# AI-Powered Mock Interview Platform

Dr. M. A. Pradhan[1], Swaraj Sawant[2], Soham Manjare[3], Shivam Wayal[4], Uday Kunjir[5]
[1]*Associate Professor, Computer Department AISSMS College of Engineering, Pune. Savitribai Phule Pune University Pune, India*
[2,3,4,5]*Student, Computer Engineering AISSMS College of Engineering, Pune. Savitribai Phule Pune University Pune, India*

***Abstract*—Securing a technical role in the modern software industry requires navigating a gauntlet of high-stakes evaluations. While algorithmic problem-solving remains central, the ability to verbally articulate complex system logic under pressure is equally critical. Unfortunately, quality preparation ecosystems are largely gated behind expensive paywalls involving human mock-interviewers, while free alternatives fail to simulate the psychological friction of a genuine interview room. We introduce InterviewAI, a highly accessible, automated SaaS platform engi-neered to democratize elite interview preparation. By coupling the semantic reasoning capabilities of Google Gemini with native browser hardware APIs (WebRTC and Web Speech), the system facilitates a rigorous, time-bound verbal assessment. Candidates must navigate a heavily curated domain-specific question bank while subjected to strict tab-switch proctoring. The underlying cognitive layer bypasses legacy keyword-matching algorithms, opting instead to evaluate the transcribed responses dynamically across technical accuracy and communication clarity. Empirical trials demonstrate that InterviewAI achieves a tight grading correlation with senior human engineers while delivering actionable, multi-rubric feedback in under four seconds.**

***Index Terms*—Mock Interview, Large Language Models, Generative AI, WebRTC, Continuous Proctoring, Next.js, Algorithmic Assessment, Natural Language Processing.**

## I. INTRODUCTION

### *A. The Reality of the Modern Hiring Gauntlet*

The mechanics of technical recruitment have undergone a massive recalibration. Driven by remote-first work cultures and an unprecedented influx of computer science graduates, hiring managers no longer rely on localized, face-to-face whiteboard sessions. Instead, candidates face distributed, multi-tier hurdles testing everything from granular database normalization to broad behavioral conflict resolution.

This environment breeds intense ”interview fatigue.” A candidate might possess the latent coding ability to construct a scalable React application, but fail an interview because they cannot verbally articulate the component lifecycle while a stranger watches them over a webcam. The ”think-out-loud” protocol is a distinct skill, and it requires targeted, pressurized practice.

### *B. The Socioeconomic Gap in Preparation*

The current market for interview preparation is sharply bifurcated. On one end of the spectrum are premium platforms that pair candidates with established engineers from major tech firms. While highly effective, these sessions are financially inaccessible to the average undergraduate, often costing hun- dreds of dollars per hour. On the opposite end are free, static repositories like HackerRank or LeetCode. These platforms excel at validating compiled code but offer zero simulation of the human element. They do not watch the user, they do not enforce a strict time limit on verbal explanations, and they cannot critique a candidate’s spoken communication style.

Attempts to bridge this gap with automation have historically disappointed. Early Applicant Tracking Systems (ATS) and rudimentary chatbots relied on brittle boolean logic. If a platform asked how to instantiate a class, and the candidate nervously explained the concept using the word ”blueprint” instead of ”constructor,” the system would unfairly penalize them.

### *C. Bridging the Gap*

InterviewAI was conceptualized to dismantle this specific barrier. We hypothesized that by layering a modern Generative Pre-trained Transformer (GPT) beneath a strict, proctored front-end interface, we could synthetically replicate both the intelligence and the intimidation factor of a human interviewer. The platform forces candidates to maintain eye

contact with their camera, locks their browser focus to prevent cheating, and listens to their verbal explanations in real-time. This paper details the architectural constraints, the mathematical evaluation models, and the performance benchmarks of the resulting system.

