June 2025 | Volume 1 | Issue 1 | Article No. 25203

Journal of Smart Sensors and Computing

Research Article | *Open Access* |

Attendie AI: A Smart and Proactive Solution for Overlapping Meetings Using Natural Language Processing and Machine Learning

Jaimeet Sarode,* Arya Teli, Nilesh Apsingkar and Miheer Abhyankar

Department of Computer Science and Engineering, MIT ADT University, Pune, Maharashtra, 412201, India
**Email:* jaimeetsarode@gmail.com (J. Sarode)

Abstract

When put on a conference table, the smart device Attendie records sessions and produces a text file that can be downloaded along with bulleted summaries. It can participate in meetings that are held online using Zoom, Google Meet, or Microsoft Teams and notify the user when their name is called during the meeting. In addition, Attendie may automatically join meetings and connect with the user's Google calendar. Attendie recognizes and transcribes audio signals using computational linguistics to translate spoken words into text. In this procedure, sophisticated machine learning models select and convert audio inputs into text that can be updated and saved on a specific device using linguistic algorithms. The words are also converted into Unicode characters during transcription to make them easier to display and compatible with a variety of hardware and software. The ability to instantly attend online meetings and a notification function that warns users when their name is mentioned are just two of Attendie's many advantages over its rivals.

***Keywords*:** Machine learning algorithms; Data analysis techniques; Artificial intelligence applications; Deep learning frameworks; Computational efficiency in AI models.

Received: 18 February 2025; Revised: 25 March 2025; Accepted: 15 May 2025; Published Online: 25 May 2025.

1. Introduction

In today's fast-paced world, especially within corporate environments and academic institutions, efficient management of meetings has become increasingly critical. Attendie AI is an innovative tool designed to enhance productivity during meetings by automating tasks such as transcription, summarization, and notifications. This tool aims to assist users by "attending" meetings on their behalf when they have overlapping schedules or cannot join due to other commitments.[1] By recording, transcribing, and summarizing the content of discussions, Attendie ensures that users remain informed on essential points and follow-up actions, minimizing the risk of missing key information.[2]

Effective meeting management is essential, particularly in collaborative workplaces where missed information or incomplete notes can result in delays for subsequent tasks.[3] In busy office settings, employees often struggle to keep up with multiple simultaneous meetings, which can lead to gaps in information and hinder productivity. Similarly, in universities, students may miss lectures due to conflicting schedules, impacting their learning experience and academic performance. A system like Attendie can help resolve these issues by enabling users to stay updated on essential discussions without needing to attend every meeting or lecture physically.[4]

The primary issue that Attendie AI addresses is straightforward yet significant: it helps users avoid missing important information from meetings they cannot

DOI: https://doi.org/10.64189/ssc.25203

attend. By providing concise summaries and notifying users when their names are mentioned or when they are assigned tasks, Attendie ensures that critical details are never overlooked. Whether applied in corporate offices, educational institutions, or even casual group meetings, Attendie is designed to save time and improve overall productivity.[5]

2. Literature review

The increasing need for efficient meeting and lecture management systems has spurred research and development in technologies like Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and artificial intelligence-driven summarization. Attendie AI aims to use these technologies to provide an innovative solution that captures audio, transcribes speech, summarizes content, and integrates with online platforms, making it an essential tool in corporate and academic settings.

2.1 Automatic Speech Recognition (ASR) technology

Automatic Speech Recognition has become a foundational technology in the development of smart meeting assistants. ASR enables the conversion of spoken language into text, which can be processed further for summarization, analysis, and storage. A notable ASR system is Whisper, developed by OpenAI, which is known for its high accuracy and capability to handle diverse accents, languages, and environments[6] is designed to perform well in challenging acoustic conditions, making it suitable for real-time meeting environments where audio clarity may vary due to background noise, speaker variability, and other factors.

