Luka Škodnik, Robert Jutreša

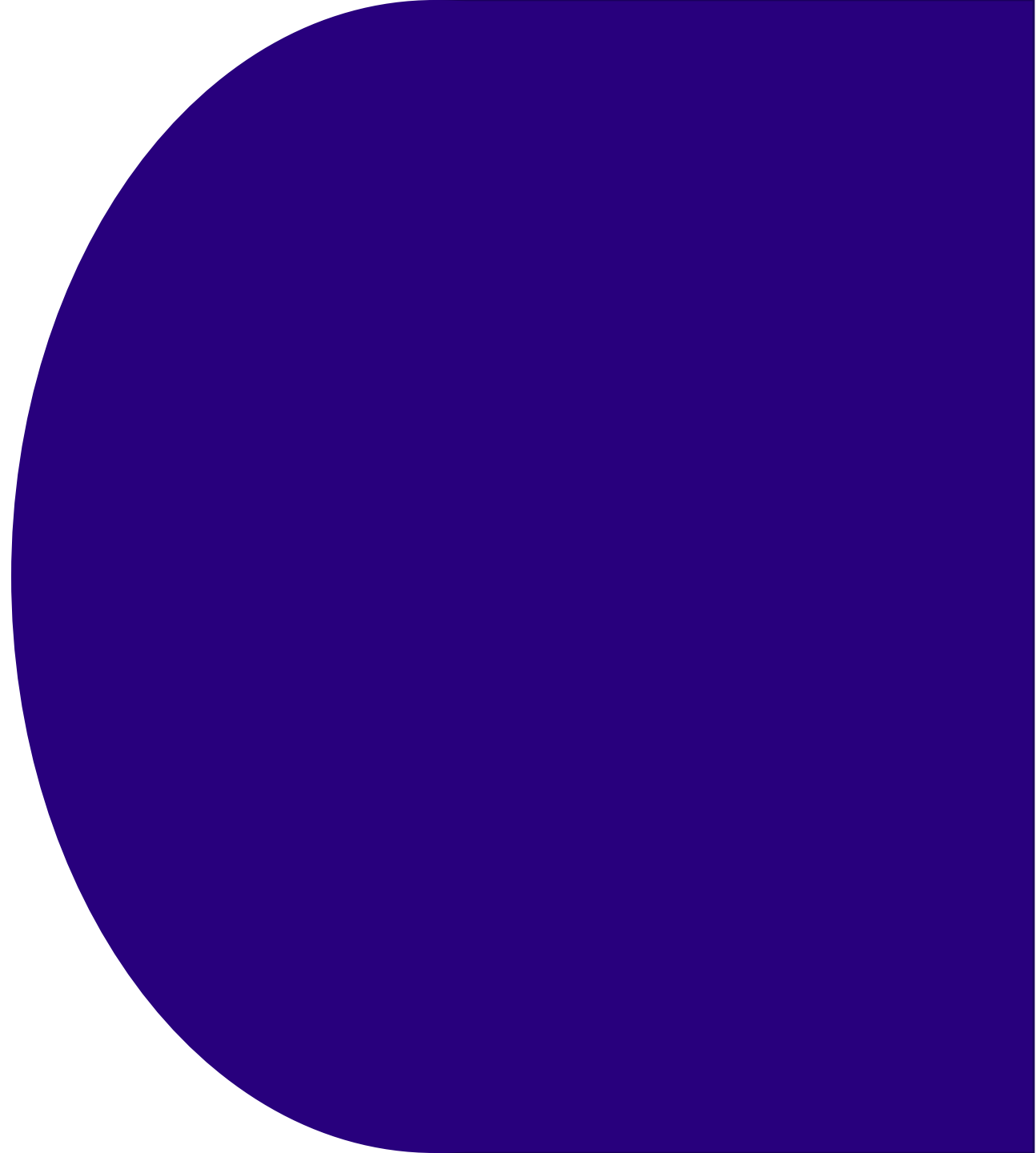# Question Answering Pipeline for Closed Domain Questions

Research work for Data Science Project competition

Presentation overview

Problem Description	Question-Answering system for the NLB Group.
Motivation	Process optimization and ESG.
Implemented Solution	Datasets, models and pipelines.
Presentation of Results	Metric scores and Qualitative Analysis for models and pipelines.
Demo	Quick demonstration of β application
Discussion	Usability and possible improvements.
Questions	For any additional information, the project repository is available on GitHub.

