International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 03 | Mar 2026 www.irjet.net p-ISSN: 2395-0072

Auralis: An AI-Driven Digital Twin-Based Virtual Personal Assistant for Meeting Intelligence and Productivity Optimization

Arpit Yadav[1], Siddhu Kumar[2], Ambika Rajpurohit[3], Harsh Yadav[4], Prof. Ashwini More[5]

[1,2,3,4] *Diploma in Computer Engineering, Thakur Polytechnic, Mumbai, India*
[5]*Professor, Department of Computer Engineering, Thakur Polytechnic, Mumbai, India*

Abstract - *Modern professionals often have to manage several digital interactions at once, such as meetings, emails, and task coordination. However, human attention is limited, which can lead to missed information and lower productivity. This paper introduces Auralis, an AI-driven Digital Twin-based Virtual Personal Assistant designed to help users by automating meeting insights and routine communication tasks.*

Unlike traditional meeting tools that mostly act as passive recorders, Auralis actively processes real-time conversation data and generates meaningful responses based on context. The system uses large language models built on Transformer architectures, combined with retrieval-augmented generation techniques that help maintain semantic memory. Real-time communication is supported through modern web-based protocols, allowing the system to capture and analyze meeting data effectively.

Beyond participating in meetings, Auralis also summarizes discussions, extracts actionable items, and helps with email management. By using a Digital Twin model, the system learns from user behavior and offers context-aware assistance over time.

Auralis aims not to replace human involvement but to lessen the burden of repetitive tasks and enable more efficient participation across various workflows.

1. INTRODUCTION

Remote collaboration is now a key part of modern work processes, especially with the rise of digital communication platforms. Individuals often find themselves attending numerous meetings, managing emails, and coordinating tasks in different settings. This growing demand can overwhelm our capacity, resulting in lower productivity and mental fatigue. While tools like video conferencing platforms improve communication, they still require constant human presence
and manual effort. The idea of a Digital Twin was first introduced to create virtual representations of physical systems for monitoring and simulation [3]. Recently, this concept has expanded to model human behavior and interaction patterns.

In this context, a Digital Twin can learn from user activities, preferences, and communication styles, allowing for personalized and adaptive support. Auralis builds on this idea by serving as an AI-driven Digital Twin-based Virtual Personal Assistant. Rather than being a passive tool, the system actively processes real- time data from meetings and user interactions.

It understands conversational context, generates relevant responses, and aids in decision-making. This is possible due to the use of Transformer-based language models [1] and retrieval-augmented generation techniques [2], which together support real-time reasoning and long-term contextual memory.

In addition to tracking meetings, Auralis also helps with task management and automating emails, functioning as a complete productivity assistant. By continuously learning from user behavior, the system aims to lessen repetitive tasks and provide support tailored to various activities. The goal is not to replace human involvement but to improve efficiency by allowing smarter and more flexible interaction with digital systems.

2. EASE OF USE

Ease of use refers to how easily users can understand and operate a system with minimal effort and technical knowledge.

In intelligent systems, usability is crucial for real-world adoption.

Even advanced technologies may fail if they are hard to use or require a lot of learning. For Auralis, ease of use is a key design focus since the intended users are not expected to have expertise in artificial intelligence or system setup. The system manages complex tasks like language understanding, contextual reasoning, and memory retrieval internally. It uses Transformer-based models [1] and retrieval-augmented generation techniques [2]. These processes remain hidden from the user, making interaction simple and intuitive. Auralis operates
with minimal user input. After an initial setup phase, where users provide basic preferences and behavioral inputs, the system begins to work as a Digital Twin. It can assist with meetings, emails, and task management. Instead of requiring

constant input, it supports semi-autonomous operation. This reduces the need for continuous user involvement while still allowing control when necessary. The interaction layer uses standard web communication technologies [4], which help it integrate smoothly with existing platforms. From the user's point of view, most actions involve straightforward triggers like enabling features or granting access permissions. The system handles underlying processes, including real-time data management and AI-based decision-making, without exposing any technical complexity. By prioritizing simplicity in interaction and minimizing manual configuration, Auralis strikes a balance between advanced functionality and usability. This approach ensures that the system can be easily adopted and used as a practical digital assistant, rather than a complicated technical solution.