Studies in ASR technology, such as by Graves *et al.* (2013), emphasize the importance of deep neural networks and recurrent architectures for accurate transcription. The introduction of Connectionist Temporal Classification (CTC) has further improved ASR accuracy by enabling networks to process variable-length sequences without requiring precise alignment between audio inputs and text outputs.[7] TransfD ASR models, such as those discussed by Dong et al. (2018), further enhance the capabilities of ASR systems by enabling parallel processing and better handling of long dependencies within audio data.[8]

Whisper's use transformer-based deep learning architectures aligns with these advances, making it an ideal choice for real-time applications. In addition, OpenAI has trained Whisper on a vast dataset that includes diverse dialects and accents, providing an ASR model that can handle global user bases in corporate and academic environments. This adaptability is critical for meeting assistants like Attendie AI, as users may come from various linguistic backgrounds and settings.

2.2 Natural Language Processing (NLP) for summarization

Once the ASR transcribes the spoken language, NLP techniques are used to summarize the content into key points. Summarization is essential for reducing information overload, especially in long meetings where not all details are relevant to every participant. NLP-based summarization techniques can be broadly categorized into extractive and abstractive summarization.

Extractive summarization selects important sentences or phrases directly from the transcript, while abstractive summarization involves generating new sentences to encapsulate the main ideas.[9] Attendie AI's summarists would benefit from abstractive models, which are more advanced but provide concise and coherent summaries that enhance readability and retain critical information.

Transformer-based models, especially those based on BERT (Bidirectional Encoder Representations from Transformers), have revolutionized NLP tasks, including summarization.[10] BERT's bidirectional approach all capture context more effectively than previous models, making it particularly useful for summarizing meeting content where understanding the relationship between ideas is crucial. Attendie AI may utilize these transformer models to extract meaningful insights from conversations and distil them into short, bulleted summaries.

A further advancement in NLP is the T5 (Text-To-Text Transfer Transformer) model by Raffel et al. (2019), which treats every NLP task as a text-generation problem. This approach is particularly beneficial for tasks that require flexibility in language generation, such as abstractive summarization.[11] With transformer-based summarization model AI can generate summaries that are not only concise but also maintain the semantic integrity of the original content, providing users with an efficient way to review essential points.

2.3 Integration with online meeting platforms

With the rise of remote work and online education, integration with online meeting platforms like Zoom, Google Meet, and Microsoft Teams has become critical for AI-driven meeting assistants. These platforms offer APIs that allow third-party applications to join meetings, capture audio, and interact with participants. Attendie AI leverages these integrations to ensure that it can participate in meetings, record audio, and notify users of key mentions and tasks, even in virtual environments.

Research on meeting assistants, such as Google's Project Euphonia, has highlighted the importance of API integration for accessibility and functionality.[12] By connecting with online meeting platforms, Attend perform real-time transcription, send notifications when specific names are mentioned, and even provide summaries after the meeting concludes. These integrations enable the tool to be versatile, functioning in hybrid environments and catering to both in-person and virtual participants.

2.4 Machine learning for personalized notifications and

task management

Apart from transcription and summarization, Attendie AI differentiates itself by notifying users when their name is mentioned or when they are assigned tasks. This feature requires machine learning models that can detect specific keywords, phrases, and user-specific identifiers. Research by Vaswani *et al*. (2017) introduced the Transformer model, which is adept at understanding sequences and is widely used for keyword extraction and notification systems.[13] These capabilities are crucial for Attendie AI, as they allow to actively monitor discussions and notify users about relevant information.

