International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
SJIF (2025): 8.036

# IntelliPrep AI: Conversational AI Voice Interview Preparation Platform

**Vanshika Gupta[1], Dr. Satish Gujar[2]**

Student of MCA (Computer Science), JSPM University, Wagholi, Pune, India
Email: *vanshika555gupta[at]gmail.com*

Professor, Head of Department of MCA (Computer Science), JSPM University, Wagholi, Pune, India

**Abstract:** ***Getting hired today requires much more than just knowing the right answers to questions. It takes being able to speak with confidence, to communicate under pressure, and to be able to deal with the uncertainty of conversing in a real-time setting, none of which can be learned from using a question bank that only has pre-answered questions in it. IntelliPrep AI is different. Rather than having candidates read example answers, it will place them into a mock interview that is conducted entirely by an AI agent developed from Google Gemini, with the candidate being interviewed using a telephone-based interface. After completing each mock interview, the candidate will receive a score and written feedback, across five different dimensions of performance. All of this can be done by using nothing but a browser-based interface, no software has to be installed, and there is no requirement for additional hardware. Additionally, all sessions are stored in a cloud-based database (Firebase) to allow users to track their progress over time. What sets IntelliPrep AI apart from earlier AI-assisted interviewing technologies, is the combination of real-time voice recognition, multiple axes of scoring, and the ability to track performance over time, all in one product.***

**Keywords:** AI Interview Preparation, Google Gemini, VAPI SDK, Voice-Based Learning, Adaptive Feedback, LLM, Firebase, NLP, Mock Interview, Career Development

## 1. Introduction

Historically, interviews have been stressful experiences for applicants. One reason is that many people view interview preparation as an exercise in memorizing the "right answer" (e.g., perfect score = get a job), when actually a successful interview requires just as much attention to how you present yourself as it does to what you say. In fact, research shows that job applicants who provide technically correct answers yet lack enthusiasm, speak in a monotone voice, or do not convey their thoughts logically are frequently passed over in favor of applicants that present their responses more naturally and casually, even if those applicants do not possess identical levels of accuracy in the content of their knowledge.

In the past, individuals who wanted to prepare for an interview through structured means (e.g., mock interviews with experienced individuals) were limited by the availability of these types of preparation sessions; accordingly, structured preparation has become associated with privilege. While practicing with peers is beneficial, consistency of quality level among participants is often not present. Most online preparation websites provide users merely with a text-based Q&A section (to enable users to remember answers through keyword matching); however, these are ineffective for assisting users with developing verbal fluency and maintaining composure, both skills commonly assessed by employers during interviews.

The availability of large language models paired with real-time speech interaction represents the sea change that allows new possibilities to be created in this area; together these technologies enable IntelliPrep AI to provide applicants with an innovative experience: conducting real-life, voice-activated interviews followed by a complete assessment of the interaction (using GM's language comprehension capabilities) and ultimately providing applicants with actual benefit (in the form of a breakdown of strengths and weaknesses) instead of a pass/fail score.
The system handles the complete workflow: users log in, specify their target role and experience level, talk through a full mock interview with an AI interviewer, then immediately receive feedback covering communication clarity, technical accuracy, problem-solving, cultural fit, and confidence. All of it runs in a standard browser. The session data persists in Firestore, so returning users can see whether they're actually improving.

The core contributions are: a functional voice-interview simulation backed by Gemini and VAPI; a five-axis scoring engine; a full-stack Next.js architecture that keeps everything serverless; and a demonstration that this kind of sophisticated tool can be built accessibly without requiring specialized hardware.

## 2. Related Work

Interest in AI-based interview preparation has grown greatly in the past 10 years, however, most of the initial prototypes developed in this area were evaluation-based only, as opposed to being simulation-based. For example, Suen, Hung, and Lin (2019) developed a system called AVPA, which used video recording technology to provide an evaluation of applicants' personality characteristics during live interactions.

NLP research conducted by Naim et al. supported what many professionals in that field already felt was true: interview success cannot be determined by verbal content alone. Prosodic features- pausing patterns, speech rate, pitch variation- carry significant signal. In the wake of the development of transformer-based language models, the breadth of feedback expanded beyond the limitations of keyword matching.