3. RELATED WORK

Recent advancements in natural language processing have been driven by Transformer-based language models. These models have improved how systems understand and generate text that makes sense in context [1]. Building upon this, retrieval-augmented generation (RAG) approaches combine language models with external knowledge sources. This allows systems to create more accurate and context-rich responses [2]. These techniques form the basis for modern intelligent assistants that need both reasoning and memory skills. The idea of a Digital Twin was first introduced to model and simulate physical systems in a virtual space [3]. Recently, this concept has been expanded to represent user behavior and interaction patterns. This change allows the development
of systems that can adjust to individual preferences and workflows. It has opened new opportunities for creating personalized, context-aware digital assistants. Real-time communication tools like WebRTC have made it possible for web applications to exchange data smoothly and with minimal delay. This makes them ideal for systems needing live interaction, such as meeting support tools [4]. Meanwhile, research on human–AI interaction emphasizes the need to design systems that are predictable, clear, and easy to use.
This is crucial for building user trust and encouraging adoption [5]. From a system perspective, ideas from autonomous agent theory outline how intelligent systems can represent users while staying focused on their goals [6]. Additionally, memory-based structures, such as memory networks, highlight the need to keep and use long-term context in conversational systems [7]. Current meeting-related solutions mainly focus on recording, transcribing, and summarizing discussions [8].

While these tools are helpful for analyzing meetings afterward, they do not engage or assist during live interactions. Auralis enhances these methods by merging real-time processing, semantic memory, and Digital Twin modeling. This allows for active participation and context-aware help during meetings, as well as additional support for productivity tasks like email management and task tracking. Related Work Influencing

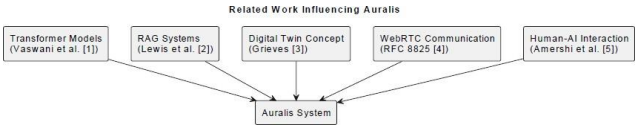

Fig -1: Related Work Overview

4. SYSTEM ARCHITECTURE

The Auralis framework is based on a modular and real-time AI system that facilitates intelligent assistance through a Digital Twin-based representation of the user. The overall system architecture is based on a pipeline-based approach, where several components of the system collectively contribute to data input processing, context analysis, and finally, response generation. The overall system architecture is based on ensuring low latency and aligning itself with the user's behavior and communication style.
The overall system is divided into seven primary components, namely, (1) Audio Capture Module, (2) Speech-to-Text Engine, (3) Context Analyzer, (4) Persona Engine, (5) Response Generator, (6) Memory Store, and (7) Communication Interface. Fig. 3 represents the overall workflow of the system. Firstly, the overall workflow of the system is initiated by
the Audio Capture Module, which collects real-time audio from meetings and/or user interaction. The collected audio is then sent for further processing by the Speech-to-Text Engine, which generates text from the collected audio. The generated text is then analyzed by the Context Analyzer component of the system.
To ensure that the response is aligned with the user's style of communication, the Persona Engine includes user-specific preferences and patterns. This allows the system to provide responses that are similar to the user's style of response in a particular situation. Then, the Response Generator uses the context and persona information, in addition to language models, to create a response that is significant and understandable.
Another significant aspect of the Auralis architecture is the Memory Store. This allows the system to retain context over a period of time, enabling it to recall past interactions. This is useful for retrieval-augmented generation, making the response provided by the Auralis system more accurate. Lastly, the

Communication Interface is responsible for facilitating interaction in real-time with external applications. This allows the Auralis system to integrate easily with a meeting system and other applications for effective communication. All these aspects of the Auralis architecture make it a powerful and intelligent assistant that can be useful in a variety of applications.

4.1 Audio Capture Module

The Audio Capture Module is responsible for continuously receiving the incoming audio stream from the meeting environment. It interfaces directly with the communication platform and performs basic preprocessing such as noise filtering and buffering to ensure stable input for downstream components. The module operates in real-time to minimize delays and maintain smooth interaction.

