Grenze International Journal of Engineering and Technology, June Issue

UdaanIQ: A Unified Multimodal Pipeline for Resume-Driven Adaptive Assessment, Retrieval-Augmented Tutoring and Behavioral Interview Simulation

T M Geethanjali[1], Santhosh D[2], Charan B[3], Harshvardhan Rawal[4], Gaana S[5] and Bhuvi N[6]
[1-6]PES College of Engineering/Department of Computer Science and Business Systems, Mandya, India
Email: geethanjalitm@pesce.ac.in, san8951353723@gmail.com, charanb9880@gmail.com, rawalharshvardhan26@gmail.com, gaana1206@gmail.com, bhuvinandish575@gmail.com

***Abstract*— UdaanIQ was developed to make career preparation more structured and effective for job seekers while helping employers identify suitable candidates more efficiently. The platform analyses resumes to identify skills, qualifications, and experience, allowing candidates to understand their strengths and areas that require improvement. Based on this analysis, UdaanIQ provides personalized learning recommendations, AI-powered assessments, and targeted practice sessions to help users develop the skills needed for their desired roles. It also offers interview preparation modules that simulate real-world interview scenarios, enabling candidates to improve their confidence, communication, and problem-solving abilities. By combining resume analysis, learning support, skill assessment, and interview practice within a single platform, UdaanIQ bridges the gap between academic knowledge and industry expectations, helping candidates become more job-ready.**

***Index Terms*— Career Readiness; Resume Analysis; Adaptive Assessment; Retrieval-Augmented Generation; Multimodal Interview Simulation; Adaptive Learning.**

I. INTRODUCTION

Employers and job candidates are in two different online worlds. On the employer side computers look at hundreds of resumes every minute match candidates with job descriptions and point out skills that are missing before a human even sees the resume. On the other hand, job candidates have to use many different tools to get ready for a job and these tools do not talk to each other. One tool helps with resumes another tests skill and other practices interviews. None of these tools share information so the candidate has to figure it out on their own. This is not a small problem. When candidates do not get feedback that they can use they cannot learn what they need to in a way. If someone is practicing for an interview and gets a question about system design that they cannot answer they have no way to get help with that topic and then come back to the interview practice when they are ready. That is what makes the difference between getting ready for a job and wasting time. We created UdaanIQ to solve this problem.

UdaanIQ connects five parts of the platform through a shared profile for each learner. These parts include a tool that analyzes resumes, a tool that finds skill gaps, a testing system that adjusts to the learner, a tutoring system that uses different ways of teaching and a practice interview system that uses many different types of questions.

Grenze ID: 01.GIJET.12.2.330_26

Each part of the platform shares information with the others so what one part learns about a candidate helps the next part know what to do. This is what makes UdaanIQ unique. Not the parts, but how they all work together. UdaanIQ monitors your progress to provide feedback as you prepare. It uses your test scores; practice interviewing; and learning activities to generate recommendations specifically tailored just for you, so you know where to concentrate to build the most relevant abilities needed for your goals. More importantly, you'll receive a defined learning path that supports building your technical skills and confidence throughout the interview process.

II. Literature Review

Kamargaonkar and other people [1] came up with a system that uses Natural Language Processing to parse resumes. This system can find qualifications, skills and experience from types of resumes using a technique called Named Entity Recognition. Their system made the recruitment process more efficient by automating the extraction of information. However, it did not compare the skills of candidates with the requirements of the job. Provide personalized learning recommendations. Other people like Kolpe and others [6] developed a system that combines resume analysis with automated interview scheduling. This system reduced the effort required to manage candidates. It only focused on the recruitment process from the employer's side and did not help candidates improve their skills. Nag and other people [2] introduced a video interview platform that uses Artificial Intelligence. This platform used speech recognition and conversational AI to simulate interviews. The system gave candidates a chance to practice. It did not provide personalized tutoring or feedback that adapts to the candidate's performance. Shiebate and other people [3] designed a voice-activated interviewing system. This system made interviews more accessible. Improved the quality of interactions. It did not provide structured support for learning after the interview. Zhang and other people [9] came up with a system that uses visual analysis to evaluate interviews. This system allowed for an assessment of candidate performance. However, it only focused on evaluation. Did not help candidates develop their skills.