## II. CORE SYSTEM FOUNDATIONS

### *A. Deconstructing the Interview Workflow*

To accurately synthesize an interview, the platform must mimic the three distinct phases of a live session:

1) Environmental Setup: The initial moments where cam- era feeds are established, creating immediate psychological accountability.
2) The Inquiry and Articulation: The interviewer presents a randomized domain question. The candidate must process the prompt and formulate a coherent vocal response without relying on search engines.
3) The Critique: Immediate, objective feedback on what was correct, what was omitted, and how the communication could be streamlined.

### *B. The Shift to Semantic Evaluation*

Older architectures failed because they read text with- out understanding meaning. Modern Large Language Mod- els bypass this by utilizing transformer architectures capable of self-attention. When InterviewAI captures a candidate's fragmented, nervously spoken answer, the Gemini inference engine maps those tokens into a high-dimensional semantic space. It understands context. It recognizes that a candidate struggling to define a "Promise" in JavaScript might still be describing asynchronous behavior accurately, allowing for a nuanced, holistic grade rather than a binary pass/fail based on keywords.

## III. REVIEW OF RELEVANT LITERATURE

The intersection of artificial intelligence and educational assessment has seen rapid iteration. By examining recent literature, we can pinpoint exactly where legacy systems faltered and how the mechanisms within InterviewAI were engineered to solve those specific shortcomings.

### *A. AIced-Prep - AI Based Mock Interview Evaluator*

In their 2025 publication, Rajbhar et al. [4] introduced AIced-Prep, a multimodal mock evaluator heavily reliant on Automatic Speech Recognition (ASR). Their architecture suc- cessfully demonstrated that candidates interacting with voice-first AI agents exhibited measurable reductions in interview anxiety over time. However, the system architecture harbored a critical flaw: severe server-side latency. By offloading raw audio streams to cloud-based processors before generating a text transcript, the platform suffered conversational delays that shattered the illusion of a live interview. Furthermore, the system struggled to maintain a coherent context window when evaluating lengthy, rambling technical explanations. In- terviewAI resolves this bottleneck by delegating the entire transcription workload to the client's local hardware via the native browser Web Speech API. This architectural pivot achieves near-instantaneous text generation. Consequently, the backend LLM only processes lightweight JSON text payloads, preserving the rapid conversational rhythm required to synthe- size genuine interview pressure while expanding the contextual understanding of fragmented speech.

### *B. An Intelligent System for Automated Interview Question Generation: A Comparative Analysis of LLM Adaptation Techniques*

Recent explorations into generative assessments, such as the comparative analysis of LLM adaptation techniques published in 2026 [5], focused heavily on dynamic inquiry. Their proposed system parsed candidate resumes to autonomously generate highly personalized interview questions on the fly. While technologically impressive, this methodology introduces a fatal flaw for standardized preparation: uncontrollable base- line variance. If every mock interview session spawns a completely unique set of questions, statistical tracking of candidate improvement becomes impossible. Furthermore, dynamic generation introduces a high risk of algorithmic hallucination, where the AI might ask nonsensical or technically flawed questions. InterviewAI circumvents this systemic instability by strictly decoupling generation from evaluation. Our platform enforces a heavily curated, static database of over 300 peer- reviewed technical questions. By applying the Generative AI exclusively to the evaluation and feedback phase rather than question creation, InterviewAI guarantees that every candidate is measured against an immutable, fair, and standardized technical baseline.

### C. Q&AI: An AI powered Mock Interview Bot for Enhancing the Performance of Aspiring Professionals

The 2024 release of the Q&AI platform [6] attempted to push the boundaries of automated assessment by integrating comprehensive emotion and sentiment analysis. Their archi- tecture continuously monitored candidate webcam feeds to evaluate facial micro-expressions, posture, and eye-contact ratios to quantify ”cultural fit” and composure under stress. The fundamental flaw in this approach is the reliance on pseu- doscientific visual metrics that inherently introduce profound algorithmic bias. Such systems routinely penalize neurodiver- gent individuals, visually impaired candidates, or users from diverse cultural backgrounds where eye-contact norms differ drastically. InterviewAI explicitly rejects visual behavioral grading on ethical grounds. Instead, our platform utilizes the WebRTC camera feed strictly as a psychological deterrent and environmental proctoring tool—verifying identity and locking browser focus to prevent cheating. The actual algorithmic grading engine remains completely blind to the user’s physical appearance or nervous tics, confining its rigorous evaluation entirely to the semantic substance and technical accuracy of the transcribed spoken response.

