Meetsum: AI-Based Multi-lingual Meeting Summarizer

Saiesh Kulkarni[1], Anant Suryawanshi[2], Atharv Rokade[3] , Rushikesh Vetal[4] , Vedant Jadhav[5]
[1,2,3,4]*Student, Department of Computer Engineering, Jaywant College of Engineering and Polytechnic K.M. Gad*
[5]*Assistant Professor,Department of Computer Engineering, Jaywant College of Engineering and Polytechnic K.M. Gad*

***Abstract*—MeetSum: Online Meeting Summarizer and Report Generation System is an Artificial Intelligence (AI)-driven web application designed to automate meeting documentation and enhance organizational productivity. The increasing volume of online meetings has created a need for efficient tools that can accurately capture, process, and summarize meeting content. The proposed system addresses this challenge by converting meeting audio and video recordings into text transcripts, generating concise summaries, extracting key discussion points, and providing multilingual translation support.The system supports both uploaded recordings and real-time meeting processing, enabling users to access structured meeting information with minimal effort. Faster-Whisper is employed for high-accuracy speech-to-text conversion, while the OpenRouter API facilitates translation and AI-based text processing tasks. Natural Language Toolkit (NLTK) is utilized for keyword extraction and text analysis, and Google Text-to-Speech (gTTS) provides audio playback functionality. The application is developed using React for the frontend, FastAPI for backend services, and MongoDB for secure data storage and management.**

***Index Terms*— Meeting Summarization, Speech-to-Text, Natural Language Processing, Artificial Intelligence, Faster-Whisper, OpenRouter API, FastAPI, React, MongoDB, Multilingual Translation, Report Generation.**

I. INTRODUCTION

Meetings are super important for things like making decisions, sharing ideas, and getting work done together. They happen everywhere—from schools to big businesses to government offices. Because more people are working from home these days, there are tons more virtual meetings.

Managing these meetings is tough though. Keeping track of all the info, like key talking points and tasks, is usually left to hand-written notes. This method takes forever, can lead to mistakes, and often skips over important stuff. That makes follow-ups way harder and less effective.

Another big issue is when meetings have folks speaking different languages. These barriers really slow things down and make collaboration a lot harder. Plus, going through long audio or video recordings to find what's important is super time-consuming.

There are solutions out there, but many aren't very helpful. You often need several different tools just to transcribe, translate, and sum up the main points. This ends up being really messy and not very productive at all.

Advancements in AI, Speech Recognition, and Natural Language Processing have led to the creation of smart systems that accurately understand human speech. We can now use tools like automatic speech recognition, machine translation, keyword extraction, and text summarization. These allow us to automate meeting notes and make info more accessible.

This paper introduces MeetSum – an online meeting summarizer and report generator. It's an AI web app designed to transcribe, summarize, and automatically generate reports from your talks. MeetSum turns speech into text, creates summaries, pulls out key words, and even does multilingual translations. Plus, it lets you convert text back to speech and safely store your meeting recordings.

Users can handle their meeting data much better with these features. The tool works great for lectures, business talks, webinars, and training sessions. To build MeetSum, the creators used React for the front end, FastAPI for backend stuff, and MongoDB for data handling. They also plugged in Faster-Whisper for recognizing speech, the OpenRouter API for language work, NLTK for text analysis, and gTTS for speech output.

By mixing all these tech pieces, the system becomes really adaptable and easy to use for managing meetings of all types.

The main goal of the system is to make meeting documentation easier, improve communication for those speaking different languages, cut down on the work of taking notes manually, and make meeting info more accessible. By creating precise transcripts, summaries, translations, and downloadable reports, users can easily get the key points without watching full recordings. Centralized data storage helps with quick information finding and managing records over time.

MeetSum does more than just transcribe; it gives a total meeting management solution. It aids multilingual conversation, auto-generates reports, extracts keywords, and stores data securely. Plus, there's room for future improvements like recognizing speakers, doing sentiment analysis, moving to the cloud, fitting into common conferencing tools, supporting mobile apps, and using AI better. In all, MeetSum shows a smart way to manage meetings that boosts productivity, teamwork, and info access in different work settings.