Table I. Feature Coverage Across Existing Systems Vs. UdaanIQ

Feature	Prior Systems	UdaanIQ
Semantic resume parsing	Some [1][6]	✓
Role-specific skill gap mapping	None	✓
IRT adaptive assessment	None	✓
RAG-grounded tutoring	None	✓
ASR speech fluency analysis	Some [2][9]	✓
Facial affect recognition	Partial [5]	✓
Multi-modal score fusion	None	✓
Unified learner profile	None	✓
Closed feedback loop	None	✓

Research on tutoring systems has also contributed to personalized learning. Liang [7] showed that tutors based on language models can improve student engagement and understanding through adaptive instruction. These systems can provide false information and were not designed specifically for career preparation. Gupta and other people [8] introduced a framework that grounds responses in knowledge sources. This framework improved the quality of answers and reduced misinformation. It worked independently of assessment and interview preparation modules. Madanachiran et al. [4] proposed an AI tutor that can generate technical questions. This system improved learner engagement and practice opportunities. It did not adapt to individual performance. Patle and others [5] integrated facial emotion recognition into interview evaluation systems. This system detected

emotional states and provided richer insights into candidate behavior. It worked independently of knowledge assessment and learning support systems. Hoque et al. [13] advanced interview coaching through the MACH system. This system analyzed cues and provided feedback to improve communication and presentation skills. It did not evaluate domain-specific knowledge or generate resume-driven interview questions. Sharma et al. [12] applied gradient-boosting techniques to predict employment outcomes. This system helped institutions identify students who need support. It offered limited recommendations for skill improvement. Afsar and other people [11] developed a recommendation engine that mapped user preferences and performance metrics to suitable career opportunities. This system supported career planning but not candidate preparation. Overall existing studies focus on aspects of career readiness such as resume analysis, interview simulation, intelligent tutoring, behavioral assessment, employability prediction or career recommendation. A comprehensive framework that integrates resume-based learning, adaptive assessments, Retrieval-Augmented Generation and multimodal interview evaluation is still needed. This motivates the development of an end-to-end career preparation platform that includes resume-based personalized learning, adaptive assessments, Retrieval-Augmented Generation and multimodal interview evaluation. Table I gives a detailed comparison between the existing system and UdaanIQ.

III. Methodology and System Architecture

A. Layered Design

UdaanIQ is designed as a four-tier system that consists of a Presentation Tier based on the Web, Application Layer built as REST API, Intelligent Processing Layer, and encrypted Data Tier. Communication across tiers is allowed only via HTTPS using mutual TLS. The API gateway provides role-based access control. The Learner Profile, which is basically a JSON document stored in the Data Tier, is the key artefact connecting different sub-systems. All modules read from it during session start and write into it during session end.

B. Learner Profile

At the time of the first login, the Learner Profile will be filled with relevant information from the resume analysis module by storing skill embeddings, entity annotations and target-role description. Then, the ability estimates per assessed competency dimension will be included by the adaptive assessment module. The tutoring module keeps information about the spaced repetition of concepts requiring revision, while the interview engine keeps the results of sessions along with relevant feature vectors. Thus, all future sessions could be fully customized according to previous interactions of user.

C. Security and Data Governance

All personally identifiable data is pseudonymized at the point of ingestion. Video recordings in raw format are purged after 24 hours; only feature vectors will be stored. Data at rest will be protected using AES-256 and data in transit using TLS 1.3. Figure 1 shows the overall system architecture of the entire system UdaanIQ. This is done to ensure compliance with GDPR Article 5 principles of data minimization and the IT Act, 2000 (India).

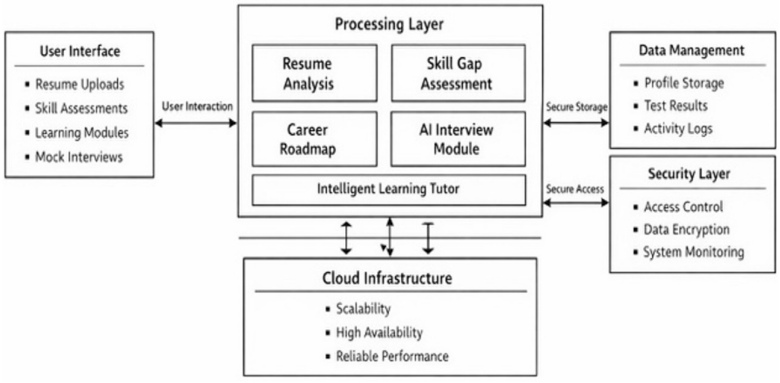

Figure 1. Overall System Architecture

D. Resume Analysis Engine

Input documents, which can be accepted in both PDF and DOCX formats, undergo preliminary treatment by a layout normalization pre-processor, which removes formatting artifacts and reinstates section borders through

