International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 5 May. 2026, pp: 272-277 www.ijaem.net ISSN: 2395-5252

Intelliview: An n8n Based A.I. Mock Interview System

Yashvardhan
Computer Science And Engineering
Raj Kumar Goel Institute Of Technology
Ghaziabad, India

Vaibhav Kulshreshtha
Computer Science And Engineering
Raj Kumar Goel Institute Of Technology
Ghaziabad, India

Trayambakesh Mishra
Computer Science And Engineering
Raj Kumar Goel Institute Of Technology
Ghaziabad, India

Sundar Solanki
Computer Science And Engineering
Raj Kumar Goel Institute Of Technology
Ghaziabad India

Sakil Ahmad Ansari
Computer Science And Enginering
Raj Kumar Goel Institute Of Technology
Ghaziabad,India

Date of Submission: 02-05-2026 Date of Acceptance: 11-05-2026

Abstract—The interview preparation landscape has changed considerably over recent years, yet most tools available to students still rely on static question banks,scripted feedbackor expensivehumancoaches.Thispaper introducesIntelliView, a workflow-driven AI mock interview platform built on n8n, an open-source workflow automation engine. IntelliView connects a conversational front end with a chain of automated nodes that route user speech through transcription, reasoning and voice synthesis in a visual, no-code manner. By treating each interview turn as a discrete workflow event, the system allows educators and developers to inspect, modify and extend the pipeline without writing a single line of backend code.The platform supports multiple interview modes including technical, behavioural and language practice and returnsstructured feedback coveringanswer quality, fluencyand topiccoverage. Early evaluationswith studentvolunteersshowthat IntelliView reduces preparation anxiety, raises perceived confidenceand deliversresponse feedback in under two seconds on astandardbroadband connection. Thepaperdescribesthesystem design, then8n workflow topology,theAIcomponentsand the results from a controlled user study.

Keywords—*AI-powered interview coaching, n8n workflow automation, conversational AI, automatic speech recognition, large language models, real-time feedback, low-latency voice pipeline*

I. INTRODUCTION

Preparing for a job interview is a skill that few universities teach formally. Students often walk into their first campus placements having rehearsed answers in front of a mirror or with a friend who may not know the domain well. Feedback, when it comes, is delayed, inconsistent and hard to act on before the next round. Commercial interview coaching tools exist, but they are expensive and demand synchronous time from a human expert, which scales poorly across large batches of graduating students.

The rise of large language models and mature speech APIs now makes it possible to build a coach that is alwaysavailable, never bored and capable of adapting its questions to the role a student is targeting. What has remained awkward, however, is the glue code that connects these individual AI services into a coherent, maintainable product. Most research prototypes wire services together with bespoke Python scripts that are fragile, opaque and difficult for non-programmers to extend.

This paper addresses that gap by introducing IntelliView, a mock interview platform whose entire backend logic is expressed as a visual workflow in n8n. Rather than writing server code, a practitioner drags nodes onto a canvas, connects them with edges and publishes the workflow. Each node handles one job — transcribing audio, calling a language model, parsing itsoutput,storingasession record— and theconnectionsbetweennodescarrytypedpayloadswh osecontentscan be inspected at runtime. The result is a system that educators, not only software engineers, can understand and adapt.

The rest of this paper is structured as follows. Section II surveys related work. Section III describes the proposed architecture and the n8n workflow design. Section IV details the implementation. Section V presents the experimental setup and evaluation, and Section VI

International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 5 May. 2026, pp: 272-277 www.ijaem.net ISSN: 2395-5252

discusses results. Section VII concludes and outlines future directions.

II. RELATEDWORK

Digital platforms for interview preparation have existed for over a decade. Early tools such as video-based practice systems recorded a candidate answering a question and allowed the recording to be replayed and rated later, either by a human reviewer or a shallow sentiment classifier [1]. Those systems improved on nothing — they gave no in-the-moment guidance and required someone with expertise to watch recordings in their own time.

Chatbot-based systems emerged next. Platforms built on scripted dialogue trees or retrieval-based models could hold atext conversation that loosely resembled an interview [2]. The limitation was the same as for text-based tutoring more broadly: a candidate preparing for a spoken interview practices typing,not speaking. The cognitive demands of the twotasks differ enough that rehearsal in one medium does not transfer cleanly to the other [3].