II. REVIEW OF LITERATURE

Table 1. Literature Review

Sr	Paper Details	Key Findings	Research Gap
1	Online Meeting Summary Generator (2023) – Yashodara P.H.E. et al.	Developed an automated meeting minutes system using NLP, Speech Emotion Recognition (SER), text summarization, action item extraction, and progress tracking. Generates structured meeting reports.	Limited focus on real-time multilingual summarization; lacks advanced LLM-based abstractive summarization and high contextual understanding.
2	SmartMeeting: Automatic Meeting Transcription and Summarization for In-Person Conversations (MM'21, 2021) – Yuanfeng Song et al.	Introduced a real-time meeting transcription and summarization system with ASR, speaker identification, diarization, and summarization. Improves meeting management and transcript enrichment.	Focused mainly on English meetings and in-person settings; lacks multilingual translation and detailed report generation.
3	Meeting Summarization: A Survey of the State of the Art (2022) – Lakshmi Prasanna Kumar & Arman Kabiri	Comprehensive survey covering extractive, abstractive summarization, datasets, evaluation metrics (ROUGE, BERTScore), and transformer models such as BERT, GPT, PEGASUS.	Survey paper only; does not implement a practical system or address multilingual meeting summarization.
4	Meeting Summarizer and Plan of Action Generator using NLP (2025) – Gajender Kumar et al.	Uses TF-IDF and TextRank for meeting summarization and scheduling. Focuses on Microsoft Teams transcripts for generating summaries.	Uses traditional NLP methods, lacks deep learning/LLM models, speaker diarization, and multilingual support.
5	AI-Powered Virtual Meeting Summarization System (2025) – Mohd Salique Khan & Rabiya Khan	Proposed a Gemini-based AI summarization framework integrating ASR, speaker diarization, NLP, and email distribution for meeting summaries.	Computational complexity and dependency on internet/cloud APIs; limited multilingual and offline capabilities.
6	Virtual Meet Summarizer (JETIR, 2024) – Ansari Fatima Anees et al.	Developed a system using NLP and Machine Learning to summarize virtual meetings and track participant	Limited use of advanced transformer-based summarization and lacks action-item automation.

		contributions.	
7	Techniques for Meeting Summarization: An Analysis and Suggestions for Improvement (2022) – Viveksheel Yadav et al.	Proposed a hybrid approach combining extractive and abstractive summarization, including speech-to-text and OCR/text extraction from meetings.	Lower performance on long sequences and lacks real-time multilingual implementation.
8	Online Meeting Summarizer and Report Generation using GenAI (2025) – Sanika Ingle et al.	Surveyed Generative AI and LLM-based meeting summarization, emphasizing context awareness, sentiment analysis, multimodal integration, and report generation.	Mostly survey-based; lacks practical implementation framework and real-time evaluation.
9	Meeting Summarizer Using NLP (IRE Journals, 2025) – Lekhana R et al.	Proposed a hybrid NLP-based system integrating speech-to-text, extractive + abstractive summarization (T5, BART), Named Entity Recognition (NER), and action item extraction for meeting summaries and task management.	Limited real-time implementation, dependency on transformer models increases computational cost, and lacks multilingual translation support.

In addition to existing research on meeting summarization, several modern frameworks and APIs have significantly contributed to the development of intelligent meeting management systems. OpenRouter provides a unified interface for accessing multiple Large Language Models (LLMs), enabling advanced text summarization, multilingual translation, and contextual language understanding. Faster-Whisper, an optimized implementation of OpenAI's Whisper model, offers efficient and accurate speech-to-text transcription while supporting local execution with reduced dependency on external speech services. Supporting technologies such as gTTS enhance accessibility through multilingual text-to-speech conversion, while FastAPI facilitates high-performance backend API development with asynchronous processing capabilities. MongoDB provides flexible and scalable storage for transcripts, summaries, reports, and user data, whereas React enables the creation of responsive and interactive user interfaces through its component-based architecture. Additionally, NLTK serves as a lightweight NLP framework for keyword extraction and text analysis, while Axios and Framer Motion improve frontend communication and user experience through efficient API handling and modern interface animations. Collectively, these technologies form a robust ecosystem for developing scalable, AI-driven meeting summarization platforms capable of supporting transcription, translation, summarization, report generation, and multilingual communication.

