I.J. Information Engineering and Electronic Business, **2024, 4, 47-65**
Published Online on August 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2024.04.04

Closed Domain Question Answering System Tailored for Crime Events Using Deep Learning for Both Statistical and Contextualized Responses

Dipti Pawade*
Department of Information Technology, K J Somaiya College of Engineering, Mumbai, India
E-mail: diptipawade@somaiya.edu
ORCID ID: https://orcid.org/0000-0003-3436-3208
*Corresponding Author

Sonali Patil
Department of Information Technology, K J Somaiya College of Engineering, Mumbai, India
E-mail: sonalipatil@somaiya.edu
ORCID ID: https://orcid.org/0000-0002-4358-4937

Chaitanya Bandiwdekar
Department of Information Technology, K J Somaiya College of Engineering, Mumbai, India
E-mail: c.bandiwdekar@somaiya.edu

Siddhesh Bagwe
Department of Information Technology, K J Somaiya College of Engineering, Mumbai, India
E-mail: siddhesh.ab@somaiya.edu

Pooja Kaulgud
Department of Information Technology, K J Somaiya College of Engineering, Mumbai, India
E-mail: pooja.kaulgud@somaiya.edu

Aditi Kulkarni
Department of Information Technology, K J Somaiya College of Engineering, Mumbai, India
E-mail: aditi.hk@somaiya.edu

Received: 06 April, 2024; Revised: 24 May, 2024; Accepted: 22 June, 2024; Published: 08 August, 2024

Abstract: The goal of the question-answering system is to respond to user queries expressed in natural language. Unlike search engines, the closed domain question answering systems are specialized to specific domains, providing concise and precise answers often derived from structured data. This paper focuses on a question-answering system tailored for crime events, capable of addressing both statistical and contextual inquiries. In terms of crime statistics, the fine-tuned GPT-3 model outperforms the USE, TAPAS, TAPEX, and GPT-3 models, while for context-based crime-related queries, the fine-tuned RoBERTa model surpasses the BERT and RoBERTa models. This system is capable of providing the responses in natural language format, supplemented with relevant data visualizations. The models are train on Q2A and NewsQA datasets while it is tested on NCRB and NewsTimes datasets. The Q2A and NCRB datasets are used for statistical queries while NewsQA and NewsTimes datasets are used for contextual inquiries. The paper presents an analysis of various models and showcases results for sample case studies. Such a system can prove valuable in applications where users seek to study criminal cases or gather pertinent insights for specific cases. Furthermore, it can assist in understanding patterns and trends in criminal events, particularly concerning geospatial information. Linking crime event-based question-answering systems to geospatial information facilitates exploration of niche areas and furnishes precise details about local crime with minimal hype and hence worth exploring.

Index Terms: Question Answer System, Statistical Questions, Contextual Questions, Criminal activity, RoBERTa, GPT-3.

1. Introduction

A question-answering system is a powerful and innovative application of artificial intelligence that aims to provide accurate and relevant answers to user queries [1]. This sophisticated technology utilizes natural language processing and machine learning algorithms to understand and interpret questions, enabling it to extract valuable information from vast databases, documents, and various online sources [2,3]. The system's capability to analyze and comprehend complex language structures empowers it to deliver concise and precise responses to a wide range of questions making it a valuable tool for individual users to seek information [4]. As this technology continues to evolve, it holds the potential to revolutionize information retrieval and reshape the way we interact with machines, opening new avenues for knowledge dissemination and problem-solving in the digital age.

A geographic question-answering system is a specialized application of question-answering technology that focuses on providing location-based information and answers to user queries [5]. This advanced system leverages geospatial data to understand and process questions related to locations and spatial relationships. Users can ask questions about directions, distances, nearby landmarks, local details and other geographical aspects, and the system can deliver accurate and contextually appropriate answers [6]. The difference between a general question-answering system and a geographic question-answering system is extremely small. To elaborate, the former deals with a wide range of questions on various topics, while the latter specializes in providing location-based information and answers related to geospatial queries [7].

Relevant and precise information retrieval systems concerning geospatial information are in high demand across various sectors, with one such prominent domain being crime-related information retrieval. In this area, the need for accurate data is paramount, as exact information is required to analyze crime patterns and make data-driven decisions [8]. Therefore, a robust and efficient crime-related information retrieval system plays a vital role in getting crime details. As per the National Crime Records Bureau, a total of 6,601,285 cognizable crimes were reported in 2020, including 4,254,356 Indian Penal Code (IPC) crimes and 2,346,929 Special and Local Laws (SLL) crimes [9]. Every day, millions of people search for information about criminal cases using search engines. Search engines give precise information for popular cases. However, many cases aren't in the spotlight. In these situations, search engines may not provide a satisfactory response to the query. Similarly for queries like "Which state has the highest murder count in 2020?" search engines often yield search results with links to relevant articles, but fail to provide the direct statistical answer sought by the user. The search engine result page may highlight the names of states in the description of the page links, but this information alone is insufficient for users looking for concrete statistical data. This underscores the need for a dedicated question-answering system viz, closed domain question answering system [10] that can precisely handle crime-related inquiries considering geo location and straightforwardly provide accurate statistical data. Such a specialized system could bridge the gap between the wealth of crime-related content available online and the user's requirement for direct, data-driven answers, significantly enhancing the accessibility and usefulness of crime-related information for various stakeholders.

This paper discusses the question-answering system dedicatedly design to answer the crime-related question considering the geospatial aspect of it. Broadly the questions are classified as statistical questions and context-based questions. The statistical questions are asked to get the statistical data related to various crimes in particular geographic location. For example "How many kidnapping cases happened in Patan in 2015?" or "What is the murder count in different states of India?". A context-based question is a type of question that requires knowledge or understanding of the surrounding context to be correctly interpreted and answered. In other words, the meaning of the question relies on the specific situation or information provided in the context. For instance, consider a question such as, "Where did the gold theft involving Priyal Saxena occur?" To respond accurately, one must first comprehend the context and then provide the location where the gold theft took place, involving the individual named Priyal Saxena. The question's answer is prompted based on an understanding of the specific circumstances surrounding the incident and the relevant details about the crime scene. The proposed question-answering system has been intricately crafted to proficiently address both statistical and context-based queries. It utilizes two distinct models, each specifically tailored to handle the different question types. The primary objectives of this research encompass:

- Objective 1: Investigating the current landscape of question-answering systems.
- Objective 2: Developing a specialized dataset tailored for addressing geography-specific inquiries related to crime.
- Objective 3: Assessing the effectiveness of cutting-edge question-answering models for crime events.
- Objective 4: Enhancing the model's performance to optimize its utility for the intended purpose.

The proposed system focuses on several key aspects to enhance existing models. It applies the advanced natural language processing techniques to capture the fine distinctions in the context of questions asked about crime and return specific and relevant answer. By using machine learning algorithms, it recognizes and analyzes the pattern of crime in an effective manner; thus, it gives insights that usually are not listed under traditional models. It can support context-based crime case queries to provide an exact answer, which otherwise requires lots of time and efforts. Besides, it deals with a remarkable absence of direct question-answering systems providing crime statistics linked to geographic

locations. The system also offers robust ways of visualization to show occurrences of the crime rate basically in different geographic locations. These features collectively demonstrate the system's ability to handle complicated crime-related queries, setting it apart from existing technologies.

With the help of this system, Law enforcement agencies can easily access detailed information about specific cases, enabling them to gain deeper insights into crime patterns and trends, which aids in strategic planning. Policymakers can utilize the system to access comprehensive crime statistics and contextual information, informing legislative and regulatory decisions. Researchers can benefit from the extensive data and advanced analytics capabilities to conduct in-depth studies on crime and its various aspects. For the general public, the system provides a transparent and accessible way to obtain information about particular cases and understand crime in their communities, thereby enhancing public awareness and safety. Individuals seeking information about a specific case can also use the system to access detailed case information, gain contextual insights, visualize crime data related to the case's geographic area, and contribute to informed decision-making and safety measures.

The significant contributions of this paper are outlined as follows:

- Exploring the current panorama of question-answering systems to understand their functionalities and constraints in addressing crime-related queries. Section 2 delineates various types of question-answering systems, providing an in-depth exploration of their capabilities and limitations.
- Discuss the various reliable and validated datasets, such as Q2A, NCRB, NewsTimes, and NewsQA, which are used to provide accurate responses to user queries. Notably, Q2A is a custom-made query-based dataset designed from scratch, while NCRB and NewsTimes are developed through web scraping techniques.
- Evaluating the effectiveness of state-of-the-art question-answering models in handling crime-related events. This involves refining and enhancing these models to optimize their performance, ensuring their efficacy and practicality for intended applications. Specifically, models such as TAPAS without NER, TAPAS with NER, TAPEX without NER, TAPEX with NER, GPT-3, and fine-tuned GPT-3 are evaluated for statistical queries. Additionally, BERT, RoBERTa, and fine-tuned RoBERTa are examined for context-based questions.

Furthermore, the structure of the paper is as follows: Section two provides a concise overview of the endeavors undertaken in the realm of question-answering systems. In Section Three, a discussion on the diverse datasets employed in this study is undertaken. The fourth section provides a comprehensive discourse on the methodology employed and the array of models utilized. The section five focuses on evaluation of the question-answering system pertaining to crime statistics, the performance of various models, namely the Universal Sentence Encoder (USE), TAPAS, TAPEX, GPT-3, and fine-tuned GPT-3 models. Among these, the fine-tuned GPT-3 model demonstrates superior performance compared to the other models. For the context-based crime-related question-answering system, an evaluation is conducted involving BERT, RoBERTa, and fine-tuned RoBERTa models. In this context, the fine-tuned RoBERTa model outperforms the other models. The sixth section culminating conclusive remarks.