Furthermore, the ability to personalize notifications based on user preferences and priorities can be implemented using reinforcement learning (RL) models. RL can optimize Attendie's notification system to reduce irrelevant notifications, focusing only on high-priority mentions. Studies on RL in notification systems, such as those by Li *et al.* (2019), demonstrate the effectiveness of this approach in improving user satisfaction and minimizing distractions.[14]

2.5 Challenges and future directions

The development of Attendis forward certain challenges, including data privacy and ethical considerations. Recording and transcribing meetings involve capturing sensitive information, which raises privacy concerns. According to Rajpurkar et al. (2018), ensuring data security in ASR and NLP systems requires robust anonymization techniques and compliance with data protection regulations like GDPR.[15] Attendie AI must implement secure data handling practices to protect user forever, advancements in sentiment analysis and emotion detection in NLP can be explored to enhance Attendie AI. Emotion detection could allow the system to provide additional insights into the sentiment of the meeting, identifying potential concerns or action points that require immediate attention. This direction aligns with recent research in affective computing, which examines the intersection of AI and emotional intelligence.[16]

3. Prior art

3.1 Otter.ai

Otter.ai offers real-time transcription for meetings, webinars, and interviews. Its standout features include AI-generated meeting summaries, speaker identification, and the ability to capture slides. The platform integrates seamlessly with tools like Zoom, Microsoft Teams, Google Meet, and Slack, allowing automated transcription and note-sharing during meetings. Otter also provides mobile apps and browser extensions for enhanced accessibility.

The platform supports live collaboration with shared meeting notes and action items, ideal for teams. Updates can be synced with tools like Slack and HubSpot for productivity. Supports platforms like Salesforce, Google Workspace, Dropbox, and Amazon S3, making it suitable for business teams and sales operations.

Otter is primarily focused on transcription and meeting-related tasks. It may lack features for industries needing advanced audio or video editing tools. Ideal for professionals, businesses, and the education sector, particularly for hybrid or virtual meetings. We have collected this information from their official website.[17]

3.2 Meeting Owl 4

The Meeting Owl 4+, released in 2024, is Owl Labs' flagship product aimed at hybrid meeting environments. It builds upon its predecessors with improved hardware, AI-driven features, and seamless integration capabilities, ensuring enhanced communication and collaboration for hybrid teams.

Features a 64 MP sensor with a 4K Ultra HD resolution for superior video quality. It offers a 360° panoramic view with auto-focus on active speakers, using the Owl Intelligence System to track visual and audio cues dynamically. It tracks motion, voice, and facial cues to focus on the active speaker.

Equipped with 8 omnidirectional smart microphones, it ensures clear audio pickup within a 5.5-meter radius. The system automatically equalizes speaker volume to amplify quiet voices. Dual integrated speakers provide 360° sound coverage with a maximum output of 79 dB SPL, ensuring clear in-room communication. Compatible with USB-C, Enterprise WIFI, and Ethernet for flexible deployment in different spaces.

Works with most video conferencing platforms such as Zoom, Microsoft Teams, and Webex. For larger spaces, the Meeting Owl 4+ can connect with additional devices like another Meeting Owl creating a multi-camera ecosystem for comprehensive coverage. Supports hybrid brainstorming with integrations like the Whiteboard Owl, which makes in-room content digitally accessible to remote participants, managed via the Meeting Owl App. Add-ons like the Expansion Mic can extend the microphone range by an additional 2.5 meters. Premium cost compared to standard video conferencing cameras, Meeting Owl 4+ costs $1,999. We have collected this information from their official website.[18]

4. Unique value proposition of attendie AI

4.1 Ease of use as a standalone device

One of the strongest advantages of Attendie AI is its simplicity and ease of use. Unlike many other meeting management solutions that may require complex setups or software installations, Attendie AI functions as a standalone device. It can be easily placed in a meeting room or lecture hall, without requiring additional hardware or technical expertise. This makes it useful for a wide range of users, from corporate professionals to university staff, ensuring that anyone can utilize the tool without specialized training.[19]

4.2 Comprehensive feature set

Attendie AI offers a well-rounded set of features that address all key aspects of meeting management. It transcribes audio in real time, summarizes lengthy discussions into key points, and integrates with Zoom, Google Meet, and Microsoft Teams. Users can also receive notifications when their names are mentioned, or when they are assigned tasks. This combination of transcription, summarization, and personalized notifications streamlines the entire meeting process, making it easy to capture important information, even after a meeting is completed.[20]

