Faculty of Computers & Artificial Intelligence
Helwan University
AI Department

VIRTUAL AI INTERVIEWER

Supervisor : Dr. Soha Ahmed Ehssan

CONTENT

1. INTRODUCTION

WHAT'S WRONG WITH TRADITIONAL INTERVIEWS?

01 We found over 60% of interviews show signs of bias , often leading to unfair hiring and increased candidate stress.

02 Limited access to realistic practice and feedback,

03 A 2022 study by SHRM and Glassdoor found that over 65% of recruiters struggle to fairly evaluate large applicant pools due to high volume and time constraints.

WHAT ARE WE TRYING TO PROVE?

01 No stress anymore

02 AI scoring can match or exceed human fairness.

03 Provide candidates with detailed, personalized reports.

04 Automation reduces recruiter workload and improves outcomes.

2. LITERATURE REVIEW

WHAT DOES THE LITERATURE SAY?

01 Most rely on text/audio only

A 2024 IEEE Access review on AI recruitment tools found that 76% of systems use only text and voice input, with no video or visual emotion integration.

02 limited feedback and no real-time adaptivity

A 2024 Springer AI & Ethics study reported that 68% of AI interview platforms provide static question sets and delayed or generic feedback

03 They lack multimodal emotion analysis

According to a 2024 ACM Transactions on HCI paper, only 12% of evaluated platforms support multimodal emotion detection

HOW DOES OUR SYSTEM PUSH BEYOND EXISTING WORK?