2. Related Work

The first objective of this study is to investigate the current landscape of question-answering systems to comprehend the existing work in this field. This section focuses on exploring the diverse facets of the question-answering system. In the thorough study, Marco et. al. [11] conducted a comprehensive examination aimed at clarifying the methodologies, technologies, measurements, domains, and principles associated with a range of Question Answering strategies. Their findings highlighted that developing Question Answering systems, which involve processing Queries, Documents, and Responses, requires integrating various principles, technologies, and frameworks from information retrieval, natural language processing, and machine learning. Additionally, the study identified key metrics used by researchers to evaluate the effectiveness of their Question Answering systems. Notably, the analysis revealed that many researchers rely on 'Precision' and 'Recall' metrics to assess their natural language processing models, while 'Elapsed time' measurement is notably missing from the evaluations of the majority. Only a mere six percent of the reviewed studies incorporated this metric. Separately, Sanjay K. Dwivedi et al. [12] presented a classification of different Question Answering approaches, outlining their advantages and disadvantages and providing a comparative table. They discussed three main approaches: linguistic, statistical, and pattern-based, including examples and subtypes. According to their research, the choice of technique depends greatly on the specific problem. The hybrid approach, which combines multiple techniques, emerged as more effective in terms of speed, relevance, and recall. Through these studies, it has been noted that researchers predominantly focused on investigating open-domain and closed-domain questions for their analyses. Through these studies, it has been noted that researchers predominantly focused on investigating open-domain and closed-domain questions for their analyses.

In a series of studies, various researchers explored innovative approaches to open-domain question answering (QA). Lin et al. [13] introduced a distantly supervised model that incorporates an information retrieval-based paragraph selector and a paragraph reader employing a multi-layer LSTM network. Dipti et al. [14] developed a system that processes audio files, using a high-parameter ALBERT model trained on SQuAD 2.0, achieving notable EM/F1 scores of 85.3/88.1. This system can swiftly extract context from media files and provide accurate answers within a minute. Yang et al. [15] demonstrated an end-to-end QA system integrating a BERT-based reader with the Anserini IR toolkit

for answer identification from extensive Wikipedia articles. Karpukhin et al. [16] focused on refining training procedures by using sparse question and passage pairs and a simple dual-encoder setup for dense retrievals. Seo et al. [17] introduced the Dense-Sparse Phrase Index (DENSPI), a model for real-time open-domain QA that combines BERT-based dense vectors and term-frequency-based sparse encoding. Qu et al. [18] proposed an open-retrieval conversational QA model using BERT and ALBERT based encoders and decoders. Zhangning [19] evaluated different models, including BERT with various configurations, highlighting EM/F1 scores, and an ensemble approach. Similarly, Sam et al. [20] found that adjusting learning rates significantly influenced the performance of the BERT base model in comparison to the BiDAF baseline model.

The closed domain question answering system is designed to furnish concise and precise answers pertaining to a specific domain. In alignment with the proposed study's objectives, diverse question-answering systems related to crime and geospatial domains are investigated and key contributions are discussed here. In their study, Wei Chen [21] analyzed 800 spatial questions, categorizing them into five main groups. Their question-answering framework incorporates four essential elements: an NLP pipeline, ML model, ontology, and GIS (utilizing SPARQL queries and geo special functions). Impressively, they achieved an accuracy exceeding 90%. The framework involves NLP pattern analysis for data extraction, integrates formal inquiries into the geographic information system using sentence-derived data, and employs Spatial SQL queries for information retrieval from spatial databases. Dharmen P. et. al. [22] enhanced the existing "Frankenstein" question-answering architecture by incorporating a geoQA system. Their geoQA pipeline consisted of various components, including a dependency parse tree generator, concept identifier, instance identifier, geospatial relation identifier, property identifier, SPARQL query generator, and query executor. Upon testing with a gold standard dataset and 201 custom-made questions. Hamzei et al. [23] converted location queries and responses into semantic encodings through tokenization, tagging, and dependency parsing, creating 1024-dimensional vectors with Jaro similarity and k-means. Questions and answers were then clustered using semantic encodings to uncover patterns via Calinski-Harabasz score. Hamzei et al. [24] then introduced a geospatial dataset with 200 questions,

Table 1. Summary of Open and Closed Domain Question Answering System

Ref No	Author	Year	Type	Method	Dataset	Performance
[13]	Lin et al.	2018	Open-domain	Multilayer perceptron, Recurrent Neural Network	Quasar-T, SearchQA, TriviaQA, WebQuestions, CuratedTREC	outperforms DS-QA models
[14]	Dipti et al.	2023	Open-domain	ALBERT model	SQuAD 2.0	EM score: 85.3, F1: 88.1
[15]	Yang et al.	2019	Open-domain	Anserini Retriever (IR based), BERT	Wikipedia, SQuAD	-
[16]	Karpukhin et al.	2020	Open-domain	Dense Passage Retriever (DPR)	Wikipedia, Natural Questions, TriviaQA, WebQuestions, CuratedTREC, SQuAD v1.1	Outperformed B25
[17]	Seo et al.	2019	Open-domain	Dense-Sparse Phrase Index- a combination of dense and sparse vectors.	Wikipedia, SQuAD	43 times faster in a real setup than DrQA and 6.4% higher EM
[18]	Qu et al.	2019	Open-domain	BERT, ALBERT Model	QuAC dataset	Outperformed DrQA and BERTserini
[19]	Zhangning	2019	Open-domain	ALBERT model	SQuAD 2.0	• Single model: 76.5% F1, 73.2% EM • Ensembled model: 77.6% F1, 74.8% EM
[20]	Sam et al.	2019	Open-domain	BERT	SQuAD	F1:76.545, EM: 73.609
[21]	Wei Chen	2014	Closed Domain (Geospatial Question Answering)	Dynamic Programming, Voting Algorithm: NLP, Supervised learning: Classification, Ontology, GIS operations.	DBpedia, GeoNames	Avg. Accuracy is over 90%
[22]	Dharmen P. et. al.	2018	Closed Domain (Geospatial Question Answering)	Dependency sparse tree, Instance identification (NER+NED), Geospatial relation identification, and concept identification	The dataset created formed by linking DBpedia, OpenStreetMap, and the GADM global administrative areas dataset.	Precision: 37.38% Recall: 41.43% F1: 35.50%
[23]	Hamzei et al.	2019	Closed Domain (Geospatial Question Answering)	Semantic encoding, Embedding, Clustering	MS MARCO V2.1	-
[24]	Hamzei et al.	2022	Closed Domain (Geospatial Question Answering)	Dynamic template-based approach	Handcrafted Dataset, extended YAGO2geo	Accuracy 78.6%
[25]	Li et. al.	2021	Closed Domain (Geospatial Question Answering)	Semantic encoding using BERT, syntactic parsing using BiLSTM	GeoData201	Outperformed over Rule-based Model

extending encoding classes for events, dates, numbers, etc., through pre-trained models for fine-grained NER and part-of-speech tagging. Constituency and dependency parsing were used to analyze sentence structure. Li et. al. [25] has designed the NeuralGQA model for geospatial question answering. It consists of a semantic encoder and a semantic dependency parser. The model employs a geospatial semantic graph to represent sentences and a rule-based GSG query generator that uses the graph to produce executable queries. Table 1 presents an overview of Open and Closed Domain Question Answering Systems.

The research discussed in this paper centers around crime statistics and context-based questions related to geographic locations. This prompts us to delve deeper into the development of a question-answering system tailored for crime-related inquiries. In the study conducted by Kamdi et al. [26], they introduced a question-answering system that focuses on the Indian Penal Code (IPC) and Amendment Laws. This system classifies questions into five distinct types and employs a heuristic technique to extract relevant keywords from these questions. These extracted keywords are then utilized as indices to retrieve pertinent passages from the knowledge base of the IPC document. The responses are presented based on the specific type of question posed. This methodology demonstrated commendable accuracy, successfully providing accurate answers for 82 out of 100 questions. Similarly, another researcher, S. Pudaruth [27], devised a knowledge-based system that is designed to recognize crime-related keywords, keyword phrases, or WH-type questions. This system offers provisions for addressing such queries within the context of the Mauritian Judiciary. In a parallel vein, Shubhangi et al. [28] proposed a comparable system for the Indian Penal Code (IPC) by utilizing information retrieval techniques based on the Jaccard similarity measure. Taking a further step forward, Dipti et al. [29] developed a system capable of accepting descriptions of crime incidents and providing information about the corresponding IPC sections related to the crime, along with the associated punishments. This is achieved through the utilization of Word Mover's Distance similarity measure.

An alternative strategy employed to enhance comprehension of criminal activities involves the utilization of GIS based data visualization and exploratory data analysis techniques. Numerous scholars [30, 31, 32, 33] have highlighted the significance of visualizing crime rates. They have developed dashboards or interfaces that exhibit crime rates and crime hotspots, contingent upon user queries or selected options. Table 2 provides an outline of the crime question-answering system and the approaches for analyzing crime rates across different geographic locations.

