iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: https://doi.org/10.22214/ijraset.2026.85343

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 X Oct 2026- Available at www.ijraset.com

AI Mock Interviewer

Surabhi K S[1], Siva A L[2]
Nehru College of management, India

Abstract: ***Interview performance is influenced by domain knowledge, the ability to organize evidence, communication clarity, and familiarity with interview formats. Access to repeated, individualized practice can be limited by the cost and availability of human interview coaches. This paper proposes an AI Mock Interviewer, a practice-oriented system that simulates role-specific interviews and generates structured, actionable feedback. The framework combines a candidate profile and job-description parser, a competency-based question engine, text or speech input, speech-to-text processing, natural-language analysis, a transparent scoring rubric, and a feedback module. Rather than predicting employability or making hiring decisions, the proposed system is designed to support low-stakes learning. The evaluation plan specifies measures for feedback agreement with trained human raters, scoring reliability, response relevance, usability, accessibility, and fairness across user groups. A weighted rubric is presented as an initial design choice and is explicitly subject to empirical validation. Prior research on automated interview assessment indicates potential benefits in standardization and scale, while also identifying concerns about reliability, validity, bias, disability access, and applicant perceptions. Accordingly, the framework incorporates consent, data minimization, uncertainty communication, user control over recordings, and human review for consequential decisions. Since no implementation data are supplied in this paper, the contribution is a research proposal and system design, not a report of experimentally demonstrated performance.***

Keywords: artificial intelligence; mock interview; natural language processing; large language models; speech recognition; formative assessment; personalized feedback; algorithmic fairness.

I. INTRODUCTION

Job interviews require candidates to communicate knowledge, explain decisions, provide relevant examples, and respond to follow-up questions under time constraints. Conventional mock interviews can provide useful practice, but access may depend on mentors, placement officers, peers, or paid coaching services. A software-based mock interviewer could make repeated practice more available and provide consistent feedback between human coaching sessions.

Recent research has examined the use of language models and machine-learning systems to assess interview responses. Evidence suggests that AI may process responses quickly, but agreement with human ratings, reliability across repeated assessments, and validity for real-world outcomes require careful testing. Studies have also highlighted potential effects on people with disabilities and fairness perceptions in AI-mediated interviews [1-4]. These concerns motivate a design that prioritizes practice and transparent feedback rather than automated selection.

This paper proposes an AI Mock Interviewer that adapts questions to a target role and candidate experience level, analyzes answers using a transparent rubric, and produces specific recommendations for improvement. The system is intended for formative self-assessment and is not proposed as a replacement for professional interviewers or as a hiring decision engine.

A. Research problem

Many candidates lack access to frequent, structured interview practice and may receive feedback that is delayed, inconsistent, or too general to guide improvement. The research problem is to design an accessible AI-assisted practice system that provides repeatable interview simulations and interpretable feedback while controlling risks associated with inaccurate scoring, privacy, and unequal performance across users.

B. Aim and objectives

To design a modular AI-based mock interview system for role-specific practice. To generate competency-aligned questions at adjustable difficulty levels.

To analyze responses for relevance, evidence, structure, domain accuracy, and clarity. To provide personalized feedback linked to observable aspects of the response.

To define an empirical evaluation protocol for reliability, usability, accessibility, and fairness.

©IJRASET: All Rights are Reserved

II. RELATED WORK

Research on automated video interviews has explored machine-learning assessment of verbal and non-verbal signals. Hickman, Tay, and Woo examined reliability, validity, and bias in automated video-interview cognitive assessments, reinforcing the need to test what a model measures rather than assuming that automated scores are inherently valid [1]. Research on language-model-based interview assessment has likewise reported uneven performance across traits, test-retest concerns, and potential bias, recommending caution in high-stakes applications [2].

Fisher, Bonaccio, and Connelly discuss how AI-based selection tools may create both opportunities and barriers for applicants with disabilities, making accessibility and applicant experience core design requirements [3]. Szanda■a compares AI and human expertise in IT recruitment and reports that AI assessment can be faster while still showing consistency limitations, illustrating why human-rater comparisons should be part of validation [4]. These findings inform the proposed practice-first approach: the system should expose its rubric, provide evidence-linked feedback, allow user correction, and avoid making hiring decisions.