4.3 State-of-the-art technology

Attendie AI leverages advanced technologies, including Whisper API for real-time transcription and Transformer-based NLP models for summarization. Whisper's high accuracy ensures that audio is transcribed correctly even in challenging environments with background noise or diverse accents. The NLP models used for summarization are trained to condense large amounts of spoken information into concise summaries, helping users to focus on key points without having to sift through lengthy transcripts. This state-of-the-art technology ensures that Attendie AI provides accurate and efficient results in both the corporate and academic environments.[21]

4.4 Advantages over other tools in the market

Attendie AI distinguishes itself from its competitors by offering a fully automated and comprehensive solution for meeting management. Other tools may focus solely on transcription or require manual intervention for summarization; however, Attendie AI handles the entire process autonomously. Its integration with multiple meeting platforms, combined with its ability to provide notifications and summaries, sets it apart as a more holistic tool. Additionally, the standalone nature of the device makes it more convenient than software-based tools that may require installation, updates, or compatibility checks.[19]

5. Methodology

Attendie AI is designed to simplify the process of managing meetings and lectures using advanced technologies for transcription, summarization, and integration with online platforms.

The system begins by capturing real-time audio using the Whisper API, an Automatic Speech Recognition (ASR) tool developed by OpenAI.[22] This tool helps convert spoken words into text with a high degree of accuracy, ensuring that even complex or technical language is captured properly. By working in real time, Whisper allows Attendie AI to provide instant feedback and transcription during meetings and lectures.

Once the transcription is complete, Attendie AI uses transformer-based Natural Language Processing (NLP) models to summarize the discussion. These models can understand long and detailed conversations and condensing the content into easy-to-read key points or bullet summaries.[23] This saves time for users who do not want to go through the entire transcription but still require important highlights.

Finally, Attendie AI seamlessly integrates with popular online meeting platforms such as Zoom, Google Meet, and Microsoft Teams.[24] This makes it useful not only for in-person meetings, but also for remote or hybrid setups, allowing Attendie to work in multiple environments and cater to a wide range of users, from students attending virtual lectures to professionals in corporate meetings. Fig. 1 shows the flowchart for Attendie AI.

5.1 Block diagram description

Attendie is placed in the center of the meeting rooms and captures the meeting's audio and video. The data is then sent to the cloud server, where it is processed to generate real-time transcription and intelligent key highlights. Attendie performs keyword tagging, which involves recognizing the topics spoken in the meeting and generating related summaries. Furthermore, Attendie includes action item tracking, highlighting the name of the person whenever he/she is given a certain task or when the person's name is called out. The processed data includes identification of the

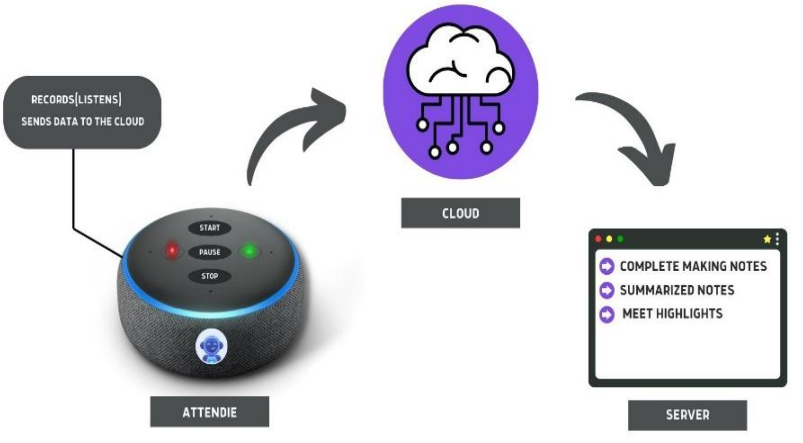

Fig. 1: Attendie AI flowchart.

speaker, keyword tagging, and action item tracking, organizing the information efficiently for easy retrieval and understanding. The processed information is sent to the client and backed up on the server.