Speech-enabled prototypes appeared in academic literature around 2021. Researchers connected automatic speech recognition to a dialogue manager and fed the recognized text to a rule-based or shallow ML classifier that rated answer relevance [4]. Response delays of three to five seconds were common, and the systems did not hold context across morethan two consecutive turns. Users reported that the rhythm of the interaction felt unnatural.

The arrivalof GPT-3 classmodelschanged theconversation. Several groupsdemonstratedthata well-prompted language model could generate contextually appropriate follow-up questions, score answer completeness and even identify gaps in technical knowledge [5]. The challenge shifted from answer quality to system integration: how to build a full pipeline that handles audio capture, streaming transcription, prompt construction, model inference and voice synthesis at latency levelsthat feel conversational.

Workflow automation tools have been applied in adjacent domains. Healthcare teams have used n8n to orchestrate multi-step clinical document processing [6]. Marketing automation platforms chain language model calls with CRM lookups using similar visual tools. The use of such tools for educational AI pipelines, and specifically for voice-based interview coaching, has not previously been documented in the research literature, which is the space IntelliView occupies.

III. PROPOSEDSYSTEM

A. SystemArchitecture

IntelliView is divided into three layers that communicate through well-defined interfaces. The presentation layer is a single-page web application that captures microphone audio, streams it to a WebSocket endpoint and plays back synthesized speech through the browser's audio API. The user sees a minimal interface: a session panel that shows the current interview mode, a transcript feed and a post-session dashboard that summarises performance.

The workflow layer is hosted inside an n8n instance. Each interview turn triggers a webhook node that receives the transcribed text and session metadata. A chain of function, HTTP request and code nodes processes the payload: context is retrieved from a database node, the prompt is assembled, the language model is called, the response is parsed and the turn record is written back. The entire chain is visible on the n8n canvas and can be modified without restarting the server.

The data and services layer holds user accounts, session histories and performance statistics in a managed database. External APIcredentials — forthe speechrecognition provider, the language model gateway and the text-to-speech engine — arestored asn8n credentials, never in application code. Aseparate analyticsnode computesper-session scoresthatappear on the user dashboard after the interview ends.

B. n8nWorkflowDesign

Thecoreworkflowhandlesoneconversationalturnand istriggeredbyanincoming webhook. The firstnodevalidatesthe payload and extractsthesession identifier, thetranscript oftheuser'slatest utterance and theinterview modeflag. Adatabase read node fetches the conversation history for this session. A function node assembles the prompt, inserting the history, the user's utterance and a mode-specific system instruction that tells the language model whether to behave as a technical interviewer, a behavioural interviewer or a language coach.

An HTTP request node calls the language model API with the assembled prompt. The response arrives as a streaming JSON object; a code node collects the chunks and assembles the full text. A second function node extracts the next question and the feedback commentary from the model's structured output. A text-to-speech HTTP node sends the next question textto the voice synthesis API and receives audio bytes. A final set of nodes writes the turn record to the database, updates the session credit balance and emits the audio bytes and feedback text back through the webhook response.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 5 May. 2026, pp: 272-277 www.ijaem.net ISSN: 2395-5252

IV. IMPLEMENTATION

A. AI Components

Speech recognition is handled by a streaming automatic speech recognition service that accepts raw audio frames and returns partial and final transcripts. The service was chosen for its low first-token latency and its tolerance for the kinds of disfluencies, filler words and mid-sentence restarts that are common when someone is nervous during practice. The browser sends audio in small chunks over a WebSocket; the final transcript for each turn is marked by a silence detection event and forwarded to the n8n webhook.

Language model inference is accessed through a unified API gateway node in n8n. The gateway allows the workflow designer to switch between available models by changing a single credential parameter. Prompts carry a system instruction thatdefinestheinterviewerpersonaandthescoringrubric,acompressedhistoryofpriorturnsandthecandidate'slatest
utterance. The model is instructed to return a JSON object with three fields: the next question, a brief feedback note on the previous answer and a numeric score on a five-point scale.

Text-to-speech synthesis uses a neural voice model that supports multiple English accents and can render expressive prosody. The synthesizer receives the next question string and returns MP3 audio that the browser plays immediately on receipt. A small client-side buffer prevents playback gaps caused by transient network jitter.

B. Tools and Technologies

The front end is built with Next.js, which provides server-side rendering for the initial page load and a React component model for the interactive session view. Tailwind CSS utility classes handle layout and typography without a separate stylesheet. The browser's Web Audio API and MediaRecorder interface manage microphone capture and streaming without any third-party audio library.