III. OBJECTIVE OF THE STUDY

This study aims to develop an AI-based meeting summarization system that enhances organizational productivity through automated meeting management. The proposed system utilizes advanced speech recognition techniques to accurately convert spoken conversations into text transcripts, ensuring reliable documentation of meeting discussions. Furthermore, natural language processing (NLP) and machine learning algorithms are employed to analyze the transcripts and extract key information, including important decisions, action items, discussion topics, and critical insights. By generating concise and meaningful summaries automatically, the system reduces the time and effort required for manual note-taking and review, thereby improving information accessibility, collaboration, and decision-making efficiency. The integration of speech-to-text conversion and intelligent summarization technologies provides a scalable and effective solution for managing meeting records in both academic and professional environments.

IV.RESEARCH METHODOLOGY

The proposed system is an AI-powered platform that automates meeting documentation, transcription,

translation, and summarization. It uses AI, speech recognition, and NLP to turn talks into useful text. To make sure it's accurate, scalable, and user-friendly, the team relies on computer-aided tools and modern web tech.

Users can either upload meeting recordings or let the system handle live conversations to get transcripts, summaries, translated captions, key terms, and reports. This cuts down on manual labor, breaks down communication barriers, and ramps up meeting efficiency.

Built with a full-stack architecture, the system lets the front end and back end work separately but talk through APIs. The front end makes things easy for users to interact with, while the back end does the heavy lifting like speech recognition, summarization, and translation. So, the setup aims to be both powerful and easy-to-use.

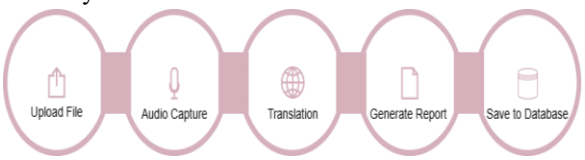

Fig 1.Flowchart Representing the Workflow of MeetSum

Step 1: Input Acquisition

To start, users load their meeting data onto MeetSum. This can be audio or video files from meetings. The system handles both live and recorded sessions too. Uploaded files then get used for the next steps - transcription, translation, and summary making, basically.

Step 2: Speech Capture and Preprocessing

After the file arrives, the audio gets plucked out and prepped for speech recognition. Besides handling uploaded recordings, the system tackles live meeting processing too. It does this by grabbing audio from microphones, browser tabs, or a mix of both. The captured speech is then preprocessed to make it better suited for transcription and other language processing tasks.

Step 3: AI-Based Language Processing

Artificial Intelligence and Natural Language Processing make the magic happen. First, the Faster-Whisper model turns speech signals into text accurately. Next, the text goes through OpenRouter API for translation and comprehension in various languages. Then, more NLP steps in to pluck out key words, whip up snappy summaries, and spot crucial points. This morphs meeting chats into solid info while handling diverse tongues too.

Step 4: Report Generation

Once the system wraps up transcription and language processing, it creates thorough meeting reports. These reports feature full transcripts, summaries, key words, discussion points, and translations when needed. Users can easily recap vital info without rewinding through the whole recording.

Step 5: Data Storage and Retrieval

In the final stage, all meeting artifacts—transcripts, summaries, keywords, translations, and reports—are safely stored in a MongoDB database. Users can access this info via a personalized dashboard to retrieve, review, and manage records easily. This system ensures data is persistent and accessible, making meeting info management a lot more effective.

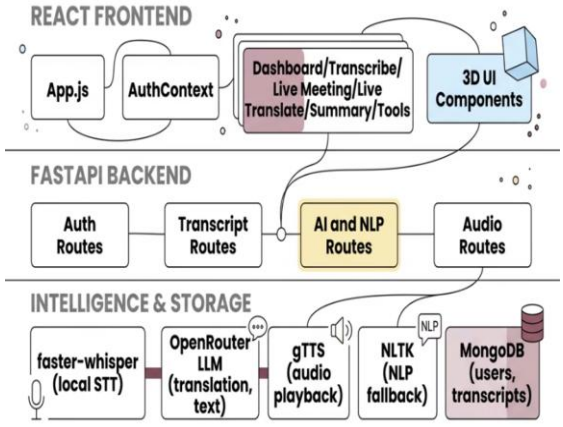

Fig 2.Overall System Architecture of MeetSum

The MeetSum system has three key parts: the Frontend, Backend, and Intelligence & Storage Layer. The React Frontend lets users upload meeting recordings, hold live meetings, and check transcripts, summaries, and translations. It also offers AI-powered tools. The FastAPI Backend handles communication, doing tasks like user auth, managing transcripts, audio processing, and AI/NLP jobs such as summarization and keyword extraction. The Intelligence & Storage Layer uses Faster-Whisper for turning speech to text

and OpenRouter LLM for summing up and multilinguall stuff. It uses gTTS for text-to-speech and keeps NLTK as a backup NLP system. MongoDB then stores everything safely—user info, meeting details, summaries, reports, and session metadata—to make sure the data works well across the whole platform.

