MOHAMED BALLOUCH

Data Scientist / AI engineer

Email:ballouch.mo@gmail.com
Linkedin:mohamed-ballouch
Phone: +212 611660019

Address: Sidi Maarouf, Casablanca, Morocco
GitHub: github.com/Mohamedballouch

About Me

Data Scientist and **ML Engineer** with **5 years of experience** in academia and industry, specializing in developing and deploying innovative AI solutions. Expertise in cloud-based machine learning, NLP, and deep learning, with proficiency in frameworks like TensorFlow, Rasa, and LangChain. Skilled in translating complex challenges into ethical AI-driven solutions across various sectors, including insurance and tech. Experience with AWS and Azure for model deployment and MLOps. Adept at implementing cutting-edge techniques such as RAG systems and prompt engineering. Combines rigorous academic methodologies with practical industry applications. Committed to continuous learning and staying abreast of the latest AI/ML advancements.

Professional Experiences

- **DXC technology** Casablanca, Morocco
  Senior Machine Learning Engineer *– Hybrid* *January 2024 – Present*
  - Spearheaded the identification and implementation of use cases for Large Language Models (LLMs) and generative AI in the insurance sector, enhancing decision-making processes and customer interactions.
  - Significantly improved RAG system performance by implementing and enhancing reranking algorithms, resulting in a 20% increase in response accuracy.
  - Developed a comprehensive evaluation tool for RAG solutions, enabling systematic assessment of question-answering quality and facilitating continuous improvement of the system's performance.
  - Optimized system speed by introducing cache databases, reducing latency by 30% for frequently accessed data.
  - Researched and implemented multimodal LLM capabilities, enabling the system to process and analyze both text and image data, expanding the AI's applicability across diverse insurance use cases.
  - Conducted extensive prompt engineering experiments and enhancements, iteratively refining LLM performance on insurance-related queries. This process significantly improved response relevance and substantially reduced hallucinations, markedly increasing the reliability of AI-generated insights in insurance applications.
  - Established robust feedback loops with key stakeholders, ensuring continuous improvement of AI solutions and alignment with business objectives.
  - Conducted regular research on advancements in LLMs and GPT models, integrating cutting-edge techniques to maintain competitive advantage in AI-driven insurance solutions.

  Languages/Technical usage: Python, Pytroch, Transformer, Hugging-face Ecosystem, LangChain , Docker, AWS, LlamaIndex, FastAPI, NLTK, OpenAI GPT, Anthropic Claude, MongoDB, Redis, Pinecone

- **BlackStone Eit** Seattle WA, USA
  Data Scientist || ***Machine Learning Engineer*** *– Remote* *December 2020 – December 2023*
  - Developed and deployed machine learning models on AWS and Microsoft Azure, demonstrating proficiency in cloud-based AI solutions and MLOps practices.
  - Developed an archive digitalization solution using Azure Video Indexer, enabling multi-modal search across video and image content by keywords, topics, and visual elements.
  - Implemented a computer vision system using pretrained CNNs for automated pizza quality assessment and personalized recommendations.
  - Implemented robust data quality checks and error handling mechanisms within Azure Data Factory

pipelines, ensuring data integrity and reliability throughout the ETL process.

- Orchestrated end-to-end MLOps pipelines, incorporating version control, continuous integration/continuous deployment (CI/CD), and automated model deployment, streamlining the machine learning lifecycle.
- Designed and implemented an internal chatbot leveraging GPT technology, integrating it with LangChain, OpenAI's APIs, and VectorDB (Pinecone) to enhance internal knowledge management and query resolution.
- Implemented advanced image segmentation techniques for satellite imagery analysis, enhancing geospatial data interpretation for urban planning and environmental monitoring.
- Conducted extensive text analytics and sentiment analysis using advanced NLP libraries and techniques, providing valuable insights for customer feedback analysis and product improvement.
- Proactively stayed abreast of the latest research and advancements in AI/ML, computer vision, and cloud technologies, ensuring the delivery of cutting-edge solutions aligned with industry best practices.

Languages/Technical usage: Python , Azure Video Indexer, SciPy , Pandas, Scikit-learn, Postgres , Azure Data Factory, OpenAI API, Docker, AWS, ML-Cycle, LangChain, Flask, GeoPandas, YOLO

- **3W Media** Casablanca, Morocco
  Data Scientist || ***Machine Learning Engineer*** *June 2020 – December 2020*
  - Scraping Data from social media using tools: Python, Selenuim, GraphQl & beautifulsoup.
  - Data processing, structuring, manipulation using SQL.
  - Generating dashboards and reports using Tableau, Qlik and PowerBI.
  - Designed Natural Language Processing (NLP) algorithms to evaluate Social Data Posts & comments using sentimentanalysis (Python, NLTK).