5.2 Real-time audio capture

Attendie AI initiates the process by capturing real-time audio from the meeting environment. This critical step is made possible through integration with the Whisper API, which facilitates seamless and high-fidelity audio capture.[22] Attendie AI is equipped with external audio input devices to ensure flexibility and adaptability to various meeting setups.

5.3 Whisper API integration

Whisper, an Automatic Speech Recognition (ASR) system developed by OpenAI, is used in Attendie AI's audio capture capability. Fig. 2 illustrate the Whisper API Working. This API is powered by state-of-the-art deep learning models that have been trained on extensive datasets of spoken language, making it highly accurate and versatile.[22] Whisper directly processes raw audio waveforms using a deep learning architecture, typically based on transformer models. This approach allows it to handle a wide range of audio quality, accents, and languages.

5.4 ASR technology

Whisper's ASR technology excels in accurately transcribing spoken words into text, even in challenging acoustic environments. It is specifically fine-tuned to handle diverse accents, languages, and speaking styles, ensuring a high degree of transcription accuracy.[22] It is trained on a large corpus of text and spoken data, which allows it to handle various languages, dialects, and accents. This transformer model is designed to predict sequences of words based on the context of the audio, ensuring more accurate transcription even in noisy environments or with various speaking styles

5.5 Real-time processing

Whisper API enables Attendie AI to perform real-time audio processing, which is crucial for capturing meetings as they happen. This real-time capability ensures that meeting content is immediately available for transcription and subsequent stages of the methodology.[22] Whisper uses a greedy search as the default decoding algorithm for transcribing audio. This means the model selects the most probable transcription at each step based on the given acoustic and language model predictions. There is an option for beam search, but it is not the default. In some cases, Whisper can utilize beam search to explore multiple potential transcriptions, selecting the one with the highest overall likelihood by considering various possible sequences.[27]

5.6 Natural Language Processing (NLP) summarization

After obtaining the real-time transcription, Attendie AI applies Natural Language Processing (NLP) techniques to generate concise and coherent summaries of the meeting's content.

5.7 Transformer-based NLP models for summarization

Attendie AI leverages Transformer-based NLP (Natural Language Processing) models to perform the crucial task of summarizing meeting content. These models represent a significant advancement in NLP and are pivotal in Attendie AI's ability to distil lengthy discussions into concise and coherent summaries.

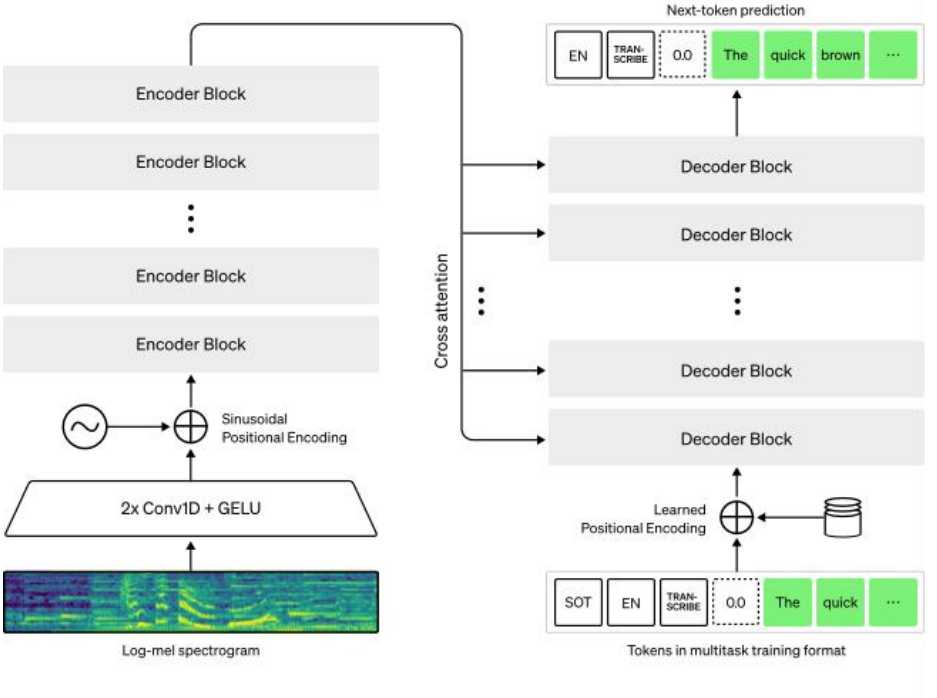

Fig. 2: Whisper API Working.[25]

