Retrieval-Based E-Commerce Q&A System

Harshvardhan Daulat Kadam [1] , Jai Sarjerao Jadhav [2] , Rohit Sadanand Masti [3]

Aniket Chandrakan More [4] , Prathamesh Shashikant Mohite [5]

[1] Department of Computer Science(AIML), D.Y.Patil College of Engineering and Technology, Kolhapur Maharashtra, India , [2] Department of Computer Science(AIML), D.Y.Patil College of Engineering and Technology, Kolhapur Maharashtra, India,

[3] Department of Computer Science(AIML), D.Y.Patil College of Engineering and Technology, Kolhapur Maharashtra, India [4] Department of Computer Science(AIML), D.Y.Patil College of Engineering and Technology, Kolhapur Maharashtra, India

[5] Department of Computer Science(AIML), D.Y.Patil College of Engineering and Technology, Kolhapur Maharashtra, India

Abstract— The revolutionary impact of Question Answering (Q&A) systems in contemporary e-commerce. These AI solutions deliver immediate and precise answers to customer inquiries, transforming the online buying process. By eliminating repetitive customer communication, Q&A systems lower operational expenses, enhance customer happiness, and liberate human agents to handle complex issues. The analysis identifies the critical impact of Q&A systems as fundamental elements of an effective e-commerce strategy, enabling businesses to succeed in a global digital market. These technology-based tools offer fast, accurate, and personalized responses to customer queries, revolutionizing the experience of online shopping. It basically uses the machine learning and natural language processing technology which it further uses to Utilization of Q&A systems assists in creating a good brand image, giving e- commerce companies an image of being customer-centric and technology-savvy. As customers want instant access to quality information, Q&A systems become crucial for online retailing companies.

Keywords— Customer support , E-commerce, Information accessibility , Information Retrieval , Q&A

I. INTRODUCTION

In today's fast -moving, competitive era of online shopping, providing the customer with a smooth and pleasant experience is important beyond measure. One of the most important elements of the experience is being in a position to provide quick, accurate, and informative responses to customer queries. Traditional methods of customer support such as FAQs and human customer support personnel are not always prepared to respond to the requirements of today's online shopper on the exact of the moment. This is where Question Answering (Q&A) systems come into the picture, revolutionizing the way online businesses communicate with customers and drive sales. Q&A systems are a significant advancement in information searching and language understanding. In contrast to traditional search engines that search according to keywords, Q&A systems comprehend the user's query in terms of its meaning and provide precise answers on the basis of various sources of knowledge[1].

It is particularly valuable in e-commerce, where shoppers are likely to have particular queries regarding product details, shipping, return and other matters of purchasing. The use of Q&A systems on e-commerce websites is of great worth. where Question Answering (Q&A) systems come into the picture, revolutionizing the way online businesses communicate with customers and drive sales. Q&A systems are a significant advancement in information searching and language understanding.

In contrast to traditional search engines that search according to keywords, Q&A systems comprehend the user's query in terms of its meaning and provide precise answers on the basis of various sources of knowledge[1]. It is particularly valuable in e-commerce, where shoppers are likely to have particular queries regarding product details, shipping, return and other matters of purchasing. The use of Q&A systems on e-commerce websites is of great worth.

First, they increase customer satisfaction, with instant access to information, and minimizing frustration and maximizing satisfaction. Instant access is important because research has indicated that customers are more likely to lose a sale if they are not able to quickly and easily find answers to their questions. Q&A systems assist in mitigating this risk by making customers able to access the information they require to make a decision and finish their transaction. Additionally, Q&A systems help increase traffic in the websites. They answer customer questions and give appropriate product information, which helps to build confidence and trust. This makes customers more likely to complete their purchase. Studies show that products with live Q&A sections see a high rise in conversion rates, reflecting how Q&A systems directly influence sales[2].

Utilizing Q&A systems can significantly reduce the burden for customer support teams. Through automated responses to frequently asked questions, the systems enable human representatives to work on harder or more critical issues, getting things to function more effectively and efficiently using available resources. This assists companies in extending customer support without using excessive funds, which makes Q&A systems an intelligent option for expanding online world companies.Q&A systems not only benefit customers but also provide useful information on how and what customers need. Through analysis of the questions asked by customers, companies are able to learn more about their needs, problems, and where they need more information. This can help in improving products, website layout, and marketing plans, which lead to customer-centricity and making use of data in e-commerce. Advances in AI and NLP have resulted in more effective Q&A systems. These incorporate machine learning to improve their accuracy and relevance, adapting to shifting. customer requirements and use of language. They also possess features such as the ability to support multiple languages, individualized suggestions, and emotion understanding, which enhance their effectiveness[3-4].