01

MULTIMODAL INPUT

Voice, face, and text fusion for deeper analysis

02

SMART QUESTIONS

RAG + LLaMA for role-based, adaptive interviewing

03

FAIR SCORINGS

Judgment agent (Mistral) with explainable feedback

04

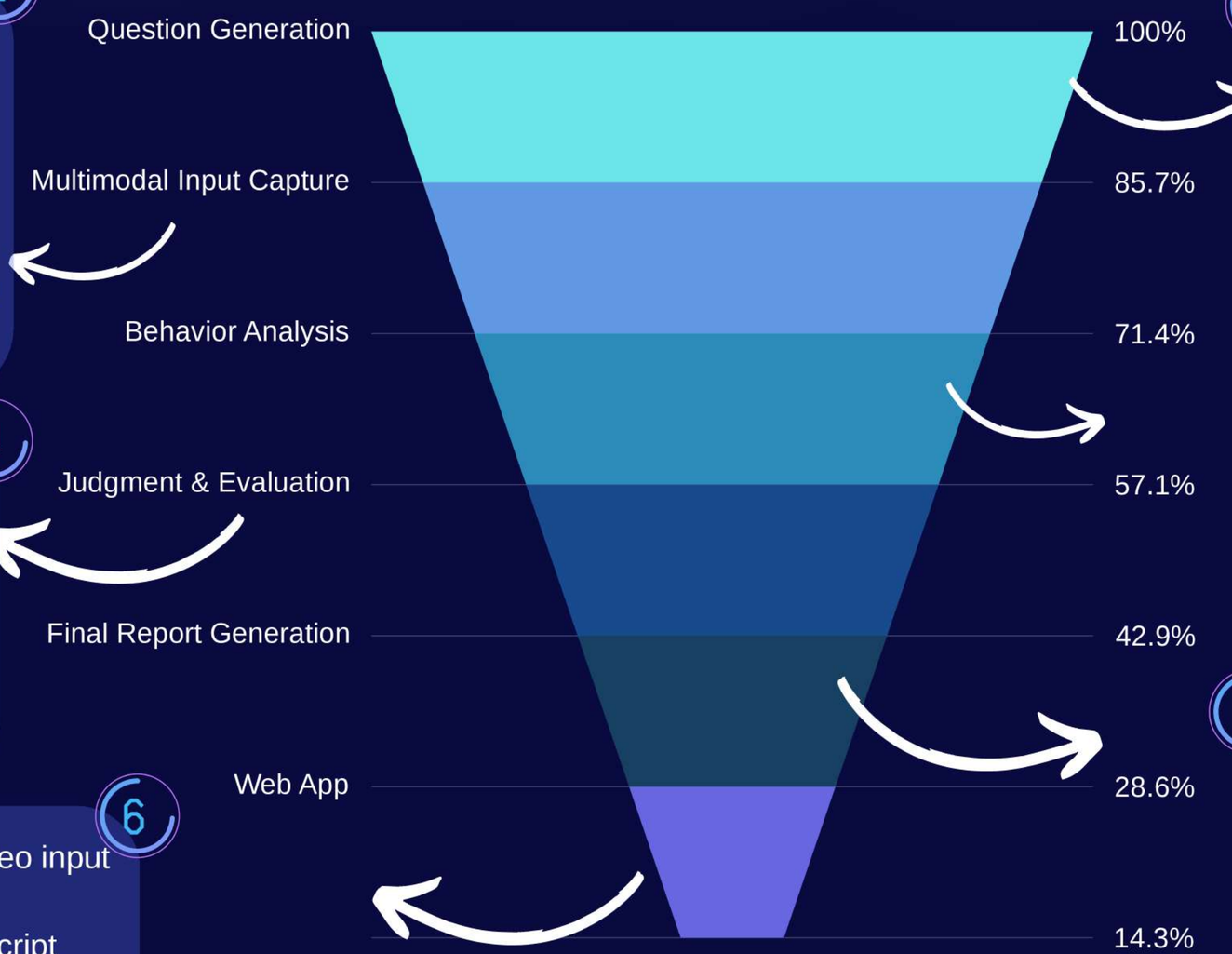
INSIGHTFUL REPORTING

Generates detailed reports with scores, emotions & a clear hire/reject verdict

3. SYSTEM ARCHITCTURE

3.SYSTEM ARCHITECTURE

AI INTERVIEWER ARCHITECTURE - END-TO-END PIPELINE



1
- Built on LLaMA 3.3 + RAG
- Tailors questions to CV + job role
- Context-aware, diverse, and grounded

2
- Voice → Processed via Whisper (STT)
- Interviewer Speech via Bark (TTS)
- Facial Video → Analyzed using DeepFace
- Emotion Detection (Wav2Vec 2.0 + DeepFace)

3
- Classifies 7 emotional states (voice + face)
- Emotion trends logged for scoring & feedback

4
- Ideal answers via LLaMA 3.3
- Scoring with Mistral 7B – relevance, clarity, completeness
- Transparent feedback: score + reasoning + improvement tips

5
- Auto-generated by LLaMA 3.3
- Includes: scores, emotional insights, strengths/weaknesses, hire/reject verdict

6
- Captures audio/video input from the user
- Displays live transcript and emotion feedback.

4. SYSTEM METHODOLOGY

1. KEY AI MODELS & THEIR ROLES

LLAMA 3.3 (70B)

01

- Generates context-aware, domain-specific interview questions
- Also used to create high-quality reference answers for evaluation

BARK (TTS)

02

- Evaluates candidate answers based on relevance, clarity, and accuracy
- Produces scores + natural language feedback for fairness and explainability

WHISPER (BASE)

03

- Converts spoken answers into accurate transcripts in real time
- Handles varied accents, informal tone, and background noise

1. KEY AI MODELS & THEIR ROLES

04

WAV2VEC 2.0

- ➤ Detects emotions in voice (e.g., stress, confidence, calmness)
- ➤ Fine-tuned on labeled datasets to recognize real-time vocal cues

05

DEEPFACE

- ➤ Analyzes facial expressions from video to classify emotional states.
- ➤ Processes sampled frames for performance efficiency and emotion trends.

06

MISTRAL-7B-INSTRUCT

- ➤ Evaluates candidate answers based on relevance, clarity, and accuracy
- ➤ Produces scores + natural language feedback for fairness and explainability

2.PREPROCESSING

AUDIO

- Resampled to standard frequencies (16kHz for Whisper, 22kHz for wav2vec)
- MFCCs (Mel-frequency cepstral coefficients) extracted for emotion features.

VIDEO

DeepFace runs on every 15th frame from webcam video for efficient facial emotion detection.