NLB

Introduction

Problem Description

Extract from the Data Science Project Competition Website

In56ight, in collaboration with the NLB banking group, is exploring the idea of developing a chatbot that would be trained on data from annual reports and would be capable of answering different questions regarding the banking group performance.
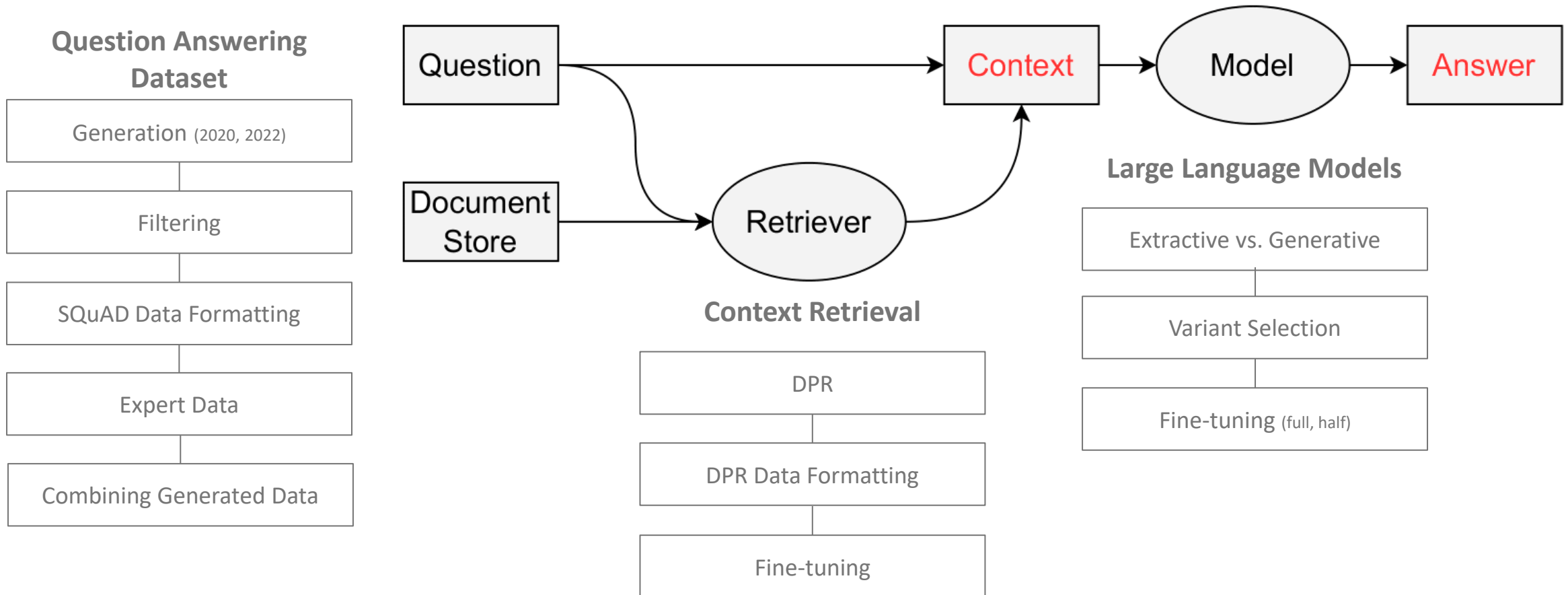
NLB

Motivation

- Thousands of employees check NLB policy documents daily.
- Saving as little as 15 minutes per employee is a big deal.
- Quickly extract relevant information from various reports.

ESG - A very important world-wide topic for companies

- Environmental, Social and Governance
a. Efficiency
b. Customer and employee relations
c. Leadership

NLB

Implemented Solution

NLB

Solution

Question Answering Dataset

Generation (2020, 2022)

Filtering

SQuAD Data Formatting

Expert Data

Combining Generated Data

Context Retrieval

DPR

DPR Data Formatting

Fine-tuning

Large Language Models

Extractive vs. Generative

Variant Selection

Fine-tuning (full, half)

NLB

Results

NLB

Results: Metrics

		Bert Precision	Bert Recall	Bert F1	Bleu	SQuAD Exact	SQuAD F1
DPR not fine-tuned	Distilbert	86.98	86.94	86.93	0	4.76	11.58
	RoBERTa	88.15	87.33	87.63	4.61	4.76	21.53
	T5-Small	88.78	87.44	88.04	13.48	/	/
	T5-Base	86.11	88.20	87.06	24.60	/	/
DPR fine-tuned	Distilbert	87.41	87.19	87.25	0	4.76	16.87
	RoBERTa	88.78	87.75	88.21	5.1	14.29	27.57
	T5-Small	86.92	85.23	86.03	0.88	/	/
	T5-Base	85.45	87.42	86.38	18.75	/	/

Extract from Table 4 of the report: Pipeline performance using fine-tuned models (expert data only)

NLB

Results: Qualitative analysis

		Correctness				Sensibility				"None"
		1	2	3	Avg[%]	1	2	3	Avg[%]	
2020/22	Distilbert	13	1	7	57	7	5	9	70	0
	RoBERTa	13	1	7	57	7	5	9	70	0
	T5-Small	13	1	7	57	7	6	8	68.3	0
	T5-Base	13	1	7	57	7	5	9	70	0
Expert	Distilbert	17	3	1	41.3	8	6	7	65	0
	RoBERTa	18	3	0	38	11	4	6	58.7	0
	T5-Small	16	5	0	41.3	12	3	6	57	7
	T5-Base	17	3	1	41.3	14	1	6	54	0

Table 5 of the report: Qualitative analysis using defined metrics

NLB

Results: Qualitative analysis

Question: How much did the bank reduce the use of paper in 2022 compared to the previous year?
Ground truth answer: 19%

Retrieved context: In 2022, another step to achieving this goal has been made as paper usage (prints) was reduced by 17% compared to 2021. Thus, we saved 866 trees. Paper usage in NLB was lower by 26%, which is the highest reduction group – wide. All banking subsidiaries experienced rediction as well ranging from 8% to 19%, respectively.
Our answer(s): 17%
ChatGPT answer: Based on the given context, the bank reduced the use of paper by 17% compared to the previous year (2021). However, the exact amount of paper usage in 2021 or any specific quantity of paper used is not mentioned.

NLB

Demo

Discussion

NLB

Findings about the Data	• More data would be required. • Fine-tuning on expert data yields the most improvements. • Specific types of questions
Model Usefulness	• Useful for more straightforward questions. • Highly depended on the DPR model. • The pipeline requires improvements before production.
Possible Improvements	• More high quality data • DPR improvements • Larger models

NLB

Thank you for your attention