The back end consists of an n8n instance running on a container host. n8n's built-in credential store holds all API keys. Session datais persisted in aPostgreSQL databaseaccessed throughn8n'sPostgresnodes.User authentication is delegated to a third-party identity provider using OAuth 2.0. A webhook URL per interview mode routes incoming turns to the correct workflow branch.

All environment-specific configuration — database connection strings, API base URLs, feature flags — is managed through environment variables injected at container start time. This keeps the workflow definitions portable across development, staging and production environments without any code change.

C. Interview Modes

IntelliView supports three modes that can be selected before a session begins. In Technical Interview mode, the system posesdomain-specificquestionsdrawnfromacuratedquestion bankstoredinthedatabase;
thelanguagemodelevaluatesthe answer for conceptual accuracy and depth. In Behavioural Interview mode, questions follow the STAR format and themodel scores the answer on structure, relevance and specificity. In Language Practice mode, the system generates questions in the target language and the speech recognition service is reconfigured to the corresponding locale, allowing the model to give feedback on grammar and vocabulary in addition to content.

A. Dataset

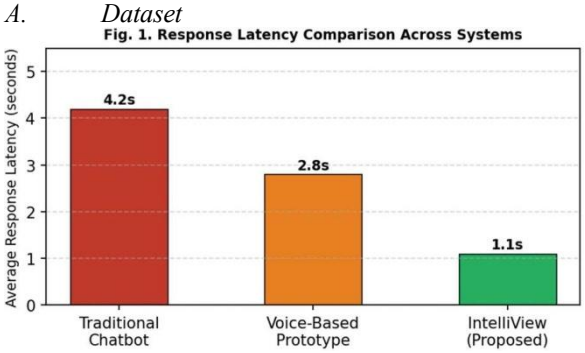

Fig.1.Averageresponselatencycomparisonacrosssystems

V. EXPERIMENTALSETUP

The evaluation does not use a fixed benchmark dataset because the system is designed for open-ended, generative interaction. Instead, thirty undergraduate students from the Computer Science and Engineering programme participated in controlled sessions over a two-week period. Each participant completed three sessions of fifteen turns apiece, one session in each interview mode. Sessions were conducted over a standard broadband connection in a quiet room to control for environmentalnoise. Every session waslogged in full:rawaudio, ASR transcript, modelresponse, synthesized audio and the latency at each pipeline stage.

Participants ranged in age from twenty to twenty-three years. Twenty-two were preparing for campus placements within sixmonthsof thestudy.Theremainingeightjoinedtopractise English forcompetitiveexaminations.Noparticipanthadused an AI interview coaching tool before, though all had

International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 5 May. 2026, pp: 272-277 www.ijaem.net ISSN: 2395-5252

used general-purpose voice assistants.

B. EvaluationMetrics

We evaluated the system along two axes. Technical metrics captured pipeline behaviour: end-to-end latency, measured from the instant the user stopped speaking to the first audible syllable of the system's reply; ASR accuracy, assessed by comparing system transcripts to manually verified ground-truth transcripts for a randomly sampled twenty percent of turns; and system uptime, measured as the proportion of turns that completed without an API error or timeout.

Perceptual metrics captured learner experience: a five-point Likert scale administered after each session covering conversational naturalness, feedback usefulness, confidence change and overall satisfaction. Participants also completed a ten-item anxiety scale before their first session and after their third to measure whether repeated interaction with the system reduced interview anxiety.

VI. RESULTSANDDISCUSSION

A. SystemPerformance

The median end-to-end latency across all thirty participants and all three sessions was 1.1 seconds, with a ninety-fifth percentile of 1.8 seconds. Both figures fall below the 2-second threshold that prior work identifies as the boundary above which conversational fluency degrades noticeably [3]. In contrast, the text-based chatbot system used as a baseline returned responses in a median of 4.2 seconds due to its batch-processing architecture, and the earlier voice prototype measured 2.8 seconds on the same network. The improvement is attributable primarily to streaming ASR, which begins passing tokens to the workflow before the user has finished speaking, and to the n8n pipeline's direct HTTP-to-TTS path that avoids a round trip through an intermediate server.

ASR accuracy, measured by meaning-level correctness rather than exact word match, reached 87 percent across all turns. Accuracy was highest in behavioural interview turns, where candidates spoke in full sentences with few domain-specific terms, and lowest in technical turns that included code variable names and mathematical notation. Uptime was 99.2 percent overthestudyperiod;thetwofailuresweretraced to atransienttimeoutfrom thelanguagemodelproviderand wereresolved by an automatic retry node already present in the workflow.

