker, Azure, AzureOpenAI

• AI Engineer

@ Pharmazie.com

📅 2023–today (2 years)

- Pharmazie.com is a Germany-based web platform that gathers data from over 25 databases to provide reliable information on pharmaceutical products to healthcare professionals.
- I led the company's transition into generative AI from scratch, providing strategic guidance and translating management's ideas into feasible products. I designed, developed, and deployed AI products end-to-end, from proof of concept all the way to production.
- I designed the company's AI assistant, including conversation workflows, widgets, and front-end interfaces, powered with an agent workflow built with LangGraph and FastAPI, as well as a series of dedicated RAG pipelines built with LangChain, to leverage pharmaceutical data from 25+ databases and allow subscribers to interact with the company platform with via natural language.
- I designed custom data extraction and chunking processes for RAG pipelines from scratch that were able to smartly extract data from PDFs and other document types into hierarchically structured and semantically sensible chunks (e.g. whole chapters, chapter chunks, etc.), as well as custom reranking strategies, that significantly improved our retrieval system before such technologies became available with external code libraries or third-party service providers.
- I designed evaluation and optimization pipelines incorporating LLM-as-judge strategies, user feedback, and domain expert advice to continuously improve each of our AI system component's performance metrics.
- I designed, developed, and deployed custom AI pipelines for third-party clients in the pharmaceutical sector to extract data from structured and unstructured documentation and databases and format it into custom templates to fill out forms tailored to our customers' needs, reducing processing time from days to minutes.
- I continue to provide maintenance and strategic advice.

Skills: Python, NLP, LLMs, OpenAI, GPT, LangChain, LangGraph, LlamaIndex, PydanticAI, Conversation Design (Voiceflow), Streamlit, FastAPI, Prompt Engineering, Chatbots, Retrieval Augmented Generation (RAG), FAISS, PGVector, PostgreSQL, Ragas, AI Agents, Unstructured, Docling, Docker, AWS

• NLP / ML / AI Consultant

@ Various private business clients

📅 2020–2025 (1- to 6-month-long projects)

- I built a lightweight LLM-powered email/calendar assistant that can filter incoming emails and provide automated replies, modify calendar events, and process invoices.
- I built customer support chatbots that leverage knowledge bases derived from internal documentation and data sources to provide grounded responses.
- I trained and optimized ML models for sentiment analysis to power a social media analysis tool.

Skills: Python, AI (OpenAI, Anthropic, PydanticAI, LangChain, LangGraph, etc.), NLP (NLTK, spaCy, Haystack, etc.), ML/DL (Tensorflow, Keras, Scikit-Learn, Huggingface, Ollama), conversation design (Voiceflow, Botpress), data science (Pandas, NumPy, Matplotlib), statistics, web development (FastAPI, Flask, etc.), SQL (PostgreSQL, SQLite, MySQL), bash / zsh, Linux, Jupyter, Git, Github, Docker

• NLP Specialist and Full-Stack Web Developer

@ Leiden University Center for Linguistics 📅 2020–2021 (6 months)

- I developed sanskritmeterscan.net, a web application geared towards academic researchers in historical linguistics and Indology that automates the metrical scansion of poetic texts composed in two ancient Indian languages (Sanskrit and Vedic), reducing the scansion time by an average of 1500 verses per month to seconds.
- I curated both the design of the Python NLP script powering the application core, the Flask backend, the frontend design, and the deployment.

Skills: NLP, computational linguistics, backend web development, frontend web development, Python, Flask, HTML, CSS, Javascript

• Academic Researcher and Lecturer (Linguistics)

@ LUCL / IIAS 📅 2014–2020

From 2014–2020, I was a research assistant, then PhD researcher, post-doctoral researcher, and lecturer at the Leiden University Centre for Linguistics (2014–19) and the International Institute for Asian Studies (2019–20).

Skills: Linguistics (general, historical, computational), academic research, academic writing, academic publishing, teaching, research project management

EDUCATION AND PRIZES

- PhD, Linguistics (Leiden University, 2019)
- Research MA, Linguistics (Leiden University, 2014; grade: 9)
- BA, Classics (University of Turin, 2011; grade: 110/110 *cum laude*)
- 2015 First Prize for Best MA Thesis by the Indogermanische Gesellschaft

CERTIFICATIONS

Machine Learning (Stanford Online, 2020) • NLP Specialization (DeepLearning.ai = DL.ai, 2022) • NLP with Transformers in Python (Udemy, 2022) • Deep Learning Specialization (DL.ai, 2022) • Statistics for Data Science with Python (IBM, 2022) • Practical Data Science on the AWS Cloud (AWS, DL.ai, 2022) • Fine Tuning LLMs (Lamini, DL.ai, 2024) • Building Agentic RAG (LlamaIndex, DL.ai, 2024) • LLMOps (Google Cloud, DL.ai, 2024) • Functions, Tools and Agents (LangChain, DL.ai, 2024)