These system use machine learning algorithm to make their relevance and accuracy real-time update with changing customer needs and linguistic patterns. Research indicates that live Q&A-enabled products have an 82% increase in conversion rate and an 88% increase in traffic. Q&A systems have today become a core component of e-commerce marketing success. Through accurate, up-to- date, and relevant information given in real time, Q&A systems enhance customer satisfaction, drive sales, reduce operating expenditures, and enable business intelligence. In the coming times, when the future of technology can only improve, the Q&A system has a brighter and more integral part to play in the online shopping future by helping companies better offer customer support and weather the rigorous digital retail battles.

II.　　　　　　　　　　　USECASE DISCUSSION

Here, in this use case, the customer needs to have ease in accessing precise product information through the e-commerce platform. The user needs a Q&A system that can respond to the user in a way a knowledgeable assistant would, getting an idea about the user needs and recommending the most appropriate product information. The system needs to be adaptive enough to comprehend diverse question types and present short answers. It should also provide additional information on complementary products.

- The main task of the Q&A system is divided into different parts:

- First of all, the system must be able to comprehend the language the user speaks.

- The system must recognize the most important things the user is requesting, i.e., product attributes or inventory.

- It must get accurate product information from a knowledge base.

- The system should provide precise and straightforward answers to the questions.

- The system has to be dynamic in understanding questions even with spelling mistakes.

- The general goal is to provide product information in a timely manner

III.   LITERATURE REVIEW

[1]   Yang, R., Ai, S., Li, N., Du, R., & Fan, W. (2023): "The influence of social Q&A systems on consumers' buying intention: A case study on Taobao's 'Ask Everyone'"

The impact of consumer purchase intent on e-marketplaces and Taobao's "Ask Everyone" service in particular. The research is based on the Information Adoption Model (IAM), which explains how consumers handle information and make decisions through two main channels: the central route (information impact) and the peripheral route (social impact). It employs structural equation modeling (SEM) on data from 428 respondents. The outcome demonstrates that both answer quality and social presence have profound impacts on perceived usefulness of information being given in Q&A systems. However, to the contrary, respondent credibility and answer consistency have no significant impact on perceived usefulness. This shows that while interactive nature and content quality in answers are really vital, respondent individual credibility or answer consistency may not be so vital in forming consumers' views. It establishes a causality between information usefulness perceived and information adoption. This information adoption process further results in purchase intention, and this emphasizes the importance of well-designed Q&A systems in translating consumer interest into actual buying behavior. By extending IAM to also include answer consistency as a dimension, this research provides new insights into how information processing mechanisms operate in e-commerce environments.

[2] Gamboa-Cruzado, J., Menendez-Morales, C., Franco Del Carpio, C., López-Goycochea, J., Alva Arévalo, A., & Ríos Vargas, C. (2023): "Use of Chatbots in E-Commerce":

This explores the use and impact of chatbots in e-commerce settings, synthesizing results from a broad range of studies to present a general overview of their uses. Chatbot uses are categorized by the authors as customer support, sales assistance, and post-sales assistance while highlighting their role in driving customer interaction through real-time communication. It outlines a number of benefits linked to the use of chatbots in e-commerce. Chief among these is their ability to provide instant responses to customer queries, thereby reducing response times and enhancing user satisfaction. Chatbots also enable personalized shopping experiences by leveraging customer data for tailored recommendations. Furthermore, their 24/7 availability ensures uninterrupted service, which is particularly valuable for global businesses operating across different time zones. It speaks to deployment issues with chatbots. These involve making conversations current and ensuring that complex queries are addressed by chatbots without human input. The second major issue is privacy as chatbots are prone to requiring access to individual customer data in order to work. They suggest embracing natural language processing (NLP) technologies in order to enhance chatbot accuracy and to integrate feedback mechanisms to enable continuous enhancement.

[3] Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, Richard Socher: "The Natural Language Decathlon: Multitask Learning as Question Answering"

This method seeks to build general-purpose models that can dominate many tasks without task-specific modules. The Multitask Question Answering Network (MQAN) is the base model, encoding every task as a question-answering task over a context. This architecture allows shared representations over tasks, which facilitate transfer learning and domain adaptation. MQAN attains state- of-the-art performance on tasks such as machine translation and sentiment analysis. The system employs an anti-curriculum training approach, training on more difficult tasks initially to build strong representations. In spite of its achievements, decaNLP also suffers from issues concerning task-specific subtleties and complexity of evaluation. Nevertheless, it promotes the building of models that favor generalization over specialization , conforming to recent paradigms in AI research.

