iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.81778

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Vocal Hire-Real-time AI Recruiter Voice Agent

Vidhi Verma, Rudra Pratap Singh, Neha Singh
Department of Computer Science (Data Science) Galgotias College of Engineering and Technology

Abstract: ***In today's world of global recruiting, we're witnessing a significant evolution from old methods of selecting candidates that are primarily reliant on human inter-vention (and therefore subject to bias) toward a fully automated means of screening applicants using a variety of sophisticated technologies designed specifically for the elimination of biasfromthe recruitmentprocess.Inthispaperwe'lldescribethe framework we created for an AI-Based Recruiting Voice Assistant, and how the implementation of this solution allows companies to communicate in real-time with applicants using the spoken word, in a highly scalable Software as a Service (SaaS) model, built with Next.js, Firebase® and Vapi Orchestration Services.Theoutcomeofoureffortsisarecruitingpipeline that will leverage the capability of speech-to-text, reasoningbased upon large language models and text-to-speech; all of this can occur within sub-second response times.Based on extensive empirical studies conducted in the field, the research finds that AI recruitment leads to more efficient hiring, with higher ratesof acceptance of job offers [12].***

Index Terms: AI Recruitment, Voice Agents, Real-Time Sys-tems, SaaS Architecture, LLM Orchestration, Human-in-loop, distributed system.

I. INTRODUCTION

The present state of the global hiring environment is at a levelofstressnotseenbefore.Viaonlineplatformswhich are increasing in popularity we see a "volume crisis" play out which in turn sees recruiters in technical fields like software engineering and distributed computing flooded with 100s of applicantsforeachopenposition.Thatlargeofanumber of applications requires very quick filtering which in turnvery often plays into human bias[1]. For years we have had eitherstaticresumescreenersorinflexiblechatbotsbased on pre-determined rules which make up what is known as Automation-as-a-Service (AaaS) solutions. But also included in that is the fact these solutions do not evaluate the human elements of a candidate's skill set and ability to problemsolveandchat.Inthisresearchweputforththeideaof anAIRecruiterVoiceAgentwhichistosay.Inthisstudy we present our put forth of an AI Recruiter Voice Agentwhich is to evaluate job candidates' problem solving skillsand conversation ability in real time and in dynamic way as a conventional human recruiter would. We do this by means of usingagenticLLMsandlowlatencystreamingpipeline.Inthis paper we present a which achieves a balance between three keyelementsoftheAItrianglelatency,costandaccuracy.We have used Next.js for front end state management,Firebase for the native cloud backend service and Vapi for the STT-LLM-TTS pipeline orchestration.

TheFirebaseplatformisbasedonseveralkeytechnologies: Firebase real-time NoSQL database, Firebase Authentication, and Firebase Functions.

II. LITERATURE SURVEY

Yuetal.[1]looked at how AI-powered recruitment systems use machine learning to automate candidate screening. Their study reports that which systems we have see to improve effi-ciency in terms of filtering and also in reducing manual work. At the same time these systems are very much into structured input and do not do well with real time conversations, hence theydofallshortindynamiccandidateevaluation.Inthework by Smith et al. [2] we see that they looked at conversationalAI in recruitment which included a study of voice based and chatbasedinterfaces.Theresultsdemonstratethatbyutilizing voice interaction to improve the level of engagement and the realism of the interview process, there are disadvantages to be foundwithinthesesystems(suchaslatency)andtheirinability toreacttocontextduringtheflowoftimewillaffectreal-time engagement throughout the interview process. Addition-ally, Paradox.ai has created quot;Oliviaquot;, a conversational recruiting assistant that automates how candidates interactwithrecruiters.AlthoughIthanhasperformedverywellon a high-volume basis, its limited reasoning abilities and rule-based framework make it less viable for complex interview processes. Similarly, HireVue has created AI-generated video interviewsthatallowrecruiterstoevaluatecandidateresponses againstpre-definedmetrics. Thistheysaydoesscalet hingsout but at the same time there is concern about transparency and biasinthesystemalsothe factthatitdoesn'tputforthadaptive questions in real time is an issue. Google Cloud [5] put out NLP based recruitment solutions built on a scalable platform.