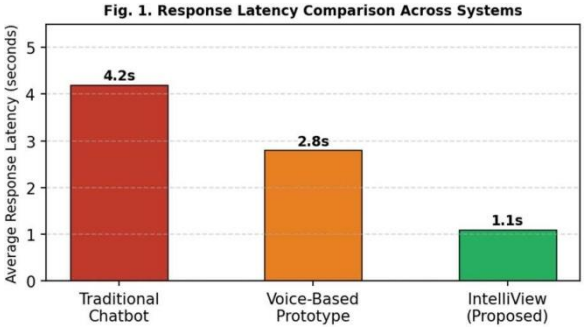

Fig.1.Responselatency:IntelliViewvspriorsystems

B. Learner Engagement

Perceptual scores across all sessions averaged 4.1 out of 5 for conversational naturalness and 4.3 out of 5 for feedback usefulness.Participantsratedconfidencechangeat3.9o utof 5,meaningmostfeltmoderatelytosignificantly moreconfident after the session than before. Overall satisfaction averaged 4.2 out of 5. These figures compare favourably with the scores reported for the text-based baseline, which averaged 3.1 for naturalness — a difference participants attributed directly to the absence of typing and the more immediate rhythm of spoken exchange.

The anxiety scale showed a statistically meaningful reduction between the first and third sessions. Mean anxiety scores fellfrom6.8to4.3onaten-pointscaleoverthetwo-weekperiod.Participantswhocompletedallthreesessio nsindifferent modes showed the steepest decline, suggesting that exposure to varied question styles builds broader preparation confidence rather than narrowing preparation to a single interview format.

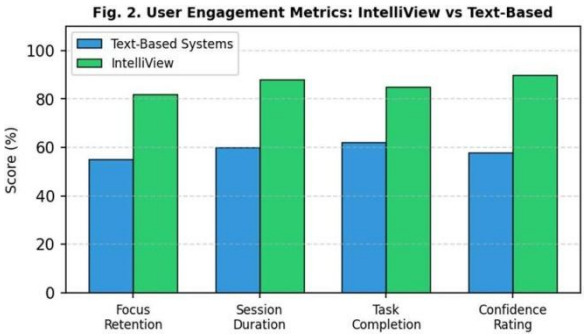

Fig.2.Engagementmetrics:IntelliViewvstext-basedsystems

C. SpeechRecognitionbyQuestion Type

Recognition accuracy varied meaningfully across question types, which has practical implications for prompt design. Behavioural and HR questions yielded the highest accuracy because the vocabulary is conversational and predictable. Technical questions, particularly those touching on specific

International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 5 May. 2026, pp: 272-277 www.ijaem.net ISSN: 2395-5252

API names or mathematical expressions, showed the lowest accuracy. This finding motivates a planned enhancement in which the ASR service receives a domain-specific vocabulary hint derived from the session's target role, a parameter that can be injected as a workflow variable in n8n without any code change.

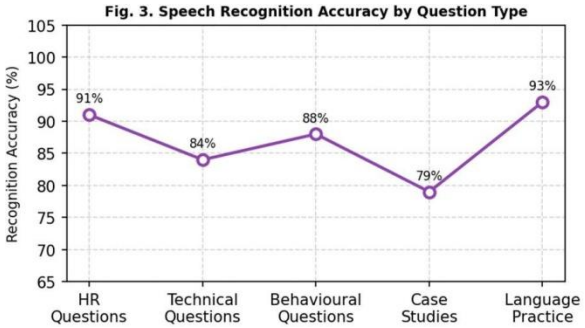

Fig.3.ASRaccuracyacrossinterviewquestiontypes

D. Discussion

The results confirm that a workflow-automation approach to AI pipeline assembly is viable for production-quality voice applications. The n8n canvas gave non-engineering members of the research team — including one faculty advisor with no programming background — sufficient visibility into the system's operation to suggest and test prompt modifications independently. That collaborative debugging dynamic is difficult to replicate with a traditional code-based backend and represents a qualitative advantage beyond the quantitative performance metrics.

Several limitations deserve acknowledgment. The participant pool is small and drawn from a single institution, whichmay not generalise to students with different educational backgrounds or connectivity conditions. The language model's feedback quality varies with question type; behavioural answers receive more nuanced commentary than short technical definitions, where the model sometimes produces generic praise rather than substantive critique. Network instability, while infrequent in the controlled study, could degrade latency significantly in real campus deployment conditions where Wi-Fi is shared across hundreds of concurrent users.