TEXT & RETRIEVAL

Semantic search using embeddings retrieves the most relevant Q&A pairs for LLaMA to ground its question generation

3.TRAINING DATA

WAV2VEC 2.0

Was fine-tuned on emotional speech datasets

- RAVDESS
- TESS
- EMO-DB
- CREMA-D

&

RAG

Job-role-specific Q&A pairs Embedded and indexed with Qdrant to enable accurate retrieval in RAG-based question generation.

- Total Questions: **3,809**
- Unique Job Roles: **26**

5. IMPLEMENTATION
WHAT HAPPENS BEHIND THE SCENES ?!

LOCAL DEPLOYMENT & DEVELOPMENT ENVIRONMENT

DEVELOPED IN PYTHON, OPTIMIZED FOR LOCAL MACHINES WITH CONSUMER-GRADE CPU/GPU.

VS CODE

development and writing code

KAGGLE / GOOGLE COLAB

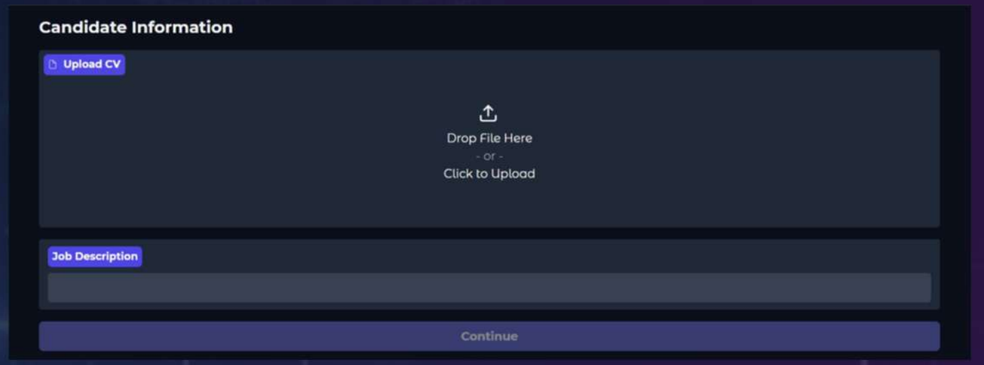
for model training and experimentation

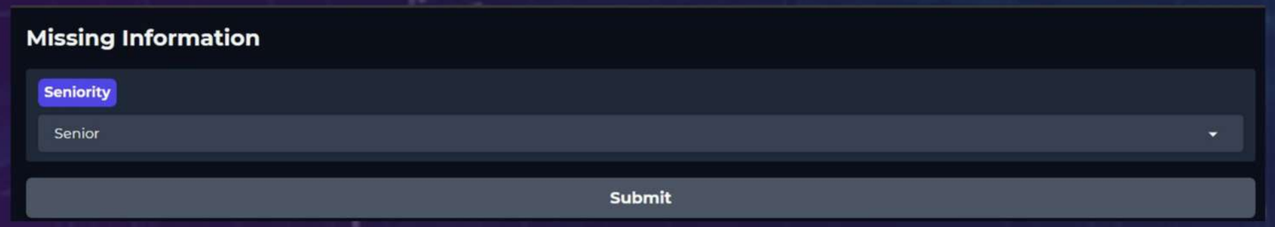
GITHUB

for version control and team collaboration

IMPLEMENTATION FLOW

THE USER ENTERS HIS/HER INFORMATION:





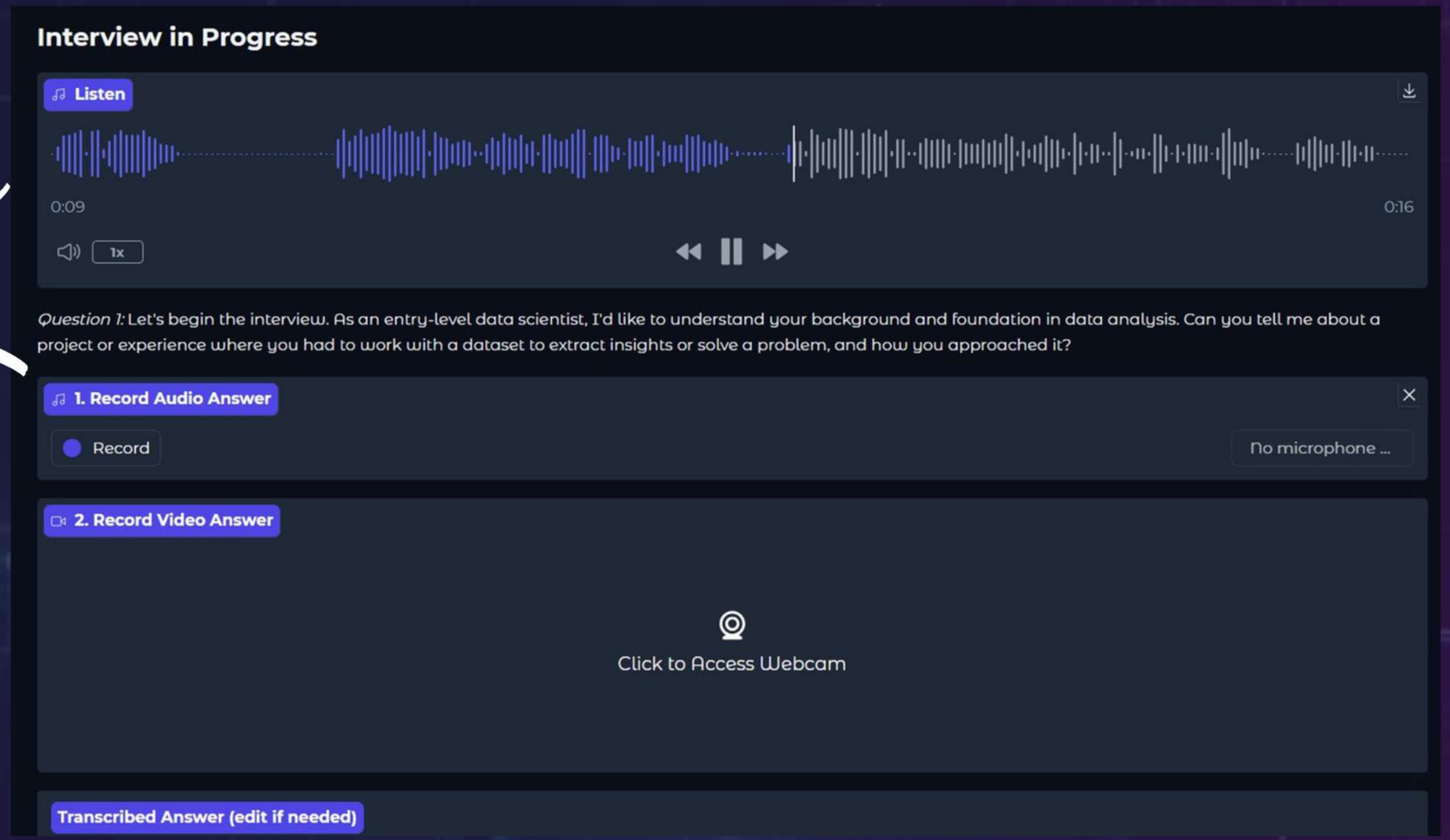
IMPLEMENTATION FLOW

INTERVIEW_LOOP() DRIVES THE SESSION:

LETS SHOW HOW IMPLEMENTATION IS DONE
IN THE FOLOWING STEPS >

IMPLEMENTATION FLOW

3. Uses Bark to vocalize it → the generated text question is passed to Bark, which transforms it into expressive audio tokens, then decodes them into natural-sounding speech using EnCodec.