©IJRASET: All Rights are Reserved

Their tools do a good job of parsing resumes and matching candidates but they don't go into real time voice interactionor streaming data pipelines. Amazon Web Services[6] looked at voice tech in the enterprise space which they stressed forits scalability and integration. The platform does deploy well but fails to put together a unified solution which includesSTT, LLM, and TTS for real time recruitment. Pymetrics Research[7] rolled out AI based cognitive assessments for hiring which included behavioral and gameified tests. While that is a novel approach of theirs' it operates independent of conversational systems which means it doesn't fit real time interviews. Young [8] disassembled conversational AI which hedidregardingdialoguemanagementandflexibility.Whatis shown is that these systems do well with varied conversations buttheydonotlookatlatencyorrealtimestreamingissues.

Brown et al. [9] put forth large scale language models with few shot learning which improved contextual reasoning. In recruitment though these models have high computational costs and are not tuned for real time tasks. Devlin et al. [10] put forth BERT for deep bi directional context understanding inNLP.Itdoesgreatfortextanalysisbutdoesn'tsupport real time conversation or voice based work flows.Kapoor and Narayanan[11] studied ethics in AI recruitment which they looked at fairness and bias mitigation. They put forward transparency as a key issue but did not present strategies for real time systems. OpenAI [12] did work on natural language understanding which improved reasoning and conversation abilities. Real- time applications often face several difficulties involving lag time amp; profitability; nevertheless, advance-ments made through AI-enabled recruitment have proven efficiencies through both time-saving and improving potential candidates39; experiences, according to an investigation by LinkedIn Talent Solutions. Unfortunately, the research does not elaborate on the details of the technologies or systems (including architecture) used to process these activities in real time. Furthermore, another paper authored by Raj et al., dealt with the subject of speech emotion recognition when applied within recruitment and demonstrated how helpful the use of vocalcharacteristicsistomakingevaluations.However,itonly focuses on identifying emotions and does not discuss how thoseresultscouldbeusedinconjunctionwithcomprehensive interview systems. According to a study by Bansal, much is still needed to address the use of analytics to improve human resourcesandrecruitingfunctions,asrecruitingofteninvolves manyinteractionsanddiscussionsinreal-timeorduringacon-versation with a potential applicant, which can be difficult for analytics to handle despite their known effectiveness. Xu and Zhaodesignedavoice-enabledassistantthatcouldassisthiring managers in their ability to communicate with job applicants, thereby improving the quality of hiring decisions. However, while they both improve the quality of communication, these twovoice-enabledassistantsalsohavelimitationswithrespect to their adaptive reasoning capabilities, limiting the overall depth of their evaluation of an applicant39;s performance. Krishnan et al. analyzed candidates39; speech patterns for sentiment analysis as a means to obtain additional insight into their behaviors while interviewing for a job. Their system was also designed to work as a standalone module, rather than as anintegratedsolutionwithanyexistinganalyticssystems.The use of AI, which is helping to increase productivity through more efficient processes and automation within human capital management,hasbeenreportedbyDeloitteInsights.However, the report did not cover any of the technical difficulties of implementingreal-timesystems.WhileMcKinseyamp;Com-pany provided a separate study on the long-term impact of AI on the recruitment process and productivity and cost savings, the report was also missing details on system design and practice.