Table 2. Overview of Crime Related System

Ref No	Category	Author	Theme	Methodology	Dataset
[26]	Question answering system: Process user queries to suggest relevant legal provisions.	Kamdi et al.	Accept Wh-type questions containing the type of crime and provide an IPC section related to crime and associated punishment	SVM and Jacard Coefficient	Curated training dataset with 200 questions on IPC sections and amendment laws
[27]		S. Pudaruth et. al.	Based on crime keyword phrase provide provision in Mauritian Judiciary	Knowledge Base	-
[28]		Shubhangi et.al.	Accept Wh-type questions containing the type of crime and provide an IPC section related to crime and associated punishment	Information Retrieval, Jaccard Coefficient, TF-IDF	Curated dataset of 300 different questions on IPC
[29]		Dipti et. al.	Based on the crime incident description provide an IPC section related to crime and associated punishment	Word Mover's Distance	Vakilbabu
[30]	Crime rate Visualization	Kim et. al.	Crime statistical analysis	Choropleth map	Neighborhood boundary dataset
[31]		Edholm	Geospatial analysis of Property crime	Kernel density function	Data from DataSF's website
[32]		Yang et. al.	Crime hotspot	Heatmap	Open data portal
[33]		Jimoh	Crime statistical analysis	Shiny app	The official crime records issued by the UK Police

Based on the examination of the literature, it can be deduced that ontology constitutes the prevailing methodology within the context of geospatial question-answering systems. However, there are shreds of evidence concerning the exploration of BERT's (the preeminent model for open domain question answering systems) performance in geospatial question answering. Thus, there exists an opportunity to assess the efficacy of deep learning models in the context of closed-domain geospatial question-answering systems. Most studies concentrate on open-domain and closed-domain QA systems without delving deeply into specialized domains like geospatial or crime-related QA, indicating a gap for the targeted applications.

Current crime-related question-answering systems are focused on recommending IPCs against a particular crime, depending on the scenario described by the user. Such systems are good at recommending legal aspects but cannot handle fine-grained queries based on the context pertaining to an individual case where the details of the event of the crime are needed. This reduces the understanding of complicated scenarios by users. It is also the case that the question-answering systems with in-depth crime statistics pinned to specific geographic locations are relatively few. Some of the methods available presently like crime rate distributions across geographic areas using heatmaps or Google Map gives basic visualizations. These approaches rarely allow dynamic querying of data or any other meaningful trend and pattern analysis. Therefore, in real-world applicability, these are immense limitations to effectively evaluate crime data in context and in relation to other variables for actionable insights. It thus opens up the opportunity for the development of

a geospatial crime-related question-answering system that may answer contextual and statistical questions. Such a system would address not only specific crime event-related queries but also a holistic view of crime trends to the user for informed decision making and resource allocation. This is, hence, a key area of innovation in crime-related geospatial question answering system.

3. Dataset Description

The second objective of this study is to developing a specialized dataset tailored for addressing geography-specific inquiries related to crime. This section gives overview of all the dataset used for this study. The proposed question answering system is designed to efficiently handle both statistical and context-based questions. To achieve this, two distinct models are employed, and they are trained on the appropriate datasets. Table 3 presents all the essential details of the datasets used. Out of the four dataset, Q2A dataset is a custom made dataset while other are generated by scrapping. These tailored dataset help to achieve better results for closed domain geography specific question. The comprehensive explanation of these datasets can be found in the subsequent section.

Table 3. Datasets considered for this study

Dataset	Dataset Name	Source	Size	Attributes	Purpose	Models
Custom made Question Dataset	Q2A	Created from scratch	2 table with around 1730 rows	Question, Query, Answer Format	Training models used for statistical question-answering system	TAPAS, TAPEX, GPT3, Fined Tuned GPT3
Crime count of India [34]	NCRB	NCRB, India	7 tables with around 2500 rows	State, District, Crimetype, Year, Cases	Testing the models used for statistical question-answering system	TAPAS, TAPEX, GPT3, Fined Tuned GPT3
NewsQA: CNN Article QnA Dataset [35]	NewsQA	Hugging Face datasets	1 table with around 90,000 rows	story_id, story_text, question, answer	Training models used for context based question-answering system	BERT, RoBERTa, Fine tuned RoBERTa
Crime articles of India Dataset	NewsTimes	Scraped from the TOI website	1 table with 1500 rows	Link, Title, Summary, Text, Keywords	Testing the models for context based question-answering system	BERT, RoBERTa, Fine tuned RoBERTa

3.1. Q2A - Custom made Question Dataset

For the statistical question-answer system, a query based approach is used, where the model converts natural language questions to SQL query, and its answer is fetched and mapped to the associated answer format. The tailored dataset named Q2A consists of around 1730 questions along with other attributes. These natural language question in English is mapped to their corresponding SQL query and the answer format. Questions in natural language, SQL query, and answer format are the attributes of this dataset. An example entry of the Q2A dataset is given in Table 4.

Table 4. Q2A dataset sample values

Question	SQL Query	Answer format
How many kidnappings took place in Patna in 2018?	SELECT Cases FROM kidnapTable WHERE District = 'Patna' AND Year = 2018	\{\} kidnappings took place in Patana in 2018.
Where was the highest murder count in 2020?	Select District, State, MAX(Cases) from murder where Year = 2018 AND Cases = (Select MAX(Cases) from murder where Year = 2018)	The highest murder count in 2020 was in \{\}.
Which district has the lowest crime?	Select District, State, Sum(Cases) from ipc where Cases=(Select MIN(Cases) from ipc) GROUP BY District ORDER BY SUM(Cases) ASC ;	\{\} district has the lowest crime.
Which was the lowest committed crime in 2018?	Select Crimetype, SUM(Cases) FROM allcrime where Year = 2018 AND Crimetype!="IPC Crime" GROUP BY Crimetype ORDER BY SUM(Cases) ASC LIMIT 1	\{\} was the lowest committed crime in 2018.
What is the information available on the crime against women rate in Madhya Pradesh in 2018?	SELECT Cases FROM state WHERE State = 'Madhya Pradesh' AND Crimetype = 'Crime committed against Women' AND Year = 2018	The crime against women rate in Madhya Pradesh in 2018 in 2018 is \{\}.

The curly braces were used as a placeholder to substitute for the actual numeric value in the answer format. This would help the system answer in natural language as well instead of a standalone numeric value. The models like TAPAS, TAPEX, and GPT3 are trained on this dataset. Figure 1 and 2 presents the dataset distribution of the question type (question starting with How, What, Where, Which) and answer type (Crime Percentage, Location, Year) respectively.

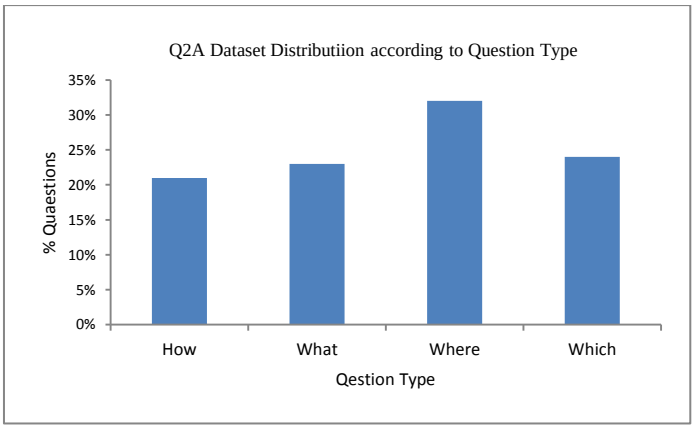

Fig. 1. Distribution of Q2A Dataset by Question Type

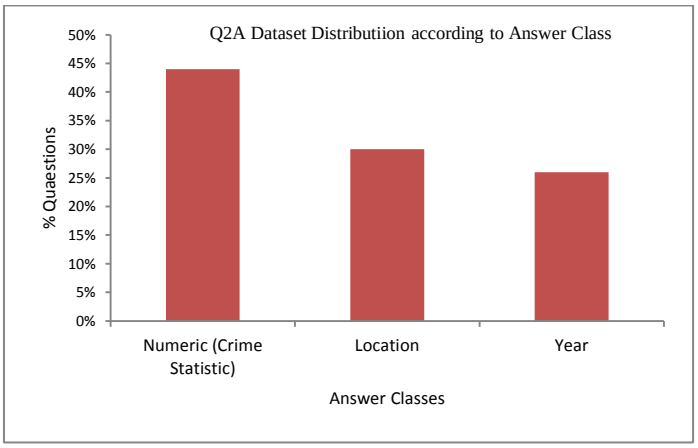

Fig. 2. Distribution of Q2A Dataset by Answer class

3.2. NCRB Dataset - Crime Count of India Dataset

This dataset is used to answer the statistical questions of the user. This dataset is acquired from the official government NCRB website. On the official website separate datasets were available for different types of crimes yearwise starting from 2016 to 2020 [34]. For example, for the crime type “Cyber crime" the datasets available were Cybercrime - 2016, Cybercrime -2017, and so on till the year 2020. Each dataset had attributes containing multiple sub-crime types. The dataset transformation was required to filter out and keep only the necessary attributes. It is carried out using a Python script. After the transformation process, the attributes present in the dataset are - State, District, Crimetype, Year, and Cases. A variant of this dataset is created for State level crime information by eliminating the column "districts" and adding the state crime count to the "count" column of the dataset. This dataset is also used for the overall data visualization of crime data in India.