### D. VoxMentor: A Voice-Driven AI Platform for Interview Sim- ulation, Company-Specific Preparation, and Student Learning

The VoxMentor platform, detailed in a 2025 IEEE confer- ence proceeding [7], successfully demonstrated the viability of voice-driven AI for simulating group discussions and technical coding rounds. Utilizing a combination of Whisper and Google Gemini, the system achieved a high degree of conversational fluidity. However, the pedagogical flaw within VoxMentor was its ”black box” grading mechanism. The system outputted a raw numerical score at the end of the session but failed to provide granular, explainable justifications for point de- ductions. A candidate receiving a low score had no action- able pathway to improve. InterviewAI tackles this deficiency through aggressive, deterministic prompt engineering designed specifically for Explainable AI (XAI). Rather than allowing the LLM to output unstructured text, our Node.js gateway wraps the candidate’s transcript in a strict JSON-schema envelope.

### E. Automated Interview through Online Video Interface

Earlier attempts at web-based interview automation, such as the architecture proposed by standard asynchronous video in- terfaces in 2023 [8], focused merely on recording answers for later human review. The primary structural flaw of these asyn-chronous platforms is the complete absence of environmental friction. Without live supervision, these systems essentially function as open-book exams. Candidates can easily navigate to external browser tabs, search for the optimal algorithmic solution, and read it directly off the screen, rendering the assessment data entirely useless for measuring true cognitive recall. InterviewAI introduces synthetic friction to solve this integrity crisis. By binding native JavaScript event listeners to the Document Object Model (DOM), our platform actively monitors the browser’s visibility state. The moment a candidate attempts to switch tabs or open a new window, the system registers a violation. This continuous proctoring engine utilizes an exponential decay penalty function, actively deducting points for prolonged cheating attempts while maintaining a realistic, high-pressure interview environment.

## IV. MATHEMATICAL EVALUATION MODELING

Understanding the leap from legacy systems to InterviewAI requires defining the underlying mathematical relationships that govern text evaluation.

Understanding the leap from legacy systems to InterviewAI requires defining the underlying mathematical relationships that govern text evaluation.

### A. The Failure of Traditional Matching

Assume a candidate provides an answer $A$ compared against an ideal rubric $I$. Older systems calculated a score $S_{legacy}$ based on exact lexical intersection:

$$S_{legacy}(A, I) = \sum_{t \in A \cap I} \frac{f_{t,A}}{|A|} \times \log \frac{N}{df_t} \quad (1)$$

Here, if the ideal term $t$ does not appear exactly in the candidate’s speech, the score drops to zero, entirely ignoring synonyms or conceptual explanations.

### B. Semantic Mapping via Transformers

Google’s Gemini operates on a scaled dot-product attention mechanism [9]. This allows the engine to evaluate the weight of every spoken word against the broader context of the sentence:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad (2)$$

This mathematical architecture translates the spoken answer into a dense vector embedding. The system then calculates the conceptual accuracy by measuring the cosine similarity between the candidate's response vector and the ideal technical vector:

$$Sim(A, I) = \cos(\vartheta) = \frac{\mathbf{v}_A \cdot \mathbf{v}_I}{\|\mathbf{v}_A\|\|\mathbf{v}_I\|} \quad (3)$$

*C. The Proctored Scoring Function*

The final score presented to the user is not a raw semantic value. It is a weighted composite, $S_{final}$, adjusted dynamically by their behavior during the session:

$$S_{final} = \alpha S_{tech} + \beta S_{comm} + \gamma S_{clarity} - P_{proctor} \quad (4)$$

The weights ($\alpha, \beta, \gamma$) shift depending on the interview domain. A coding logic question demands a high $\alpha$ value, while a behavioral HR question shifts weight toward $\beta$.