Recent studies have begun to develop systems using multiple modalities. (Patil, Hole, Patil, 2024) developed InterviewEase which included the recognition of gestures and emotions, and analysis of sections in resumes. The approach was certainly innovative, but the reliance on computer vision made the tool less accessible, as users with older or lower quality cameras/lights were basically excluded. Koshti et al (2025) attempted HR simulation, in addition to resume analysis and technical assessment, but again, no real time voice input was integrated in that system.

The most relevant works before now were likely NexInterview, a project of Sri Shakthi Institute of Engineering and Technology. It used very similar technologies, such as Next.js, Google Gemini, and a voice API (VAPI), as did IntelliPrep. NexInterview demonstrated that users are more engaged in practicing with a voice-activated interface rather than with a text-only interface. The area where NexInterview did not perform was in scoring; there was no multi-dimensional, structured evaluation. Furthermore, the type of questions asked would not necessarily adapt to the level of difficulty based on how many times they were answered.

The AI Voice Interviewing Agent developed by Yadav & al. (2025) uses prosodic score to help give the applicant confidence about how to respond to questions asked in the voice-only format by XGBoost. This system evaluated the responses of the applicants using a multitude of measuring tools, such as measuring a participant's pitch, detecting pauses and the level of fluency. Additionally, this scoring model is interpretable, making it valuable to users. Unfortunately, delivery of feedback and persistent infrastructure for providing feedback were not fully developed.

IntelliPrep takes this research as its basis for developing an adaptive learning system. The foundational research for this system is the development of adaptive learning based on the research of Chrysafiadi and Virvou (2013), generation of few-shot questions using LLMs (Brown et al., 2020), and applying the principles of Bloom (Anderson & Krathwohl, 2001) for immediate feedback, all of which indicate that an adaptive learning system, using voice updates that synthesize feedback in real time, is possible. While there is a conceptual divide in the uses of the SDK, the current difficulties associated with these differences are associated with their implementation.

## 3. System Design and Architecture

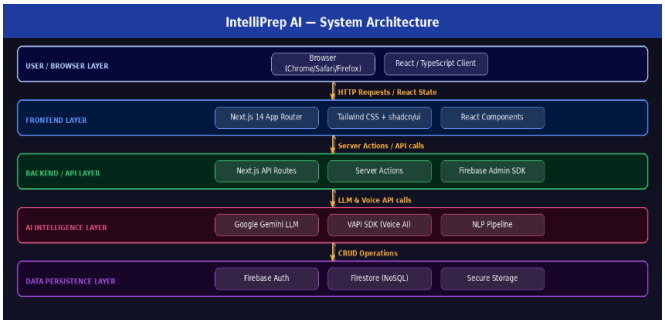


**Figure 1:** IntelliPrep AI- Five-Layer System Architecture

### a) Overview

IntelliPrep is built using Next.js 14 with an App Router architecture. Next.js allows the IntelliPrep application team to consolidate server-side rendering and client-side logic into a single codebase IntelliPrep has four main components for its functionality: the React-Typescript frontend is the portion of the application that the user sees; the Next.js API Routes and Server Actions are the components that provide the logic of the application; the Google Gemini and VAPI SDK components are the components that provide artificial intelligence for the components; and Firebase platform is used to authenticate and store user data. The Firebase platform is used to authenticate and store all user data and the application authenticates all user data through Firebase platform.

### b) Frontend

The IntelliPrep UI is built entirely in React using TypeScript with a Tailwind CSS styling layer, and shadcn/ui component styling. The interface is broken down into four separate points throughout the user experience: the authentication and profile setup screens, interview configuration screen, the live voice conference screen and finally the feedback and historical session data review screen. A significant design consideration in the development of IntelliPrep was to ensure all sections of the application from the user interface to the voice processing module were responsive and provided low-latency voice/audio processing regardless of user operating from a mobile device or desktop computer.

### c) Backend

To make everything work on the backend, we utilize Next.js. API Routes are used for the VAPI webhook and the Gemini API. Server Actions are used to read and write authenticated Firestore with type-safe validates. There is no need for a backend service which makes deployment a lot easier, and we can still keep everything on the server.