3.3 NewsQA: CNN Article QnA Dataset

The NewsQA dataset is a large-scale question-answering dataset that was created for research purposes in the field of natural language processing and machine learning [35]. It was developed by researchers at the University of Washington and the Allen Institute for Artificial Intelligence (AI2). The dataset consists of approximately 100,000 question-answer pairs that are derived from a set of news articles from CNN. Each question is accompanied by a short paragraph that serves as the context for answering the question. The dataset contains the following fields:

- 'story_id': An identifier of the story
- 'question': The question or query presented to the model.
- 'story_text': The news article that contains information relevant to the question.
- 'answer': The correct answer to the question.
- 'answer_token_ranges': A list that provides the start and end token positions of the answer within the 'story_text'.

For this study, the 'answer_token_ranges' are pre-processed to extract the 'start_positions' and 'end_positions' of the answer, along with the actual answer strings ('answers'). This pre-processing task was accomplished using a tailor-crafted script. The final training dataset has the following columns:

- 'story_id': An identifier of the story
- 'story_text': text of the story
- 'question': A question about the story.
- 'answer_token_ranges': Word-based indices to answers in story_text. E.g. 196:202,217:228. Multiple selections from the same answer are separated by, The start is inclusive and the end is exclusive. The end may point to whitespace after a token.
- 'start_positions': Starting index of the answer to the question.
- 'end_positions': Ending index of the answer to the question.
- 'answers': Actual answer string of the question.

BERT, RoBERTa, and fine-tuned RoBERTa models are trained on this dataset.

3.4. NewsTimes: Crime Articles of India Dataset

The NewsTimes dataset serves the purpose of answering contextual user questions. This dataset is developed from scratch by scraping 1500 of the latest newspaper articles up to February 8th, 2023, from the TOI (Times of India) website. The scraping process involved accessing each Times Of India page. On average, it took around 1 minute to scrape each page. During the scraping process, the Python script accessed each anchor tag to open the related article. Utilizing the Newspaper3k package, essential information such as article title, link, content, summary, and keywords were extracted and stored in a dataset. To further enrich the dataset, Spacy was employed to extract keywords from each article. These extracted keywords were compared to the keywords obtained earlier. Spacy performed part-of-speech tagging and assigned weight values to the context. The article context was then split into tokens, with each token consisting of words with part-of-speech tags such as "PROPN," "ADJ," or "NOUN". Next, the most common 'T' number of hot words were identified from the list and stored as Spacy Keywords in the dataset. Subsequently, a Python script was created to select the most relevant context from this dataset based on the user's question. The selected context, along with the user's question, is then fed into a question answering model. This model generates the answer to the user's question and also provides the title and link of the TOI article that was used to address the user's input question. This process ensures that the answers are well-informed and contextually relevant to the user's queries. BERT, RoBERTa and fine-tuned RoBERTa models used this dataset to answer the user query.

4. Methodology

The objective 3 and 4 of this study focuses on assessing the effectiveness of cutting-edge question-answering models for crime events and enhancing the model's performance to optimize its utility for the intended purpose respectively. This section gives overview of different models used and how they are optimized to get more accurate results. The performances of these models are analyzed in section 5.

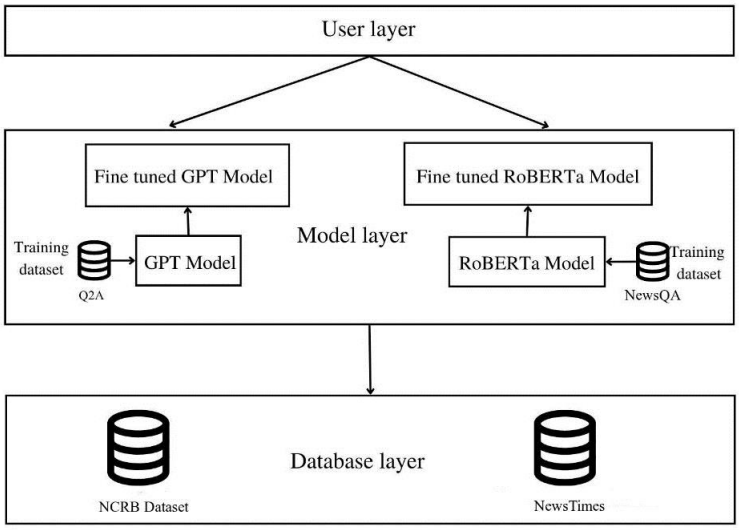

Fig. 3. Workflow diagram

The envisioned Question Answering system comprises three distinct layers: the User layer, the Model layer, and the Database layer. This system leverages two specialized models: fined-tuned GPT-3 for handling statistical data and fine-tuned RoBERTa for processing context-based data. Figure 3 presents the workflow diagram, illustrating the sequential steps of the system's functioning. The initial layer is the user layer, where users interact with the system by posing relevant questions. The subsequent layer is the model layer, housing two models: the fine-tuned GPT model and the fine-tuned RoBERTa model. The final layer consists of datasets used to retrieve answers for the user's queries. As

mentioned before, distinct models are employed to handle statistical questions and context-based questions. The details of these models are elaborated in the subsequent section of the paper.

4.1. Model for Handling Statistical Questions

The primary goal of the model is to deliver accurate and relevant answers, specifically focusing on locations and their related statistics, when confronted with questions expressed in natural language. To achieve this, various models like USE, TAPAS, TAPEX, GPT etc. were explored.

4.1.1 Universal Sentence Encoder(USE)

The initial approach involved utilizing a Universal Sentence Encoder [36] to encode the questions in the training dataset and subsequently classify them into specific output classes. Its primary purpose is to transform sentences and short text into fixed-length, dense vector representations known as embedding. This embedding encodes the semantic meaning of the input text and can be used further. The USE encodes both the input question and potential answer into embedding. The model assesses the similarity between the question and each answer to determine the most suitable response for the given query. Regrettably, the accuracy of the test dataset was unsatisfactory (39.94%), prompting us to explore alternative approaches. Hence it is not considered for further discussion.

4.1.2 Table Based Models

The next approach was using the TAPAS model. TAPAS is a BERT-based model with a specialized focus on answering questions related to tabular data [37]. Unlike standard BERT, TAPAS incorporates relative position embedding and includes 7 token types specifically designed to encode the structure of tabular data. For question-answering tasks, TAPAS employs two distinct heads: a cell selection head and an aggregation head. The cell selection head is responsible for identifying relevant cells in the table that correspond to the question. The aggregation head allows the model to perform aggregations such as counting or summing on the selected cells, thereby providing comprehensive answers to queries involving tabular data. To delve further into exploration, another table-based model called the TAPEX is also employed [38]. The TAPEX utilizes a BART model for pre-training, enabling it to tackle synthetic SQL queries effectively. Subsequently, the model is fine-tuned to address natural language questions concerning tabular data and perform table fact checking. Both the TAPAS and TAPEX models demonstrated accuracy in answering simple questions involving single column single row responses. However, for the complex queries containing multiple SQL queries or various aggregation operations, their outputs of model were not appropriate. Additionally, the time required to respond to simple queries, involving one-word answers with straightforward SQL queries, ranged from 30 seconds to over 2 minutes, with exponential increases in response time relative to the dataset size. To mitigate the time issue, the implementation of Named Entity Recognition (NER) was adopted. This involved filtering out rows based on specific criteria like location (e.g., Mumbai) and type of crime (e.g., kidnapping) mentioned in the question. As a result, the table size was reduced significantly from 20,000 rows to approximately 15-20 rows. However, despite this optimization, the accuracy problem persisted in both models.

4.1.3 Generative Transformer

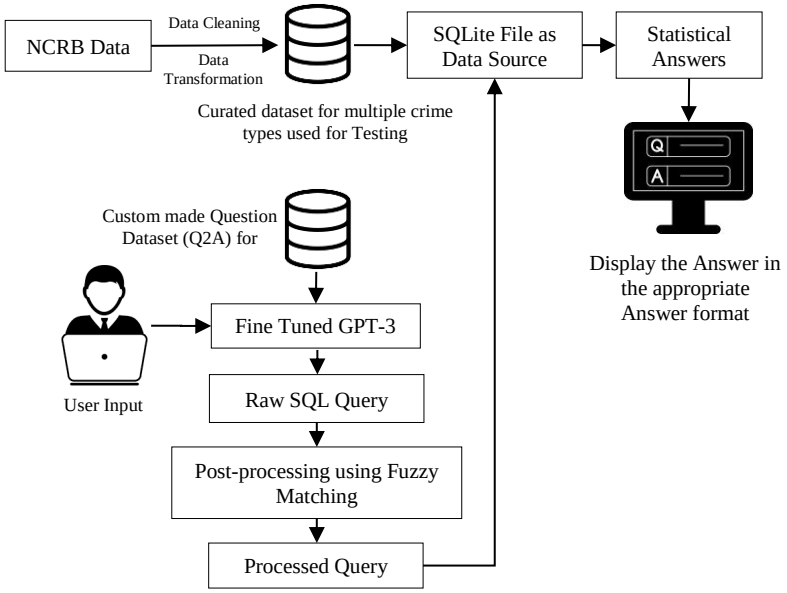

Fig. 4. Workflow diagram for GPT-3 model

After careful consideration, Generative Pre-trained Transformer-3 (GPT-3) [39] is explored to generate the answer. The GPT-3 model as shown the satisfactory results over above discuss methods. To further enhance the performance of the GPT-3 model, fine-tuning is employed using task-specific custom made Q2A datasets of approximately 1730 questions from scratch. The workflow of the statistical question-answering system using the GPT-3 model is illustrated in Figure 4.