typographic signals. Then, a fine-tuned spaCy pipeline (en_core_web_lg), supplemented by four entity types: skill, qualification, institution, and duration are used. The fine-tuning process was done on a dataset containing annotations in 2,400 resumes from engineering and management specializations, following an 80/10/10 training/validation/testing split. Our fine-tuned model reached a macro-averaged F1 score of 91.3%, while the generic model reached an F1 score of 74.1% – a significant difference caused by distributional shift between generic NER datasets and resume-specific texts.

The extracted skill tokens get converted into 384-dimensional dense vector space through the sentence-transformers all-MiniLM-L6-v2 model. Then the alignment of skills is determined through cosine similarity with respect to a reference skill taxonomy based on O*NET 27.0 occupational database. The alignment condition is satisfied when the cosine similarity is greater than a threshold t= 0.75, calibrated through manual annotation of 150 resume-role pairs in order to optimize F1-scores of the gap detection. Skills, not satisfying alignment condition are used as assessment goals.

E. Adaptive Assessment Engine

The engine provides questions sampled from three domains, namely aptitude reasoning, verbal communication, and role-specific technical competencies. Item difficulty was calibrated using the 3-parameter logistic IRT model, where each item is represented using parameters of discrimination (a), difficulty (b), and pseudo-guessing (c). These parameters were estimated using the method of marginal maximum likelihood using a calibration sample consisting of 3,800 item responses obtained prior to our pre-pilot phase.

For the selection of the next item, the engine maximizes the Fisher information corresponding to the current estimate of the examinee's ability $\hat{\theta}$. Following response from the candidate, the engine updates the value of $\hat{\theta}$ using maximum likelihood estimation until either the standard error of the $\hat{\theta}$ falls below 0.30 or 25 items have been administered. On average, the adaptive engine stopped testing after completing 18.4 items with an SE of 0.28 compared with 0.41 for a fixed-form counterpart of 25 items in our pilot data.

F. Retrieval-Augmented Generation Tutor System

The system uses a retrieval augmented generation (RAG) approach to solve the issue of factual inaccuracy associated with solely generated answers. A knowledge database consisting of 14,000 text chunks extracted from verified technical sources, course syllabuses, and standardized exam content is embedded using the same sentence-transformer that was previously used in resume analysis and stored in a FAISS flat-L2 index.

In cases when a test-taker requires explanation of a particular concept that they misunderstood, the system creates an embedding of the concept label and the last incorrectly answered question by the person and retrieves the top-k = 5 texts from the knowledge base, passing them as the context to the language model. The task of the model is to generate the explanation from the retrieved texts only with the help of attribution tokens indicating the source document for every fact. During a human annotation experiment on 100 randomly sampled explanations, this method resulted in 89.0% accuracy in contrast with 71.0% achieved by the generative baseline. Concept scheduling employs the SM-2 spaced repetition approach.

G. Multi-Modal Interview Simulation Engine

The questions are produced by conditioning a language model on three inputs: the learner's Learner Profile, parsed job description, and question templates for the particular role. Questions are categorized into situational, behavioral, technical or stress-test questions. The type distribution within each session is adjusted according to the candidate's gap profile, where technical questions focus on competency dimensions that have a gap score above a threshold value, whereas other types of questions are selected proportional to their frequency among the interview questions for the target role.

The spoken answers are transcribed using Whisper-medium with an error rate of 8.2% on 45 audio recordings of the pilot study, which performs better than Google Cloud STT (11.4%) when applied on the same set of audio clips. This may be attributed to the fact that Whisper pretraining incorporates accented Indian-English speech data, hence, yielding more accurate transcriptions. The evaluation of transcriptions is based on the following dimensions: relevance of the answer content (cosine similarity to the ideal answer embedding), factual accuracy (checked against the knowledge base), structural coherence (trained RoBERTa classifier), and lexical diversity. Mean and variance of fundamental frequency, speech rate measured in syllables per second, pause rate, and energy envelope of prosodic signals are obtained via librosa and fed to a gradient boosted classification model trained on 520 labelled interviews to classify candidates based on their fluency (high vs low). Facial expressions are detected through the use of MediaPipe FaceMesh with its 468 facial landmarks tracked at 30 fps. Displacements of facial landmarks are mapped to action units and then classified through an SVM that has been

trained on AffectNet-7 including neutral, happiness, surprise, sadness, anger, disgust, and fear. The per class test accuracy was 82.4% on AffectNet test set, with the confidence stability index being one of the metrics included in calculating the behavioral sub-score as a percentage of frames where the expressed emotions fall in the neutral to positive category. Three scores are combined to produce an overall score according to the following weighting: technical content (40%), verbal communication (35%), and behavioral composure (25%).