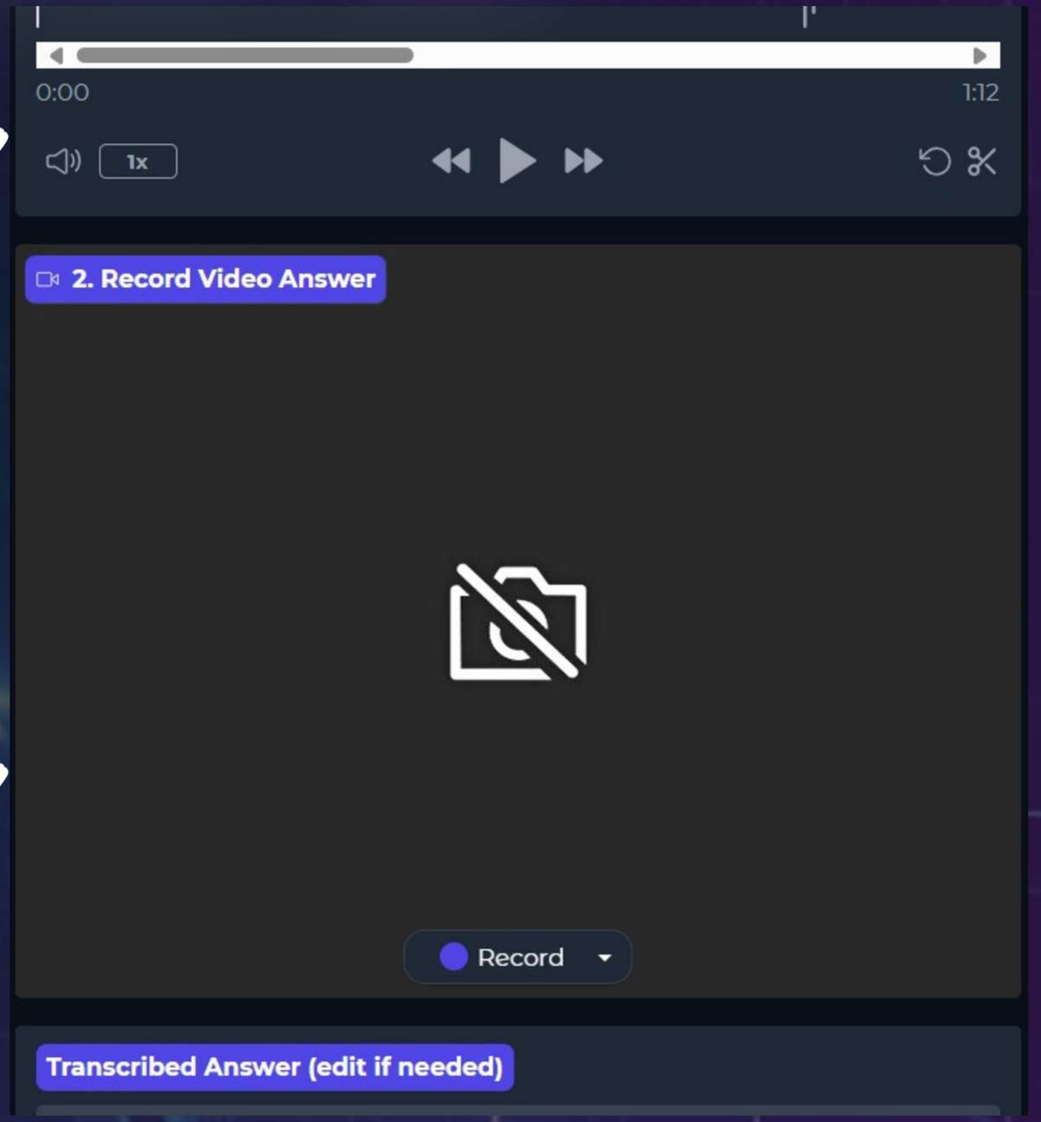


1. Retrieves contextual data from Qdrant
2. Builds a prompt for LLaMA to generate a tailored question

IMPLEMENTATION FLOW

4. Waits for verbal user response → transcribed using Whisper → audio is captured via microphone, resampled to 16kHz using Torchaudio, and processed by the HuggingFace.

5. Records Video Answer → The user's webcam is accessed via OpenCV, capturing frames in real-time. Every 15th frame is passed to DeepFace, which analyzes the facial expression and extracts the dominant emotion.



WHAT HAPPENS NEXT ?