Evaluation metrics of Audio /Video Length vs Processing Time

Table 2. Audio/Video length and Time required

Audio/ Video Length (Minutes)	Processing Time (Seconds)
5	18
15	50
20	67
30	110
45	165
60	210

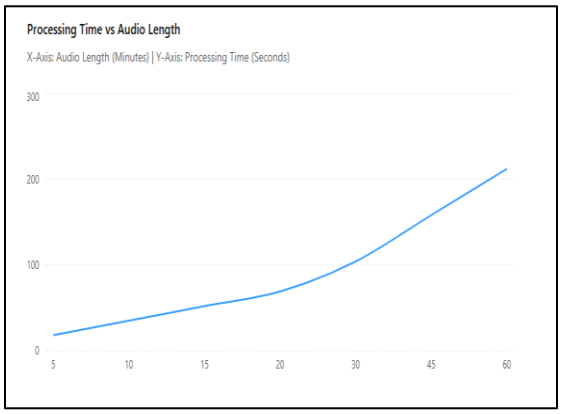

Fig 3.Processing Time vs Audio Length

The graph shows the relationship between the length of the audio recording and the processing time required by the Multi-Language Meeting Summarizer. As the audio duration increases from 5 minutes to 60 minutes, the processing time rises from approximately 18 seconds to 210 seconds. This increase occurs because longer audio recordings contain more speech data that must be transcribed, translated, analyzed, and summarized. Although the processing time grows with audio length, the system remains capable of handling extended meetings efficiently. The results indicate that the proposed solution scales effectively for both short and long meetings while maintaining acceptable processing performance for real-world applications.

Table 3. Comparison with existing systems

Feature	Zoom AI	Otter.ai	MeetSum	Project advantage
Live translated captions	Yes	Limited	Yes	Works through browser tab/mic capture for meetings.
Multilingual summaries	Yes	Limited	Yes	Summaries plus translation flow inside one dashboard.
Local Whisper processing	No	No	Yes	faster-whisper can run speech-to-text locally.
Open project source	No	No	Yes	Customizable full stack React and FastAPI codebase.
Cross-platform meeting capture	Zoom only	Meet/Zoom/ Teams	Any shared tab	Designed for Google Meet, Zoom, Teams, and browser audio.
Specialized AI outputs	Limited	Limited	Yes	Business minutes, student notes, and tweet/thread tools.

V. RESULT

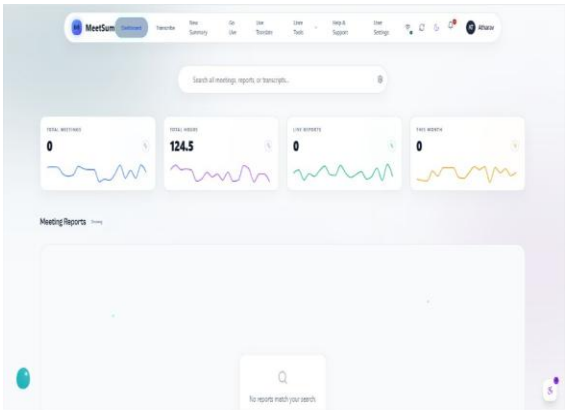

Fig 4.Dashboard

The Dashboard is the main screen for the MeetSum platform. It gives users an overall view of their meeting activities and system stats. Key metrics, like processed meetings, transcriptions, reports, and monthly trends, are shown here. Search and filter options make finding stored transcripts and reports easier. The Dashboard also helps usability and workflow by letting users manage and keep track of meeting info all in one spot.