### d) AI Intelligence Layer

Two components are responsible for AI. Gemini, via the @ai-sdk/google adapter, and can also evaluate post-session transcripts. Gemini generates interview questions based on a user's prompt including the user's target role, experience level, and their preferred tech stack. The post-session transcript evaluation is handled by VAPI SDK. VAPI SDK is responsible for the live voice interaction. Two assistants are deployed, one that collects post setup preferences through natural conversation and one that conducts the interview.

### e) Data Layer

The Firebase platform provides email/password authentication in addition to Firestore data storage of user data. The data associated with a user is stored in a document structure made up of individual users, and a collection of each user's interview configurations, transcripts, Gemini generated feedback, and metadata about each user's interview sessions. The Firebase platform leverages security rules to restrict access to users granting authenticated users access to the collections of user data. Further, communication between the client application (i.e. the IntelliPrep user interface) and the Firebase platform is encrypted.

## 4. Methodology

### a) Authentication and Profile Setup

Users access the system via Firebase Authentication. Upon first login, Firestore generates a profile document including their name, email, and interview history. The authentication status is consistent through the use of Next.js context providers & session state. It travels through each page. The system presents roles and configurations that users have established in the past on the interview setup screen which prevents redundant setup for users in the past.

### b) Interview Configuration and Question Generation

Once the interview begins, the user converses with the VAPI preference-collection assistant. During the dialogue, the assistant rapidly captures the target job role, years of experience, preferred tech stack, and experience domain. The assistant also captures the interview type preference. The parameters are fed into a Next.js Server Action that builds a structured Gemini prompt. The question set is stored in Firestore in the session document. Prompt engineering ensures that the questions set the bar to be as high as the expectations of the industry for that position.

### c) Voice Interview Execution

The interviewer assistant retrieves and presents the pre-generated questions from Firestore. During the session, the assistant adds natural speech pacing, pauses, and contextual follow-up prompts. The user responds to each prompt through the browser mic. The response is then processed and converted to text by VAPI's proprietary speech-to-text solution. All user speech is collected by VAPI as a structured conversation. The system is able to handle accents disparities. This is particularly important to optimize the solution's accessibility. When the session completes, the text response is sent to the backend where it is used by Gemini to conduct an assessment.

### d) Feedback Generation Across Five Dimensions

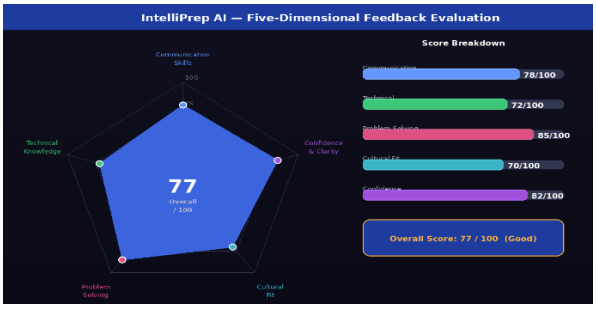


**Figure 2:** Multi-Dimensional Feedback Scoring Framework

Gemini receives the full transcript along with a structured evaluation prompt and scores the conversation across five axes: Communication Skills (vocabulary, grammar, clarity), Technical Knowledge (accuracy and depth of domain-specific responses), Problem Solving (reasoning approach to open-ended questions), Cultural Fit (enthusiasm and professional alignment), and Confidence and Clarity (structure, assertiveness, conciseness). Each of the dimensions is given a score out of 100, along with a qualitative assessment and specific suggestions on how to improve. An overall score is also given. The full assessment is saved to Firestore and then displayed in the application's review interface.

### e) Session History and Longitudinal Tracking

Every completed session — configuration metadata, transcript, and feedback report — is stored under the user's Firestore profile. The dashboard view organizes sessions by creation date and role, displaying summary cards with the core metrics at a glance. Users can revisit any past session's full feedback, making it practical to track whether scores on specific dimensions are actually moving in the right direction.

