

Jean Bianchi

jpbianchi@gmail.com | [LinkedIn \(200 endorsements\)](#) [references](#)

I help companies improve their processes using SOTA ML, AI & Blockchain technologies

SKILLS

Programming Languages

Python, Solidity, SQL, Scala, R

LLM Models

BERT, GPT2-3-4-4V, (Code)Llama, Falcon, Phi, Mixtral

Frameworks & Libraries

Transformers, Diffusers, OpenAI, Langchain, Lora, LlamaIndex, Accelerate - DeepSpeed, Cleanlab, Argilla, Pytorch, Nltk, Spacy, Scikit-Learn, Dask, Spark

Data Bases

Weaviate, Redshift, Snowflake, PostgreSQL, Neo4J, GraphQL

Deployment

AWS Lambda, Azure with Terraform, Docker, Vercel, Streamlit, FastAPI, Modal, Kubernetes

Tools & Platforms

HuggingFace, DBT, Dagster, Github Actions, Celery, Neo4J

EDUCATION

2022-

MS Blockchain technologies
University of Cyprus at Nicosia

2019-2021

MS Machine Learning
Georgia Institute of Technology

MS Electronics
University of Illinois at U-C

CERTIFICATIONS [\(full list\)](#)

LLM apps with Langchain
LLMs in production
Productizing LLMs
Architecting AI products
Image generation with Diffusion
Advanced Prompt Engineering
Recommender systems
Kubernetes
MLOPs
Data-Centric Deep Learning
Forecasting
Data Science for Security Fraud
Full Stack ML
Applied Statistics for DS
Advanced SQL
Data Engineering with Dagster

INTERESTS

Hackatons:

PEARC HPC 21: [winner](#)

Falcon LLM 23: [winner](#)

Multimodal LLM: [2nd place](#)

RELEVANT EXPERIENCE [\(more details on next page\)](#)

Liquidity Group – Senior ML / AI engineer

June 2024 -

- Financial forecasting
- Hybrid quantitative & qualitative RAG-based chatbot
- Deployment with RabbitMQ / Lambda / fast API

iLiv – CTO

Mar 2024 – June 24

- MVP development: deep integration of novel AI features in website
- Hybrid vector database / Mongo search [RAG architecture](#)
- Multi-agent planner system

Maastif AI – Generative AI engineer

Oct 2023 – Mar 2024

- Development of AI cognitive hierarchies using multi-agent system
- Great improvement of code bugs repair with LLM's ([results](#))

Omdena – Generative AI team lead

Oct 2022 – Oct 2023

- Barcelona team lead, specialized in Blockchain – AI fusion
- Development of a crypto currencies price prediction algorithm, [deployed on HuggingFace](#)
- RAG system based on a Weaviate vector database using fine-tuned Sentence Transformers embeddings, and GPT4 & Llama, deployed on [Streamlit \(Huggingface\)](#) & Modal (backend)
- Development of an AI psychiatrist based on Sentence Transformers and deployment with FastAPI and React on GCD

BitMEX – Data Scientist

Nov 2021 – Oct 2022

- Support of internal customers (marketing, fraud, stakeholders)
- Airdrops (free coins) development for BitMEX new crypto currency
- Traders behaviour modelling (FFT & proprietary methods)

AITech Engineering – CEO / CTO

2008 – Aug 2019

- Machine Learning applied to industrial applications
- Hired, managed and scaled company from 1 to 24 people

PROJECTS [\(my complete portfolio\)](#)

[RAG for financial reports \(code\)](#)

This RAG analyses financial reports in PDF and stores the information in a Weaviate database using Sentence transformers embeddings.

[Lora fine tuning of PHI-1 5, Llama2](#)

Parameter efficient fine tuning (PEFT) low rank adaptation (LORA) to fine-tune the Phi-1.5 model for summarizing scientific papers on Colab T4's GPUs

Models fine-tuning for data analysis visualization

Comparison between fine-tuned 4bit Falcon 7B Instruct, CodeLlama and Mistral 7B to generate visualization code

[MLOPs & Data-centric Deep Learning](#)

Full pipeline implementations of various models (kNN, SVM, Boosting, Random Forest, Bayes, NN), including monitoring, data and label drift detection, outliers rejection, performance metrics and deployment on various clouds

ARTICLES [\(my publications\)](#)

Blockchain and AI: A powerful synergy ([Medium](#))

[Vectorized Convolutions](#): how to use Numpy vectorization and broadcasting to execute the convolutions of a CNN in parallel

Languages

English, French, Italian

EXPERIENCE DETAILS

Liquidity (financial forecasting & asset management): June 2024 –

Senior Machine Learning & Generative AI engineering

I work on the development of a quantitative (financial data) / qualitative (companies' reports) RAG that spearheads the company's new chatbot designed for financial analysts.