4.2 Speech-to-Text Engine

Upon completion of basic audio preprocessing, the processed audio stream is passed seamlessly to the Speech-to-Text (STT) Engine, which converts spoken input into textual form. Since all higher-level processing depends on text, the accuracy of this component is critical. The generated transcripts are timestamped and forwarded to the Context Analyzer for further interpretation.

4.3 Context Analyzer

The Context Analyzer interprets the conversational flow of the meeting, identifying intent, tracking ongoing topics, and evaluating the relevance of different speakers. Using transformer-based language models to capture contextual relationships within the dialogue [1], it determines whether a response is required or if information should be stored for future reference.

4.4 Persona Engine

To ensure that system responses remain consistent with the user, Auralis includes a Persona Engine that models communication style, preferences, and behavioral patterns. The persona is initialized during the setup phase and refined over time using interaction history. This enables the system to generate responses that reflect the user's tone and decision making approach, which is crucial for maintaining trust in human–AI interactions [5].

4.5 Response Generator

The Response Generator produces replies by combining the current context with relevant past information. It uses a retrieval-augmented generation (RAG) approach [2], where important data is retrieved from memory and incorporated into the response. The underlying model is based on the Transformer architecture [1], enabling coherent and context-aware language generation

4.6 Memory Store

The Memory Store maintains both short-term and long-term information, including conversation history, key decisions, and user preferences. Inspired by memory network architectures [7], this component enables the system to retain context across interactions and continually improve its performance over time. It also supports post-meeting review by providing structured access to past discussions.

4.7 Communication Interface

The final output is delivered back to the meeting environment through a real-time communication interface based on WebRTC technologies [4]. This enables seamless integration with existing platforms and ensures low-latency interaction. Within the meeting, the Digital Twin appears as an active participant representing the user.

Overall, this modular design separates perception, reasoning, and communication processes, allowing the system to operate efficiently while remaining scalable and extensible for future improvements.

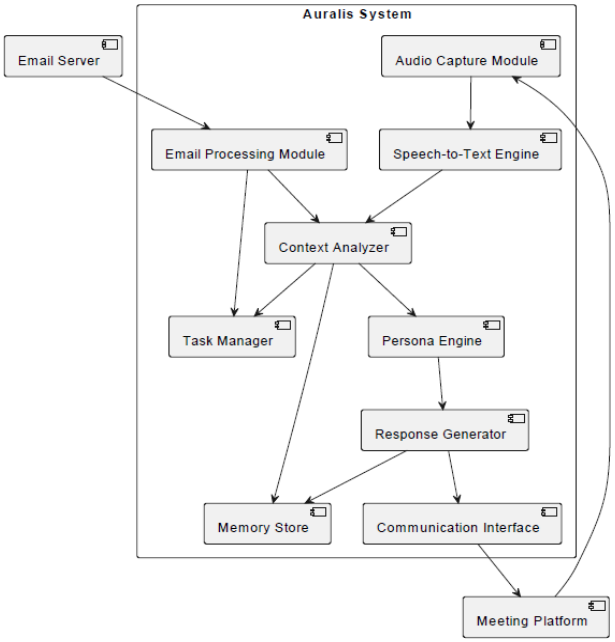

Fig -2: System architecture

5. METHODOLOGY

The development of Auralis follows a pipeline-based approach designed to support real-time interaction and intelligent assistance through a Digital Twin model. The methodology integrates multiple stages, including speech processing, contextual understanding, persona alignment, response generation, and task management, into a unified workflow. The overall design focuses on maintaining low latency while ensuring that the system remains consistent with user behavior. The process begins with capturing audio input from the meeting environment, which is then converted into text using a speech-to-text engine. This textual data serves as the primary input for further analysis. The system then performs contextual understanding by analyzing the conversation to identify intent, key topics, and relevant information. Transformer-based models are used at this stage to capture relationships within the dialogue and generate meaningful representations of the conversation [1].

Once the context is established, the system incorporates user-specific information through a persona modeling mechanism.

This step ensures that any generated response reflects the user's communication style and preferences. The persona is continuously refined using past interactions, allowing the system to adapt over time.