## 5. System Workflow

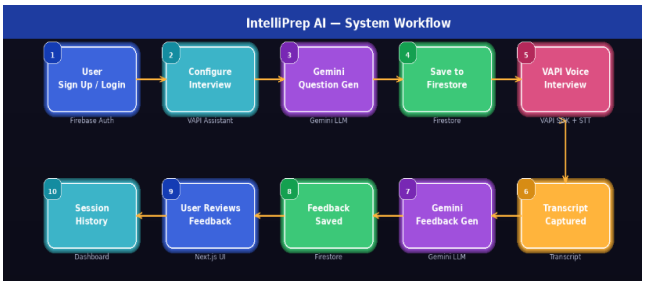


**Figure 3:** IntelliPrep AI- End-to-End Session Workflow

IntelliPrep consists of six stages that build on one another, without exception:

| Phase | What Happens |
|---|---|
| 1. Auth | User registers or logs in via Firebase; profile created in Firestore. |
| 2. Configure | VAPI preference assistant collects job role, experience level, tech stack, and interview type through voice conversation. |
| 3. Generate | Gemini API constructs a role-specific question set; stored in the Firestore session document. |
| 4. Interview | The VAPI conducted the interview and the session was voice-based. The conversation was structured and the transcript was captured in real time. |
| 5. Analyse | Transcript sent to Gemini; multi-dimensional feedback report generated and stored. |
| 6. Review | Feedback displayed in a structured report; session saved to user history for longitudinal tracking. |

## 6. Results and Discussion

The IntelliPrep application has been built and tested through validation as a fully functional web application. There is no requirement for special hardware; the only requirements are a web browser and a microphone. Many of the system features that previous systems were either unable to provide or only delivered partially were validated with respect to their performance.

Simulated interviews powered by VAPI lead to noticeable improvements in interaction quality as compared to interviews which utilize text as a medium. Users have reported a lower anxiety level during this form of a session compared to written Q&A formats. This aligns with Kumar et al. (2023) who found that environments which leverage voice-based learning have the ability to lessen the impact of performance anxiety.

Feedback from the engine presents a report that is easily

readable and structured. It arrives only seconds after the session is over. The five-axis breakdown provides feedback which is constructive so that the candidate is not left with a deep feeling of being lost. Thompson et al. (2024) determined that feedback which is constructive and specific helps to foster a level of improvement which is not likely going to be established by feedback which is general or by a lack of feedback. This is deeply reflected by the design of IntelliPrep. First, in terms of feedback, the activity-based feedback that would otherwise have disappeared after an existing session or industry-related user, can now be obtained after an IntelliPrep session to build a user's detailed and permanent record of previous performance data over multiple sessions. Next, in the validation of the functionality of the IntelliPrep web application on a browser-based system, the employment of the IntelliPrep web application on a browser based application had unexpected limitations on user experience and performance due to the fact that existing platforms may have used VR headsets and or would require users to have excellent or high-end systems in order to use those same platforms in the past.