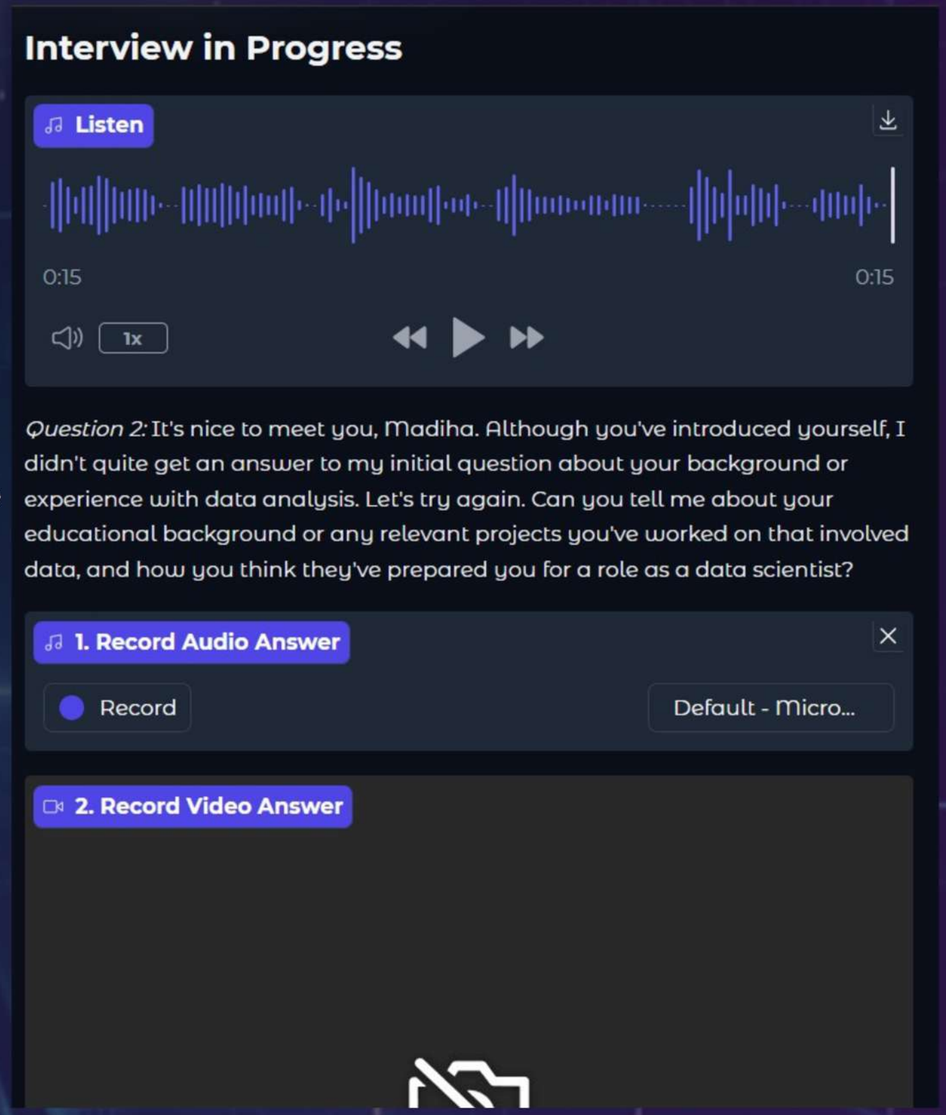
6.EVALUATE THE ANSWER

- The user's response (text from Whisper + emotion from video/audio) is passed to Mistral-7B-Instruct.
- It scores the answer on 5 dimensions: Relevance, Correctness, Clarity, Completeness, and Emotion.
- The system also calculates an effective confidence score using:

```
0.5 * answer_score + 0.22 * voice_emotion + 0.18 * facial_emotion + 0.1 * control_bonus
```

7.GENERATE FOLLOW-UP QUESTION (E.G., Q2)

- Based on the score, It then builds a fresh prompt (with feedback + context) and asks a new tailored question using LLaMA 3.3.
- This question is again converted to speech using Bark, and the loop continues.



8. SMART INTERVIEW SUMMARY - FULLY AUTO-GENERATED

- Every Q&A pair is logged with the user's exact answer.