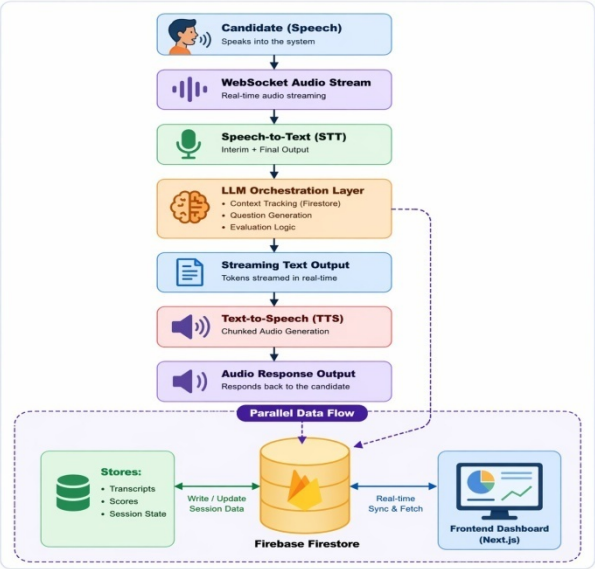


Fig.1.Workflow

Joshi et al's investigation into the use of automation for early- stage recruitment found significant reductions in time spent screening candidates, but their method does not supportinteractiveevaluationsofthefullrangeofcandidate skills before selection. Lastly, Accenture's research into AI and talent acquisition describes how scaling and efficiency poseadditionalchallengestothesuccessfuluseofAIin the talent acquisition process. The research did not focus on the depth of conversation or whether or not the interactionwasinreal-time.AcasestudyperformedbyIBMlooked into how conversational interfaces can be used to automate human resource tasks and found that there was a significant improvementintheoverallqualityofinteraction.Thatsaidin-tegration with real time streaming architectures is still limited. Sharma and Gupta [23] pushed for bias reduction in AI hiring systems which they focused on fairness metrics like Disparate ImpactRatio.ForbesInsights[24]reportedonRecruitment 4.0 which is how AI and analytics come together in hiring. They put out strategic insights but did not go in to technical rigor or system evaluation. Zhao[25]looked at multi lingualAI agents for global recruitment which he reported to have issues of scalability across languages. Real time latency and orchestration remain unsolved problems. Kumar [26] lookedat cloud-native HR systems with AI which he put forward in terms of scalability and modularity. His work did not get into low latency streaming issues. NASSCOM[27] reported on AI inIndianHRtechwhichtheysawplayoutinanincrease inadoption.Thereporthoweverlackedintechnicaldepth and empirical evidence. World Economic Forum[28] lookedat AI's role in recruitment which they put forth in terms of efficiencyandaccess.Alsotheydidnotaddressissuesofbias, transparency and real time performance.

III. EVOLUTION OF AI IN RECRUITMENT

TheevolutionofAIinrecruitmentishappeningveryfast which we can categories into three phases.

A. TheResumeParsingPhase

Early output was seen in Natural Language Processing (NLP) based techniques for keyword extraction. We saw use of Named Entity Recognition (NER) to map resumes to job descriptions. While these approaches were computationary efficient they had issues with what we term "keyword stuff" and also were not able to verify the authenticity or depth of a candidate's claims [2]. As a result, they mainly functioned as filter tools rather than true evaluation systems.

Thefirstgenerationofsystemsthatprocessedresumesused Natural Language Processing (NLP) to find the keyword intheresumeandmatchedthatkeywordfromthejobdescription usingNamedEntityRecognitiontechniques.Whiletherewere many benefits to the above approach like being inexpensive andeffective,therewerealsoseveralmajorflawsorlimitations to arise as a result of this approach [2]. Candidate can manipulate the system.

B. TheIntent-BasedChatbotPhase(2015–2022)

Afterward, there were conversational artificial intelligence (AI)platformsandintent-basedchatbotsemerging,wherethey were based on rigid decision trees. If a candidate engagedwith those systems in a manner that conformed to what was beingexpectedbythedecisiontree,thesystemwouldfunction adequately; if the candidate deviated from the decision tree, that system (was not able to) become confused and used a 'catch- all' response which resulted in less than an optimal experience for the candidate and also provided less than optimal information gathered by the company.

C. TheAgenticLLMPhase(Present)

Nowadays,transformer-basedAImodelsarebeingused to reason. An example of this is Vocal Hire. Rather than followingascript,theAIhasspecificgoalstoaccomplish.For example, it listens to candidate responses and creates follow-upquestionsbasedonhowmuchtheyknowaboutthesubject. Italsoasksadditionalquestions,whenappropriatebasedupon how in-depth their answers were.

IV. SYSTEM ARCHITECTURE OVERVIEW

The process is in this loop. You take the input audio which could be Vapi and run it through a module which turns it into text.Transcriptiontakesplaceatthesametimeasadiscussion ratherthanhavinglargeperiodsoftimebetweenwhencoming totheend;itisdevelopedovertimeratherthanattheveryend. After transcribing, the text is reviewed for the implications of the answer, the present point in time in relation to the inter-view,andhowcloselythecandidate39;sanswerscorrespondto the job requirements. Once generated, a response is converted into audio that is sent in small pieces, to give the user a responsive experience.