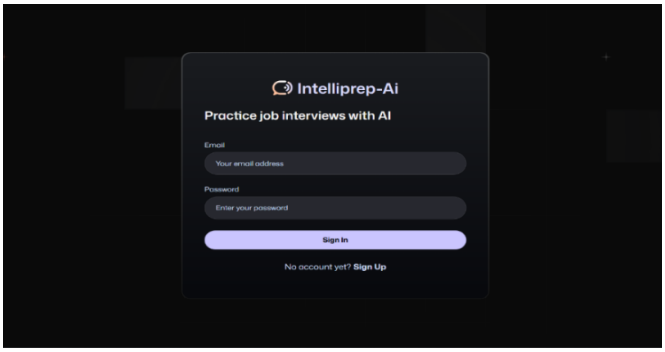


**Figure 4:** IntelliPrep AI- Sign-In Page

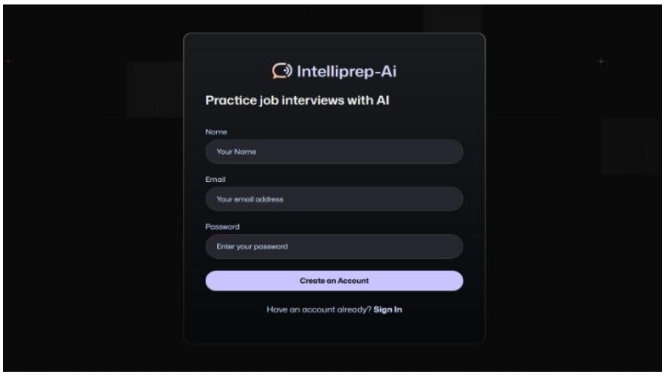


**Figure 5:** IntelliPrep AI- Sign-Up Page

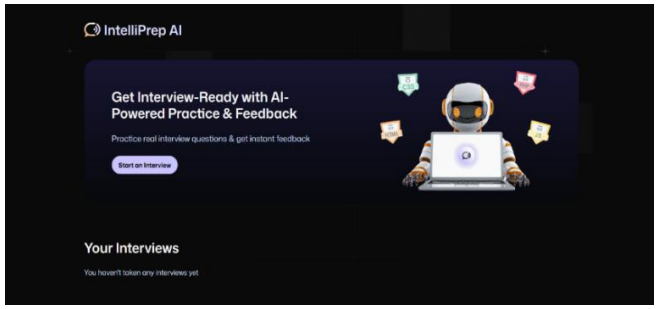


**Figure 6:** IntelliPrep AI- Front Page

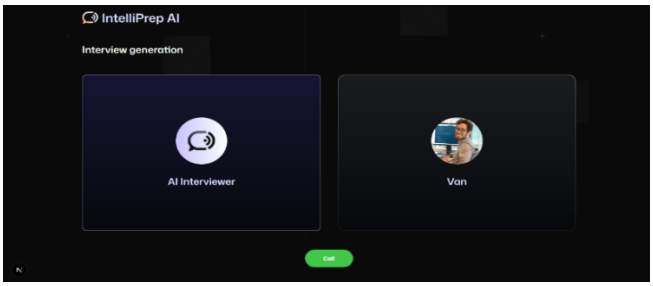


**Figure 7:** IntelliPrep AI- Interview Pannel

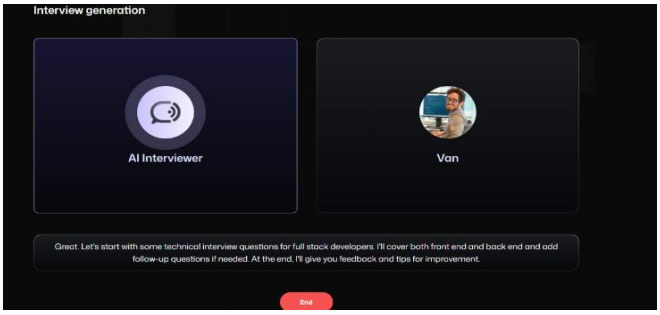


**Figure 8:** IntelliPrep AI- Interview Happening

## 7. Comparative Analysis

However, IntelliPrep will operate on a mid-range laptop according to the features of the IntelliPrep AI as shown in Table 2 below across the features of the IntelliPrep AI as compared to the six other existing industry related systems that have failed to provide the features necessary to enable all users to perform effectively in order to be properly prepared for in person interviews:

| Feature | IntelliPrep AI | NexInterview | AI Voice Agent | InterviewEase | AIDA |
|---|---|---|---|---|---|
| Voice Interaction | Yes (VAPI) | Yes (VAPI) | Yes (VAPI) | No | No |
| LLM Engine | Gemini | Gemini | Open LLM | Custom | GPT/Gemini |
| Multi-dim. Scoring | 5-axis | No | XGBoost | Partial | Partial |
| Firebase Backend | Yes | Yes | Yes | No | No |
| Session History | Yes | Yes | No | No | No |
| Adaptive Questions | Yes | Partial | Yes | No | Yes |
| Browser Accessible | Yes | Yes | Yes | Yes | Yes |
| Resume-Based QGen | Partial | No | Yes | Yes | No |

## 8. Conclusion and Future Directions

### 8.1 Conclusion

IntelliPrep AI is working to bridge the gap between being prepared for an interview and the actual interview process with the technology we have available. Voice simulation, multi-dimensional feedback from artificial intelligence, and tracking your progress over time are all individually successful technologies, and now we're creating a single product that utilizes them all together on a browser platform. While each of these capabilities can help you prepare for an interview, the reason for using the IntelliPrep AI platform is due to the workflow. The candidate doesn't simply practice but they will receive feedback from their practice sessions in the form of scores and the ability to determine what area(s) they underperformed, and then return the next day to see if any progress has been made. Existing tools do not provide this feedback loop and therefore do not offer the opportunity for real preparation.