The proposed contribution is a modular, low-stakes practice framework that combines adaptive question selection, multimodal response capture, rubric-grounded evaluation, and a defined validation protocol. The contribution should be interpreted as a design proposal until implemented and tested.

III. PROPOSED SYSTEM

A. *System overview*

The candidate selects a target role, experience level, interview type, preferred response mode, and optional job description. The system builds a competency profile and selects questions from a curated question bank or generates draft questions that are checked against the role profile. The candidate responds through text or speech. Speech responses are transcribed, and the response is analyzed against a question-specific rubric. The system returns criterion-level feedback, examples of stronger response structure, and a recommended next practice activity.

Figure 1. Proposed system architecture for the AI Mock Interviewer.

B. *Functional modules*

C. *End-to-end workflow*

Figure 2. Workflow from candidate setup to feedback and repeated practice.

IV. METHODOLOGY

A. *Research design*

A design-science approach is proposed. First, requirements are derived from candidate and placement-coach needs. Second, a prototype is developed using modular services. Third, its outputs are compared with human ratings and user feedback. Finally, the system is revised based on error analysis, usability findings, and fairness audits. The implementation should be documented sufficiently for replication, including model/version information, prompts, rubric versions, transcription settings, and evaluation data.

B. *Question generation and selection*

Questions should be mapped to job-relevant competencies such as technical knowledge, problem solving, collaboration, communication, and situational judgment. The system may select questions from a reviewed question bank and optionally use a language model to draft role-specific variants. Generated questions should pass checks for relevance, clarity, difficulty, duplication, and inappropriate personal content. Adaptive selection may increase or decrease difficulty based on the candidate's prior answers, but the selection rule should be recorded and auditable.

C. *Response processing*

Text answers can be analyzed directly. For speech responses, an automatic speech-recognition component produces a transcript. Users should be able to correct transcription errors before evaluation because transcription quality can affect downstream scores. Audio-derived measures, if included, should be limited to task-relevant communication features and validated for accessibility and language variation. Facial-expression, gaze, accent, emotion, or personality inference should not be treated as reliable proxies for competence without strong independent evidence.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Rubric-based answer assessment

The proposed system uses a criterion-based rubric rather than a single opaque prediction. The illustrative weights below are design hypotheses, not validated constants. Each criterion is scored on a 0-4 scale: 0 = absent or incorrect; 1 = limited; 2 = partially adequate; 3 = strong; 4 = excellent. For technical questions, domain accuracy should be assessed against reviewed reference material. For behavioral questions, the system may assess whether the answer gives a clear context, describes the candidate's actions, and explains an outcome; the STAR structure is a coaching aid, not a mandatory format for every response.

Figure 3. Illustrative scoring criteria and weights for formative feedback.

A weighted score can be calculated as: Overall Score = □(w■ × s■ / 4), where w■ is the criterion weight expressed as a percentage and s■ is the criterion score from 0 to 4. The resulting score ranges from 0 to 100. The system should also display criterion-level evidence and uncertainty, rather than presenting the aggregate score as an objective measure of a candidate's worth.

E. Feedback generation

Feedback should contain (i) two or three specific strengths, (ii) missing or weak elements linked to the rubric, (iii) a short example of how the response could be organized, and (iv) one or two next actions. The system should distinguish between content correctness and communication style. It should avoid inventing facts about the candidate, penalizing disability-related communication differences, or recommending that users imitate a single accent or personality.

V. PROPOSED EVALUATION PLAN

Because no prototype or dataset was provided, this section defines a study to conduct rather than reporting results. A pilot study should be completed before any claims are made about effectiveness.

F. Participants and procedure

Recruit a diverse sample of students or job seekers who consent to participate. A sample size should be determined through an a priori power analysis based on the primary outcome and anticipated effect size. Participants complete a baseline mock interview, use the system for a defined practice period, and complete apost-practice interview with comparable difficulty. A subset of responses should be independently rated by at least two trained evaluators who are blinded to the AI scores. If feasible, use a counterbalanced or comparison-group design to separate system effects from repeated-practice effects.