To generate responses, Auralis uses a retrieval-augmented generation (RAG) approach [2], where relevant information is retrieved from a memory store and combined with the current context. This enables the system to produce responses that are both contextually appropriate and consistent with previous interactions. The generated output is then delivered back to the meeting through a real-time communication interface [4].In addition to real-time response generation, Auralis includes a task management component that identifies and extracts actionable items from conversations and emails. Using natural language processing techniques, the system detects tasks, deadlines, and responsibilities, and organizes them into a structured format. These tasks are stored and can be reviewed or updated by the user, enabling better tracking of commitments discussed during meetings. The system also maintains a memory component that stores conversation history, key decisions, extracted tasks, and user preferences. This allows Auralis to improve its performance over repeated interactions and provide more accurate and context-aware assistance. Overall, the methodology ensures a balance between real-time responsiveness, contextual accuracy, and adaptive behavior.

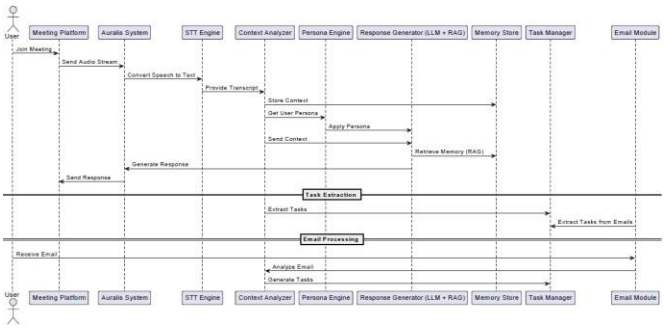

Fig -3: Auralis methodology

5.1 System Workflow

The operational workflow of Auralis begins when the Digital Twin connects to the user's working environment, including online meetings and communication platforms. For meeting scenarios, the system joins through a WebRTC-based interface [4], where incoming audio streams are continuously captured and buffered by the Audio Capture Module. The buffered audio is then forwarded to the Speech-to-Text engine for transcription.

Once converted into text, the conversation is processed by the Context Analyzer, which determines semantic intent, topic relevance, and conversational flow using Transformer-based representations [1]. Based on this analysis, the system decides whether to generate a response, store the information, extract tasks, or remain silent.

If a response is required, the Persona Engine conditions the system using user-specific behavioral parameters. The Response Generator then produces a context-aware reply using a retrieval-augmented generation pipeline [2]. The generated output is finally transmitted back into the meeting through the communication layer, enabling real-time participation.

In addition to meeting interaction, Auralis extends its workflow to email and task management. Incoming emails are analyzed using natural language processing techniques to identify important information, categorize content, and detect actionable items. Similarly, during meetings, the system extracts tasks, deadlines, and responsibilities from conversations.

These tasks are structured and stored for user review, allowing efficient tracking of commitments across different workflows.

5.2 Persona Modeling

Auralis employs a structured persona modeling approach to maintain alignment with the user's communication style. During the initialization phase, users provide preference data such as tone (formal or informal), response length, domain expertise, and meeting priorities.

These attributes form the base persona profile.

The persona is continuously refined using interaction history and feedback. This adaptive approach ensures that system responses remain consistent with user expectations and improve personalization over time, following established human–AI interaction principles [5].

5.3 Retrieval-Augmented Response Generation

To improve response quality and contextual grounding, Auralis uses a Retrieval-Augmented Generation (RAG) mechanism [2]. The process involves three main steps:

1) Formulating a query based on the current context
2) Retrieving relevant information from memory
3) Generating a response using a Transformer-based model

This approach helps reduce incorrect or irrelevant outputs and ensures that responses are aligned with both current and past interactions [1].

5.4 Memory Management Strategy

Auralis maintains both short-term and long-term memory layers. Short-term memory stores the ongoing meeting or interaction context, while long-term memory maintains historical data such as previous conversations, decisions, extracted tasks, and user preferences.

The memory system is inspired by neural memory network concepts [7], enabling efficient retrieval of relevant information when needed. After each session, important discussions are summarized using meeting summarization techniques [8] and stored for future reference. This allows the system to maintain continuity across meetings and communication channels.

5.5 Task Extraction and Management

Auralis includes a task management mechanism that automatically identifies actionable items from both meetings and emails. Using natural language processing, the system detects tasks, deadlines, and assigned responsibilities within conversations and written communication.