### 8.2 Future Directions

There are many opportunities for more enhancements and features in the near future. The ability to parse a candidate's resume will provide the ability to ask interview questions that will be tailored to their actual work experience rather than just what they list on their resume; this will take personalization to a completely different level. The addition of machine learning models like XGBoost for prosody analysis (pitch, pauses, speech rate) will also expand the evaluation of confidence to beyond just the analysis of the candidate's spoken answers.

Incorporating multilingual support into this platform will significantly improve its ability to serve many different non-English speaking job markets; as currently, there are very few tools to assist with access to these communities. Gemini will also have the ability to use recombinant question difficulty whereby the agent can alter the level of difficulty for the upcoming questions based on how accurately the applicant answered previous questions. This will mimic the structure of how many "real life" interviews work; when applicants are correct in their answers they tend to be asked more difficult questions.

At present, there is no proper explanation regarding the feedback loop from the evaluation process. For example, informing a candidate that they ranked 68/100 on technical knowledge is less beneficial than providing them with an explanation of how their answers impacted their final score. Hence, incorporating this reasoning into the Gemini evaluation prompt, and clearly indicating that explanation in the UI, is likely the most valuable enhancement available for the system as a whole.

## References

[1] S. Patil, V. Hole, and S. Patil, "InterviewEase: AI-powered interview assistance," ResearchGate, 2024.
[2] H. Koshti, P. Gosavi, R. Pagar, P. Khairnar, and S. Talekar, "AI-Powered Interview Preparation System: Integrating Resume Analysis, HR Simulation, and Technical Skill Assessment," Journal of Engineering Research and Reports, vol. 27, no. 5, pp. 21 – 33, 2025.
[3] Dr. Prakash S, Siranjeevi K, Vinoth Kumar N, Pathmesh G, and Ponnarasu A, "NexInterview - AI-Driven Mock Interview Preparation Platform," IJARSCT, vol. 5, no. 7, pp. 279 – 288, May 2025. DOI: 10.48175/IJARSCT-26835.
[4] S. Yadav, C. Dalvi, S. Taksal, D. Bhaganagare, and Y. Patil, "AI Voice Interview Agent for Real-Time Personalized Mock Interviews," IJSAT, vol. 16, no. 4, pp. 1–8, 2025.
[5] A. Kr. Singh and A. Kumar, "Interview Assistant: AI Powered Interview Preparation and Assessment Platform," IJSREM, vol. 09, no. 10, Oct. 2025. DOI: 10.55041/IJSREM53352.
[6] T. Brown et al., "Language Models are Few-Shot Learners," Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 1877–1901, 2020.
[7] H.-Y. Naim, M. E. Tanveer, D. Gildea, and M. E. Hoque, "Automated Analysis and Prediction of Job Interview Performance," IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 191–204, 2018.
[8] L. Hemamou et al., "HireNet: A Hierarchical Attention Model for the Automatic Analysis of Asynchronous Video Job Interviews," AAAI, vol. 33, pp. 653–661, 2019.
[9] K. Chrysafiadi and M. Virvou, "Student Modeling Approaches: A Literature Review for the Last Decade," Expert Systems with Applications, vol. 40, no. 11, pp. 4715–4729, 2013.
[10] L. W. Anderson and D. R. Krathwohl (Eds.), A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Addison Wesley Longman, 2001.
[11] H.-Y. Suen, K.-E. Hung, and C.-L. Lin, "AVPA: An automated video-based personality assessment and interview training system with feedbacks," International Journal of Human–Computer Studies, vol. 129, pp. 18–29, 2019.
[12] J. Morris and Y. Zhang, "Context-Aware Question Generation Using Large Language Models for Interview Preparation," IEEE Transactions on Learning Technologies, vol. 16, no. 4, pp. 573–588, 2023.
[13] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of the 22nd ACM SIGKDD, pp. 785–794, 2016.
[14] H. Li, J. Chen, and Y. Wang, "Prosodic Features for Speech-Based Confidence Estimation," IEEE Transactions on Audio, Speech and Language Processing, vol. 27, no. 6, pp. 1054–1065, 2019.