LLM EVALUATION:

- Each response is scored by Mistral-7B, including:
- A qualitative score (Excellent, Good, Medium, Poor)
- Reasoning behind the score
- Concrete improvement suggestions

EMOTION DATA IS INTEGRATED:

- facial & vocal tones are recorded per answer.

EFFECTIVE CONFIDENCE SCORE

- is calculated and logged.

Interview Summary

Q1: Let's begin the interview. As an entry-level data scientist, I'd like to understand your background and foundation in data analysis. Can you tell me about a project or experience where you had to work with a dataset to extract insights or solve a problem, and how you approached it?

- *Answer*: hey, my name is madiha
- *Q Eval*: {'Score': 'Good', 'Reasoning': 'This question is somewhat relevant as it asks for a practical example of data analysis, which aligns with the role of an entry-level data scientist. However, it could benefit from more specificity regarding the type of datasets, problems, or techniques involved.', 'Improvements': ['- Specify the types of datasets (structured/unstructured) and their sources', '- Provide examples of specific problems or business cases', '- Request details on the analytical methods used (e.g., statistical tests, machine learning models)']}
- *A Eval*: {'Score': 'Medium', 'Reasoning': 'The answer provides some context and mentions a project, but lacks detail in describing the approach taken, the specific insights extracted, or the methods used.', 'Improvements': ['Describe the specific steps taken in the analysis process', 'Provide more details about the insights extracted and their significance', 'Explain the specific methods used for data cleaning, exploration, and modeling']}
- *Face Emotion: no_face, **Voice Emotion*: sad
- *Effective Confidence*: 0.429
- *Time*: 245.33s

Q2: It's nice to meet you, Madiha. Although you've introduced yourself, I didn't quite get an answer to my initial question about your background or experience with data analysis. Let's try again. Can you tell me about your educational background or any relevant projects you've worked on that involved data, and how you think they've prepared you for a role as a data scientist?

6.EVALUATION AND RESULTS

READY TO EVALUATE?
LET'S BREAK IT DOWN

EACH MODEL WAS EVALUATED BASED ON ITS CONTRIBUTION TO REALISM, ADAPTIVITY, AND HUMAN-LIKENESS, NOT ONLY TECHNICAL ACCURACY.

HOW DO WE KNOW EACH MODEL PERFORMS WELL?

Model	Function	Evaluation Method	Key Insights
LLaMA 3.3	**Question + Reference Answer Generation**	**Manual quality checks + GPT-4-based rubric (Q relevance, specificity)**	**Strong role-specific adaptation; high diversity & clarity**
Mistral-7B	Answer Scoring + Feedback	Scored responses on 5 dimensions; GPT-4 behavioral realism review	Consistent scoring; slight generosity; improvement suggestions
Whisper	Speech-to-Text Transcription	Compared against ground truth and manual edits	Accurate with noisy input; editable for fairness
Bark	Text-to-Speech Synthesis	Subjective human review of audio output + expressiveness	Natural-sounding, emotion-rich speech with prosody
wav2vec 2.0	Voice Emotion Detection	Accuracy measured on RAVDESS, TESS, EMO-DB (85%+)	Robust across 7 emotions; strong generalization
DeepFace	Facial Emotion Detection	Qualitative testing under lighting, camera angles	Reliable dominant emotion detection; consistent session trends

HOW DID THE SYSTEM ACTUALLY PERFORM?

KEY EVALUATION RESULTS (5-POINT SCALE) USING A GPT-4 RUBRIC:

Category	Score	Insight
Question Relevance	4.9	Strong alignment with job roles
Contextual Responsiveness	4.7	Adaptive follow-ups, slight transition rigidity
Scoring Consistency	4.8	Fair and well-reasoned, though slightly generous
Language Coherence	4.6	Formal and clear, lacks conversational warmth
Behavioral Realism	4.4	Professional tone, needs emotional flexibility

Our Team

MOHAMMAD TAREK

MUHAMMAD YASSER

FARIDA KHALED

HANIA RUBY

ESRAA MOHAMED

MADIHA SAIED


THANKS FOR YOUR TIME