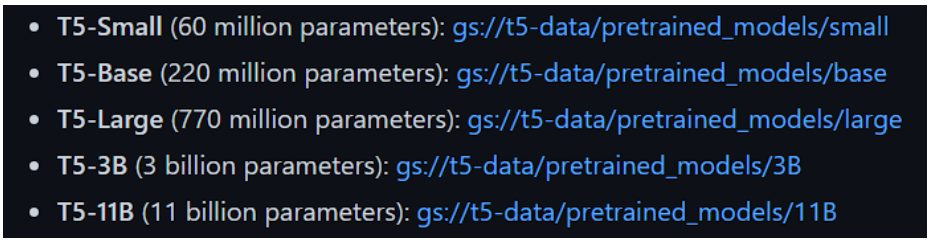

Fig. 3: Hugging face transformer models.[28]

5.8 Model preparation

When using models like T5 for text summarization, you first tokenize the input text using the corresponding tokenizer (AutoTokenizer). For example, when using T5, a common practice is to prepend a task-specific prefix (e.g., "summarize:") to the input text. This helps the model understand the task it needs to perform. Fig. 3 shows the Hugging face transformer models.

5.9 Adaptation to meeting context

Fine-Tuning to make the summarization process effective in the context of meetings, Attendie AI fine-tunes these Transformer models. Fine-tuning involves training the models on meeting-specific data, including transcripts from various types of meetings and discussions. During fine-tuning, the models adapt to meeting-specific terminology, phrases, and nuances. This ensures that the generated summaries are not only concise but also highly relevant to the content discussed in meetings.

5.10 Continuous improvement

Attendie AI's use of Transformer models is not static. The system continually learns and adapts based on user interactions and feedback. This iterative learning process contributes to improving the quality of the summaries generated over time.