G. Evaluation metrics

H. Analysis plan

Report descriptive statistics, confidence intervals, missing-data handling, and pre-specified primary outcomes. Compare AI scores with human ratings using agreement statistics appropriate to the data and design. Use paired or mixed-effects models for repeated measurements where assumptions are satisfied. Report effect sizes and uncertainty rather than relying solely on p-values. Investigate disagreement cases qualitatively to identify whether errors arise from transcription, question ambiguity, domain knowledge, rubric design, or model generation. Any subgroup analysis should be planned in advance, adequately powered, privacy-preserving, and interpreted cautiously.

VI. EXPECTED CONTRIBUTIONS

The expected contribution is a reusable design blueprint for a practice-oriented AI interview system. The framework combines role-specific question selection, optional speech interaction, rubric-based analysis, transparent feedback, and progress tracking. If validated, it may support repeated low-cost practice and help users identify specific areas for improvement. These are anticipated benefits and must be tested empirically; they are not findings established by this paper.

VII. ETHICAL, PRIVACY, AND ACCESSIBILITY CONSIDERATIONS

Informed consent: explain what is recorded, how it is processed, and whether external model providers receive data.

Data minimization: avoid collecting sensitive personal data unless strictly necessary; do not retain raw audio or video by default.

User control: provide transcript correction, deletion, export, and a text-only alternative.

Accessibility: support keyboard navigation, captions, screen readers, adjustable time limits, and alternatives to speech-based interaction.

Bias monitoring: evaluate errors across language backgrounds and other relevant groups where ethically and legally appropriate; do not infer sensitive traits.
Transparency: display rubric criteria, limitations, model version, and uncertainty; identify AI-generated suggestions as suggestions.
Non-consequential use: do not use practice scores as automatic hiring, admission, or promotion decisions without separate validation, legal review, and meaningful human oversight.
Security: encrypt data in transit and at rest, restrict access, set retention limits, and document incident response.

VIII. LIMITATIONS

The proposed design has not been implemented or empirically evaluated in this manuscript. Large language models can produce plausible but incorrect feedback, and semantic similarity does not guarantee factual correctness. Speech transcription may perform unevenly across accents, languages, speaking conditions, and assistive speech patterns. A rubric may omit context-specific competencies or over-reward a preferred communication style. Results from a limited pilot would not automatically generalize to other occupations, languages, institutions, or user groups. These limitations require iterative validation and transparent reporting.

IX. CONCLUSION

This paper presents a proposed AI Mock Interviewer framework for accessible, personalized interview practice. Its core components are role-aware question selection, text or speech response capture, response analysis, transparent rubric scoring, tailored feedback, and progress tracking. The paper emphasizes that formative coaching is a more defensible initial use than automated hiring decisions. Future work should implement a prototype, evaluate agreement with trained human raters, measure learning outcomes and usability, and conduct systematic accessibility, privacy, and fairness assessments. Claims of effectiveness should be made only after these studies are completed and their results reported.

REFERENCES

[1] Hickman, L., Tay, L., & Woo, S. E. (2025). Are automated video interviews smart enough? Behavioral modes, reliability, validity, and bias of machine learning cognitive ability assessments. Journal of Applied Psychology, 110(3), 314-335. https://doi.org/10.1037/apl0001236
[2] Li, et al. (2024). Can Large Language Models Assess Personality From Asynchronous Video Interviews? A Comprehensive Evaluation of Validity, Reliability, Fairness, and Rating Patterns. IEEE Transactions on Affective Computing, 15(3), 1769-1785. https://doi.org/10.1109/TAFFC.2024.3374875
[3] Fisher, S. L., Bonaccio, S., & Connelly, C. E. (2024). AI-based tools in selection: Considering the impact on applicants with disabilities. Organizational Dynamics, 53(1), 101036. https://doi.org/10.1016/j.orgdyn.2024.101036
[4] Szanda■a, T. (2025). ChatGPT vs human expertise in the context of IT recruitment. Expert Systems with Applications, 264, 125868. https://doi.org/10.1016/j.eswa.2024.125868
[5] Biswas, S., Jung, J.-Y., Unnam, A., Yadav, G., Gupta, S., & Gadiraju, U. (2024). "Hi. I'm Molly, Your Virtual Interviewer!" Exploring the Impact of Race and Gender in AI-Powered Virtual Interview Experiences. Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 12(1), 12-22. https://doi.org/10.1609/hcomp.v12i1.31596
[6] Poe, R. L., & El Mestari, S. Z. (2024). The Conflict Between Algorithmic Fairness and Non-Discrimination: An Analysis of Fair Automated Hiring. Proceedings of the ACM conference. https://doi.org/10.1145/3630106.3659015