The architecture allows for scalable performance; since there are multiple components instead of onebigcomponent,thisallowsforeitheradjustingthespeedof processing the components or using various service providers dependingonthestateofthenetwork.Forexample,ina low-bandwidthscenario,DeepgramperformsmuchfasterthanWhisper, while when there is a stable connection, Whisper performsbetter.ThefrontendwascreatedinNext.jsusing the App Router and Server Components for faster page loads. Also, due to the use of server-side rendering, recruiters will have up-to-date information on their dashboards. All of the live interviews take place locally on the user39;s end through WebSockets for the purposes of communication and for live updates of statistics with the goal of achieving an engaging experience. The backend architecture uses Firebase as the basis for all features including authentication, database, and real-timeupdates,whichsimplifiestheentireprocess.For the database, Cloud Firebase is the NoSQL choice, scalable and flexible. Now worries about data sets growing later. Security comes in with multi tenant set up through strict rules. Firebasesecurityrulesmakesurerecruitersonlyseetheirown organization's data.
In this app the whole system is divided into different modules and those module help in completing the process. EachusergetsauniqueIDbasedonfromwhichEmailaddress it login and on the basis of that it get a unique ID as well as unique session. When anyone done with there interview after the interview a report is generate which is generated through the Gemini 2.0 Flash which reads the transcript and generate the scorecard of the user and it include also key area's to improve for candidate.

V. END-TO-ENDLOW-LATENCY SPEECH INTERACTION PIPELINE

The real issue arises when you attempt to put together a smooth running streaming setup without any hiccups. We go with speech to text first. We hook into Deepgram's Nova 2 model via a WebSocket connection. What we did was to skip the usual wait for quiet moments. Instead we went with an interim results which we built up. Also the model hits 90 on that.OntheLLMsidewedividedtasksbetweentwomodelsto keepthingsefficient.Thepresentliveinterviewflowishandled by GPT-4o-mini. We tried out the very fast response time whichisakeyfeatureinrealtimeinteraction.Oncethechatis over we move to Claude 3.5 Sonnet or GPT-4 for the in depth analysis. We go over the full transcript, put together a report and present scores which go straight to the recruiter. /section Latency Modeling and Optimization AI interaction should be very quick. Any delay of a second or more is noticeable and out of place. The best setups kick back in like half a second [3]. /sectionLatencyModelingandOptimization TalkingtoanAIshouldfeelquick.Ifthewaittopsasecond, it just drags and seems wrong. The best setups kick back in like half a second [3].

A. Latency Equation

Thetotalsystemlatencyismodeledas:

$$L_{\text{total}} = L_{\text{net}} + L_{\text{STT}} + L_{\text{LLM(TTFT)}} + L_{\text{TTS(TTFB)}} + L_{\text{buffer}} \quad (1)$$

Where:

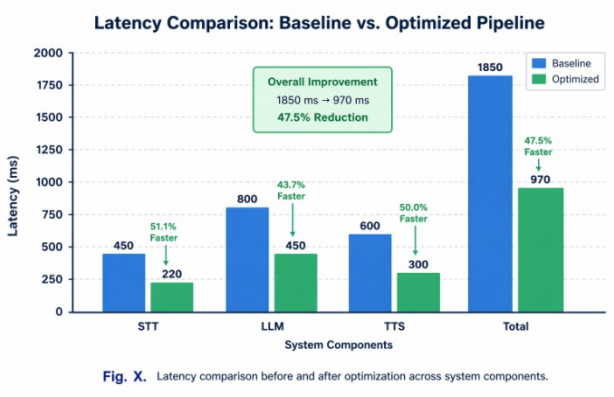


Fig.2.Latencycomparisonbeforeandafteroptimization

- L_{net}:Round-triptime(RTT)betweentheclientandthe nearest edge node.
- L_{STT}: Time for the Speech-to-Text (STT) engine to pro-vide a stable transcript.
- $L_{\text{LLM(TTFT)}}$:TimetoFirstToken(TTFT)fromthelan-guage model.
- $L_{\text{TTS(TTFB)}}$:TimetoFirstByte(TTFB)ofsynthesized audio.

- L_{buffer}:Additionalbufferingdelayintroducedduring streaming.

