1 RAG Projects

1.1 Ngenux Leave Policy

Description: This Leave Policy outlines Ngenux's comprehensive guidelines for employee time-off, supporting work-life balance. It details eligibility, entitlements, and procedures for various leave types, including Casual, Sick, Privilege, Maternity, Paternity, Adoption, and Sabbatical Leave. Additional provisions cover Loss of Pay (LOP), Compensatory Offs, Recall on Duty, and Work from Home (WFH) policies. The policy defines rules for accrual, carry-forward, encashment, and documentation, emphasizing managerial approval and compliance via KEKA. It ensures flexibility while maintaining operational needs and accountability. Leave is managed transparently by HR and governed under standardized rules, promoting employee welfare and organizational efficiency.

Sample Questions:

- What is the total annual leave entitlement for permanent employees at Ngenux, and how is it divided between Privilege Leave and Casual/Sick Leave?
- Under what conditions can an employee avail Paternity Leave, and what documentation is required if the leave is for a legal adoption?
- What are the requirements for an employee to apply for Maternity Leave, including the advance notice period and any restrictions on combining it with other types of leave?
- How does the company handle unavailed Privilege Leave when an employee leaves the organization, and what is the maximum accumulation limit for such leave?
- What are the eligibility criteria and approval process for availing Sabbatical Leave, including the minimum service period and the purpose for which it can be granted?

1.2 Tesla's Transforming ESG into Impact Measurement

Description: Current ESG methodologies prioritize investment risk over real-world impact, often misrepresenting companies' environmental and social contributions. This document critiques these flaws, advocating for a shift to measuring tangible positive outcomes, like accurate use-phase emissions in automotive industries. It highlights Tesla's approach, emphasizing sustainable products, transparent data, and equitable workplaces. By 2030, Tesla aims to sell 20 million EVs annually, significantly reducing CO2e emissions. The report calls for stakeholder collaboration to redefine ESG as Impact, focusing on lifecycle emissions, employee welfare, and ethical supply chains to drive meaningful global change.

Sample Questions:

- What are the main flaws in current ESG evaluation methodologies?
- How does Tesla's approach to measuring vehicle emissions differ from traditional ESG reporting?
- What are Tesla's goals for electric vehicle sales and energy storage by 2030?
- How does Tesla ensure pay equity and competitive compensation for its employees?
- What measures does Tesla take to uphold human rights in its supply chain?

1.3 AutoML-Agent Framework Paper

Description: The ”AutoML-Agent: A Multi-Agent LLM Framework for Full-Pipeline AutoML” paper introduces a groundbreaking framework that automates the entire machine learning pipeline, from data retrieval to model deployment, using large language models (LLMs). Unlike traditional AutoML systems requiring technical expertise, AutoML-Agent leverages a multi-agent architecture with specialized LLMs for tasks like data preprocessing, model selection, and hyperparameter optimization. It employs retrieval-augmented planning and multi-stage verification to enhance efficiency and accuracy. Extensive experiments across seven tasks and fourteen datasets demonstrate its superior success rate and performance, making it accessible for non-experts and efficient for diverse ML applications.

Sample Questions:

- What is the main goal of the AutoML-Agent framework?
- How many downstream tasks were used to test AutoML-Agent?
- What are the names of the five data modalities covered in the experiments?
- Which LLM model is used as the backbone for most agents in AutoML-Agent?
- What is the purpose of the multi-stage verification process in AutoML-Agent?

2 Text2SQL Projects

2.1 HealthQuery

Description: HealthQuery is a Text2SQL project designed to streamline querying of a PostgreSQL database for pharmaceutical operations. It leverages a schema with tables like patients, visits, diagnoses, treatment_plans, doctors, visit_diagnosis, and treatment_doctors to enable natural language queries for healthcare insights. Using advanced LLMs, HealthQuery translates user inputs into accurate SQL queries, prioritizing precision for tasks like patient visit tracking and treatment analysis. Deployed on AWS with FastAPI, it ensures scalability and speed, supporting medical staff in retrieving critical data efficiently for improved decision-making and operational efficiency in healthcare settings.

Sample Questions:

- List all patients’ full names and their date of birth.
- Get the details of all doctors who specialize in cardiology.
- Show all visits that happened after January 1, 2024.
- How many diagnoses are recorded in the database?
- List all unique diagnosis codes in the system.
- Find the number of visits each patient has had.
- List patients along with their latest diagnosis date.
- Get all treatment plans along with the doctors assigned to them.
- Which patients have been diagnosed more than 3 times?
- For each doctor, count how many treatment plans they are involved in.

2.2 ACME Chatbot Database Schema

Description: The `acme_chatbot` schema in PostgreSQL supports ACME's pharmaceutical operations with nine tables. `demand-forecast` predicts product demand for inventory planning. `material-packing-tracker` manages component inventory. `otif` and `kpi_invoice_to_order_calc` track order fulfillment and delivery KPIs. `manuf-delay` and `batch-and-packing-tracker` monitor production delays and batch timelines. `manufacturing-` assesses equipment utilization. `doh` calculates inventory days-on-hand for stock management. `process-area-details` evaluates manufacturing efficiency. Together, these tables enable supply chain optimization, production planning, and performance analysis for ACME's pharmaceutical manufacturing and distribution processes.

Sample Questions:

- What is the current OTIF (On-Time In-Full) score by region/product?
- Which SKUs are consistently missing OTIF targets?
- What are the top 5 reasons for OTIF failure this month?
- Can you show fulfillment gaps by plants?
- Which 3PL providers have the highest SLA breach rates?
- What is the forecast vs. actual demand delta by product family?
- Are current inventory levels aligned with next month's demand?
- Are there any material shortages impacting production?
- What is the current Days on Hand (DOH) for product SKU123?
- Are there any SKUs at risk of stockout this week?
- Which SKUs are showing abnormal demand spikes?
- Are current inventory levels aligned with next month's demand?
- What were the DOH trends for the past quarter?
- Which batches are delayed in manufacturing right now?