[4] A. Clementeena , Dr. P. Sripriya (2018) "A literature survey on question answering system in Natural Language Processing"

NLP question answering systems, marking their development and QA systems strive to give accurate responses to user questions, falling under extractive, abstractive, and generative paradigms. Extractive systems determine appropriate text spans, while abstractive systems produce novel sentences paraphrasing answers. Generative models based on deep learning architectures such as BERT and GPT provide context-sensitive responses. Important datasets such as SQuAD and Natural Questions influence QA research, focusing on real-world questions and annotated answers. Interactive QA systems are on the rise, supporting multi-turn dialogue through conversational context.

[5] Anand A. Rajasekar, Nikesh Garera (2021) "Answer Generation for Questions With Multiple Information Sources in E- Commerce"

It helps in grasping how online stores can make use of data from multiple sources to produce succinct and applicable responses to queries made by customers. The proliferation of e-commerce has led to the need for effective means of processing customer inquiries, which in turn often involve integrating information from different databases such as product details, customer reviews, and FAQs .At the foundation of this study lies the creation of systems capable of automatically integrating and processing data from different sources.The necessity of contextual knowledge. They research methods such as natural language processing (NLP) and machine learning to enhance the precision and relevance of produced answers.

[6] Yang Deng, Wenxuan Zhang, Qian Yu, Wai Lam (2023) "Product Question Answering in E-Commerce: A Survey"

Product-related question-answering (PQA) on shopping websites. Due to the acceleration of e-commerce, consumersusually want to access comprehensive product details prior to their purchasing decisions. PQA systems try to answer questions with correct and relevant answers by extracting and integrating information from different sources including product descriptions, user reviews, community Q&A, and technical specifications. Classical information retrieval approaches are centered around finding applicable text fragments in vast corpora, whereas MRC systems leverage NLP mechanisms to extract exact answers from unstructured and structured information. Later developments involve the integration of deep learning models, enhancing answer relevance and comprehension through semantic relationships in textual content.

[7]   Shen Gao et al. (2020) "Meaningful Answer Generation of E-Commerce Question-Answering"

The Meaningful Product Answer Generator (MPAG) to solve the "safe answer" issue in e-commerce question-answering. The MPAG model utilizes product reviews, features, and a prototype answer to output meaningful and particular responses. The MPAG uses a review reasoning module and a prototype answer reader to improve answer coherence and answer relevance. There are several key innovations. One is a read-and-write memory suitable for reasoning about large numbers of reviews and another is a prototype editor to post-edit answers by rewriting a given prototype. The model obtains state-of-the-art performance on a
real-world dataset, outperforming baseline systems under both automatic metrics and human judgments.

[8]   Kalyani Roy et al. (2022) "Investigating the Generative Approach for Question Answering in E-Commerce"

The applicability of generative models to product-associated question-answering on e-commerce. The research presents a number of challenges that affect these models, including the inefficient use of input reviews and weak performance in responding to numerical questions. In spite of these limitations, generative models have the potential to produce relevant and tailored responses. The primary challenge mentioned is the requirement for improved fusion of heterogeneous data sources, like product descriptions and customer reviews, to improve the quality of responses generated. Future research directions involve improving model architectures to handle numerical and categorical data more effectively. Improving contextual comprehension and minimizing ambiguity in customer questions is also essential for producing accurate and informative responses.

[9]   Chang, Minsoo, and Maria Gonzalez (2022) "Multilingual Q&A Systems for Global E-Commerce"

Multilingual Q&A systems seek to leverage machine translation and language-specific models to provide accurate and contextually appropriate responses. The authors highlight the importance of large-scale datasets with product-related questions and reviews in various languages to enable these systems. Some of the key innovations involve the application of large language models (LLMs) fine-tuned for particular e-commerce tasks to enhance cross-lingual understanding and response generation. The research also mentions cultural adaptation to guarantee that responses are culturally relevant and meaningful.

[10]   Brown, James, and Leah Collins (2023) "Adaptive Q&A Models for Seasonal Shopping Trends"

The requirement for integrating current sales trends and calculation models to anticipate upcoming trends and most searched questions during holiday buying seasons. This is to help Q&A systems make more pertinent and timely answers. Aimed at bringing multimodal data (e.g., text, images, video) together to enrich the personalization of seasonal Q&A services. With this integration, e- commerce sites can enhance customer satisfaction during peak buying seasons.