B. EmpiricalLatencyMeasurements

TABLEI
EMPIRICALLATENCYMEASUREMENTS

Component	Baseline(ms)	Optimized(ms)	Improvement(%)
STT	450	220	51.1%
LLM	800	450	43.7%
TTS	600	300	50.0%
Total	1850	970	47.5%

WeoptimizedusingSileroVAD(VoiceActivityDetection) toknowexactlywhensomeonestopstalking.Thiscutsout the awkward pauses where nothing happens while waiting for the LLM to spring into action.

VI. SAASMONETIZATION AND COSTMODELING

Inorderfortheplatformtocontinuetoremainprofitable, we operate on a credit-based billing model. Conducting a 15-minute interview costs significantly more than a text chat becausebothspeechsynthesisandspeechrecognitionare costlyservicesthatyoupay forpercharactergeneratedand per minute of audio.

A. CostOptimizationFormula

Thetotalcost(C_{total})forasingleinterviewsessionofduration Tminutes is:

$$C_{total}=(C_{STT}\cdot T)+C_{LLM}\cdot\frac{N_{tokens}}{1000}+(C_{TTS}\cdot N_{chars})$$

Where:

- C_{STT}≈$0.01perminute.
- C_{LLM}≈$0.015per1ktokens.
- C_{TTS}≈$0.0003percharacter.

B. SaaSPricingStrategy

To achieve a 70deduction logic is set at:

$$\text{Credits}_{\text{deducted}}=\frac{C_{total}\cdot(1+\text{Margin})}{\text{Price}_{\text{percredit}}}$$

VII. EMPIRICAL EVALUATION AND RESULTS

We circulate our feedback form to approx 200 users who has used our platform across the colleges in our near by to get thefeedbackabouthowsystemisworkingandwhat'sneed to be improved and what was the best part of it.

A. ConversionEfficiency(OutcomeEffectiveness)

This shows whether people genuinely see value in the platform and would suggest it to others—not just toss out aso-so rating. Formula:

$$\text{ConversionEfficiency}=\frac{\text{NormalizedNPS}+\text{EngagementProxy}}{2}$$

SinceyouonlyhaveNPS:

- $\text{NPS}=+39.5$
- $\text{NormalizedNPS}=\frac{39.5+100}{200}\times 100=69.75\%(\approx 70\%)$

B. EvaluationBreadth(SystemCapabilityCoverage)

This tells you how well your system judges different things liketechnicalskills,clearthinking,andcommunication.**Prac-tical Proxy (based on your data):**

$$\text{EvaluationBreadth} = \frac{\text{AverageRating}}{\text{MaxRating}} \times 100$$

$$= \frac{3.94}{5} \times 100 = 78.8\%$$

C. CandidateSentiment

It's less about numbers and more about how users actually feel. Track both ratings and NPS. **Formula:**

$$\text{CandidateSentiment} = \frac{\text{RatingEfficiency} + \text{NormalizedNPS}}{2}$$

$$= \frac{78.8+70}{2} = 74.4\%$$

Metric	Score	Interpretation
Conversion Efficiency	~ 70%	Moderate–
GoodEvaluation Breadth	78.8%	Strong
CandidateSentiment	~74.4%	Moderate–Strong

VIII. ETHICAL CONSIDERATIONS AND BIAS MITIGATION

AIrecruitingisheavilyscrutinized,andforgoodreason.We have a human-in-the-loop system so the AI provides a recom-mendation but a human makes the final decision.

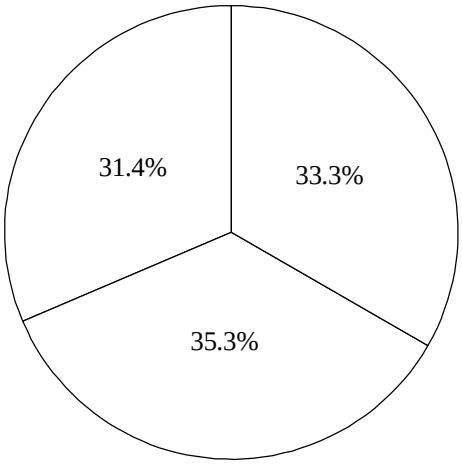


Fig.3.DistributionofPerformanceMetrics(GrayscaleOptimized)

A. BiasQuantification

WemeasurebiasusingtheDisparateImpactRatio(DIR):

$$DIR = \frac{SelectionRate_{Protected}}{SelectionRate_{Majority}}$$

B. MitigationStrategies

Accent bias was an issue. We attacked it in two ways. First we trained on a variety of accents. Second we inserted a layer called ”Acoustic Normalization” that disregards how some-thing is said–tone, accent, speech patterns–and only focuseson the words and intended meaning.

