

UNAI GARAY

SENIOR AI ENGINEER



- +34 662128333
- unaigaraymaestre@gmail.com
- Alicante, Spain
- [Personal Website](#)
- [LinkedIn](#)

PROFILE

As a Senior AI Engineer specializing in Image and Natural Language Processing, I am dedicated to advancing AI solutions through innovative research and development. I thrive in collaborative environments, contributing significantly to team success and fostering a positive work culture.

SKILLS

Programming & Tools: Python, C++, TensorFlow, PyTorch, Transformers, Flask, FastAPI, Docker, GCP, SQL, Git

AI/ML: NLP, Computer Vision, Deep Learning, MLOps, Neural Search, vLLM

Soft Skills:

Design Thinking, Problem-Solving, Project & People Management, Communication

EDUCATION

BSC COMPUTER SCIENCE

University of Alicante

2014 - 2018

MSC ARTIFICIAL INTELLIGENCE

Valencian International University

2018 - 2019

PATENTS

- [Feature Engineering With Question Generation](#) [Sorcery]
- [Ontology Integration for Document Summarization](#) [Sorcery]
- [Cross-context Natural Language Model Generation](#) [Sorcery]

PROJECTS

Guexit - AI Image Generation Service

- Technologies:** DeepSpeed, Huggingface, Python, Docker, C#, TypeScript
- Description:** Built an AI-driven card gaming platform with custom backend, infrastructure, and web frontend. Developed and deployed scalable AI image generation pipelines.
- Image Generation:** Fine-tuned diffusion models using DeepSpeed to produce highly targeted images (people/pets) for card game use cases.

EXPERIENCE

SENIOR AI ENGINEER

2026 - Present

Inulti

- Built production voice AI sales agents (LiveKit/SIP, STT-LLM-TTS, PCI-safe payments) and an end-to-end LLM retention platform (FastAPI + React, AWS), with LLM-as-judge eval harnesses gating every model swap.

SENIOR LLM ENGINEER / DATA SCIENTIST

2025 - 2026

Tekever

- Engineered an end-to-end data pipeline that automatically creates Jira tickets from Signal chats, implementing a conversation clustering algorithm using semantic similarity and temporal analysis, and leveraging an LLM to generate structured ticket data, all orchestrated with Python, Airflow, Docker, and GCP

SENIOR AI ENGINEER

2020 - 2025

Sorcero

- Led AI R&D for a Natural Language Understanding platform for unstructured medical data.
- Key contributions: generative finetuned AI theme tagging, NER, summarisation, recommendations, custom neural search engine with RAG (generative AI), Agents, etc.

DATA SCIENTIST

Accenture

2019 - 2020

- Development of a tool that uses Data Analysis, Clustering, Topic Modelling and other NLP algorithms on problem descriptions from ticketing systems to improve ticket resolution.

AI ENGINEER

Smartloto

2018 - 2019

- Development of an end-to-end image validation system of lottery tickets using Deep Learning at scale on Google Cloud Platform, Android and IOS.

RESEARCH

Data Augmentation via Variational Auto-Encoders

Publication at CIARP 2018 congress in Madrid

- Author of a research paper about overcoming lack of data when training a neural network by using Variational Auto-Encoders to generate new varied data.

Spiking Neural Networks: A Review

Research Publication

- Author of a research paper where Spiking Neural Networks are reviewed and put into practice by simulating them using Nengo library with different approaches.