Ali El Baha

Portfolio **(click here)** | Profile **(click here)**

(33) 605739544 | ali.elbaha@hotmail.com

I am an AI engineer. My passion for Data extends to Statistical Analysis, Inferential Statistics and Deep Learning.

Education

CentraleSupelec — Paris, France
Applied Mathematics and Computer Science — *2023 - 2024*

- Stochastic Optimization, Statistics, Deep Learning, Image Processing, Computer Science.

Centrale Casablanca — Casablanca, Morocco
Engineering Degree — *2022 - 2023*

- Applied Mathematics, Numerical Methods, Statistics, Optimization, Machine Learning, Quantum Physics, Statistical Physics, Energy Systems.

CPGE — Khouribga, Morocco
Preparatory Classes MP — *2020 - 2022*

- Two-year post-baccalaureate program in advanced mathematics and physics leading to entrance exams for *Grandes Écoles*.

Experience *[Analytical and synthesis skills - Autonomy - Communication]*

Research Data Scientist | *Société générale* — Oct.24 – La Défense, Paris

- Conduct of a comprehensive literature review on fuzzy matching, phonetic transcription algorithms, and graph-to-phoneme conversion using state-of-the-art deep learning techniques.
- Enhancement of review protocols for name matching models to identify and mitigate risks associated with model limitations.
- Implementation of deep learning-based algorithms for performing relevant stress tests, including phonetic-based evaluations, on name matching models.
- Production of synthetic technical documentation and the presentation of results.

Deep Learning Researcher | *Laboratoire de Mécanique Paris-Saclay (LMPS)* — Gif-sur-Yvette, Île-de-France, Feb.24 – June.24

- Study of recurrent neural network (RNN) architectures, with particular emphasis on Long Short Term Memory (LSTM) networks, in the context of processing time series data.
- Evaluation of generalized adversarial representation networks (RepGAN) in order to test their data generation and signal reconstruction capabilities.
- Implementation of a new approach integrating the LSTM model with the RepGAN model, with the aim of predicting the seismic response of historic monuments
- Comparative performance analysis between the RepGAN model incorporating an LSTM and one incorporating a Transformer.

Data Scientist | *NORMA* — Oct.23 – Jan.24, Casablanca

- Development of "Juridico", a legal assistant that meets a major market need for automated analysis and contract management.
- Design of a scalable AI solution using the "Falcon" language model and RAG technology with the LangChain framework, tailored to the needs of business leaders and legal professionals.
- Fine-tuning of *Falcon* LLM to enhance the model's ability to interpret legal documents with precision.

Software engineer Intern | *OCP Group* — July.23 – Sept.23, Khouribga

- Development of an interface, programmed in Python, over a complex VBA-based Excel spreadsheet, to facilitate essential data collection and entry tasks.

Volunteering activities *[Pedagogy - Rigor]*

Centrale Tech Association

* Preparation of theoretical courses on AI for first-year students.
* Organization of workshops on solving real industrial problems using AI techniques.

Technical Skills

Languages: French (C1), English (Professional, Toeic: 920/990), Arabic (native), Japanese (beginner).

Tools: Python, MATLAB, R, SQL, Power BI, Office Suite, Git, VS Code, Tensorflow, PyTorch, Scikit-learn, Pandas, Seaborn, NumPy, Flask, Streamlit, Dash.

Projects *[Teamwork - Project Management]*

Data Analytics for the Supply Chain | *Power BI, Power Query, scikit-learn* Jan. 2024

- Cleaning and preprocessing a dataset comprising 2 million parcel deliveries, ensuring data reliability by removing anomalies and correcting missing values.
- Using Power BI to create interactive visualizations, identify seasonal trends, peak delivery periods, and operational improvement areas.
- Conducting statistical analyses to understand the impact of different variables on delivery success rates and identify obstacles in the delivery process.
- Applying clustering techniques to segment delivery areas based on demand, optimizing route planning and reducing delivery times.

Computer Vision for the Visually Impaired | *Tensorflow, Computer Vision, LLMs, TTS, Streamlit* Dec. 2023

- Designing smart glasses capable of real-time environment detection and description, providing users with helpful assistance for navigation
- Implementation of DETR-ResNet-50 for precise object detection in near real-time scenarios.
- Integration of the Global-Local Path Networks (GLPN) model, trained on NYUv2, for accurate monocular depth estimation.
- Use of the LLM for real-time environment description and user guidance towards their desired destination.

Predictive Maintenance for PV Panels | *Tensorflow, Data analysis, Power BI* Oct. 2023

- Used a deep learning approach for forecasting PV system power output, using data from a 9 MWp grid-connected plant.
- Built and trained separate LSTM models for forecasting power output and for anomaly detection using autoencoder architecture.
- Achieved a high level of explanatory power in power output forecasting ($R^2 = 0.98$) and demonstrated effective signal reconstruction with low mean squared error ($MSE = 0.02$)

NLP for UX | *Python, SST, Flask, Generative AI* Sept 2022 – June 2023

- Design and development of a system for managing feedback from public transportation users.
- Use of Language Models (LLMs) to process and classify passenger feedback based on urgency.
- Support for text and audio formats in three languages: English, French, and Arabic.
- Implementation of Flask for the system's backend to manage requests and responses between the user interface and the server.

Certificates

Project Management certification | *Google Project Management*

Scrum Certification | *Scrum Master*

Modeling Software Systems using UML | *The Hong Kong University of Science and Technology*

Client Needs and Software Requirements | *University of Alberta*