The proctoring penalty, $P_{proctor}$, utilizes an exponential decay function based on how long the user lost window focus ($t_{loss}$). Brief, accidental misclicks are forgiven, while prolonged cheating attempts stack aggressive point deductions:

$$P_{proctor} = \lambda \sum_{i=1}^{n} 1 - e^{-k \cdot t_{loss,i}} \quad (5)$$

## V. SYSTEM ARCHITECTURE

We constructed InterviewAI using a highly decoupled micro-architecture. By separating the client-side audio capture from the heavy inference logic, we maintain high availability and prevent browser thread blocking.

*A. The Presentation Layer (React & Vite.js)*

The frontend is the sole point of user contact, engineered with React 18 and Vite.js . We aggressively utilized Tailwind CSS to build a "Glassmorphism" interface. This design choice is deliberate; a clean, premium UI significantly reduces the cognitive load on an already anxious candidate. It handles all client-side state management, ensuring the live transcription stream does not cause excessive DOM repaints.

*B. The API Gateway (Node.js & Express.js)*

Sitting between the user and the AI is a robust Express.js server. Exposing direct LLM API keys to a client browser is a severe security risk. Therefore, the Node layer acts as a secure reverse-proxy. It intercepts the user's transcript, authenticates their JSON Web Token (JWT), and dynamically wraps their answer into a heavily engineered prompt payload before pinging Google's servers.

*C. Data Persistence (MongoDB)*

A highly available MongoDB Atlas cluster anchors the system. It handles user demographic profiles, tracks histori- cal metric arrays for Recharts dashboard visualizations, and houses the static 300+ question bank, segmented by difficulty and domain tags (e.g., Data Science, Frontend, DevOps).

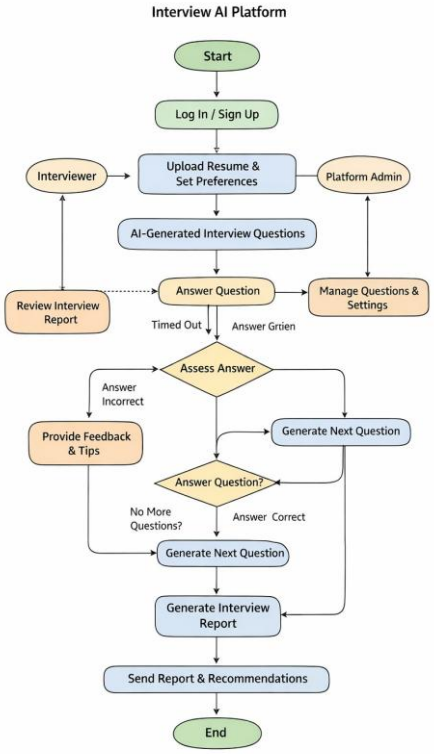


Fig. 1. A comprehensive Activity Diagram mapping the parallel execution threads of the candidate's UI experience, the background proctoring engine, and the asynchronous API evaluation routing.