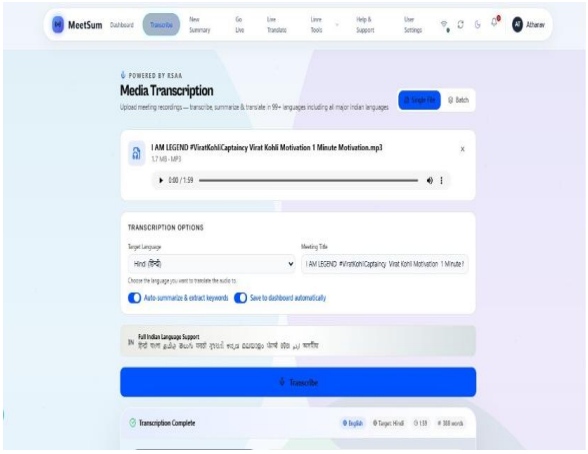

Figure 5. Transcription Page

The screen shows your media file along with the transcription settings before processing starts. Here, you can check file info, pick an output language, save space by enabling dashboard storage, and turn on AI summarization. This makes sure everything is set up just how you want it before it gets started.

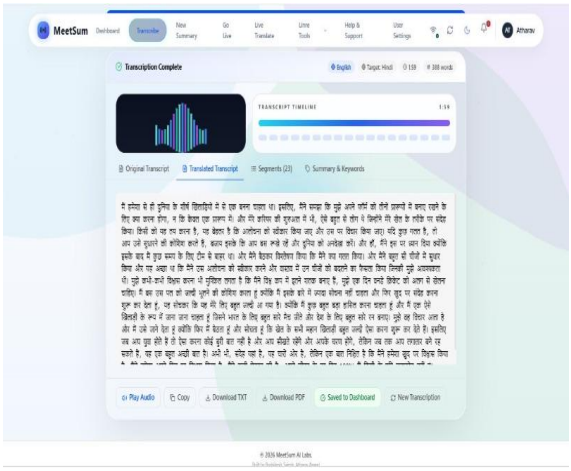

Figure 6. Transcription Result Page

After the audio is processed, users see the transcript on the Transcription Result page. They can listen to the audio, check translations, copy text, and download reports in TXT or PDF forms. Users can also save stuff to their dashboard. This page makes it easy to manage and review transcripts.

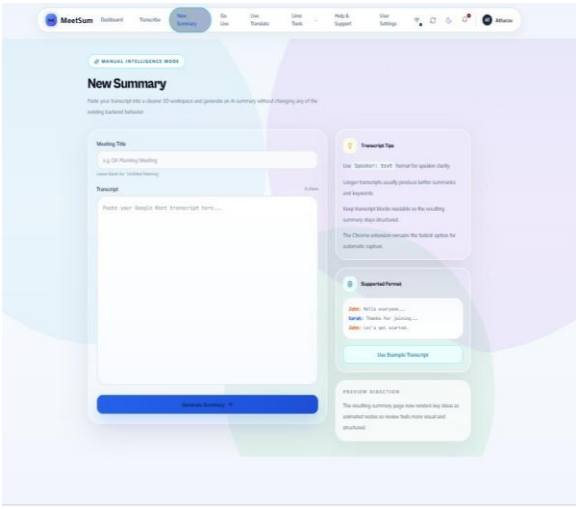

Figure 4.9.7: New Summary Module

The New Summary module lets users create AI-powered summaries from meeting transcripts. You just put the transcript text in the workspace, and the system does the rest – it scans the discussion and pulls out key points, action items, decisions, and conclusions. This makes it easier to grasp long conversations without going through everything.

VI. CONCLUSION

This paper introduced MeetSum: Online Meeting Summarizer and Report Generation System, an AI-based platform designed to simplify meeting documentation. It combines speech recognition, natural language processing, multilingual translation, and text summarization to turn meeting recordings into useful info.

MeetSum does speech-to-text transcriptions, auto summary generation, keyword extraction, multi-language translation, text-to-speech conversion, and secure report storage. With tech like React, FastAPI, MongoDB, Faster-Whisper, OpenRouter API, NLTK, and gTTS, it's efficient and easy to use. The system should lower manual doc efforts, boost productivity, and make meetings more accessible in different languages.

Though performance can dip with poor audio or big files, accuracy is generally good. MeetSum helps with quick info retrieval and breaks down language barriers during meetings.