IX. LIMITATIONS AND FUTURE DIRECTIONS

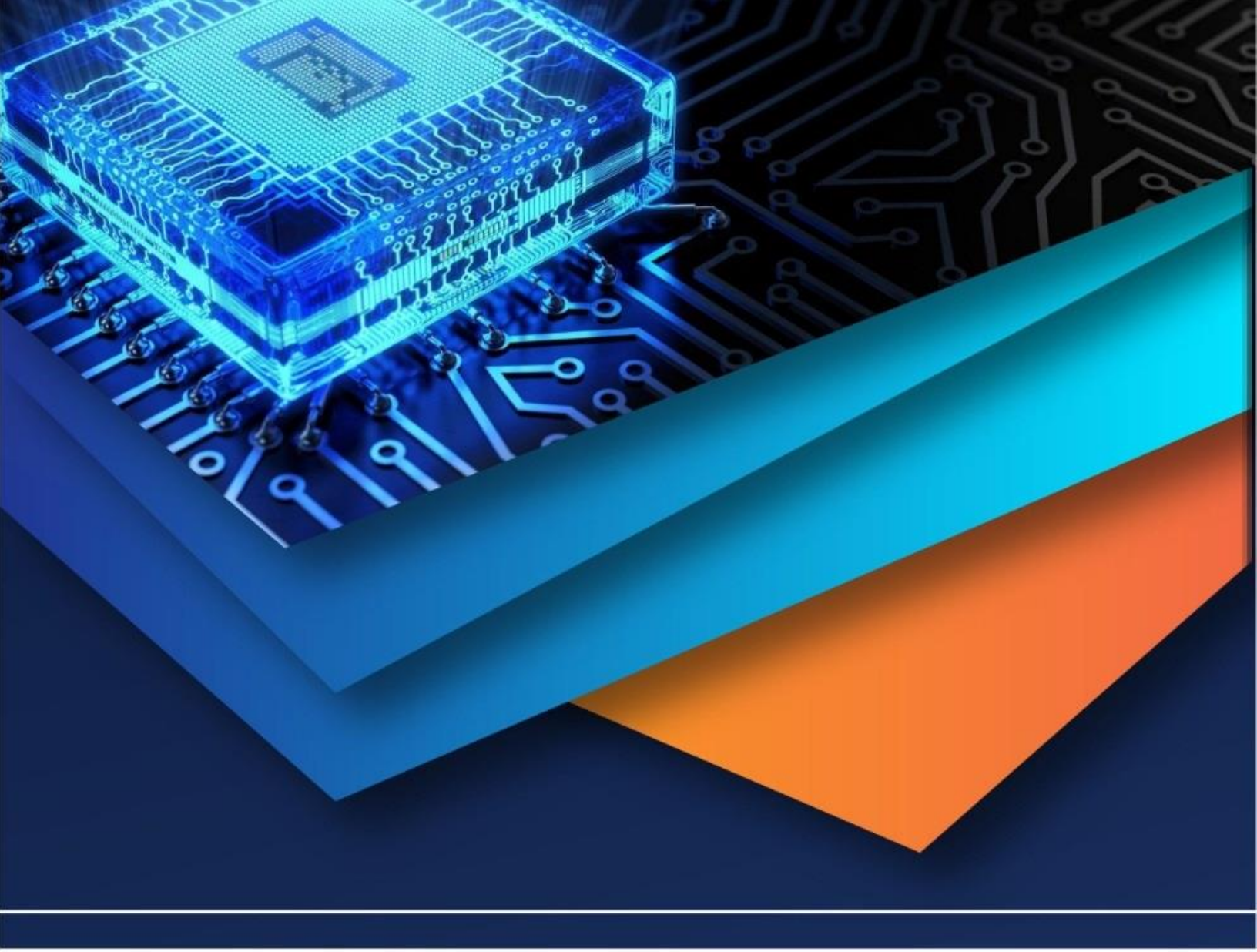
Though this process shows high levels of efficacy, it still has its shortcomings. For example, background noises and simultaneous speaking can impair the process. Currently, the processuseaudioalone,meaningthatthereisnochance for assessing the body language and facial expressions of candidates. Areas of future research can include: 1) Multi-modal Evaluation: Utilizing video input to-assess the level of engagement of the candidate. 2)Edge AI:Employing STT and TTS using WASM with latency under 200 ms. 3) End-to-end Speech Models: Using audio capabilities in models like GPT-4o directly.