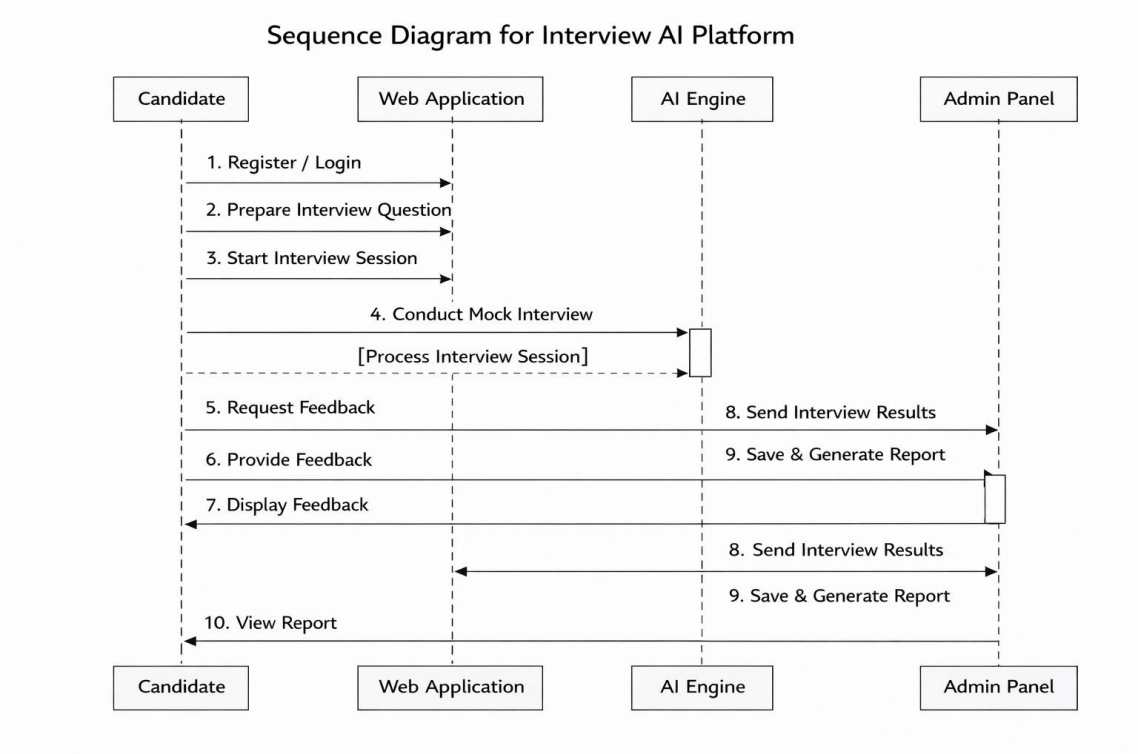


Fig. 2. The UML Sequence Diagram detailing the strict handshake protocols and data payloads passed between the Client Browser, the Node Gateway, and the Gemini Cognitive Engine.

