EPHRON MARTIN

ephronmartin2016@gmail.com| +91 7025960589 |www.linkedin.com/in/ephron-martin
EphronM (Ephron Martin) · GitHub | https://medium.com/@ephronM

SUMMARY

Passionate data scientist proficient in Python and analytics, skilled in ML algorithm development and deployment. Seeking dynamic projects to drive growth and contribute effectively.

PROFESSIONAL EXPERIENCES

★ **Data Scientist Engineer | Canwill Technologies** May 2024 - present

Client Canopyco.io | AI solutions for data breach response and privacy audits

- ➢ Government ID (SSN) card extraction by fine tuning Paddle OCR for custom data, with image preprocessing pipelines. Performance comparison with paid OCR with inhouse fine tuned OCR.
- ➢ Fine Tuning vision multi-model for documents intelligence and relationship mapping

★ **Research Engineer | Buddi.AI (A Claritrics Company)** April 2022 - May 2024

- ➢ Specialize in NLP for PHI masking, data extraction, and sentiment analysis. Expertise in reverse engineering and fine-tuning Transformer-based models.Develop end-to-end machine learning solutions using Scala and Python.Proficient in agile project management methodologies.

KEY CONTRIBUTIONS:

NER Parser for EHR chart for Entity extraction and PHI masking

- ➢ Enhanced NER model performance from 75% to 92% F1 score. Combined pre-trained DistilBert with CRF.
- ➢ Used ONNX for model quantization and CPU server deployment.Implemented BiLSTM classifier for chunk classification, reducing latency to 1 second per document. Integrated with Scala Play Framework API.
- ➢ Utilized MLflow for experiment tracking and ZenML for CI/CD pipeline with DVC for version control.

Tools Used: Python, PyTorch, pandas, Azure OCR, DataBricks, MLflow, ZenML, HuggingFace

Sentence Similarity (Linguistic as well as contextual)

- ➢ Proposed research insight: Developed a system to prioritize medical and linguistic terms in medical documents. Validated contextual terms with Bio-Bert and linguistic terms with Meteor score for enhanced accuracy.

Tools Used: Scala, Docker, Play-Framework, MySql, Bio Bert, Tensorflow

★ **Service Engineer | Thyrocare Technologies Pvt Ltd** March 2021 - April 2022

- ➢ Coordinated and managed operations of a regional processing laboratory, overseeing administrative tasks including procurement and tendering for equipment maintenance.
- ➢ Collaborated with vendors to ensure efficient lab services and support.

SOFTWARE/TECHNOLOGY SKILLS

- ➢ Retraining LLM models for custom chatbots using **LangChain** and **RAG** for better document retrieval
- ➢ **Machine Learning** and **Deep Learning** algorithms, **Data Cleaning, Integration, Transformation**, and **Dimensional Reduction. Statistics** and **Graphical tools** to analyse data representation and distribution.
- ➢ Programming And DB tools**: Python, scala, dockers, NumPy, Pandas, Scikit-learn, git version control, DVC, Azure, AWS, Snowflakes, data bricks, MySQL**

BLOGS PUBLISHED

★ **Question Answering with RetrievalQA Chain - Blog Link**

➢ Explored Question Answering using RetrievalQA Chain methodology, focusing on Map-Reduce, Refine, and Map-Ranking chain types to improve answer accuracy and relevance.

➢ Highlighted the significance of Conversational Memory in ConversationalRetrievalChain for context retention and tracking past conversations.

★ **A Beginner's Guide to Retrieval-Augmented Generation (RAG) - Blog Link**

➢ Investigated document retrieval methods, emphasizing the Lang Chain and Llama2 LLM models to improve search capabilities. Explored VectorDB with Chroma for streamlined document indexing and retrieval

➢ Covered key concepts including Similarity Search, MMR, Contextual Compression, and Self Querying to optimize search results and relevance.

MACHINE LEARNING PROJECTS

★ **Ricky - Interdimensional Personal Assistant Chatbot - App Link** Repo Link

➢ Developed an advanced personal assistant chatbot with Mistral LLM, embodying the character and persona of Rick Sanchez from the famous TV series "Rick and Morty."

➢ Containerized the chatbot with Docker, establishing a CI/CD pipeline via GitHub Actions for automated deployment. Utilized AWS ECR to host Docker images on EC2 instances for public access. Integrated continuous monitoring with Comet ML for real-time chatbot performance and interaction analysis.

★ **Invoice Extraction System using Gemini Vision Pro - App Link** Repo Link

➢ Developed an Invoice extractor using Gemini Vision Pro to extract vital invoice details from images and text, Dockerizing it for streamlined deployment. Integrated GitHub Runners for automated builds and AWS Elastic Container Registry for hosting on EC2 instances.

★ **Custom Dataset Fine-tuning of Llama 2 Model with PEFT - QLoRA -** Repo Link

➢ Retrained Llama 2 model on custom dataset using parameter-efficient fine-tuning (PEFT) techniques like QLoRA, quantizing the 7b parameters model to 4-bit precision using T4 GPU on Colab. Published the fine-tuned model on the Hugging Face repo for public access.

AREAS OF INTEREST

➢ LLM models and RAG

➢ Natural language processing (NLP)

➢ Optimizing application designs for Business models

CERTIFICATIONS

➢ Data Science with Python, Simplilearn	2021
➢ IBM Data Science Professional Certification	2021

EDUCATION

- **Bachelor of Technology (B Tech) in Applied Electronics and Instrumentation** 2016- 2020