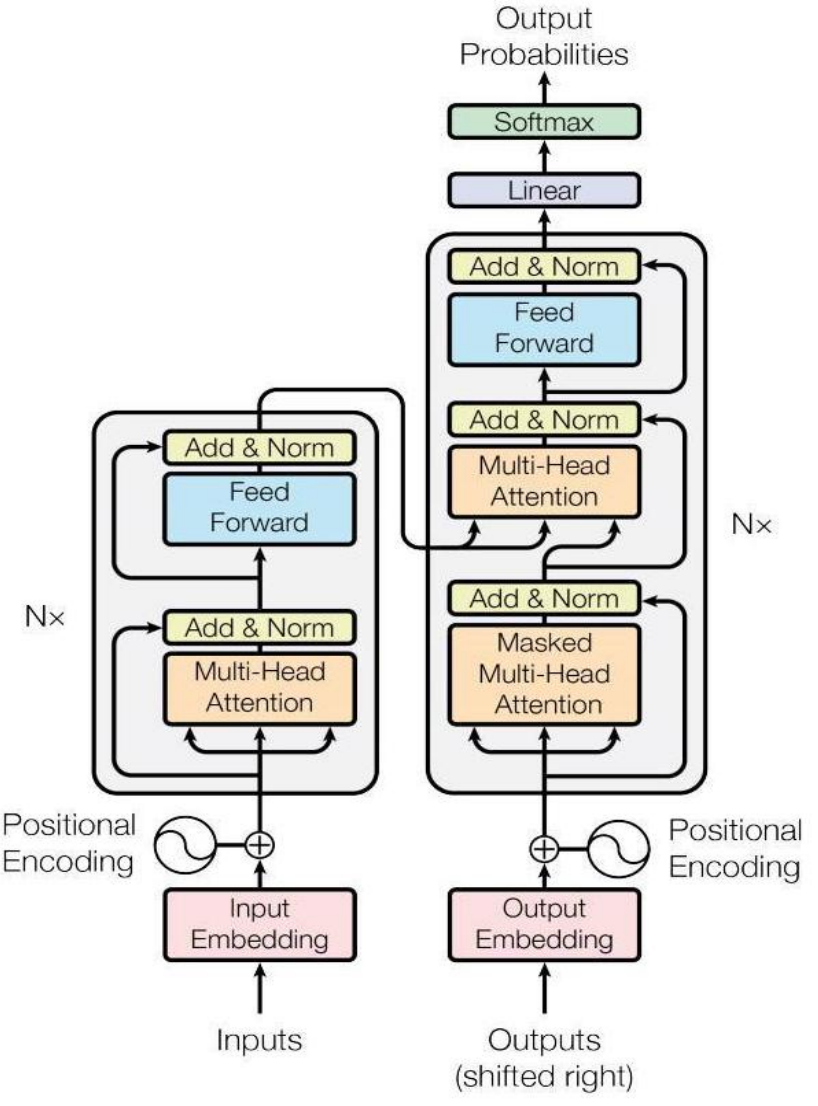

Fig. 4: Transformer Architecture. Reproduced with the permission form [26].

5.11 Multilingual support

Transformer-based models are versatile and can be extended to support multiple languages. Attendie AI can be fine-tuned to provide meeting summarization in various languages, increasing its accessibility to a global user base. Fig. 4 shows Transformer Architecture.

5.12 Software system description

The system requirements for Attendie AI can vary depending on the scale of deployment and the transcription model being used. Here are the general minimum system requirements for running Attendie AI on medium transcription model:

1. Minimum system requirements:

CPU: Hexa-core processor (Example: - Ryzen 5 4600h or equivalent)

RAM: 8gb or more

GPU: Nvidia GPU for hardware acceleration. Preferably RTX 3050 or its AMD equivalent.

Operation System: Windows 10 or higher / Ubuntu server LTS

Internet connection: High speed internet connection

2. Additional considerations:

GPU Acceleration: Using a high-end GPU can significantly accelerate certain tasks, such as deep learning-based transcription and NLP processing.

Multithreading Support: Ensure that the hardware supports multithreading, as Attendie AI benefits from multithreading for parallel processing.

6. Conclusion

In conclusion, Attendie AI provides a comprehensive solution to the challenges faced by modern management. Offering real-time transcription, intelligent summarization, and integration with popular meeting platforms ensures that no critical information is lost, even in overlapping meetings or absences. Its ease of use, as a standalone device, coupled with cutting-edge technology such as Whisper API and Transformer-based NLP models, makes it a valuable tool for both corporate and academic settings. Attendie AI not only boosts productivity but also improves decision-making by ensuring that all meeting participants stay informed and up-to-date. Looking ahead, Attendie AI holds the potential to further enhance workplace and educational efficiency, making it an indispensable asset for individuals and organizations.

Acknowledgment

We would like to express my sincere gratitude to everyone who contributed to the successful completion of this project. First and foremost, we are deeply grateful to our advisors Dr. Suvarna Pawar and Dr. Santosh Darade for their invaluable guidance, insights, and support throughout the research process. Their expertise and encouragement played a crucial role in shaping this project. We would also like to extend our thanks to our colleagues and friends who provided helpful feedback and encouragement, making the journey both collaborative and inspiring.

Conflict of Interest

There is no conflict of interest.

Supporting Information

Not applicable

Use of artificial intelligence (AI)-assisted technology for manuscript preparation

The authors confirm that there was no use of artificial intelligence (AI)-assisted technology for assisting in the writing or editing of the manuscript and no images were manipulated using AI.

