Christian **NGUIEPE SEGNOU**

AI & ML & LLMs Engineer,

Data Scientist, Trainer

L'Haÿ-les-Roses, Île-de-France

Phone : 0666484774

Email: nguiepemarius@gmail.com

LinkedIn: https://www.linkedin.com/in/christian-segnou-ph-d-4461b2102/

Github : https://github.com/ChrisDMI

Portfolio: https://chrisdmi.github.io/Portfolio-christian-segnou-2023/

CV RAG LLM approach: https://christian-segnou-cv.streamlit.app/

PROFESSIONAL SUMMARY

I am a **LLM Engineer** with extensive experience in developing and deploying **Large Language Models** and advanced **NLP** solutions. Specializing in model fine-tuning, deployment, and optimization, I leverage deep expertise in **Python**, **PyTorch**, and state-of-the-art transformer architectures. My industry background enables me to solve complex AI challenges, including intelligent extraction and intelligent conversational agents. As an educator in advanced NLP and LLM technologies, I have successfully guided professionals in applying cutting-edge techniques to real-world applications. I am passionate about driving innovation in AI and eager to contribute my skills to impactful projects.

SKILLS & TECHNICAL PROFICIENCIES

- Natural Language Processing (**NLP**)
- Computer Vision (**CV**)
- Large Language Models (**LLMs**) and Transformers
- Deep Learning and Neural Networks
- **Python**, **PyTorch**, **TensorFlow**, **Hugging Face Transformers**
- Fine-tuning and Deployment of **LLMs**
- Vector Databases and Retrieval-Augmented Generation (**RAG**)
- Machine Learning: Supervised & Unsupervised Learning
- Data Pipelines and ETL Processes (**Airflow**)
- Real-Time Monitoring and Debugging (**Kibana**)
- Data Visualization: **Pandas**, **Matplotlib**, **Seaborn**, **Plotly**
- CI/CD (**GitHub Actions, Docker, Kubernetes, MLFLOW**)
- Cloud Services: **AWS** and **Azure**
- Teaching & Mentoring: AI, NLP, Computer Vision, Machine Learning, Deep Learning

PROFESSIONNAL EXPERIENCES

Expleo Group, Montigny-le-Bretonneux, IDF 09/2022 –

NLP Engineer | Computer Vision Engineer R&D

Project 1: LLM Chatbot for Document Interaction

Situation: Develop a chatbot powered by Large Language Models to interact with Airbus concession document content efficiently.

Task: Develop and deploy an LLM-based chatbot capable of understanding and responding to user queries based on document content.

Actions:

- **Object Detection + OCR applied on PDF document for text extraction:**
  - Developed and optimized a **YOLO**-based model for detecting key areas in Airbus PDF documents, improving the accuracy and efficiency of information extraction.
  - Applied advanced **OCR** models to pre-process image crops and perform precise text recognition from scanned documents, including handwritten text.

- **Developed LLM-Based Chatbot:**
  - Fine-tuned Large Language Models using domain-specific data to improve chatbot relevance.
  - Implemented Retrieval-Augmented Generation (RAG) techniques to enhance the chatbot's ability to provide accurate information from documents.
- **Integrated with Existing Systems:**
  - Deployed the chatbot using Streamlit for UI and FastAPI for seamless integration.
  - Ensured scalability and reliability using AZURE services and containerization with Docker

- **Implemented Monitoring and CI/CD:**
  - Set up real-time monitoring using Kibana to track chatbot performance and user interactions.
  - Established CI/CD pipelines using GitHub Actions for automated testing and deployment.

Results:

- Delivered a responsive chatbot that improved user engagement by 40%.
- Streamlined access to document information, reducing response times by 50%.
- Enhanced system reliability through effective monitoring and continuous integration.

Technologies: Python, PyTorch, Hugging Face Transformers, Streamlit, FastAPI, Docker, Azure, GitHub Actions, Langchain, AzureOpenAI

Expleo Group, Montigny-le-Bretonneux, IDF 09/2022 –

NLP Engineer | Computer Vision Engineer R&D

Project 2: Automated Equipment Management using OCR and Computer Vision

Situation: The logistics department faced significant challenges in managing and tracking equipment borrowed and returned by technicians over the past six years. They had accumulated 11,000 scanned PDF files of signed equipment loan forms, making it nearly impossible to manually inventory the equipment associated with each technician.

Task: Develop an automated solution to extract and structure key data from the 11,000 PDF files to efficiently inventory all equipment borrowed by each technician since joining the company.

Actions:

- **Object Detection with YOLOv8:**
  - Developed an object detection model using YOLOv8 to identify areas of interest in the scanned documents, including technician names, manager names, intervention sites, lists of borrowed equipment, and signatures.
- **Text Extraction with Tesseract OCR:**
  - Extracted textual information from the detected areas using Tesseract OCR.
  - Employed advanced image processing techniques with OpenCV to handle variations in table formats and layouts, enabling accurate extraction of equipment lists.
- **Automation of PDF Processing:**
  - Implemented an automated pipeline using **Airflow** to process all 11,000 PDF files, extract the information, and structure it in a relational database.
- **Deployment on a Streamlit Application:**
  - Developed and deployed a user-friendly web application using **Streamlit**, allowing users to upload additional PDF files.
  - Enabled real-time processing of new files, automatically updating the database with newly extracted information.
- **Database Management :**
  - Created a relational database to store and manage the extracted data, facilitating quick and efficient searches by technician.