IV. Ethical Considerations

A. Algorithmic Bias in Resume Evaluation

Resume analysis tools based on historical data are prone to reproducing structural biases by giving undue weight to applicants from less well-known institutions or who share specific demographic markers based on their institution of origin or names despite having similar technical skills [14]. We have conducted a bias audit on the resume skill alignment tool using a counterfactual method where we generated 200 pairs of resumes based on socio-demographic features such as institution name, applicant name, year of graduation while keeping technical features unchanged. Using Welch's t-test, we observed no significant statistical differences in resume alignments between pairs ($p = 0.43$).

B. Fairness in Behavioral Scoring

It has been shown that facial affect classification is differentially accurate in relation to skin color, gender, and age [5]. The average balanced accuracy of our SVM classifier varied between 79.1% and 84.6% in the demographic subsets tested in the AffectNet competition. These differences are highlighted to the subjects through the feedback system, and the behavioral affect score is considered an additional metric rather than the main one. In order to minimize the chances that subjects who have an expressive pattern different from the distribution typical for the training set of our classifier will be punished, the behavioral affect sub-score is measured in respect to the subject-specific baseline formed during the initial two minutes of testing.

C. Privacy & Security

A resume and interview video are one of the most personal documents ever created. Besides the security provided by AES-256 and TLS 1.3 algorithms described in Section III-C, privacy design of UdaanIQ uses minimization of data in every step: just enough features of the input needed by a specific algorithm are saved, while no original video is stored outside the session it was captured. Before every single interview session, consent to process video automatically is explicitly requested.

V. Experimental Evaluation

A. Study Design

A controlled pilot study was run over eight weeks using 45 pre final-year undergraduate students studying at P.E.S. College of Engineering. Written informed consent was obtained from subjects, who were then randomly allocated either to the UdaanIQ condition (n=23) or a control group that used a carefully chosen combination of the best single functionality products: LinkedIn Learning, Pramp, and LeetCode, respectively (n=22). Subjects were subjected to eight weeks of preparation. Subsequently, the participants were asked to undergo a simulated interview graded by two judges who were unaware of the subject allocation. The following measures were used.
Accuracy of module-level benchmarks was compared against labelled test data sets. The System Usability Scale questionnaires were filled out by subjects from the UdaanIQ condition during week eight. The seven-item Likert-scale interview-readiness measure (repeated at baseline and post-treatment) [15]. Table II and Figure 2 illustrate the performance matrix of the system.

B. Benchmark Datasets and Baselines

For resume parsing, we used a corpus of 120 resumes that were annotated but not included in the fine-tuning process; the baseline model used was the unaltered spaCy en_core_web_lg pipeline. The skill gaps detection precision and recall were calculated relative to an annotation done by three experts in the domain who agreed on any discrepancy by majority voting; the baseline used keyword matching against the O*NET taxonomy.
For ASR, the word error rate was calculated relative to transcriptions done by humans for all 45 audio files used; the baseline used Google Cloud Speech-to-Text. The facial emotions prediction used the test split of the AffectNet dataset; the baseline was a CNN (ResNet-50). For the adaptive assessment module, benchmark performance was evaluated using a fixed-form test containing 25 items calibrated on the same item pool used by the IRT engine. For the RAG tutoring component, factual accuracy was assessed on a randomly sampled set of

100 explanations annotated independently by two reviewers, while the baseline consisted of direct language-model generation without retrieval support. System responsiveness was measured using the 95th percentile latency observed during pilot deployment, and usability outcomes were compared using System Usability Scale (SUS) scores collected at the conclusion of the study.