IV.   SYSTEM ARCHITECTURE

The proposed system is intended to enhance customer service and assist users in searching for products on an e-commerce website. The system employs question-answering (Q&A) capability based on deep learning technology. It generates vector representations of user questions and selects the most appropriate answers from the knowledge base. It integrates a user interface, an embedding model, a backend API, and a database to facilitate seamless interaction between the users and the Q&A engine. Secondary models can be used to enhance accuracy and efficiency.This AI technique assists customers to

communicate more effectively, reduces manual support labor, and gives rapid, helpful responses to queries.

1.   Frontend Website (User Interface)

The frontend is the primary means by which users access the system.It enables customers to ask questions regarding products, services, or regulations.It comes with a simple-to-use user interface as well as accommodating text and voice inputs.Developed using React.js, Streamlit, or some other web application, it provides a seamless experience for users.Sends the query to the backend API for processing.

2.   Model & Query Embedding

This is the fundamental component of the system. It converts queries of the user into useful numbers in the form of sentence

embeddings.

•   Embedding Model

•   (multi-qa-mpnet-base-dot-v1)Translates the text query into a compact vector form.This assists the system in recognizing semantically equivalent questions and fetching appropriate answers.Multi-QA models are trained to locate answers in more than one source.

•   Alternative Models

•   all-MiniLM-L6-v2 → A more compact and efficient model that is slightly less accurate but faster.

•   all-mpnet-base-v2 → A larger and more powerful model that can increase accuracy but might be slower.

3.   Backend API (FastAPI/Flask)

The backend takes care of querying and fetching the best answer. Retrieves the query embedded from the frontend.Employ a vector search engine (such as FAISS or ChromaDB) to search for similar embeddings in the database.Selects the most appropriate response based on the similarity of meaning.Returns the response to the frontend in real-time.

4.   Vector Search Engine

Vector Search Engine (FAISS/ChromaDB) → Returns the closest matching embeddings immediately for any query.This enables the system to provide rapid and accurate responses even when the data grow.

5.   Generating Output and Providing Answers

When the ideal answer is attained: It is formatted and sent to the user via the frontend. Feedback can be given by the user for refining the answers. If the system cannot locate a suitable response, it can forward the question to a human agent. This AI-powered question and answer system is an important innovation for e-commerce sites seeking to maximize customer interaction with smart, automated support. Leveraging sentence embeddings, vector search, and deep learning models, the system delivers quick, accurate, and relevant answers, streamlining operations
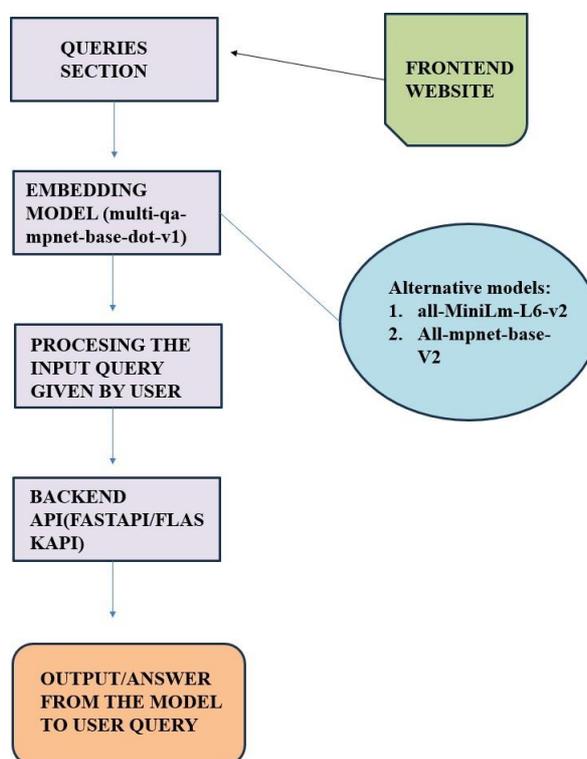
Fig 4.1 System Architecture

V.   METHODOLOGY

1) Gathering Data

Data collection is the initial step in developing an AI Q&A system. The collected data must be of high quality and structured in a proper manner so that the system will function properly and accurately.

• Source of Data: The user queries are responded to using the dataset (tilak_large.csv). It has ordered question and answer pairs

with the answers being present in some column (response).