The user provides a question to the model, which generates a query to retrieve the answer and subsequently parsed the query result into the answer format. However, the SQL query provided by the model may not always be accurate due to possible typographical errors made by the user. Unfortunately, the model does not validate the correctness of the question asked, leading to these errors being carried forward to the final output. For instance, if a user asks, "What is the murder count in nashik," the model might not recognize the correct district name in the dataset (i.e., Nasik), resulting in an inaccurate SQL query with no matching entry. To address this, a post-processing step was introduced using "Fuzzy matching". This concept relies on the Levenshtein distance, measuring the number of substitutions, insertions, or deletions needed to transform one word into another. A shorter Levenshtein distance indicates greater word similarity and helps find the most relevant dataset entry for a more accurate output. The Levenshtein distance between two strings S1 and S2 is given by using equation 1.

$$lev\ (s1, s2) = \begin{cases} |s1| & \text{if } |s2| = 0 \\ |s2| & \text{if } |s1| = 0 \\ lev(tail(s1), tail(s2)) & \text{if s1[0] = s2[0]} \\ 1+\min\begin{cases} lev\ (tail(s1), s2) \\ lev\ (s1, tail\ (s2)) \\ lev\ (tail\ (s1), tail\ (s2)) \end{cases} & Othervise \end{cases} \quad (1)$$

Where S1 and S2 are the two strings, lev() gives Levenshtein distance two strings, and tail() gives the strings excluding the first character.

Using "Fuzzy matching", the SQL output is compared with a list of district names from the dataset. If the matching percentage exceeds a predetermined threshold, set at 0.85, the corresponding dataset entry is considered, and the SQL query is adjusted accordingly. Once the correct location was substituted in the SQL query, it was fired on the dataset after which the numeric answer to the question was fetched. This numeric value is then displayed to the user in a natural language format.

The GPT3 model was finetuned on the Q2A dataset. Table 5 gives the details of hyperparameters for training GPT3. The training token accuracy of the model was 0.9513, the testing token accuracy was 0.9474 and the average training loss was 0.141834.

Table 5. Hyperparameters for training GPT3

Hyperparameter	Parameter Description	Values
batch_size	The count of samples handled prior to updating the model	1
learning_rate_multiplier	Regulates the rate at which an algorithm updates or learns parameter estimates	0.1
n_epochs	Indicate the iteration count across the training dataset	4

4.2. Context Based Question Answering System

If user want to seek information about a specific crime incident at a particular location, or if a user is interested in the location of a particular crime then in such situations understanding the context of the question becomes crucial. To address such context-based inquiries, the utilization of the BERT and RoBERTa models is investigated. At the start, the BERT model is examined to ascertain its literal capabilities. The BERT model processes both the question and context to produce an answer based on the provided context [40]. The BERT model demonstrated precise conciseness in short answer responses. However, not all of the generated responses were sufficiently detailed or in exact wording. Subsequently, the RoBERTa model [41] was employed for Context Based Questions. This model yielded considerably more detailed answers. To enhance the suitability of this pre-trained model for News article-based question and answering tasks, a fine-tuning process was undertaken. The RoBERTa model underwent fine-tuning using the NewsQA training dataset, which consisted of 10,000 entries. Table 6 shows the hyperparameters for fine-tuning RoBERTa respectively.

Table 6. Hyperparameters for training RoBERTa

Parameter	Values
evaluation_strategy	“epoch”
learning_rate	2e-5
num_train_epochs	5
weight_decay	0.01

Among the investigated models – BERT, RoBERTa, and the fine-tuned RoBERTa – the latter exhibited the most promising outcomes, leading to its selection for further consideration. Figure 5 illustrates the workflow of question and answer system tailored for Indian crime news-related inquiries using fine-tuned RoBERTa model. To determine the appropriate article for addressing user input, a matching score is computed between the user's question and each

keyword list stored NewsTimes dataset. A script is devised to select the suitable article context according to the user's question. The context selection process involves calculating the matching score between the keywords present in the context based dataset and the keywords in the question. The context, title, and link of the row with the highest matching score are then forwarded to the NLP pipeline. Following are the steps to identify the most pertinent context that can be used for further processing.

Step 1. Compute the relevance of the user input question to different contexts using equation 2.

$$f(q) \xrightarrow{yields} context \tag{2}$$

where q-> user question and context -> relevance of q to different context

Step 2. Using equation 3, compute the index ci corresponding to the maximum value of the function f(score) for scores ranging from 1 to n. It helps to identify the most pertinent context by finding the context with the highest relevance score.

$$ci = index(\max\{f(score): score = 1, \ldots . n\}) \tag{3}$$

where ci -> context index. The function f(score) represents a scoring mechanism used to evaluate the relevance or importance of a given context in the context-selection process.

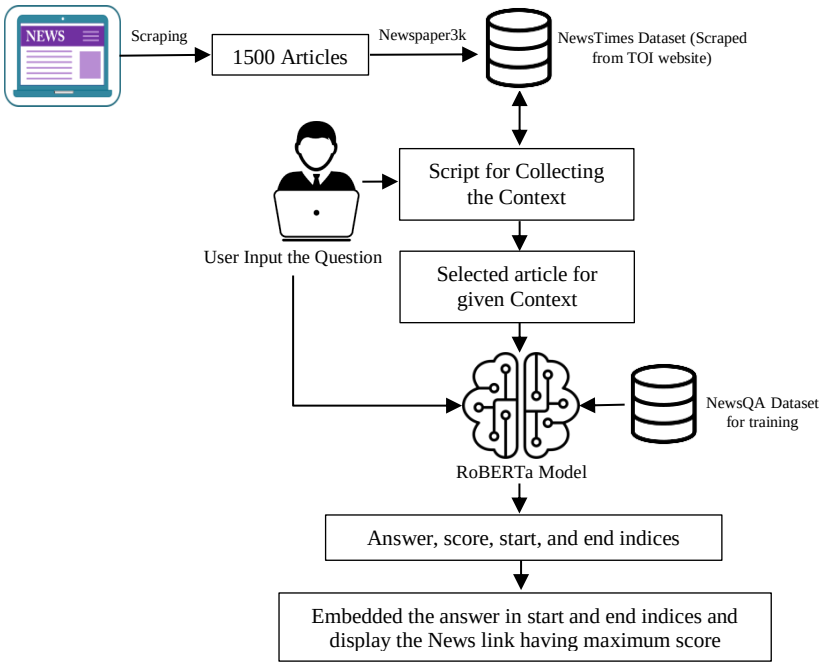

Fig. 5. Workflow diagram for Question Answering System using Fine-tuned RoBERTa model

Step 3. Once the most relevant context index ci is determined, equation 4 retrieves the corresponding context data. This context data likely includes information such as article text, title, and link.

$$context\_inf = context\_data(ci) \tag{4}$$

Step 4: Calculate the sum over keywords (kw) and question words (qw) using equation 5. It indicates the relevance of a context is determined by the presence and alignment of keywords and question words within it.

$$final\_score = \sum[\forall kw == \forall qw] \tag{5}$$

The user input question and the extracted context are fed to fine-tuned RoBERTa model to generate the answer. The model generated answer, score, start and end indices. This answer along with the selected context's article link and title is then displayed to the user in a suitable format.

5. Results and Discussion

Objectives 3 and 4 of this study are dedicated to evaluating the effectiveness of advanced question-answering models specifically for crime-related queries and enhancing the model's performance to maximize its practical utility. To substantiate these objectives, this section provides a detailed performance evaluation of the various models discussed in the preceding sections. This section is divided into two parts: the first part focuses on the performance assessment of models employed for statistical question-answering systems, while the second part presents the results of models used in context-based systems.

5.1. Discussion on Statistical Question Answering System

As discussed in the previous section, to handle the statistical questions system USE, TAPAS, TAPEX, GPT-3, and fine-tuned GPT-3 models are evaluated. Table 8 shows the training accuracy, test accuracy, precision and f1 score comparison for these models. The evaluation of all the mentioned models was conducted on the NCRB dataset. It was observed that USE, TAPAS and TAPEX demonstrated poor performance for long answer. In contrast, the fine-tuned GPT-3 model exhibited remarkable superiority, achieving an test accuracy of 0.9474. Table 7 gives the training and testing accuracy for the models under consideration.

Table 7. Accuracy Comparison Table

	USE	TAPAS without NER	TAPAS with NER	TAPEX without NER	TAPEX with NER	GPT-3	Fine-tuned GPT-3
Training Accuracy	0.3994	0.4251	0.4837	0.6412	0.6946	0.8765	**0.9513**
Test Accuracy	0.3241	0.3809	0.4682	0.6180	0.6891	0.89002	**0.9474**
Precision	0.3856	0.4034	0.4696	0.5782	0.6828	0.8294	**0.9342**
F1 Score	0.3048	0.4376	0.5061	0.5971	0.6645	0.8363	**0.9217**

To get the further inside, question typewise accuracy of each model is presented in Figure 6. This graph depicts the same conclusion that the fine-tuned GPT-3 model exhibits better performance over others.