Overall, the system hits its goal of offering an automated solution for meeting tasks. It's super helpful for orgs and businesses looking to handle meetings more efficiently. Plus, it lays groundwork for future AI-based meeting analysis improvements.

VII. FUTURE WORK

MeetSum aims to boost its transcription, user experience, and overall function with top-notch AI tech. They plan on improving it with things like better speaker ID, smoother integrations with platforms like Zoom and Microsoft Teams, support for mobile apps, multilingual features, deep analytics, and AI task tracking. With these upgrades, MeetSum will rock as an ultra-efficient tool for handling meetings in the modern world.

REFERENCES

[1]S. Ingle, K. Todkar, S. Kapure, G. Yadav, and V. Kokane, "Online meeting summarizer and report generation using GenAI," Challenges in Information, Communication and Computing Technology, pp. 576–580, Nov. 2024, doi: https://doi.org/10.1201/9781003559092-99.

[2]R. Lekhana, N. Anjum, S. Saqeeb, and I. Khan, "Meeting Summarizer Using NLP," IRE Journals |, vol. 9, no. 4, pp. 2456–8880, 2025, doi: https://doi.org/10.64388/IREV9I4-1712104.

[3]Yashodara P.H.E, Ranepura R.S, Bhashitha W.P.K.K, Thilakarathne M.H.K.T.S, Anuradha Karunasena, and Pradeepa Bandara, "Online Meeting Summary Generator," International journal of computer applications, vol. 185, no. 16, pp. 33–39, Jun. 2023, doi: https://doi.org/10.5120/ijca2023922864.

[4]V. Yadav, F. Ahmad, and A. Singh, "View of Techniques for Meeting Summarization: An Analysis and Suggestions for Improvement," Philstat.org, 2026. https://www.philstat.org/index.php/MSEA/article/view/688/376

[5]A. Anees, A. Furqan, S. Adnan, S. Rahil, and S. Lubna, "Issue 3 www.jetir.org (ISSN-2349-5162)," JETIR2403880 Journal of Emerging Technologies and Innovative Research, vol. 11, no. 3, 2024, Available: https://www.jetir.org/papers/JETIR2403880.pdf

[6]R. Khan and S. Khan, "AI-Powered Virtual Meeting Summarization System," SSRN (Social Science Research Network), Jan. 2025, doi: https://doi.org/10.2139/ssrn.5319091.

[7]Y. Song et al., "SmartMeeting: Automatic Meeting Transcription and Summarization for In-Person Conversations," ACM Digital Library, Oct. 20, 2021. https://doi.org/10.1145/3474085.3478556 (accessed May 24, 2026).

[8]L. P. Kumar and A. Kabiri, "Meeting Summarization: A Survey of the State of the Art," arXiv.org, 2022. https://arxiv.org/abs/2212.08206

[9]O. Router, "OpenRouter," OpenRouter, 2023. https://openrouter.ai/

[10]SYSTRAN, "GitHub - SYSTRAN/faster-whisper: Faster Whisper transcription with CTranslate2," GitHub, Oct. 31, 2025. https://github.com/SYSTRAN/faster-whisper? (accessed May 27, 2026).

[11]Pierre Nicolas Durette, "gTTS — gTTS documentation," gtts.readthedocs.io. https://gtts.readthedocs.io/en/latest/ [12]FastAPI, "FastAPI," fastapi.tiangolo.com, 2023. https://fastapi.tiangolo.com/

[13]MongoDB, "MongoDB," MongoDB, 2024. https://www.mongodb.com/

[14]Meta Open Source, "React," react.dev, 2025. https://react.dev/

[15]NLTK, "Natural Language Toolkit — NLTK 3.4.4 documentation," Nltk.org, 2009. https://www.nltk.org/

[16]Axios Contributors, "axios | Promise based HTTP client," Axios.rest, 2026. https://axios.rest/ [17]Motion Division, "Motion," Motion.dev, 2025. https://motion.dev/

[18]A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision," arXiv:2212.04356 [cs, eess], Dec. 2022, Available: https://arxiv.org/abs/2212.04356

[19]OpenAI, "Whisper," GitHub, Oct. 09, 2022. https://github.com/openai/whisper

[20]MongoDB, "MongoDB Documentation," https://github.com/mongodb/docs-bi-connector/blob/DOCSP-3279/source/index.txt, 2024. https://www.mongodb.com/docs/