Leon Doungala

AI & ML Engineer | Data Scientist

Portfolio : github.com/LeonDoungala22 | leondoungala22.github.io/doungala.leon.github.io/

Viale Marconi, Tradate, Italy • +39 3293670086 • doungala.leon@gmail.com • linkedin.com/in/leon-doungala

PROFILE

AI/ML Engineer - Data Scientist with an M.Sc. in Computer Engineering (AI specialization) and a Master's in Data Science, dedicated to delivering innovative solutions for healthcare and industry. Proficient in predictive and generative AI, with expertise in LLMs, RAG, Transfer Learning, and Vector Databases. Preparing for registration with the ***Italian Order of Engineers (Ordine degli Ingegneri), Section A, in 2025.***

EDUCATION

- **eCampus University, Come, Italy** **March 2025**
  M.Sc. in Computer Engineering and Automation, Artificial Intelligence
- **Rome Business School, Rome, Italy - University of Valencia, Valencia, Spain** **April 2022**
  Double Master's in Data Science
- **Ferrara University, Ferrara, Italy** **September 2020**
  Bachelor's Degree in Electronic and Computer Engineering

SKILLS

- Generative AI & LLMs
- Retrieval-Augmented Generation (RAG)
- Chat Search Systems & Autonomous Agents
- Machine Learning & Deep Learning
- Python & Predictive Modeling & Data Science
- AI Research & Development

PROFESSIONAL EXPERIENCE

AI / ML and Data Science Engineer **Mar 2024 – Present**
People Lab , *Milan, Italy*

- Developed AI solutions for 300+ clinics and companies, including a RAG-based Clinical Semantic Search Engine and an ATS system for HR recruiters supporting 200+ companies.
- Designed AI projects with Generative AI, RAG, Transfer Learning, and LLMs, building predictive systems and scalable solutions using ML, DL, MongoDB, MySQL, Azure Cosmos, Python, and Vector Databases.
- Engineered scalable AI systems for semantic search, autonomous agents, and workflow automation, driving innovation in healthcare and industry.

Software Developer | Data Scientist **Apr 2022 – Dec 2023**
Italian Stock Exchange (Euronext Group) , *Milan, Italy*

- Developed and maintained software applications, optimized SQL DataMarts (Workbench, Athena), and automated processes with Python.
- Collaborated within agile teams, managed AWS data ingestion (Glue, Python Spark), and enhanced machine learning models.
- Applied data science techniques to support decision-making and drive actionable insights.

Developer Analyst Salesforce **Jun 2020 – Dec 2021**
Lutech group , *Milan, Italy*

- Salesforce: Skilled in development, automation, deployment, error analysis, Apex, VisualForce, and data management.

HCM Technical Consultant - Intern **Sep 2019 – Feb 2020**
Oracle , *Milan, Italy*

- Gained hands-on experience in Oracle SQL, Fusion Reporting, BI, integrations, HCM cloud (BI Publisher, OTBI, HCM extracts), web services, and Taleo Client

LANGUAGES

- French - Native (C2 level)
- Italian - Proficient (C1 level)
- English - Intermediate (B1 - B2 level)

TECHNICAL PROFICIENCIES

- Programming Languages: Python, Java, SQL, Shell
- Libraries & Frameworks: Scikit-learn ,TensorFlow, LangChain, Pandas
- APIs & AI Models: OpenAI API, Anthropic API , LLM integrations
- Web Development: HTML, CSS, JavaScript, Flask
- Databases: MongoDB, MySQL
- Cloud : AWS, Azure , Salesforce
- Tools: Git, Docker, Tableau, Jupyter Notebook, GitHub
- AI Expertise: LLMs, RAG, Predictive modeling

THESIS AND PROJECTS

https://leondoungala22.github.io/doungala.leon.github.io/ml_and_ds_portfolio.html

M.Sc. Thesis

Artificial Intelligence and Machine Learning Techniques for the Characterization and Prediction of Diabetes

- **Focus:**
  Applied AI and ML techniques, including **RAG, semantic search**, and **generative AI,** to address diabetes prediction and characterization challenges. Developed supervised (e.g., XGBoost, Random Forest) and unsupervised (e.g., K-Means, PCA) models for actionable medical insights.

  Key Points:
  - Designed integrative dashboards and **predictive analytics** for healthcare.
  - Developed advanced **predictive models** for diabetes risk and characterization.
  - Implemented LangChain and **Vector Databases** for **RAG workflows** and **semantic search**.
  - Leveraged **Generative A**I for data augmentation and handling imbalanced datasets.

  Repository: Explore Thesis Repository

Highlighted Projects

Virtual Doctor AI Assistant

- **Description**: The AI Health Assistant is an AI-powered system designed to assist individuals by providing helpful, reassuring, and medically sound guidance based on their symptoms.
- **Key Features:** Dynamic medical data retrieval, conversational flows, and task automation.
- **Repository Link:**Virtual Doctor AI Assistant

Human Resources Analytics

- **Description:** Applied PCA and K-Means clustering for employee segmentation and HR data analytics.
- **Algorithm:** PCA, K-Means
- **Repository Link:** View Project on GitHub

Exploring Customer Segmentation

- **Description:** Segmented customer data using clustering techniques for market analysis.
- **Algorithm:** K-Means
- **Repository Link:** View Project on GitHub

Heart Disease Prediction

- **Description:** Built predictive models to analyze heart disease risks using ML and NLP techniques.
- **Algorithm:** SVM, NLP
- **Repository Link:** View Project on GitHub

Sentiment Analysis on IMDb Movie Reviews

- **Description:** Performed sentiment analysis to classify reviews using Naive Bayes techniques.
- **Algorithm:** Bernoulli Naive Bayes
- **Repository Link:** View Project on GitHub

Breast Cancer Classification

- **Description:** Developed a classification model for breast cancer using K-Nearest Neighbors.
- **Algorithm:** KNN
- **Repository Link:** View Project on GitHub

ATS System for CV Filtering and Job Matching

- **Description**: Developed an AI-powered ATS system to align candidates with job descriptions or assist HR with advanced semantic search. Utilized RAG, LangChain, Anthropic APIs, OpenAI, and Azure Cosmos DB for precision matching and intelligent filtering.
- **Algorithm**: Retrieval-Augmented Generation (RAG) with semantic search and conversational AI.
- **Repository Link**: View Project on GitHub

ADDITIONAL INFORMATION

- **Driving License:** Category B (with personal vehicle)
- **Soft Skills:** Problem-solving, communication, adaptability, teamwork, proactivity
- **Interests:** AI for Healthcare, bass guitar, running
- **Recommendations**: Available upon request
- **Data Privacy**: I consent to the processing of my personal data included in this CV in compliance with the GDPR (EU Regulation 2016/679).