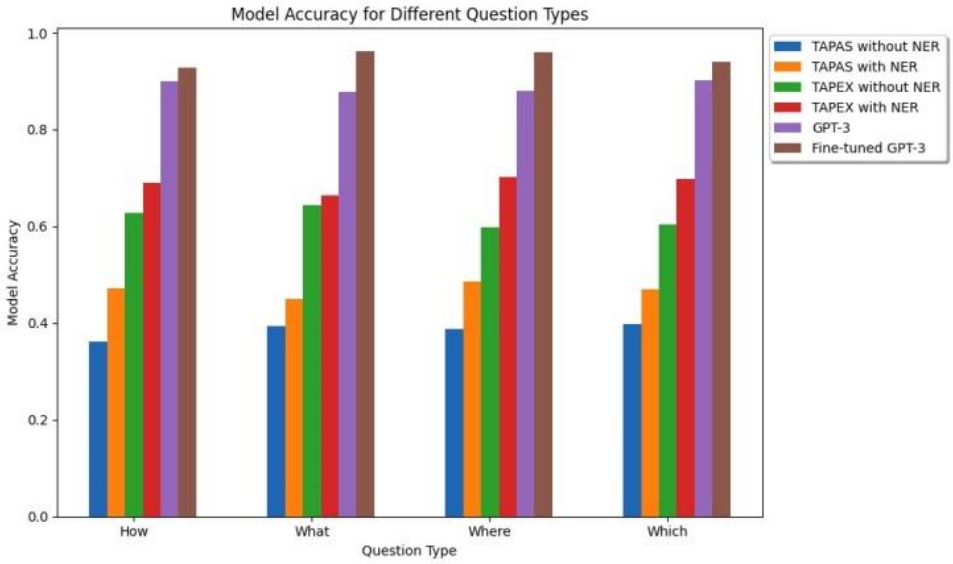

Fig. 6. Workflow diagram for RoBERTa model

Presented below is a case study showcasing the operational process demonstrating how the statistical answers are derived using fine-tuned GPT-3 model. For the case study a Wh type of question is considered and how it is processed at every stage is elaborated. To showcase the precision of the suggested question-answering system, identical questions are presented to the Google Search Engine, and the results are observed. It's notable that the search engine doesn't directly provide the exact answer but rather furnishes a handful of relevant links.

Case Study 1: Operational workflow from system input to output for statistical question

User Input:

- Query: Which state has a higher murder count maharastra or gujrat?

Expected Output:

- Maharashtra has a higher murder count.
- A pie chart for comparison between the murder counts of Gujarat and Maharashtra.

The stepwise processing of the user's input query is given as follows

Step 1: Converting Question to SQL Query:

1.1 Input Query: Which state has a higher murder count maharastra or gujrat?

1.2 Input validation: Check if each word in the question is a valid English word, a stopword, or a valid location present in the dataset. Table 8 gives the results for the input query input validation.

Table 8. Validation Result for Input Query

Word	Status
Which	Valid
State	Valid
Has	Valid
Higher	Valid
Murder	Valid
Count	Valid
Maharastra	Valid (Matched location with percentage: ('Maharashtra', 95))
Or	Valid
Gujrat	Valid (Matched location with percentage: ('Gujarat', 92) gujrat)

1.3 Raw SQL Query generation: Using fine-tuned GPT-3 model, a raw SQL query is formulated. Presented below is the SQL query generated using model.

SELECT State, SUM(Cases) FROM murder WHERE (State = 'maharastra' OR State = 'gujrat') GROUP BY State ORDER BY SUM(Cases) DESC;

1.4 Replacing locations with valid names: The name of the location in the input question has spelling mistakes and the same is replicated in the SQL query. So "maharastra" need to replace with "Maharashtra" and "gujrat" need to replace with "Gujarat". This can be done using following steps

1.4.1. As a Perquisite, list of the unique district from the curated NCRB dataset is generated. (One time Task)

district_names = dataset["District"].unique().tolist()
state_names = dataset["State"].unique().tolist()

1.4.2. The raw SQL query is parsed to identify instances of the words like "State =" and "District=". The words following these instances would be the location names. It is represented as s*str and dstr* for state and district respectively. (If the query has multiple locations then it is stored list sstr and dstr. If either State or District is not available then return NULL)

1.4.3. The extracted location names are not proper. Hence for each location, calculate the similarity ratio of the word with the list of location using equation 6.

$$f\ (string, list\ of\ strings)\ =\ min(lev(string, x)\ for\ all\ x\ in\ list\ of\ strings)\ ) \quad (6)$$

where,

- *f (string, list of strings)= closest match for the string in a list of string*
- *string = each word of the input prompt stored in sstr and/or dstr*
- *list of strings = district_names list and/or state_names list*
- *lev(string, x) = Levenshtein distance*

(ssrt is compared against state_names list and dstr is compared against district_names list)

1.4.4. The closest match location is stored in *floc*

1.4.5. Finally, the Location correction is done using fuzzy matching:

- *Location in question- maharastra ; Location in dataset- Maharashtra*
- *Location in question- gujrat; Location in dataset- Gujarat*

1.4.6. After fetching *floc*, the raw SQL query is split into 3 parts around each location name using the partition function f(x) as given in eq 7.

$$f\ (x)\ =\ [x[0:i], x[i:i+L], x[L:E] \quad (7)$$

where,

- *i = index of the location name*
- *L = Length of the location string*
- *E = End index of the query*

1.4.7. The second element of the partitioned query is replaced with *floc*. This process is done for every location to form the final query.

1.4.8. The partitioned query is then joined back which includes the changed location name.
SELECT State, SUM(Cases) FROM murder WHERE (State= 'Maharashtra' OR State = 'Gujarat') GROUP BY State ORDER BY SUM(Cases) DESC
(Replacing locations with valid names: "maharastra" is replaced with "Maharashtra" and "gujrat" is replaced by "Gujarat".)

Step 2. SQL Query to answer generation:

2.1 The final SQL query is fired on the NCRB dataset and the answer is generated in the required format using GPT-3 model. The output of the SQL query mentioned in step 1.4.8 is given below

	State	*SUM(Cases)*
0	*Maharashtra*	*8709*
1	*Gujarat*	*4398*

The model processes the query and provides the answer object, embedding it in the proper answer format at the placeholder represented as {}.
{} has a higher murder count where Maharashtra is parsed in a placeholder.

2.2 Presenting Answers and Data Visualization:
The answer object is sent to the frontend. Additionally, the answer format and crime count of each state are also fetched which is further used for data visualization.

Step 3 Actual Output:

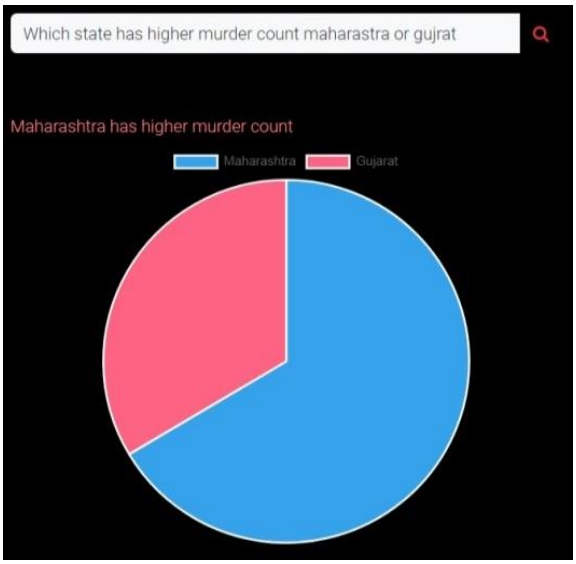

Fig. 7. Answer to a user query using the proposed system

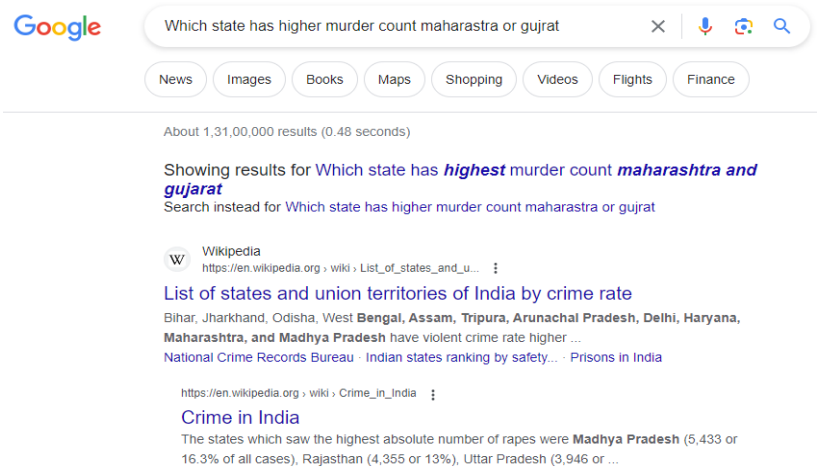

Fig. 8. Answer to a user query using Google Search Engine

As shown in figure 7 and 8, gives the results for the sample input query by the model and by Google search engine. The search engine gives a link of the related information wherein the model gives a direct answer along with the graphical representation.

5.2. Discussion on Context Based Question Answering System

For context based question-answering systems, the performance of models like BERT, RoBERTa, and Fine-tuned RoBERTa are evaluated. Table 9 indicates that RoBERTa model exhibits the most promising results over the BERT model. Hence RoBERTa is fine-tuned on the NewsQA dataset.

Table 9. Evaluation Metrics Comparison for models evaluated for Context Based Question Answering System

Model	BERT	RoBERTa	Fine tuned RoBERTa
Training Accuracy	0.8904	0.9321	0.9516
Test Accuracy	0.8749	0.9289	0.9476
Precision	0.8841	0.9098	0.9168
F1	0.8822	0.9081	0.9134
Recall	0.8816	0.9079	0.9113

Figure 9, shows the epoch vs loss graph for fine-tuned RoBERTa model. The Statistics in Table 9 prove that RoBERTa model outperformed over other models.

