

Pablo Valdunciel Sanchez

AI / Machine Learning Engineer - Karlsruhe, Germany

pablo.vs@etik.com | +49 152 07181484 | github.com/pabvald | linkedin.com/in/pabvald/

Experienced AI Engineer (M.Sc.) with focus on GenAI (LLMs, RAG, fine-tuning) & NLP. AI solutions successfully implemented with Python & Azure Cloud Services.

Professional experience

Software and AI Engineer @ DOCYET GmbH – Remote, Germany

since 10/2024

- **RAG Solution:** Architected and implemented a Retrieval Augmented Generation (RAG) pipeline over a medical knowledge graph using LlamaIndex, Ollama, and Docker.
- **Release Management:** Orchestrated end-to-end software releases and CI/CD workflows with CircleCI, deploying to Telekom Cloud.
- **Project Planning & Leadership:** Led a small team to design, plan, and deliver a Ruby on Rails–based admin panel that enables one-click chatbot creation, reducing setup time from several hours to a few minutes.

Junior Software and AI Engineer @ DOCYET GmbH – Remote, Germany

10/2023 - 10/2024 (1 y.)

- **Interactive Triage App:** Architected and developed a Ruby on Rails–based triage app for hospitals, integrating external APIs and a state-machine workflow with TurboStream and Docker. Reduced average patient anamnesis time by 25%, enhancing clinical intake efficiency.
- **NLP Automation & LLM Fine-Tuning:** Streamlined language-related workflows by fine-tuning domain-specific LLMs in the Azure AI Portal, automating routine text-generation tasks.
- **Database Architecture & Performance:** Revamped the DOCYET PostgreSQL schema using UML-driven data modeling, normalizing tables and optimizing indexes to cut query latency by 30% and improve long-term maintainability.

Machine Learning Working Student @ DFKI GmbH – Remote, Germany

06/2021 - 09/2022 (1 y. 6 mo.)

- **Master's Thesis:** Development and optimization of a Relation Extraction (RE) model (ClinicalBERT) within an Active Learning (AL) framework, with comparisons to Random Forest and BiLSTM models, focusing on performance, feature engineering, and annotation costs. Technologies: PyTorch, Hugging Face Transformers. (See Publication [1])
- **Model Interpretation:** Visualization of explanations for named entity recognition (NER) models to enhance the interpretability of transformer models. Technologies: Hugging Face Transformers, Captum. (See GitHub profile: [transformers-interpret](#))

NLP Working Student @ Universidad de Valladolid – Valladolid, Spain

11/2019 - 05/2020 (7 mo.)

- **Semantic Similarity:** Comparison of methods based on pre-trained Word2Vec, GloVe and FastText vectors to measure the semantic similarity between sentence pairs. (See GitHub profile: [semantic-similarity](#))

Full-Stack Developer (Internship) @ DOCYET GmbH – Leipzig, Germany

06/2019 - 09/2019 (4 mo.)

Technical Skills

- **Machine Learning:** Python, Scala, PyTorch, Word Embeddings, LlamaIndex, Hugging Face Transformers, spaCy, RAG, Large Language Models (LLMs), scikit-learn, Snowflake, Apache Spark, MLflow, Pandas
- **Backend & API:** Flask, FastAPI, REST APIs, Docker, SQL (PostgreSQL), Ollama, Ruby, Ruby on Rails
- **Cloud & DevOps:** Microsoft Azure, CI/CD (CircleCI), AWS, Terraform, Azure OpenAI
- **Collaboration & Methodologies:** Agile/SCRUM, Jira, Workshops, Release Management, UML

Education

M.S. in Data Science & Artificial Intelligence @ Universität des Saarlandes – Germany **10/2020 - 05/2023**

- **Grade:** 92/100 (master's thesis: 100/100)
 - **Focus:** Deep Learning for NLP
-

B.S. in Computer Science* @ Universidad de Valladolid – Valladolid, Spain **09/2016 - 05/2020**

- **Grade:** 87.8 / 100 (bachelor's thesis: 98/100)
 - * Graduated top of the class, 12 courses passed with honours
 - 240 ECTS (4-year program)
-

Languages

- **German:** Professional proficiency (*Goethe-Zertifikat C1*)
 - **English:** Professional proficiency (Cambridge Certificate of Advanced English, C1)
 - **Spanish:** Native speaker
-

Publications & References

- Publication [1]: Optimizing Relation Extraction in Medical Texts through Active Learning: A Comparative Analysis of Trade-offs (UncertainNLP-WS, 2024; Archive: aclanthology.org/2024.uncertainlp-1.3/)
 - Publication [2]: Interactive Assessment Tool for Gaze--based Machine Learning Models in Information Retrieval (CHIIR, 2022; DOI: [10.1145/3498366.3505834](https://doi.org/10.1145/3498366.3505834))
 - GitHub Profile: www.github.com/pabvald
-

Awards

Special Award (*Premio Extraordinario de Fin de Grado*) **15/02/2021**

- Prize awarded by the University of Valladolid for achieving the highest GPA in the B.S. Computer Science program (class of 2016–2020).
-

Volunteering

Erasmus Student Network (ESN) **10/2021 - 02/2022**

- Member, Partnership Manager
-

Board of European Students of Technology (BEST) **09/2018 - 05/2020**

- Member
-

Scout Group Gilwell **07/2018 - 09/2018**

- Scout Monitor