Appendix A. Example Interview Session

Target role: Graduate Software Engineer (example only)
Question: Describe a technical problem you solved, the steps you took, and the outcome. Illustrative feedback template:
Relevance: Did the answer address the question directly? Evidence: Was the candidate's own contribution identifiable? Reasoning: Were the steps and trade-offs explained?
Outcome: Was the result stated and, where possible, quantified?
Next step: Re-answer in 60-90 seconds with clearer context, action, and outcome.
This example is a template only and does not represent an evaluated participant response or a measured system result.

Appendix B. Suggested Prototype Technology Stack

A possible implementation could use a browser-based interface, a Python API service, a relational database, a speech-to-text service for opt-in audio responses, and a language model for question drafting and response feedback. The stack should be selected based on institutional requirements, budget, data residency, and privacy controls. Store rubric versions and model metadata alongside each evaluation to support auditability. This is an implementation option, not a claim that a prototype has already been built.

©IJRASET: All Rights are Reserved

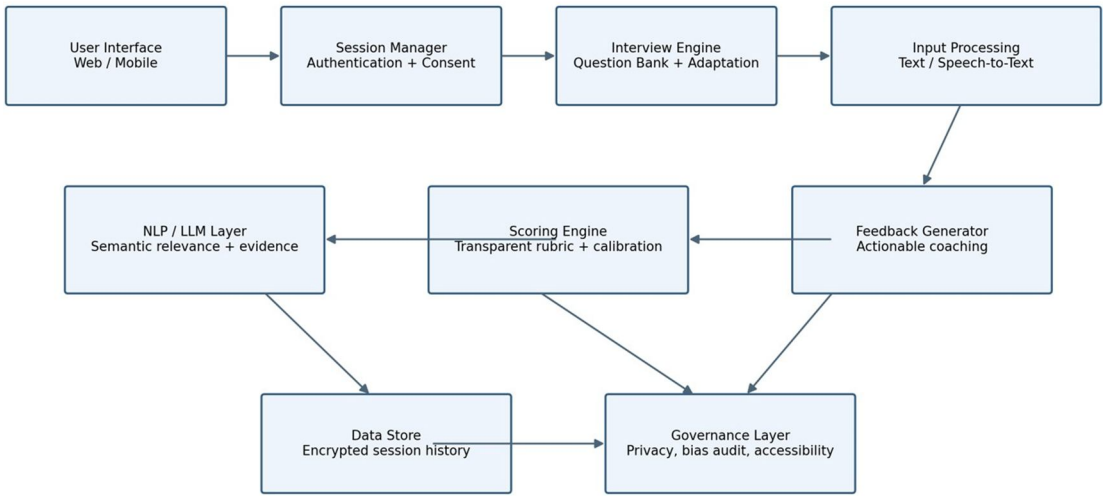
Module	Function	Design consideration
Candidate interface	Collect role, experience, preferences, consent, and responses.	Keyboard accessibility, captions, clear controls.
Profile and job-description parser	Extract skills, responsibilities, and target competencies.	Show extracted items for user confirmation.
Module	Function	Design consideration
Question engine	Select or generate questions by competency and difficulty.	Prevent irrelevant, discriminatory, or overly personal questions.
Speech processing	Convert audio to text where the user opts into voice mode.	Display transcript and allow correction.
Answer analysis	Assess relevance, evidence, structure, and domain-specific content.	Use question-specific criteria; avoid unsupported personality inference.
Scoring engine	Combine criterion scores into an overall formative score.	Expose weights and uncertainty; calibrate against human raters.
Feedback generator	Give strengths, gaps, examples, and next steps.	Ground feedback in response evidence; avoid generic praise.
Progress dashboard	Compare practice sessions over time.	Show trends without presenting scores as employability predictions.
Governance and privacy	Manage consent, retention, access, and audits.	Minimize data; allow deletion and opt-out.