iLiv (startup in pre-funding stage): March 2024 – June 2024

CTO

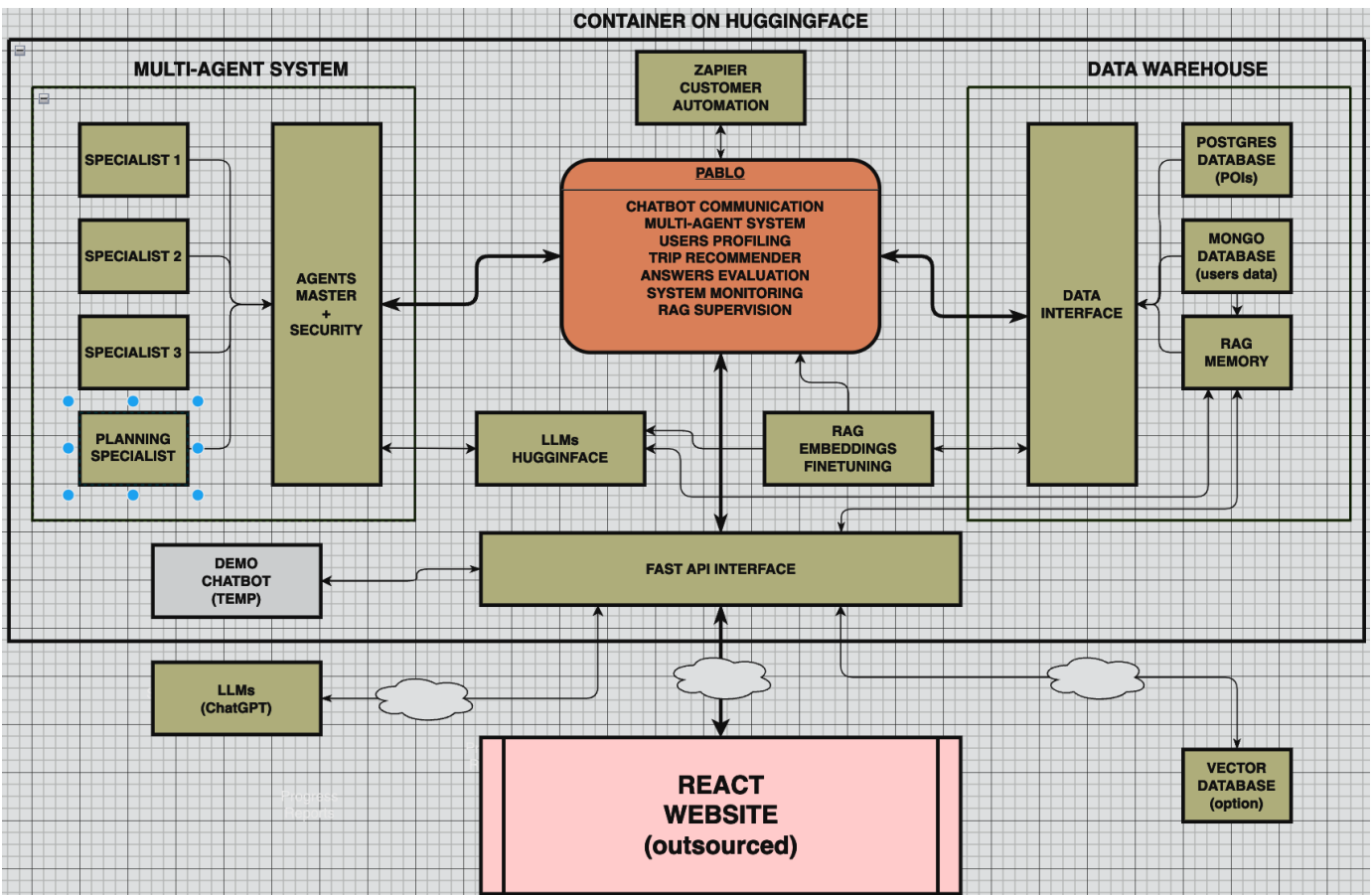
iLiv is a revolutionary short-term rental platform, designed from the ground up to integrate generative AI deeply into all possible interaction aspects, with users and between them, but also use a new business and marketing model to the market.

I am in charge of **designing the whole MVP** (see diagram), except the website, i.e. the data backend, the multi-agent AI intelligent system that will appear in users' chatbots but also in other functions across the website, and the core system that implements such functions and communication interfaces between all the systems.

In particular, I will **implement my own state-of-the-art model of human behavior** that will dramatically increase users' engagement with the platform due to a superior conversational quality. It will be used by the recommender system to create the best possible trips and position ads optimally.