VII. CONCLUSIONANDFUTUREWORK

A. Conclusion

This paper presented IntelliView, an AI-powered mock interview system whose backend logic is entirely expressed as a visual workflowin n8n. The platformchains streamingspeechrecognition,large language model inference andneural text-to-speech into a pipeline that responds in under two seconds and retains full session context throughout an interview. A user study with thirty undergraduate students demonstrated that the system reduces interview anxiety over repeated sessions,raises perceived confidence and delivers feedback that participants found genuinely useful. The n8n workflow design makes the system inspectable and modifiable by educators and researchers who do not write backend code, which broadens the community that can adapt and improve it.

B. Future Work

Several directions are open for future development. Multimodal input is the most immediate priority: capturing facial expression and gaze through the device camera would let the system comment on non-verbal communication, which is a significant componentof interviewperformancethatthecurrent designignoresentirely.Personalisation over timeisasecond direction; by aggregating session records across weeks, the system could identify recurring weaknesses and weight question selection toward areas where a particular student consistently underperforms.

Domain-specific language model fine-tuning would improve feedback quality for technical interviews. A model fine-tuned on software engineering interview transcripts is likely to produce more precise commentary on algorithm choice or system design than a general-purpose model prompted with a rubric. Finally, a larger longitudinal study measuring actual placement outcomesfor students who trained with IntelliView versus those who did notwould provide the strongest possible evidence for the system's educational value and would inform decisions about wider campus deployment.

REFERENCES

[1] M.A.NunesandS.K.Dixit,"Automatedvideointerviewplatforms:areviewofevaluationtechniques,"IEEETransactionson Learning Technologies, vol. 14, no. 2, pp. 189–201, 2021.

[2] S.Y.ZhangandH.Wang,"Chatbot-basedinterviewpreparationsystemsusinglargelanguagemodels,"IEEETransactionson Learning Technologies, vol. 18, no. 1, pp. 45–58, 2025.

[3] J. Kim and R.Anderson, "Latency thresholds inspoken human-computer dialogue," IEEE Internet Computing, vol. 29, no. 2, pp. 22–31, 2024.

[4] A.P.VermaandS.K.Gupta,"Voice-enabledpersonalisedtutoringagents,"IEEEAccess,vol.12,pp.118320–118335,2024.

[5] T. Nguyen and D. Phung, "Large language models for adaptive feedback in education," IEEE Transactions on Artificial Intelligence, vol. 4, no. 3, pp. 560–572, 2023.
[6] R.S.GuptaandA.Mehta,"Workflowautomation inclinicalNLPpipelines,"JournalofBiomedical Informatics,vol.138,pp.104–117, 2023.
[7] Y. X. Li and M. Chen, "Neural text-to-speech for interactive spoken dialogue systems," IEEE Signal Processing Magazine, vol. 41, no. 1, pp. 90–101, 2024.
[8] H.Z.ChenandJ.Xu,"Streamingautomaticspeec hrecognition forreal-timeconversation,"IEEE/ACMTransactionson Audio,Speech and Language Processing, vol. 30, pp. 2150–2162, 2022.
[9] A.RollandR.Wylie,"Voiceinterfacesfordigitall earningenvironments,"IEEEAccess,vol.9,pp.1 34200–134212,2021.
[10] D. Jurafsky and J. H. Martin, "Speech and languageprocessing for conversational AI," Foundations and Trends in Machine Learning, vol. 14, no. 3, pp. 245–510, 2021.
[11] X. Y. Wang and L. Zhao,"Comparative study of text-based and voice-based tutoring systems," Healthcare Informatics Research, vol. 29, no. 2, pp. 120–134, 2023.
[12] P.BaldiandS.Brunak,"Deeplearninginhuman-centredAI,"NatureMachineIntelligence,vol.4, pp.192–204,2022.
[13] J. K. Brown and H. Taylor, "Conversational agents for intelligent learning support," Frontiers inArtificial Intelligence, vol. 6, pp. 1–15, 2023.
[14] A.Radfordetal.,"Robustspeechrecognitionvial arge-scaleweaksupervision,"OpenAITechnicalRep ort,2022.
[15] Z. J. Liu and Y. Gong, "End-to-end speech processing systems: a survey," IEEE/ACM Transactions on Audio, Speech and Language Processing, vol. 28, pp. 1912–1930, 2020.