These extracted tasks are organized into a structured format and stored in the system, allowing users to review, update, or prioritize them. This feature reduces the need for manual note-taking and ensures that important actions are not missed.

5.6 Autonomous Decision Policy

Auralis does not respond to every input. Instead, it follows a decision policy to determine when a response is appropriate.

The system evaluates factors such as:

- Relevance of the speaker
- Directness of the query
- Importance of the topic
- User-defined priorities

This selective response strategy helps maintain natural interaction behavior and avoids unnecessary interruptions. The decision-making process is influenced by agent-based system principles, where the Digital Twin operates as an autonomous assistant [6].

5.7 Implementation Environment

The prototype implementation of Auralis is designed for deployment in a cloud-enabled environment with real-time processing capabilities. The system integrates:

- Transformer-based language models for reasoning and generation [1]
- Retrieval-augmented pipelines for contextual grounding [2]
- WebRTC-based communication for real-time interaction [4]
- Persistent memory storage for long-term context retention

The modular design ensures scalability and allows future extensions, such as multimodal inputs including video and screen context.

Overall, the proposed methodology enables Auralis to function as an intelligent Digital Twin capable of supporting meetings, managing communication, and organizing tasks, while maintaining user-specific behavior and interaction quality.

6. RESULTS AND DISCUSSION

This section evaluates the performance of Auralis across multiple functionalities, including meeting participation, task extraction, and email assistance. The objective is to assess response relevance, latency, persona alignment, and overall effectiveness of the Digital Twin in reducing user workload.

6.1 Experimental Setup

Auralis was tested in controlled virtual environments simulating real-world usage. These included online meetings conducted using WebRTC-based

communication [4], along with sample email datasets and task-oriented discussions. Meeting sessions ranged from 10 to 45 minutes and included common professional interactions such as status updates, question-and-answer sessions, and decision-making discussions.

The system utilized Transformer-based language models for reasoning [1] and a retrieval-augmented generation (RAG) pipeline [2] for context-aware responses. All interactions were logged, including response time, extracted tasks, and generated outputs for evaluation.

6.2 Evaluation Metrics

The following metrics were used to evaluate system performance:

- **Response Relevance**: accuracy and contextual correctness of generated responses
- **Latency:** time taken to generate a response after detecting a query
- **Persona Consistency**: alignment with user-defined communication style
- **Participation Accuracy**: correctness in deciding when to respond
- **Task Extraction Accuracy**: correctness of identified tasks and deadlines
- **User Effort Reduction**: reduction in manual effort for meetings and communication These metrics evaluate both technical performance and usability aspects of the system [5].

6.3 Quantitative Results

The system demonstrated stable performance across different scenarios. The average response latency ranged between 25 and 80 seconds, depending on model load and context complexity, which is acceptable for semi-autonomous meeting assistance.

Response relevance remained high in structured discussions, especially when supported by memory retrieval. The RAG-based pipeline improved contextual grounding and reduced irrelevant responses compared to standalone generation.

Task extraction achieved consistent results in both meeting transcripts and email inputs, successfully identifying key actions, deadlines, and responsibilities. The persona modelling mechanism maintained a stable communication style across sessions.

The decision policy effectively reduced unnecessary responses, allowing the system to participate only when required, which improved overall interaction quality

6.4 Qualitative Analysis

User-level observations indicate that Auralis behaves more like an intelligent assistant rather than a passive tool. Key strengths observed include:

- Context-aware participation during meetings
- Consistent and personalized communication style
- Automatic extraction of tasks from conversations and emails
- Ability to recall previous discussions through memory

The integration of meeting intelligence with email and task management significantly reduced manual workload, especially in scenarios involving follow-ups and action tracking.

However, certain limitations were observed. In meetings with heavy cross-talk or poor audio quality, transcription accuracy decreased, affecting downstream processing. Additionally, highly technical or domain-specific queries sometimes require more specialized knowledge than is available in the current system.

6.5 Comparative Discussion

Compared to traditional meeting tools and assistants, Auralis provides a more comprehensive solution. Existing systems primarily focus on transcription or summarization [8], whereas Auralis actively participates in conversations, manages tasks, and assists with communication.