Dimension	Suggested metric	Purpose
Human-AI agreement	Intraclass correlation (ICC) for suitable continuous scores; weighted kappa for ordinal criteria.	Measure agreement, not merely correlation.
Scoring consistency	Test-retest reliability and inter-rater reliability.	Check stability across equivalent answers and raters.
Response analysis	Criterion-level precision/recall or expert-rated relevance where applicable.	Identify systematic content-analysis errors.
Learning outcome	Pre/post rubric score; report effect size and confidence interval.	Assess whether practice is associated with improvement.
Usability	System Usability Scale (SUS) or validated equivalent; task completion and dropout.	Assess ease of use and friction.
Feedback quality	Blinded expert rating of specificity, correctness, and actionability.	Ensure recommendations are useful and grounded.
Fairness and accessibility	Error rates and feedback-quality comparisons across appropriately consented groups; accessibility testing.	Identify disparities and barriers.
Privacy and safety	Consent completion, deletion success, incident logs, retention compliance.	Verify responsible data handling.

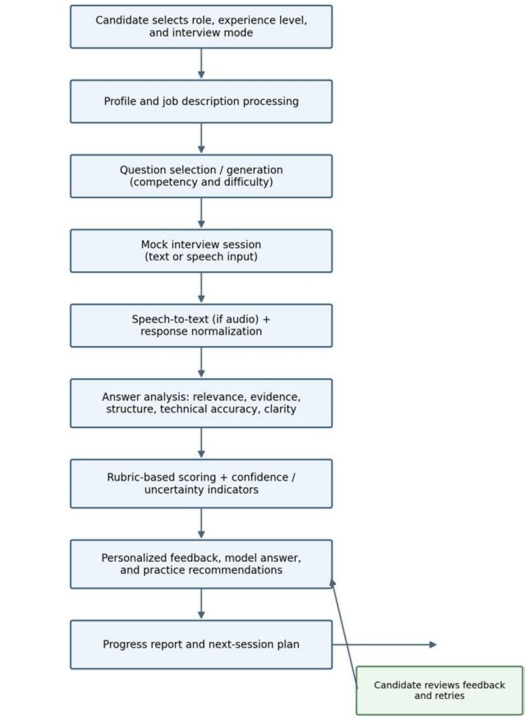
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Figures and Diagrams