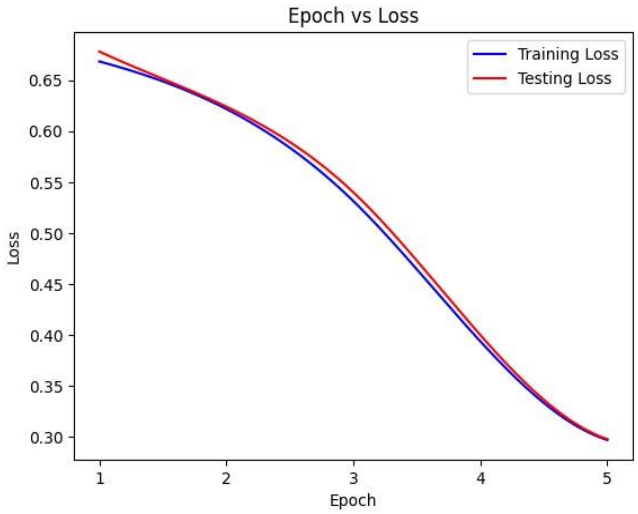

Fig. 9. Epoch vs Loss Graph

Presented below is a case study showcasing the operational process of the fine-tuned RoBARTa model as it handles input queries to generate accurate responses. To showcase the precision of the suggested question-answering system, the identical question is presented to the Google Search Engine, and the results are observed. It's notable that the search engine has highlighted the accused name “Mithu Singh" but does not directly provide any information about another accused Jabbar Ansari.

Case Study 2: Operational workflow from system input to output for context based questions
User input: Who killed swadichcha?
Step 1: Pre-process the question to get the keyword : ['killed', 'swadichcha']
Step 2: Calculated maxscore : 1 (This is the match score between keyword in user question and keyword list present in NewsTimes dataset)
Step 3: Calculate context index: 122 (Calculated based on the relevance)
Step 4: Select the article link based on context based maxscore and context index calculated in previous steps: https://timesofindia.indiatimes.com/city/mumbai/cops-hire-private-marine-firm-to-scour-sea-for-medicos-body-in-mumbai/articleshow/97335102.cms?from=mdr.
Step 5: Extract the exact answer from the article using fine-tuned RoBERTa model:
Mithu Singh and his assistant Jabbar Ansari
Step 6: Embed answer object to frontend: The following snippet is used to send the answer object to the frontend

```
{
'answer':
        'Mithu Singh and his assistant Jabbar Ansari',
'Link':
```

'https://timesofindia.indiatimes.com/city/mumbai/cops-hire-private-marine-firm-to- scour-sea-for-medicos-body-in-mumbai/articleshow/97335102.cms',
'title':
"Cops hire private marine firm to scour sea for medico's body in Mumbai",
'contextFound':
True
}

Figure 10 represents the response to the same question by using the proposed system while figure 11 depicts the Google Search engine response to the same user query. For the user query "Who killed swadichcha?" Google search engine provide a search engine result page by highlighting the name "Mithu Singh" from News article. Here the most popular News article is taken as a reference for answer prediction. In case of the proposed system based on fine tuned RoBERTa model, all the articles related to this case are considered, and a precise answer is provided as "Mithu Singh and his assistant Jabbar Ansari" (figure 10). Like search engines here also link to the article is provided for the reference of the user.

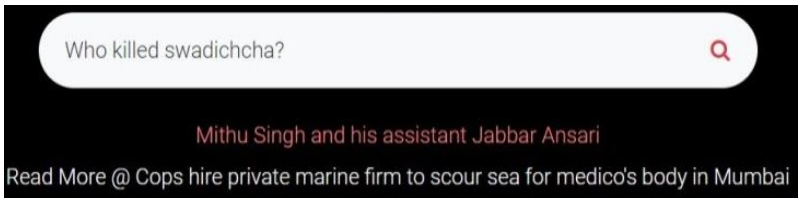

Fig. 10. Answer to user query using fine tuned RoBERTa model based system

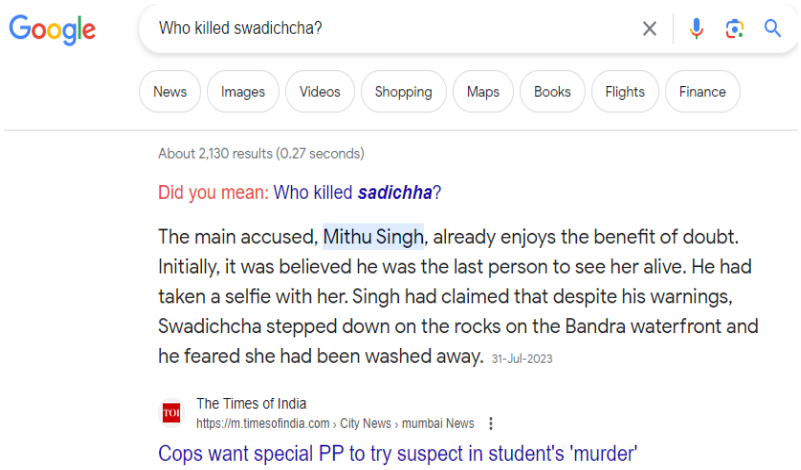

Fig. 11. Google Search Engine response for user query

To sum up, the study's findings, fine-tuned GPT-3 model-based system performs admirably for the statistical question-answering system while fined tuned RoBERTa model based system excels in the context based question answering system.

6. Conclusion

The proposed question-answering system is tailored for information related to the criminal event with respect to geographic location. Relating the crime event based question answering system to the geospatial information facilitates exploration of the niche areas and provides precise information about the local crime with subdued emphasis. The first objective of this study is to investigate the current landscape of question-answering systems. To achieve this thorough background study is carried out for different types of question answering systems like general, open ended, close ended, geospatial, and crime event related question-answering systems. The findings indicate that while the general-purpose question-answering landscape is well-developed, crime-based closed-ended question-answering methods are primarily reliant on visualization for crime statistics. Textual inquiries, on the other hand, tend to focus on recommending actions based on the IPC sections and punishment for specific crimes or legal advice. So to fill this gap the study discussed in this paper is carried out which considered both statistical and context based question-answering systems. To train and test the models datasets like Q2A, NCRB, NewsQA, and NewsTimes are used. Out of which Q2A and NewsTimes are custom made specialized datasets tailored for addressing geography-specific inquiries related to crime. This serves the

second objective of the study. To build an efficient system capable of handling the statistical questions, the performance of USE, TAPAS, TAPEX, and GPT-3 models are analyzed. Even after the application of NER, the TAPAS and TAPEX models fall short as compared to the GPT-3 model. The fine-tuned GPT-3 model achieved an accuracy of 94.74%. On the other hand, the performance of the BERT and RoBERTa models is assessed for context-based questions, where the RoBERTa model demonstrates superior results. As a result, fine-tuning is applied to the RoBERTa model to achieve an accuracy of 92.04%. Thus objectives 3 and 4 are achieved by assessing the effectiveness of cutting-edge question-answering models for crime events and enhancing the model's performance to optimize its utility for the intended purpose.

Even though the model exhibited acceptable accuracy still the model has shown poor responses for the complex user query comprising of many joins and exclusion criteria. This system does not perform satisfactorily if query has any reference of location which falls under toponymic duplication. In the future model need to improvise to handle complicated user queries. Additionally model should be improvised to handle the crude user query with a lower semantic relationship. Currently this system is able to handle specific type of 'Wh' questions only. In future user query reformulation can be implemented in order to handle any type of questions.

References

[1] Allam, Ali Mohamed Nabil, and Mohamed Hassan Haggag. "The question answering systems: A survey." International Journal of Research and Reviews in Information Sciences (IJRRIS) 2, no. 3 (2012).

[2] Diefenbach, Dennis, Vanessa Lopez, Kamal Singh, and Pierre Maret. "Core techniques of question answering systems over knowledge bases: a survey." Knowledge and Information systems 55 (2018): 529-569.

[3] Bouziane, Abdelghani, Djelloul Bouchiha, Noureddine Doumi, and Mimoun Malki. "Question answering systems: survey and trends." Procedia Computer Science 73 (2015): 366-375.

[4] Ojokoh, Bolanle, and Emmanuel Adebisi. "A review of question answering systems." Journal of Web Engineering (2018): 717-758.

[5] Chen, Wei. "Developing a Framework for Geographic Question Answering Systems Using GIS, Natural Language Processing, Machine Learning, and Ontologies." PhD diss., The Ohio State University, 2014.

[6] Ferrés Domènech, Daniel. "Knowledge-based and data-driven approaches for geographical information access." (2017).

[7] Mai, Gengchen, Krzysztof Janowicz, Rui Zhu, Ling Cai, and Ni Lao. "Geographic question answering: Challenges, uniqueness, classification, and future directions." AGILE: GIScience series 2 (2021): 8.

[8] Kabiraj, Pintu. "Crime in India: A spatio-temporal analysis." GeoJournal 88, no. 2 (2023): 1283-1304.

[9] Das, Upal, and Amit K. Biswas. "Incidence of Crime and its Determinants: A Study on Indian States." Available at SSRN 4472648 (2023).

[10] Kia, Mahsa Abazari, Aygul Garifullina, Mathias Kern, Jon Chamberlain, and Shoaib Jameel. "Adaptable closed-domain question answering using contextualized CNN-attention models and question expansion." IEEE Access 10 (2022): 45080-45092.

[11] Soares, Marco Antonio Calijorne, and Fernando Silva Parreiras. "A literature review on question answering techniques, paradigms and systems." Journal of King Saud University-Computer and Information Sciences 32, no. 6 (2020): 635-646.