  Languages/Technical usage: Python, NLTK, Postgres, Pandas, GraphQl, beautifulsoup

- **BADEM AHBS** Istanbul, Turkey
  Machine Learning Engineer *November 2019 – April 2020*

  Developed a Web App for Sentiment Analysis using ML model on Twitter.
  - Collect the data for training and testing the model.
  - Developed a machine learning model : XGBoost for sentiment analysis.
  - Developed a user interface using Flask framework, Html, Css and JavaScript

  Languages/Technical usage: Python, Matplotlib, NLTK, SciPy, Postgres, Pandas, Keras, Flask

- **JÜPITER YAZILIM** Istanbul, Turkey
  Data Scientist Intern (PFE) *February 2019 – July 2019*

  Developed a software to forecast the electricity market clearing price using ML and DL models.
  - Data collection, preprocessing & feature selection (Sliding window method).
  - Developed the forecasting models (MLP, LSTM, CNN, CNN+LSTM hybrid) with Keras.
  - Developed a graphical user interface using Tkinter.

  Languages/Technical usage: : Python , Keras, Pandas, Scikit-learn, Matplotlib

Projects

- **Multimodal LLM for Chart Analysis and Interpretation**:
  - Developed a multimodal Large Language Model capable of analyzing and interpreting complex charts and graphs, providing detailed textual descriptions and insights.**[**Github**]**

  Languages/Technical usage: Python, Numpy, Anthropic Claude, LangChain , ChromaDb

- **Telcom Customer Churn Prediction and Analysis**:
  - Developed ML models to predict customer churn and identify key attrition factors for a telecom company, using a dataset of 7043 customers.[Gihtub]

    Languages/Technical usage: Python, Pytorch, Numpy, Keras, Neural Network, Pandas
- **Facebook Page Scraping Tool**:
  - Developed a scraper to extract comprehensive data from Facebook pages, including post details, engagement metrics, and media content.[Github]

    Languages/Technical usage: Python, Selenium, BeautifulSoup, Tkinter
- **Advanced Topic Modeling using BERTopic**:
  - Implemented BERTopic for unsupervised topic discovery in large text datasets.
  - Utilized BERT embeddings and clustering techniques to identify and visualize coherent topics.[Github]

    Languages/Technical usage: Python, BERTopic, BERT, UMAP, HDBSCAN
- **Time Series Forecasting for Electricity Data**:
  - Developed deep learning models (LSTM, GRU) to forecast short-term and long-term electricity consumption patterns.

    Implemented a sliding window approach for sequence prediction, optimizing for various time horizons.

    Languages/Technical usage: Python, TensorFlow/Keras, LSTM, Pandas, Scikit-learn, Matplotlib/Seaborn

Skills

- **Programming Languages**
  - Proficient in: Python.
- **Libraries and Frameworks**
  - Proficient in: TensorFlow/Keras, NumPy, Huggingface.
  - Experience with: LangChain, LlamaIndex, LangGraph, Gradio, Matplotlib, Scikit-learn, Pandas, W&B. FastAPI
- **Machine Learning**
  - Distributed Training & Inference, Generative AI, Natural Language Processing, Computer Vision, Unsupervised Representation Learning, Audio Processing, Multimodal Models, Video Understanding Models, Large Language Models, Parameter Efficient Fine-tuning, Reinforcement learning, Classification, Regression.
- **Software**
  - Containerization: Docker
  - CI/CD: Jenkins, GitLab CI/CD
  - Cloud Platforms: AWS, Azure, GCP (Basic knowledge)
  - MLOps Platforms: Amazon SageMaker
  - Version Control: Git, GitHub
  - Development Environments: Linux, Bash, Vim, VSCode, Jupyter
- **Project Management**
  - Agile Methodologies: Scrum
  - Project Management Tools: Jira

- **Languages**
  - Native: Arabic; Amazigh
  - Proficient: English, French

Education

- **PhD Candidate in Data Science and Machine Learning** — January 2023 – Present
  INPT (Institut National des Postes et Télécommunications) — Rabat, Morocco
- **Engineering Degree in Computer Science** — September 2016 – December 2019
  INPT (Institut National des Postes et Télécommunications) — Rabat, Morocco
- **Exchange Semester in Computer Engineering** — September 2018 – February 2019
  ÇUKUROVA UNIVERSITY — Adana, Turkey

Certifications

- **Aws Cloud Practitioner (In Progress)**
- **Rasa Developer Certification | RASA**
- **Intro to Amazon Web Services(AWS) Machine Learning | COURSERA**
- **IBM Data Science Professional | IBM - COURSERA**
- **Big Data-Level I | IBM - COGNITIVE CLASS**
- **Data Analysis Using Python | IBM - COGNITIVE CLASS**
- **Spark Machine Learning Library (MLlib) | IBM - COGNITIVE CLASS**
- **Neural Networks and Deep Learning | COURSERA**

Researches and Publications

[1] M. Ballouch et al., "**Enhancing Control in Manufacturing and Microgrid Systems: Deep Reinforcement Learning with Double Q-Learning**," 2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA), Casablanca, Morocco, 2023, pp. 1-7, doi: 10.1109/SITA60746.2023.10373737.

[2] M. Ballouch, et al., "**Forecasting Call Center Arrivals Using Machine Learning**", Osmaniye Korkut Ata University Journal of Natural and Applied Sciences, vol. 4, no. 1, pp. 96–101, 2021, doi: 10.47495/okufbed.824870.