

# Joel Dsouza

Mumbai, India | [joeldzs44@gmail.com](mailto:joeldzs44@gmail.com) | +91-7506433898 | [LinkedIn](#)

## EXPERIENCE

### GEP

#### Data Scientist

Mumbai, India

Oct 2024 – Present

#### Senior Software Engineer

Jul 2024 – Sep 2024

- Built a config-driven Named Entity Recognition (NER) solution using pre-trained **SpaCy** models to extract contract types and other metadata.
- Created an annotated multilingual contract dataset for benchmarking precision scores, utilizing an internally hosted **LLaMA3** model with high-quality prompts for accurate annotations.
- Engineered impactful prompt designs for metadata extraction and clause identification in contracts, contributing to a high-stakes client demo delivered under tight deadlines using **LLMs**.
- Enhanced rule-based metadata extraction systems for contract documents by debugging pattern failures, designing new regex patterns, and updating existing ones to improve accuracy across diverse client contracts.
- Debugged and improved chatbot functionalities using **RAG**, **LangChain**, and **QdrantDB** to enhance natural language interaction capabilities for contract-related queries.
- Developed Proof-of-Concept **Streamlit** apps for **OCR** use cases, evaluating various extraction methods such as table-GPT, **transformer** models from **Hugging Face**, **LLMs**, and **Azure Vision Services**.
- Increased code reliability by writing comprehensive test cases using the **PyTest** framework, achieving over 95% code coverage.
- Collaborated with cross-functional teams, product managers and other stakeholders to ensure quality and timely deliverables, following standard Agile practices with tools like JIRA, Confluence, Git, and Azure DevOps pipelines.

#### Software Engineer

Jul 2022 – Jun 2024

#### Associate Software Engineer

Jun 2021 – Jun 2022

- Engineered an end-to-end data ingestion solution using Azure cloud, increasing dashboard refresh frequency by up to 8x.
- Automated ETL processes with Azure Functions timer triggers and Python scripts and Power Automate, eliminating manual data capture and saving up to 10 hours a week.
- Developed and maintained over 10 Power BI Reports & Dashboards with 500+ unique views weekly.
- Managed a shared capacity Power BI workspace and Semantic Models' Data Refresh and configuration.
- Led a 3-member team of developers, fostering a collaborative team environment, encouraging knowledge sharing and skill development thus boosting overall productivity by 20% through effective backlog management and daily reviews.
- Collaborated with stakeholders to understand and meet their data analysis requirements, ensuring project alignment and success demonstrating strong communication skills.
- Documented comprehensive Wikis and Confluence for organization wide KPIs and dashboards.

### Anaxo Tech

Mumbai, India

#### Web Development Intern

Aug 2020 – Jan 2021

### CMP Infotech

Mumbai, India

#### Web Development Intern

Dec 2019 – Dec 2019

## SKILLS

**Programming & Development:** Python, FastAPI, NER, SpaCy, GLiNER, OCR, Streamlit, LLM, Pandas, NumPy, PyTest, Transformer Models, Azure Vision, Asyncio, Multiprocessing, Docker.

**Data Engineering & ETL:** SQL, Azure Functions, Azure Data Factory, MongoDB, ETL, Data Modelling

**Data Visualization & Analytics:** Power BI, Data Visualization, Excel

**Other Tools:** Power Automate, Git, JIRA, Confluence

---

## EDUCATION

---

### Bachelor of Engineering in Computer Science

Xavier Institute of Engineering • Mumbai, India • 2021

---

## CERTIFICATIONS

---

[\[PL-300\] Microsoft Certified: Power BI Data Analyst Associate](#)

Microsoft, June 2024

[Introduction to NoSQL Databases](#)

Coursera, May 2024

[Core Data Concepts in Microsoft Azure](#)

Coursera, Apr 2024

[Microsoft Azure SQL](#)

Coursera, Apr 2024

[Python for Everybody Specialization](#)

Coursera, June 2021

[MTA: Software Development Fundamentals](#)

Microsoft, Oct 2019