Tech stack:

Python, LlamaIndex, OpenAI, Llama, Claude, Mixtral, CrewAI, Mongo (RAG), PostgreSQL, FAST API, Pytorch, fine-tuning, Docker, HuggingFace (deployment)



Maastif.ai (startup): Oct 2023 – March 2024

Generative AI engineer

I contributed to a DARPA grant application for software vulnerabilities repair using several LLMs by doing the experiments. The SOTA is this paper "[Examining Zero-Shot Vulnerability Repair with Large Language Models](#)". I managed to **generate 13 working patches** to some of the hardest type of vulnerabilities vs 5 from the best coding model 'Codex' (used in Github copilot).

I also worked on 'Cognitive hierarchies', i.e. collections of agents that can operate autonomously by calling other agents ('multi-agent' systems), writing code, and calling exterior libraries to bridge the gap between increasingly powerful LLMs and true Artificial General Intelligence in order to solve tough problems. An in-house agent orchestrator was (and still is) being developed, **implementing the latest strategies such as 'Everything of Thought' (XoT)** which will increase the 'thinking power' of the system.

XoT allows to go beyond current multi-agent systems, and allow to make an AI system process information and 'think' more like a human, and is, imo, one of the best strategies to model behavior, and forecast any activity heavily influenced by human psychology, such as trading for instance.

Tech stack:

Python, OpenAI, Llama, Azure

Omdena: Oct 2022 – Oct 2023

Generative AI engineer – Barcelona Team Lead

The year 2023 was a bit special because I started a Master's Degree in Blockchain and cryptocurrencies. It was a year of learning, in which I also I took 20+ courses designed by industry experts (Uplimit.com).

I started the Omdena Barcelona chapter to develop, with people from around the world, 'Fusion' projects, ie the use of both AI and blockchain technologies, which is the future of technology (along with Web3). See my [article on Medium](#) about this.

One such project was the creation by the participants of [Fusion projects for their own field](#).

I also developed on my own a cryptocurrency price prediction model, hosted on [HuggingFace Spaces](#).
Tech: Tensorflow, Scikit-learn, Pandas, Numpy, Streamlit, Time-series, LSTM

I [hosted on HuggingFace](#) a RAG system with a reranker (cross-encoder) using the episodes of the Impact Theory show.

Tech: Python, Weaviate, OpenAI, Llama, SentenceTransformer, LlamaIndex, Streamlit, HuggingFace, Modal

BitMEX: Nov 2021 – Oct 2022

Data Analyst / Scientist

BitMEX is a cryptoexchange. I did data analysis for various departments such as marketing, fraud, stakeholders.
Tech: Redshift, scikit-learn

I also developed the code for the Airdrops of BitMex' new cryptocurrency BMEX.

Finally, I worked on modelling traders' behavior using several techniques (some proprietary). The model was left unfinished due to BitMEX disbanding the team due to the crypto winter in 2022.

Tech: FFT, time-series, auto-encoder, deep-clustering, Pytorch, SageMaker

AITech Engineering: 2008 – 2019

CTO & Machine Learning Engineer

AITech is my engineering services company in Belgium. I assembled and managed a team to develop the hardware and software of an innovative Voice over IP phone, with an embedded smart card reader to provide encryption-authentication services to what was actually a service platform with telephony.

The project failed due to lack of funding (Belgium is notoriously risk-averse) despite having patented the technology and having distribution agreements with Cisco's biggest distributor. So I continued solo as a contractor to provide hardware and software services. I slowly veered towards Machine Learning as the technology improved, but since I come from electronics, the implementations were low-level, ie running on electronic boards with DSPs. This is where my interest in ML started, and culminated with a Master's degree at Georgia Tech.

REFERENCES & CERTIFICATIONS

You will find ~ **200 endorsements** of my technical skills from people in various industries (including OpenAI, Microsoft, Facebook) in pretty much all AI/ML related topics.

It is not always easy to get a proper recommendation due to several reasons, notably when I am the one leaving a project, but in recent years, since I switched to programming completely, I paid more attention to this and managed to gather a handful of signifiant [recommendations on LinkedIn](#).

I mentioned above getting **20+ certifications** in 2023, some of which are listed on the first page. I have all the certificates if necessary, but I leave you with the words of Ms Shreya Vora, program manager at [Uplimit](#).

If that doesn't describe you, please let me know and I'll help you find a course that is a next week!



Shreya (Uplimit) 8:34 PM

Hey @JP Bianchi! Hope this helps:

PDF ▾



JP Bianchi - Letter of Recommendation.pdf

PDF

To whom it may concern,

JP Bianchi has been part of the Uplimit community for a couple of years. In that time, he has taken at least two dozen courses and helped the Operations team here immeasurably by providing constructive feedback on projects and course content.

JP has also helped our learners by sharing resources (blog posts, articles) widely in cutting-edge technical courses and reviewing others' projects and providing them with insightful feedback on what they did well and/or could've done better. These courses were some of our more technically challenging ones and were created by industry professionals who were experts in their field. JP completed so many courses within the deadline, and has usually been able to give us fantastic feedback thanks to his technical depth.

JP has been a helpful addition to our community, and we really appreciate his presence at our courses live and async on Slack.