X. CONCLUSION

The current research has managed to outline the archi-tecturaldesignforVocalHireatscale.Theimplementation of Next.js, Firebase, and Vapi in the form of an end-to-end pipelinehasprovedthatsub-secondconversationallatencycan be achieved through a SaaS architecture. Studies suggest that such technologies can improve organizational processes while providing a smoother experience for candidates. Inlight of de-velopments in LLMs, the shift from automation to partnership is an unavoidable step.

REFERENCES

[1] C.Yu*etal.*,“AI-poweredrecruitmentsystems,”*IEEEIntelligentSys-tems*, 2019.
[2] A.Smith*etal.*,“ConversationalAIinrecruitment:Voicevschat,”in*Proc. ACM*, 2021.
[3] Paradox.ai,“Olivia–conversationalrecruitingassistant,”2021.
[4] HireVue,“AIinvideorecruitment,”WhitePaper,2020.
[5] GoogleCloud,“NLPinrecruitmentapplications,”2022.
[6] AmazonWebServices,“Voicetechnologyinbusinessapplications,”2021.
[7] PymetricsResearch,“AI-basedcognitiveassessmentforhiring,”2019.
[8] S.Young,“ConversationalAIsystems,”*ACMComputingSurveys*,2020.
[9] T.Brown*etal.*,“Languagemodelsarefew-shotlearners,”in*Proc.NeurIPS*, 2020.
[10] J.Devlin*etal.*,“BERT:Pre-trainingofdeepbidirectionaltransformers,”in *Proc. NAACL*, 2019.
[11] A.KapoorandS.Narayanan,“EthicalconsiderationsinAIrecruitment,”*Journal of AI Ethics*, 2021.
[12] OpenAI,“Advancesinnaturallanguageunderstandingmodels,”2022.
[13] LinkedInTalentSolutions,“Thefutureofrecruiting:AItrendsandinsights,” 2022.
[14] R.Raj*etal.*,“Speechemotionrecognitionforrecruitment,”*Interna-tional Journal of Computer Applications*, 2021.
[15] N.Bansal,“AIandHRanalytics:Areview,”*JournalofManagementSystems*, 2020.
[16] L.XuandQ.Zhao,“Voice-enabledassistantsforcorporatehiring,”*IEEEAccess*, 2021.
[17] P.Krishnan*etal.*,“Evaluatingcandidatesentimentviaspeechpatterns,”*Springer AI Review*, 2022.
[18] Deloitte Insights, “AI in human capital management,” Deloitte Review,2023.
[19] McKinsey&Company,“FutureofworkandAIrecruiting,”2021.
[20] M.Joshi*etal.*,“Automationinearly-stagerecruitment,”*IJIRSET*, 2020.
[21] Accenture,“TheroleofAIinenhancingtalentacquisition,”2021.
[22] IBMResearch,“ConversationalinterfacesforHRautomation,”IBMTechnical Journal, 2020.
[23] P.SharmaandR.Gupta,“AI-basedhiringbiasreduction,”*JournalofAI Ethics*, 2023.
[24] ForbesInsights,“Recruitment4.0:AIandanalytics,”ForbesTechnologyCouncil, 2021.
[25] K. Zhao, “Multilingual AI agents for global recruitment,” *IEEE Trans-actions on AI*, 2023.
[26] S. Kumar, “Cloud-native HR systems with AI integration,” *Springer AIApplications*, 2022.
[27] NASSCOM,“AItrendsinIndianHRtechnology,”2023.
[28] WorldEconomicForum,“ShapingrecruitmentthroughAI,”2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)