aarungade@gmail.com
+31 649284755
+91 7382584969

Arun Gade

Hyderabad
India

linkedin.com/in/arun-gade-490448b7/ | github.com/running-machin

Experience

AI Developer **Dataorbit.ai** London (Remote)**| Dec 2024–Present**

- Tech Stack: Python, LangChain, LangGraph, LLM
- The **Company** consults AI and data solutions to Finance sectors (especially Asset Management firms)
- Developed and designed LLM-powered tools for Fintech, implementing robust architectures to handle increasing data volumes, significantly streamlining finance analysts' workflows
- Leveraged KG/vector retrieval and agentic AI for automated financial and ESG report creation, resulting in a 30% reduction in report generation time and a 10% improvement in data accuracy.

Project Developer **L&T labs (Law and Tech Labs)** Maastricht, NL (Hybrid) **| Sep 2022–Jun 2023**

- The **Company** intends to integrate AI and ML technologies into legal processes and documentation.
- Managed the ETL pipeline for the EU-funded Case-Law Explorer, improving data processing efficiency by 20% and reducing data errors by 15% through close collaboration with data researchers.
- Ensured and enhanced pipeline efficiency and reliability through close collaboration with data researchers

Research Intern **L&T labs** Maastricht, NL (Hybrid) **| Jun 2021–Jun 2022**

- Tech Stack: Python, Pandas, NLP, PyTorch, spaCy, Deep Learning, Web Scraping, Information Retrieval
- Developed a system for automated retrieval of relevant statutory provisions
- Trained a novel legal domain model from an extracted, expert-labeled dataset of Dutch legal questions and articles, infused with French legal data.

Technologies

- **Programming & ML Stack:** Python, SQL, PySpark, TensorFlow, PyTorch, LangChain, LangGraph, Pandas, OpenCV, SpaCy, Java, JavaScript, Amazon S3, Amazon DynamoDB, Docker, Git
- **AI & Data Engineering:** ETL pipelines, Data Analysis, Data Visualization, Apache Airflow, Data Gathering and Labeling, Data Masking, Machine Learning, Deep Learning, NLP, Computer Vision, Generative AI, LLM

Project Experience [Please check my GitHub link for more (and personal) projects]

ESG and SFDR Solution Platform London (Remote) **|Dec 2024 - Present**

AI developer, Dataorbit.ai

- Tech Stack: Python, LangChain, LangGraph, LLM
- Spearheaded the implementation of a Retrieval-Augmented Generation (RAG) system for financial document search and analysis, combining vector similarity search with knowledge graph traversal to deliver accurate, context-aware insights.
- Built a unified knowledge representation layer by transforming unstructured financial documents into a structured knowledge graph and high-fidelity vector embeddings using state-of-the-art language models (e.g., BERT, Sentence-Transformers).
- Designed and deployed autonomous AI agents for mission-critical financial workflows, including:
  - Automated business and financial report generation
  - ESG (Environmental, Social, Governance) screening and compliance analysis
  - GRI (Global Reporting Initiative) and SFDR (Sustainable Finance Disclosure Regulation) report generation
  - Document intelligence agent for content extraction, classification, and summarization

Case Law Explorer Maastricht **|2024**

Project Developer, L&T Labs

- Tech Stack: Python, Boto, Apache Airflow, Docker, Pandas, AWS S3, AWS DynamoDB, SPARQL, ETL
- Designed and implemented scalable ETL pipelines to extract legal data and associated metadata from government websites and APIs, ensuring comprehensive data collection and traceability.
- Orchestrated pipeline workflows using Apache Airflow to automate scheduled data ingestion, improving reliability and reducing manual intervention by 90%.
- Developed and published three internal Python packages to standardize and streamline data extraction from diverse legal sources, enhancing reusability and team productivity.
- Engineered data cleaning and transformation processes to normalize unstructured legal data, enabling consistent loading into AWS DynamoDB

- Structured and stored processed data in AWS S3 as graph-ready datasets, supporting downstream training of Graph Neural Network (GNN) models for legal relationship analysis.

Statutory Article Retrieval Dataset in Dutch and French Maastricht **|2024**

Intern, L&T Labs

- Created BSARD_v2, a Dutch statutory article retrieval dataset with 598 expert-annotated legal questions linked to 20,576 Belgian legal articles, enabling citizen-focused legal information access.
- Built and evaluated dense retrieval models (Siamese, Dual Tower) using Dutch BERT variants (BERTje, RobBERT), outperforming BM25 with 64.9% Recall@100 and 0.388 MRR@100.
- Engineered a multilingual legal retrieval system by combining Dutch (BSARD_v2) and French (BSARD_v1) datasets, leveraging mBERT and DistilBERT for cross-lingual performance.
- Published and open-sourced the dataset and models, contributing to legal NLP research in low-resource, multilingual settings.

Fairness And Bias in Multimodal Summarization (Thesis) Maastricht University**|2023**

Masters Student, Maastricht University

- Tech Stack: Object detection, Gender estimation, Image captioning, Multi-Modal Summarization, Scene detection
- Evaluated fairness in multimodal video summarization models (Transformer, VASNet, SUMGAN, PGL-SUM) across gender and individual representation using the FVS dataset, identifying systematic biases in scene and object selection.
- Developed and applied SumBal, a quantitative metric to measure fairness in summary outputs, revealing near-zero gender balance violations in some models while exposing significant disparities in individual representation.
- Analyzed bias propagation from datasets (TVSum, SumMe) to models using grounded WEAT tests, showing minimal effect size ($d = 0.027$, $p = 0.448$), indicating limited but detectable gender bias in embeddings.
- Built a comprehensive bias audit framework combining computer vision models (YOLO, Mivolo) for scene, object, and gender annotation, enabling fine-grained fairness evaluation in visual summarization.
- Demonstrated model-specific biases : e.g., VASNet over-represents “cabins/farms” for women; Transformer amplifies male dominance in transportation; SUMGAN shows more balanced but still skewed distributions.

Education

Masters of Science(M.Sc.) **Maastricht University** Maastricht, Netherlands **| 2021–2024**

- Major: Artificial Intelligence - NLP and Computer Vision specialist

Bachelors of Technology(B.tech) **SRM University** Chennai, India **| 2014–2018**

- Electronics And Communication Engineering

Awards

- **Audience Winner:** Most Voted Solution for Data hackathon Healthy Brabantine City 2022
- **O(DACS) hackathon series :** 16th place out of 50+ for O(DACS) hackathon 2023 series organized by FieldLab Robotics, Maastricht University and CoRosect Project
- **Cancer Detection Hackathon:** 15th place out of 50+ for WIDS Maastricht Conference 2024, Independent event organized by the Institute of Data Science at UM