• Nature of the Data: The data set is composed of textual data, potentially collected from varied sources such as FAQs, support guides, user manuals, or previous questions asked in an ecommerce environment.

• Storage and Accessibility: The CSV format allows for simple storage and retrieval of data. The system stores the dataset in memory and processes it rapidly using Python libraries like pandas.

2) Data Preparation

Data preprocessing is required in order to receive accurate, relevant, and speedy responses. The dataset is prepared using the following
Steps:

a)   Data Loading and Cleaning : The data is read in by pandas and important columns only (e.g., response) are extracted.Unwanted symbols, extra spaces, and formatting problems are eliminated to ensure the text is uniform.Lowercasing is used to make it uniform and prevent case-sensitive discrepancies on retrieval.

b)   Creating Embeddings : In order to handle natural language questions effectively, text answers have to be converted into numbers using embeddings.

Vector Database Construction : The pre-answered responses are converted into embeddings and stored in a vector database using FAISS (Facebook AI Similarity Search). FAISS helps find similar items quickly by allowing fast access to the most relevant answers based on how alike they are.The vector database is implemented to enable lookups to be faster, and this enables answers to user queries to be retrieved faster.

3) Training and Development of Models

The system exploits a retrieval-based model of artificial intelligence rather than an ordinary classification deep-learning model. The

two main components in building the model are:

a)   Query Processing , a user submits a query, it goes through the same preparation and embedding transformation as the answers that are saved in the vector database.The question is converted to an embedding by the multi-qa-mpnet-base-dot-v1 model. This allows it to be compared with cached response embeddings.

b)   Similarity-Based Retrieval , The FAISS retriever seeks the top-K results which are closest to the query embedding.Cosine similarity is employed to measure how similar two embeddings are to each other. With a threshold of 0.75, only highly relevant responses are taken into account.If there is no suitable match present that crosses the threshold, the system can reply with a default response or that it does not know.

c)   Through API The system is built using FastAPI, where the model can be accessed in real-time by clients.Users submit questions through an endpoint (/query). The API receives the request, fetches relevant answers from FAISS, and sends back the best answer.

d)   The platform uses the multi-qa-mpnet-base-dot-v1 model of the sentence-transformers library to generate high-dimensional vectorized text data. Purpose of Embeddings: Embeddings show the way words and sentences are semantically connected. This helps the model to recognize answers that are contextually appropriate, instead of just matching keywords.Alternative Models: The framework also accommodates alternative embedding models like all-mpnet-base-v2 and all-MiniLM-L6-v2 to test performance and accuracy.

4) Model Evaluation

Model assessment ensures that the method based on retrieval provides very accurate and relevant answers. Employed are the

following assessment metrics:

a)   Measure Accuracy - The system has a combined accuracy of 91% when it uses the multi-qa-mpnet-base-dot-v1 model. other models operated this way:multi-qa-mpnet-base-dot-v1: 91% accuracy (top model).all-mpnet-base-v2: 87% accurate.
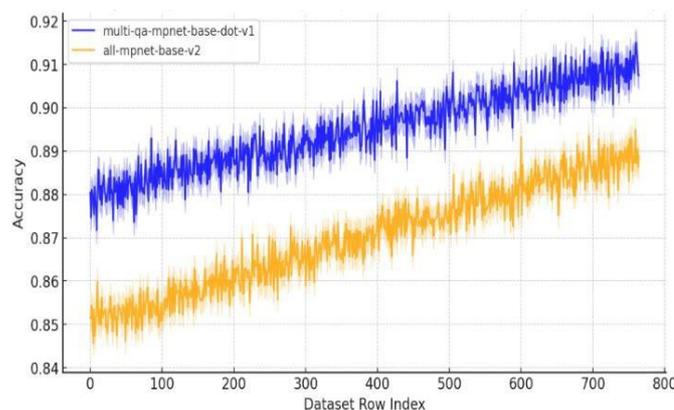