Results:

- Successfully processed all 11,000 PDF files in record time, populating a comprehensive database for all technicians.
- Enabled the logistics department to easily query the database for accurate and detailed inventory of equipment borrowed by each technician.
- Streamlined the process of adding new records, ensuring continuous database updates without manual intervention.
- Improved equipment management efficiency by **40%**, reducing time spent on manual tracking.

Technologies: YOLOv8, Tesseract OCR, OpenCV, Python, Streamlit, Airflow, Relational Databases, Docker, Git, FastAPI

University of Paris-Saclay, Orsay, IDF 09/2019 – 09/2021

Python Instructor

Educator in Python and Machine Learning Foundations

Designed and delivered comprehensive Python programming courses for undergraduate students (L1 to L3), covering key programming concepts, data manipulation, and algorithms.

Introduced and explained foundational concepts of machine learning and deep learning, facilitating the transition from basic programming to advanced AI topics.

Mentored students on their final projects, providing guidance on problem-solving, coding best practices, and project management.

Evaluated student performance through assessments and offered constructive feedback to deepen their understanding and improve their skills.

Promoted active learning through the development of interactive course materials, enabling students to engage with practical coding challenges and real-world applications.

Technical Environment: Python, Machine Learning, Deep Learning, Pedagogy

C2N, Palaiseau, IDF 10/2018 – 06/2022

PhD Candidate – Project Lead

Lead Researcher in Magnetic Domain Dynamics and Machine Learning

Led a research project aimed at understanding the dynamics of magnetic domain wall motion to improve the performance of hard drives.

Developed and implemented image processing techniques using magneto-optical Kerr microscopy for precise measurement and analysis of domain wall movement.

Conducted advanced data analysis to study the physical behavior of magnetic domain walls, leveraging Python for data handling and visualization.

Built mathematical models to simulate and predict the dynamics of domain wall motion, contributing to the development of next-generation magnetic storage solutions.

Applied machine learning algorithms to predict domain wall velocities, improving the accuracy of dynamic simulations and expanding the understanding of material properties.

Technical Environment: Python, Scikit-learn, Pandas, OpenCV, ImageJ, Git, Numpy, Scipy, Matplotlib, Sympy

EMPLOYEES TRAINING EXPERIENCES

Orinoko, Remote 09/2024 –

NLP and LLM Instructor

- Delivered comprehensive training programs focused on Large Language Models (**LLMs**) and **NLP** using **PyTorch**, teaching employees how to fine-tune and deploy models.

- Integrated the **OpenAI API** to demonstrate the use of **GPT-4** in real-time **NLP** applications, enabling learners to leverage advanced language models for dynamic content generation.

- Developed curriculum covering key aspects such as **LLM** architecture (**Transformers**, **Attention Mechanisms**), data preprocessing for **NLP**, and model quantification.

- Guided participants through practical projects including dataset construction, **fine-tuning LLMs**, and **deploying models** in local and **cloud environments**.

- Introduced advanced techniques such as Retrieval-Augmented Generation (**RAG**) and vector databases to enhance **LLM** capabilities for real-time data generation and retrieval.

- Technologies: PyTorch, TensorFlow, Python, Hugging Face Transformers, Vector Databases

AbilyCare, Remote 03/2024 –

Machine Learning and Deep Learning Instructor

- Designed and led **machine learning** training programs, including supervised and unsupervised learning, focusing on the application of decision trees, random forests, support vector machines, and clustering techniques.

- Taught practical skills in data preprocessing, visualization, and model evaluation using Python libraries like Pandas, Seaborn, Matplotlib, and Scikit-learn.

- Delivered hands-on lessons in **deep learning**, including neural networks and transfer learning with TensorFlow and Keras, and advanced topics like RNNs, LSTMs, and Transformers.

- Technologies: Python, Scikit-learn, TensorFlow, Pandas, Matplotlib, Seaborn, Plotly.

EDUCATION

Year	Degrees	School	Speciality
2022	PHd	University of Paris-Saclay	Material Sciences
2021	Master degree	Jedha	Data Science Developer and Designer
Languages	French (Native speaker), English (Professional)		

CERTIFICATIONS

- **Data Science Developer and Designer** – https://www.credential.net/8ae1ca89-eab2-44a6-9bed-98fe27562a77#gs.5fw763
- **Hands-on to Linux Commands and Shell Scripting** - https://coursera.org/verify/AUH4LGP2NCT4
- **Relational Databases (RDBMS) Essentials** - https://www.credly.com/badges/4ae1e959-dbfd-4518-9178-73cea91b9edf/linked_in_profile
- **Introduction to Relational Databases (RDBMS)** - https://coursera.org/verify/2B8WSU3F7V74
- **Relational Database Administration (DBA)** - https://coursera.org/verif y/MSBMTEKGBTHC
- **Databases and SQL for Data Science with Python** - https://coursera.org/verify/8EGZSDB8QRK3
- **Data Engineering Essentials** - https://www.credly.com/badges/2c1c40e5-8e58-4cbd-bb61-fc0cf2d975ee/linked_in_profile
- **Introduction to Data Engineering** - https://coursera.org/verify/WKEGFMRJXYBK
- **Python for Data Science, AI & Development** - https://www.coursera.org/account/accomplishments/verify/FFK2VE8FG4QN
- **Python Project for Data Engineering** - https://coursera.org/verify/BG8EXTHBUBSP
- **ETL and Data Pipelines with Shell, Airflow and Kafka** - https://coursera.org/verify/ZWPY485NAFQH
- **HTML, CSS, and Javascript for Web Developers** - https://www.coursera.org/account/accomplishments/verify/3MFJVEVM8SX7