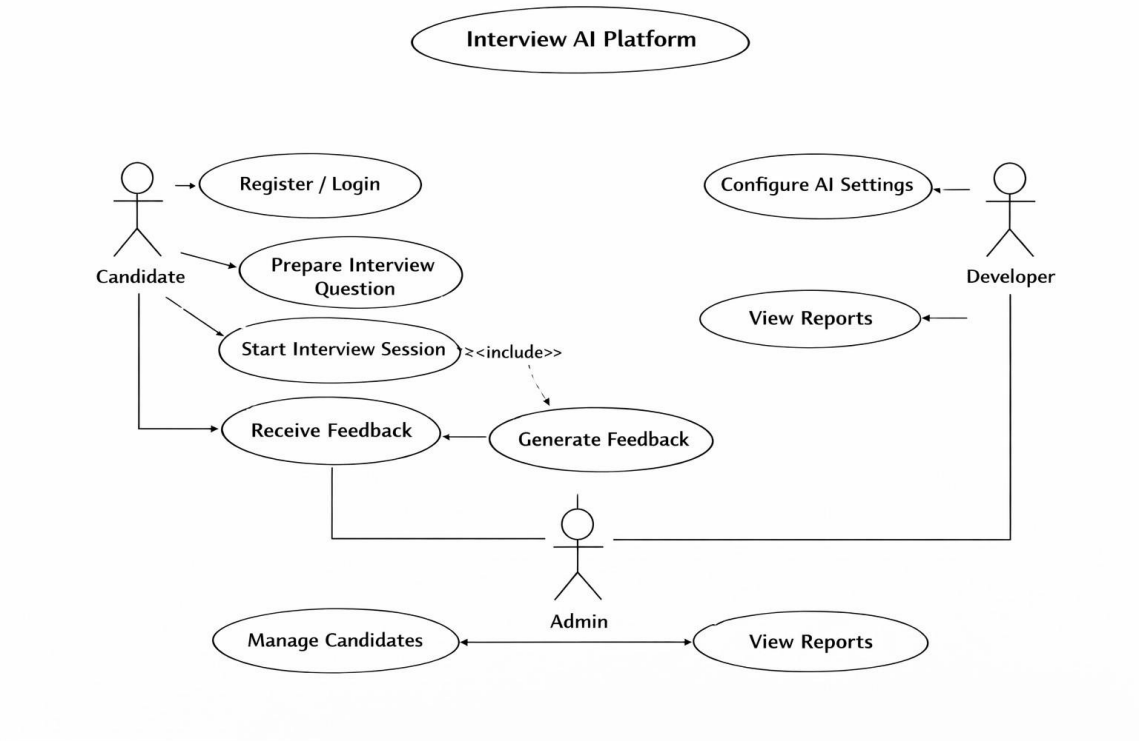


Fig. 3. System Use Case Diagram illustrating the isolation between the candidate's actions, the automated system triggers, and the external API dependencies.

## VI. CORE ALGORITHMIC ENGINES

### *A. Strict Environmental Control*

The proctoring engine operates entirely on the client side to avoid network latency. The moment an interview initiates, the browser binds native event listeners to the window object.

**Algorithm 1** Continuous Browser Integrity Monitoring

```
Initialize User Media Stream(Audio/Video)
if Stream access denied then
    Halt process. Render hardware warning.
else
    Attach visibilitychange and blur listeners.
    Set violationTracker ← 0
end if
while Session Timer > 0 do
    if document.hidden or Window Focus Lost then
        Increment violationTracker
        Dispatch WebSocket alert to Node server.
        Flash UI Warning: "Unauthorised navigation detected."
    end if
end while
if violationTracker exceeds global threshold then
    Invalidate session results.
end if
```

### B. Prompt Construction and XAI

Explainable AI (XAI) is the cornerstone of this platform. A numerical grade is useless without context. The Node server injects the candidate's transcript into a strict system envelope designed to force the Gemini model into a deterministic output format.

The invisible payload sent to the LLM reads: *"Assume the persona of a senior FAANG engineering manager. The candidate was asked: [Injected Question]. They responded verbally with: [Injected Transcript]. Evaluate the technical merit and clarity. Return a strict JSON response containing numerical scores and a concise, three-sentence justification highlighting specific missing technical concepts. Avoid hallucination."*

## VII. PERFORMANCE BENCHMARKING AND RESULTS

To validate the architecture, we subjected the platform to rigorous stress testing and grading accuracy audits.

### A. AI versus Human Baseline

We compiled a dataset of 50 recorded responses spanning various computer science domains. These transcripts were independently graded by three senior software engineers. We then fed the exact same transcripts into the Interview AI engine. The resulting delta was remarkably narrow, proving the LLM is capable of matching expert human intuition. Crucially, the system delivered this highly accurate critique in under four seconds, completely eliminating the asynchronous waiting period inherent to human-led platforms.

TABLE I: CORRELATIVE GRADING ANALYSIS (HUMAN EVALUATORS VS. SYSTEM)

| Metric Assessed | Human Mean | System Mean | Delta (Δ) |
|---|---|---|---|
| Algorithmic Accuracy | 7.24 / 10 | 7.41 / 10 | +0.17 |
| Verbal Clarity | 6.80 / 10 | 6.55 / 10 | -0.25 |
| Contextual Depth | 8.10 / 10 | 8.25 / 10 | +0.15 |
| Response Latency | ≈ 5.5 minutes | ≈ 3.2 seconds | −99% |

### B. Latency Under Load

Because the platform relies on external inference, network latency is the primary failure point. We hit the API gateway with 100 concurrent mock interview submissions to observe the degradation in response times.