References

[1] Z. Ren, S. Lv, S. Zhao, Y. Pang, Automatic Meeting Summarization with Speaker Role Classification Using Deep Learning Techniques. Journal of Computa-tional Linguistics, 2021, **47**, 345-357.

[2] D. Jurafsky, J. H. Martin, Speech and Language Processing (3rd ed.). Pearson, 2020.

[3] L. Chen, J. Zhang, Context-aware notification systems in intelligent assistants, *Journal of Human-Computer Interaction*, 2021, **37**, 512-528.

[4] S. Patel, K. Singh, A. Sharma, Machine learning-driven automation in digital assistants: applications in corporate settings, *International Journal of Artificial Intelligence*, 2022, **14**, 124–138.

[5] M. Becker, M. Esser, Unicode as a universal standard for text representa-tion in digital transcriptions, *Digital Scholarship in the Humanities*, 2020, **35**, 287–299.

[6] OpenAI. Whisper: OpenAI's Automatic Speech Recognition (ASR) Technology, OpenAI, 2023.

[7] A. Graves, A. Mohamed, G. Hinton, Speech recognition with deep recur-rent neural networks, IEEE International Conference on Acoustics, Speech and Signal Processing, 2013.

[8] L. Dong, S. Chen, B. Xu, Speech-Transformer: A no-recurrence sequence-to-sequence model for speech recognition, IEEE International Conference on Acoustics, Speech and Signal Processing, 2018.

[9] R. Nallapati, F. Zhai, B. Zhou, Abstractive text summarization using sequence-to-sequence RNNs and beyond, 2016, arXiv preprint arXiv:1602.06023.

[10] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv preprint arXiv:1810.04805, 2018.

[11] K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, Journal of Machine Learning Research, 2019, doi: 10.48550/arXiv.1910.10683.

[12] Y. Huang, B. Baker, S. Rybintsev, Project Euphonia:

Helping people with atypical speech be understood using speech recognition technology, Google Research, 2020.
[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is All You Need, Advances in Neural Information Processing Systems, 2017.
[14] R. Li, X. Wang, F. Yang, Y. Hu, Personalized notification systems using reinforcement learning for user satisfaction optimization, 2019.
[15] P. Rajpurkar, R. Jia, P. Liang, Know What You Don't Know: Unanswera-ble Questions for SQuAD, 2018.
[16] R. W. Picard, Affective Computing, MIT Press, 2015.
[17] https://otter.ai/, accessed in March 2025
[18] https://owllabs.com/, accessed in March 2025
[19] B. Smith, C. Roberts, L. Thomas, Efficiency in workplace ai: case studies on notification and summarization features, *International Journal of Productivity Science*, 2021, **32**, 73-89.
[20] M. Johnson, P. Roberts, The role of AI in hybrid meeting management, *Technology and Human Productivity Review*, 2022, **41**, 112-125, doi:
[21] Y. Li, A. Smith, R. Johnson, Advances in Standalone AI for Business Ap-plications, *Journal of AI Research*, 2020, **15**, 245-260.
[22] OpenAI. Whisper: OpenAI's Automatic Speech Recognition (ASR) Technology, OpenAI, platform.openai.com/docs/guides/speech-to-text.
[23] M. Arsalan, Transformers in natural language processing: a comprehensive review, *International Journal for Research in Applied Science and Engineering Technology*, 2024, **12**, 5591-5597, doi: 10.22214/ijraset.2024.62863.A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is All You Need, Advances in Neural Information Processing Systems, 2017.
[24] Zoom Video Communications, Inc., Google LLC, Microsoft Corporation, Integration with Zoom, Google Meet, and Microsoft Teams.
[25] openai.com/index/whisper/.
[26] ai.stackexchange.com/questions/32236/is-there-a-notion-of-location-in-transformer-architecture-in-subsequent-self-att.
[27] Robust Speech Recognition via Large-Scale Weak Supervision.
[28] github.com/openai/whisper.

Open Access