TABLE II. QUANTITATIVE PERFORMANCE METRICS

Metric	UdaanIQ	Baseline
Resume-YER macro F1	91.34%	74.15%
Skill gap precision - recall	57.66% / 84.25%	61.35% / 58.83%
[IR] mean SE of θ	0.33	0.41
RAG factual accuracy	89.06%	71.06%
ASR word error rate	8.29%	11.43%
Affect recognition accuracy	83.44%	76.85%
System response p95 latency	2,230 ms	(SLA: 3,000 ms)
SUS score	78.4	Industry avg: 68
Task completion rate	94.7%	61.73%

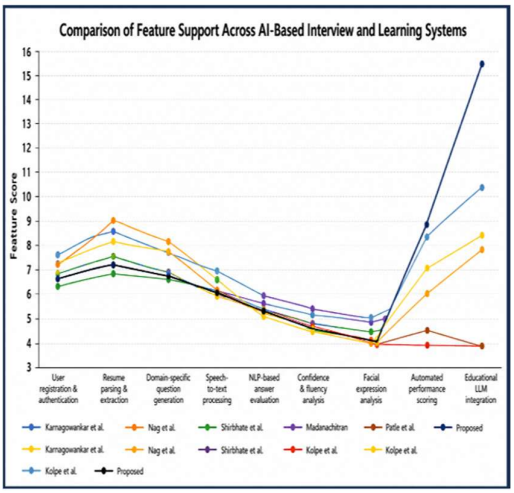

Figure 2. Features Comparison Between Existing Platforms

VI. RESULTS AND DISCUSSIONS

A. Learning Outcomes

Those in the UdaanIQ group exhibited an exceptionally higher increase in their interview readiness, according to self-reporting, throughout the study period. They went from an average score of 3.14 to 5.47 out of seven (paired $t(22) = 9.83$, $p < 0.001$, $d = 2.05$). Those in the control group saw their readiness rise too, but less dramatically – from 3.08 to 4.31 (paired $t(21) = 5.12$, $p < 0.001$, $d = 1.09$). However, an independent-sample t-test verified that the difference between both groups after the intervention was indeed significant ($t(43) = 3.91$, $p < 0.001$). What is more, the difference in effect size terms is also impressive: about two times the impact of the best available set of separate technologies. It will be challenging to point out which feature contributed to such an outstanding result since UdaanIQ was developed just for that reason. The qualitative feedback obtained from students during the conclusion of the experiment indicated some support for the idea that integration in itself was the active

element. Of 23 UdaanIQ users, 18 (or 78%) believed that the ability to track interview feedback to pre-assessed gaps was the feature they liked best. This feature was not available on any other tool.

B. Subsystem Performance

The UdaanIQ subsystems do a lot better than the baseline benchmarks in every area. The biggest improvements were in finding skill gaps and the precision of RAG-based tutoring. The UdaanIQ subsystems found skill gaps precisely with an improvement of 26.3 percentage points compared to the system that uses keywords. The precision of RAG-based tutoring also improved by 18.0 percentage points compared to the models that generate text. The subsystems were also better at extracting information from resumes. This is because the UdaanIQ subsystems were trained on datasets that're relevant to the domain. Although the quality of the speech recognition is not very high the error rate decreased from 11.4 percent to 8.2 percent. This makes a difference for the subsequent operations that use natural language processing because errors can affect how well the text makes sense and how relevant it is. The UdaanIQ subsystems met the service level agreement for 95 percent of the responses. This means that the time it took to respond was than 3,000 milliseconds, with an actual time of 2,280 milliseconds. The users were able to complete tasks at a rate of 94.7%. This shows that the complex workflow of the UdaanIQ platform was not difficult for the users to navigate. The subsystems also got a score for user satisfaction. The score was 78.4 and this score is better than the score, in the industry, which is 68. The users were able to use all the features of the UdaanIQ subsystems by themselves after they were shown how to use them.

C. Limitations

There are a number of limitations that reduce the generality of the results obtained. First, the experiment was conducted only at one university using a relatively homogeneous set of participants, and it is yet to be verified whether or not the difference would hold in case of other subjects, experience levels or languages. Second, the facial affect module shows a rather small gap in terms of accuracy among the sub-groups of population; its use in generating the final score requires further research once retraining on a larger data set is performed.