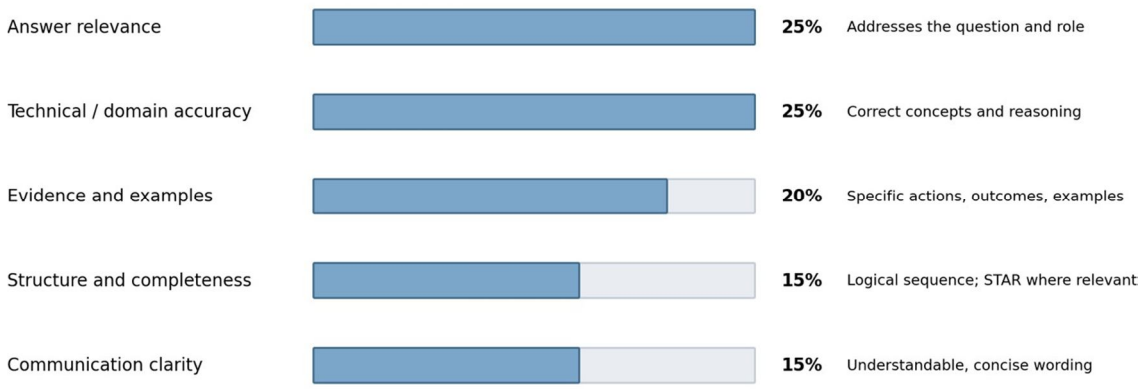
Proposed System Architecture



AI Mock Interviewer: End-to-End Workflow

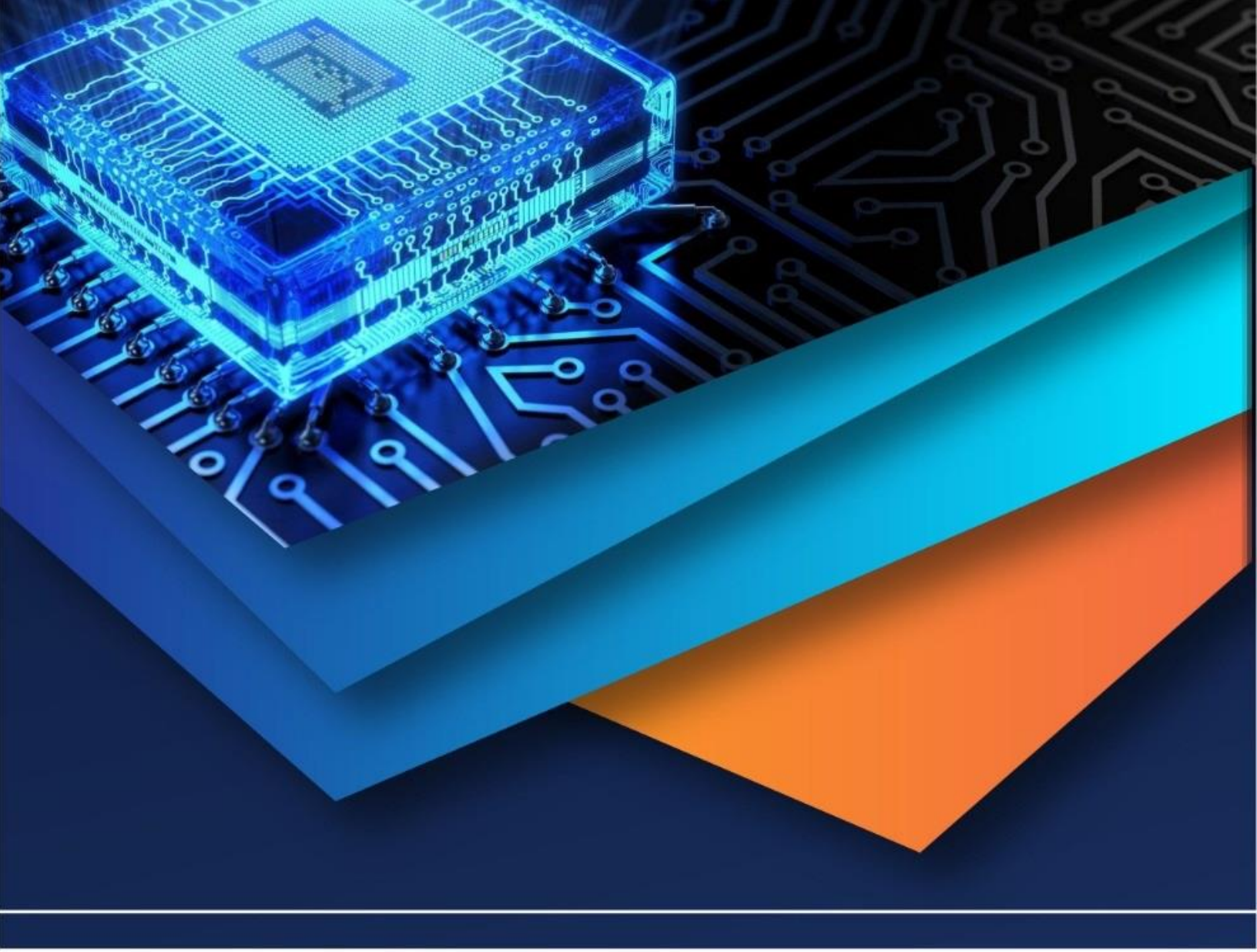


Illustrative Interview Feedback Rubric



Weights are a proposed starting point; validate and adapt them for each role and rubric.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)