The combination of Digital Twin modeling [3], Transformer-based reasoning [1], and RAG-based memory [2] enables more adaptive and personalized behavior. Unlike basic assistants, the system also incorporates an autonomous decision policy [6], which prevents excessive or irrelevant responses.

6.6 Limitations and Future Improvements

Despite promising results, several areas can be improved:

- Reducing response latency for real-time interaction
- Improving performance in noisy or overlapping speech conditions
- Enhancing domain-specific knowledge through better data integration
- Expanding email automation with advanced classification and response generation
- Strengthening long-term learning for improved personalization Future work will explore multimodal inputs (audio, video, and screen context) and more efficient model deployment techniques to improve real-time performance.

Overall, the results indicate that Auralis is a practical

and scalable system for intelligent assistance, capable of supporting meetings, managing communication, and organizing tasks while reducing user effort.

7. CONCLUSIONS

This paper presented Auralis, an AI-driven Digital Twin-based Virtual Personal Assistant designed to support modern professional workflows. The system addresses the growing challenge of managing multiple digital interactions, including meetings, emails, and task coordination, by reducing the need for continuous human involvement. Unlike traditional tools that focus only on recording or summarization, Auralis provides active assistance through real-time participation, contextual understanding, and intelligent task management. The proposed system integrates Transformer-based language models [1], retrieval-augmented generation mechanisms [2], Digital Twin principles [3], and real-time communication technologies [4] to enable context-aware reasoning and adaptive behavior. In addition to participating in meetings, Auralis extends its functionality to extracting actionable tasks, managing communication, and maintaining long-term contextual memory. This combination allows the system to function as a comprehensive productivity assistant rather than a single purpose tool.

Experimental evaluation demonstrates that Auralis can generate contextually relevant responses, maintain consistent persona alignment, and effectively extract tasks from both meetings and emails. Although the response latency varies depending on system load, the overall performance is suitable for semi-autonomous assistance. The integration of memory and decision-making mechanisms further improves interaction quality and reduces unnecessary system responses. Despite these promising results, certain limitations remain.

The system performance can be affected in scenarios involving overlapping speech or poor audio quality. Additionally, domain-specific queries may require more specialized knowledge sources to improve accuracy. Future work will focus on reducing response latency, improving robustness in complex meeting environments, enhancing email automation capabilities, and incorporating multimodal inputs such as video and screen context.

Overall, Auralis demonstrates the feasibility of combining Digital Twin modeling with modern AI techniques to create an intelligent assistant capable of supporting meetings, communication, and task management. The proposed approach highlights the potential of human–AI collaboration systems to improve productivity while reducing cognitive load in real-world environments.

TABLE I
PERFORMANCE EVALUATION OF AURALIS

Metric	Value
Average Latency	25–80 sec
Response Relevance	85–90%
Persona Consistency	High
Task Extraction Accuracy	Moderate to High

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Prof. Ashwini More for her valuable guidance, continuous support, and encouragement throughout the development of this project. Her insights and suggestions played a crucial role in shaping the design and implementation of Auralis.

REFERENCES

[1] Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), 2017.
[2] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge- Intensive NLP Tasks," in Advances in Neural Information Processing Systems (NeurIPS), 2020.
[3] M. Grieves, "Digital Twin: Manufacturing Excellence through Virtual Factory Replication," 2017.
[4] H. Alvestrand, "WebRTC: Real-Time Communication in Browsers," RFC 8825, 2021.
[5] S. Amershi et al., "Guidelines for Human-AI Interaction," in Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), 2019.
[6] M. Wooldridge, An Introduction to MultiAgent Systems, 2nd ed. Wiley, 2009.
[7] J. Weston et al., "Memory Networks," arXiv preprint arXiv:1410.3916, 2014.
[8] G. Shang et al., "Unsupervised Abstractive Meeting Summarization," in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2018.
[9] M. Hoy, "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants," Medical Reference Services Quarterly, vol. 37, no. 1, pp. 81–88, 2018.
[10] S. Kiritchenko et al., "Email Classification Using Natural LanguageProcessing," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018.