VII. CONCLUSION

Career preparation is fundamentally a problem of feedback. It involves learning what deficiencies one has, giving the possibility to fill in the gap, and putting it to test in circumstances simulating the real conditions of the professional job interview. The currently available solutions only cover part of this cycle individually. Our solution UdaanIQ covers the entire loop of career preparation. The key technological contribution of the system, therefore, is not a separate functionality component, but rather the architecture for integration of all components using an up-to-date Learner Profile. Every system component resume evaluation using BERT, assessment using IRT-adaptive testing, tutoring using RAG, and interview simulation performed significantly better than the baseline version of the technology. Controlled pilot involving 45 participants confirmed that the system provides substantially larger increases in job interview preparedness than comparable standalone tools ($d = 2.05$ vs. 1.09, $p < 0.001$), with a SUS score of 78.4 indicating good usability. The present project also incorporates certain ethical aspects related to the design of AI-based career preparation systems which are frequently ignored. This approach will include features such as the reduction of potential algorithmic biases while assessing candidates, keeping balance in behavioral assessment so as to prevent normalization of responses, and implementing data governance strategy which is consistent with the principles of the GDPR. This way, one can guarantee that the application is going to remain secure, reliable, and unbiased for all users regardless of their background. In terms of future research, it is planned to investigate how this solution works in practice with the help of placement outcome studies performed on a long-term basis. Some enhancements will be made by fine tuning the facial expression model with help of more demographically diverse data samples, and designing institutional dashboards.

REFERENCES

[1] A. Kamargaonkar, S. Kulkarni, R. Patil, and P. Deshmukh, “Automated resume analysis framework using natural language processing," Int. J. Adv. Inf. Technol., vol. 6, no. 2, pp. 45–52, 2024.

[2] A. Nag, S. Das, and P. Banerjee, "Automated interview system using speech recognition and conversational artificial intelligence," in Proc. Int. Conf. Artif. Intell. Mach. Learn. Appl., 2025, pp. 112–118.

[3] S. Shirbhate, A. Kulkarni, and R. Joshi, "Voice-based interview automation using artificial intelligence," Int. J. Compute. Sci. Eng., vol. 13, no. 4, pp. 89–96, 2025.

[4] R. Madanachitran, S. Kumar, and V. Raghavan, "Domain-specific question generation for technical assessment using artificial intelligence," in Proc. Int. Conf. Smart Compute. Technol., 2025, pp. 201–207.
[5] P. Patle, N. Mahajan, and S. Patil, "Facial emotion recognition for interview performance analysis," Int. J. Eng. Res. Technol., vol. 9, no. 6, pp. 310–316, 2024.
[6] S. Kolpe, A. Jadhav, and K. Pawar, "Artificial intelligence-based recruitment platform with resume analysis and interview automation," Int. J. Adv. Res. Compute. Sci., vol. 14, no. 1, pp. 55–62, 2025.
[7] Y. Liang, "Intelligent tutoring framework using large language models for personalized learning," IEEE Trans. Learn. Technol., vol. 15, no. 3, pp. 410–421, 2022.
[8] R. Gupta, A. Verma, and S. Mehta, "Retrieval-augmented generation-based intelligent tutoring architecture," Int. J. Artif. Intell. Educ., vol. 5, no. 2, pp. 67–75, 2024.
[9] Y. Zhang, L. Chen, and H. Wang, "Multimodal interview interaction system using speech and visual feedback," IEEE Access, vol. 11, pp. 98 421–98 432, 2023.
[10] F. H. K. T., D. K. C., H. S. T., and S. R. A. P., "Web-based academic support system for performance tracking and administration," Int. J. Compute. Appl., vol. 176, no. 18, pp. 12–18, 2020.
[11] P. Afsar et al., "Intelligent career guidance and recommendation system," in Proc. Int. Conf. Compute. Inf.,2021, pp. 94–100.
[12] A. Sharma, R. Iyer, and S. Kulkarni, "Machine learning based employability prediction using academic and behavioral parameters," Int. J. Artif. Intell. Appl., vol. 7, no. 2, pp. 23–31, 2025.
[13] M. E. Hoque, R. W. Picard, and M. S. Goodwin, "MACH: My automated conversation coach for public speaking and interview training," in Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Compute., 2018, pp. 697–706.
[14] A. K. Mishra, S. Patel, and R. Verma, "An interview system using artificial intelligence technology," Int. J. Compute. Appl., vol. 182, no. 29, pp. 15–21, 2018.
[15] A. Harry, "Role of AI in education," Interdiscip. J. Humanity, vol. 2, no. 3, pp. 260–268, 2023.