TABLE II: API STRESS BENCHMARKS (100 CONCURRENT SUBMISSIONS)

| Percentile Threshold | Resolution Time (ms) | UX Impact |
|---|---|---|
| p50 (Median) | 2,150 ms | Seamless |
| p90 | 3,420 ms | Noticeable Delay |
| p99 | 5,100 ms | High Friction |

The median resolution time sits comfortably around two seconds, ensuring the conversational rhythm of the mock interview remains fluid and engaging.

## VIII. CURRENT LIMITATIONS AND FUTURE HORIZONS

Deploying artificial intelligence into high-stakes educational environments requires acknowledging systemic blind spots.

Currently, the platform relies on the browser's native Web Speech API. While incredibly fast, its Word Error Rate (WER) degrades noticeably in

noisy environments or when parsing thick regional accents. A future iteration of this system will likely migrate the transcription duty to a dedicated acoustic neural network, such as OpenAI's Whisper model, ensuring maximum fidelity of the text payload before it reaches the reasoning layer.

Additionally, LLMs inherently carry the risk of hallucination. Despite strict system prompting, a model might occasionally invent a technical standard and penalize a candidate unfairly. To mitigate this in enterprise deployment, we propose a multi-agent validation step, where a secondary, smaller LLM cross-references the primary model's verdict for logical consistency before rendering the final dashboard to the user.

## IX. CONCLUDING REMARKS

The barrier to entry for top-tier engineering roles remains artificially high, driven more by a lack of accessible preparation tools than by a lack of raw talent. Interview AI demonstrates that by orchestrating modern web frameworks, strict browser- level proctoring, and the semantic depth of Generative AI, we can effectively digitize the human interviewer.

This platform moves assessment away from the brittle constraints of keyword matching, embracing a nuanced understanding of how candidates actually communicate complex logic. The empirical data confirms that the system operates with near-human accuracy at a fraction of the time and cost. As inference models become faster and context windows expand, tools like Interview AI will inevitably standardize how global engineering talent prepares, competes, and ultimately succeeds in the industry.

## REFERENCES


[1] Google. (n.d.). *Gemini API Documentation*. Retrieved from https://ai.google.dev/

[2] Mozilla Developer Network (MDN). (n.d.). *Web Speech API*. Retrieved from https://developer.mozilla.org/en-US/docs/Web/API/Web Speech API

[3] WebRTC.org. (n.d.). *WebRTC: Real-time Communication for the Web*. Retrieved from https://webrtc.org/

[4] S. Rajbhar, S. Shelke, A. Singh, Y. Singh, and P. Shinde, "AIced- Prep - AI Based Mock Interview Evaluator," *2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, Tirunelveli, India, 2025, doi: 10.1109/ICDICI66477.2025.11135157.

[5] "An Intelligent System for Automated Interview Question Generation: A Comparative Analysis of LLM Adaptation Techniques," *IEEE Xplore Conference Publication*, Feb. 2026. [Online]. Available: https://ieeexplore.ieee.org/document/11374807/

[6] "Q&AI: An AI powered Mock Interview Bot for En- hancing the Performance of Aspiring Professionals," *IEEE Xplore Conference Publication*, 2024. [Online]. Available: https://ieeexplore.ieee.org/iel8/10547720/10547724/10547951.pdf

[7] "VoxMentor: A Voice-Driven AI Platform for Interview Simulation, Company-Specific Preparation, and Student Learning," *2025 International Conference on Intelligent Computing, Information and Control Systems (ICOIICS)*, Nov. 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11390337/

[8] "Automated Interview through Online Video Interface," *IEEE Xplore Conference Publication*, Aug. 2023.

[9] A. Vaswani et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems (NIPS)*, vol. 30, 2017.

[10] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *NAACL-HLT*, 2019.

[11] J. Brooke, "SUS-A quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4-7, 1996.

[12] S. Young et al., "The HTK Book," *Cambridge University Engineering Department*, 2015.


– –