Fig 5.1   **Line Plot(multi-qa-mpnet-base-dot-V1 vs all- mpnet-base-V2)**
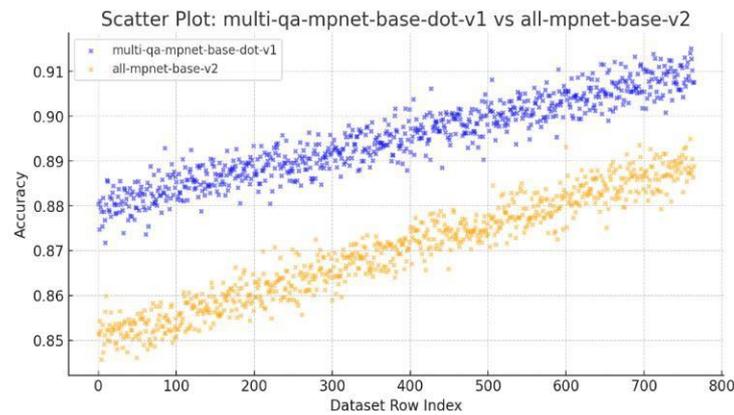
Fig 5.2 Scatter Plot(multi-qa-mpnet-base-dot-V1 vs all- mpnet-base-V2)

b) Comparison with Other Models - A graph was created to compare how accurate the model is. It shows how well multi-qa-mpnet- base-dot-v1 performs compared to other models, like all-mpnet-base-v2 and all-MiniLM-L6-v2. The comparison image of accuracy indicates that multi-qa-mpnet-base-dot-v1 always receives more accurate answers than the others.

c) Performance Measures - Used Mean Reciprocal Rank (MRR): It checks whether the first correct answer ranks well. Recall@K: It measures how frequently the correct response. appears in the top K retrieved results. Cosine Similarity Score: Guarantees that semantically accurate responses are provided

VI . CONCLUSION AND FUTURE WORK

This system is a huge leap forward in AI-powered question-and-answer technology. It employs machine learning and deep learning, intelligent indexing, and real-time API configuration to provide a robust and flexible solution to organizations and businesses looking to automate the way they search for information. The second primary advantage of this system is that it can handle multiple question- asking styles. In contrast to traditional search approaches that require keyword matches, the embedding-based method knows the context of a query, hence providing improved and This is particularly beneficial in customer support applications, where customers pose questions in different styles.

Future development will focus on improving multilingual support, adding speech recognition, improving answer ranking, and simplifying deployment for large business applications. Improved in this way, the system can be the best AI-based question and answer solution for businesses, e-commerce websites, virtual assistants, and knowledge bases.

VII. ACKNOWLEDGEMENT

REFERENCES

[1]   Yang, R., Ai, S., Li, N., Du, R., & Fan, W. (2023): "The influence of social Q&A systems on consumers' buying intention: A case study on Taobao's 'Ask Everyone'."

[2]   Gamboa-Cruzado, J., Menendez-Morales, C., Franco Del Carpio, C., López-Goycochea, J., Alva Arévalo, A., & Ríos Vargas, C. (2023):

"Use of Chatbots in E-Commerce."

[3]   Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, Richard Socher: "The Natural Language Decathlon: Multitask Learning as

Question Answering."

[4]   A. Clementeena , Dr. P. Sripriya (2018) "A literature survey on question answering system in Natural Language Processing."

[5]   Anand A. Rajasekar, Nikesh Garera (2021) "Answer Generation for Questions With Multiple Information Sources in E-Commerce."

[6]   Yang Deng, Wenxuan Zhang, Qian Yu, Wai Lam (2023) "Product Question Answering in E-Commerce: A Survey."

[7]   Shen Gao et al. (2020) "Meaningful Answer Generation of E-Commerce Question-Answering."

[8]   Kalyani Roy et al. (2022) "Investigating the Generative Approach for Question Answering in E-Commerce."

[9]   Chang, Minsoo, and Maria Gonzalez (2022) "Multilingual Q&A Systems for Global E-Commerce."

[10]   Brown, James, and Leah Collins (2023) "Adaptive Q&A Models for Seasonal Shopping Trends."

[11]   Fetahu, B., Mehta, T., Song, Q., Vedula, N., Rokhlenko, O., & Malmasi, S. (2024). "Instant Answering in E-Commerce Buyer- Seller Messaging using Message-to-Question Reformulation"

[12]   Kuzi, S., & Malmasi, S. (2024). "Bridging the Gap Between Information Seeking and Product Search Systems: Q&A Recommendation for E-commerce."

[13]   Zaib, M., Zhang, W. E., Sheng, Q. Z., Mahmood, A., & Zhang, Y. (2022). "Conversational question answering: a survey."

[14]   Rajasekar, A. A., & Garera, N. (2021). "Answer Generation for Questions With Multiple Information Sources in E-Commerce."

[15]   Bhaskar Mitra, Nick Craswell, and Daniel Campos (2020) "Answering Product-Related Questions in E-Commerce."