[12] Dwivedi, Sanjay K., and Vaishali Singh. "Research and reviews in question answering system." Procedia Technology 10 (2013): 417-424.

[13] Y. Lin, H. Ji, Z. Liu, and M. Sun, ``Denoising distantly supervised open domain question answering," in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics, 2018, pp. 1736_1745.

[14] Pawade, Dipti, Avani Sakhapara, Isha Joglekar, and Deepanshu Vangani. "Implementation of Open Domain Question Answering System." In International Conference on Data Management, Analytics and Innovation, pp. 499-507. Springer, Singapore, 2023.

[15] Yang, Wei, Yuqing Xie, Aileen Lin, Xingyu Li, Luchen Tan, Kun Xiong, Ming Li, and Jimmy Lin. "End-to-end open-domain question answering with bertserini." arXiv preprint arXiv:1902.01718 (2019).

[16] Karpukhin, Vladimir, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. "Dense passage retrieval for open-domain question answering." arXiv preprint arXiv:2004.04906 (2020).

[17] Seo, Minjoon, Jinhyuk Lee, Tom Kwiatkowski, Ankur P. Parikh, Ali Farhadi, and Hannaneh Hajishirzi. "Real-time open-domain question answering with dense-sparse phrase index." arXiv preprint arXiv:1906.05807 (2019).

[18] C. Qu, L. Yang, C. Chen, M. Qiu, W. B. Croft, and M. Iyyer, ``Open retrieval conversational question answering," in Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Jul. 2020, pp. 539-548.

[19] Hu, Zhangning. "Question answering on SQuAD with BERT." CS224N Report, Stanford University. Accessed (2019): 12-01.

[20] Schwager, Sam, and John Solitario. "Question and answering on SQuAD 2.0: BERT is all you need." ArXiv e-prints of (2019).

[21] Chen, Wei. "Developing a Framework for Geographic Question Answering Systems Using GIS, Natural Language Processing, Machine Learning, and Ontologies." PhD diss., The Ohio State University, 2014.

[22] Punjani, Dharmen, Kuldeep Singh, Andreas Both, Manolis Koubarakis, Iosif Angelidis, Konstantina Bereta, Themis Beris et al. "Template-based question answering over linked geospatial data." In Proceedings of the 12th workshop on geographic information retrieval, pp. 1-10. 2018.

[23] Hamzei, Ehsan, Haonan Li, Maria Vasardani, Timothy Baldwin, Stephan Winter, and Martin Tomko. "Place questions and human-generated answers: A data analysis approach." In International Conference on Geographic Information Science, pp. 3-19. Springer, Cham, 2019.

[24] Hamzei, Ehsan, Martin Tomko, and Stephan Winter. "Translating Place-Related Questions to GeoSPARQL Queries." In Proceedings of the ACM Web Conference 2022, pp. 902-911. 2022.

[25] Li, Haonan, Ehsan Hamzei, Ivan Majic, Hua Hua, Jochen Renz, Martin Tomko, Maria Vasardani, Stephan Winter, and

Timothy Baldwin. "Neural factoid geospatial question answering." Journal of Spatial Information Science 23 (2021): 65-90.
[26] Kamdi, Rohini P., and Avinash J. Agrawal. "Keywords based closed domain question answering system for indian penal code sections and indian amendment laws." International Journal of Intelligent Systems and Applications 7, no. 12 (2015): 54.
[27] S. Pudaruth, R. P. Gunputh, K. M. S. Soyjaudah and P. Domun, "A Question Answer System for the Mauritian Judiciary," 2016 3rd International Conference on Soft Computing & Machine Intelligence (ISCMI), Dubai, 2016, pp. 201-205. doi: 10.1109/ISCMI.2016.47.
[28] Shubhangi C. Tirpude, Dr. A.S. Alvi , "Closed Domain Keyword based Question Answering System for Legal Documents of IPC Sections & Indian Laws", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 6, June 2015, Pp 5299-5311.
[29] Pawade, Dipti, Avani Sakhapara, Hussain Ratlamwala, Siddharth Mishra, Samreen Shaikh, and Dhrumil Mehta. "Implementation of smart legal assistance system in accordance with the Indian Penal Code using similarity measures." In Advances in Computing and Data Sciences: Third International Conference, ICACDS 2019, Ghaziabad, India, April 12–13, 2019, Revised Selected Papers, Part II 3, pp. 440-449. Springer Singapore, 2019.
[30] Kim, Suhong, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri. "Crime analysis through machine learning." In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 415-420. IEEE, 2018.
[31] Edholm, Emma. "Property Crime in The City and County of San Francisco 2016-2017: Applying GIS to Crime Pattern Theory." (2019).
[32] Yang, Dingqi, et al. "CrimeTelescope: crime hotspot prediction based on urban and social media data fusion." World Wide Web 21 (2018): 1323-1347.
[33] Jimoh, Fatai. "Real time crime prediction using social media." PhD diss., 2023.
[34] Das, Upal, and Amit K. Biswas. "Incidence of Crime and its Determinants: A Study on Indian States." Available at SSRN 4472648 (2023).
[35] Adam Trischler, Tong Wang, Xingdi Yuan, Justin Harris, Alessandro Sordoni, Philip Bachman, and Kaheer Suleman. 2016. Newsqa: A machine comprehension dataset. arXiv preprint arXiv:1611.09830 (2016).
[36] Deepthi, Godavarthi, and A. Mary Sowjanya. "Query-Based Retrieval Using Universal Sentence Encoder." Revue d'Intelligence Artificielle 35, no. 4 (2021).
[37] Herzig, Jonathan, Paweł Krzysztof Nowak, Thomas Müller, Francesco Piccinno, and Julian Martin Eisenschlos. "TaPas: Weakly supervised table parsing via pre-training." arXiv preprint arXiv:2004.02349 (2020).
[38] Liu, Q., Chen, B., Guo, J., Ziyadi, M., Lin, Z., Chen, W. and Lou, J.G., 2021. TAPEX: Table pre-training via learning a neural SQL executor. arXiv preprint arXiv:2107.07653.
[39] Yang, Z., Gan, Z., Wang, J., Hu, X., Lu, Y., Liu, Z. and Wang, L., 2022, June. An empirical study of gpt-3 for few-shot knowledge-based vqa. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 3, pp. 3081-3089).
[40] Zaib, M., Tran, D.H., Sagar, S., Mahmood, A., Zhang, W.E. and Sheng, Q.Z., 2021. BERT-CoQAC: BERT-based conversational question answering in context. In Parallel Architectures, Algorithms and Programming: 11th International Symposium, PAAP 2020, Shenzhen, China, December 28–30, 2020, Proceedings 11 (pp. 47-57). Springer Singapore.
[41] Pearce, K., Zhan, T., Komanduri, A. and Zhan, J., 2021. A comparative study of transformer-based language models on extractive question answering. arXiv preprint arXiv:2110.03142.

Authors' Profiles

Dipti Yogesh Pawade is currently pursuing a Ph.D. in Information Technology from Somaiya Vidyavihar University, Mumbai. Since 2012, she has been serving as an Assistant Professor in the Department of Information Technology at K J Somaiya College of Engineering, Mumbai. She has received a research grants from the University of Mumbai and has published papers in various journals and conferences. Her research interests include Geographic Information Systems, Machine Learning, Computer Vision, Natural Language Processing and Application Development.

Dr. Sonali Patil is working as a Professor in the Department of Information Technology at K.J. Somaiya College of Engineering, Somaiya Vidyavihar University, Mumbai (India). Presently, she also owns the responsibility of Associate Dean, Academic Programmes. She received her Ph.D. in technology from the Shivaji University, Kolhapur (India) and M.E. degree in Electronics from Walchand College of Engineering, Sangli (India). She has a rich experience of over 19 years in Teaching and 1 year in Industry. She is recognized PG teacher of University of Mumbai and Somaiya Vidyavihar University. She has supervised many M.Tech students PG students in the area of information security and computer engineering. Her doctoral research was in the area of Medical Image Processing and Analysis. She is also a recognized PhD supervisor/guide of University of Mumbai and Somaiya Vidyavihar University. She has publications in peer-reviewed International Journals, Book chapters and conferences in the areas of information Security and image processing.

Chaitanya Vinay Bandiwdekar received his bachelor's degree in information technology from K. J. Somaiya College of Engineering in 2023. He is currently working at J.P. Morgan Chase & Co. as a Software Developer. His research areas include generative AI, ML, NLP and Blockchain technology.

Siddhesh Ajit Bagwe received his bachelor's degree in information technology from K. J. Somaiya College of Engineering in 2023. He is currently employed at Forcepoint Software Solutions as a Software Engineer. His research areas include generative AI, blockchain, and data analytics.

Pooja Sachin Kaulgud received her bachelor's degree in information technology from K. J. Somaiya College of Engineering in 2023. She is currently pursuing her Master's in Computer Science from the University of Southern California. Her research areas include digital forensics, blockchain, and full-stack development.

Aditi Hemchandra Kulkarni received her bachelor's degree in information technology from K. J. Somaiya College of Engineering in 2023. She is currently working at Barclays as a Software Developer. Her research areas include full stack development, UI UX designing and data science.

How to cite this paper: Dipti Pawade, Sonali Patil, Chaitanya Bandiwdekar, Siddhesh Bagwe, Pooja Kaulgud, Aditi Kulkarni, "Closed Domain Question Answering System Tailored for Crime Events Using Deep Learning for Both Statistical and Contextualized Responses", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.16, No.4, pp. 47-65, 2024. DOI:10.5815/ijieeb.2024.04.04