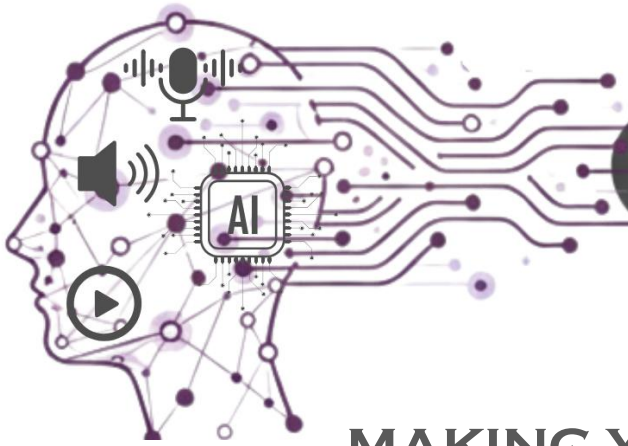


CogniHire

MAKING YOUR HIRING PROCESS SIMPLER

AI INTERVIEWER & INSIGHT GENERATOR

- USING AI TO BUILD A FAIR, FAST, AND INSIGHTFUL HIRING PROCESS.

- Project by:

Aneesh Varla
Anugya Anantapur
Bhuvan C Naik
Juhitha Radha Vegesna

THE BROKEN HIRING PROCESS

It's Too Expensive ₹2,000 to ₹8,000 per hire, or even higher for specialized roles This doesn't even include "soft costs" like the engineering hours lost when your senior devs have to stop coding to conduct interviews.	**It's Too Slow** **44 Days** = Average time to fill a position The best candidates are off the market in 10 days. A slow process means you lose top talent to faster competitors.
It's Biased **75% Rejected** - Resumes rejected by ATS robots or unconscious bias before a human conversation happens. "Affinity Bias" means interviewers unconsciously prefer candidates who are "like them," hurting diversity and fairness.	**It's Inconsistent** **Subjective** - "Gut feeling" vs. Data. Without standardized AI scoring, Candidate A might get easy questions while Candidate B gets hard ones. There is no fair baseline for comparison.

INTERVIEW STRUCTURE

Role based Round

- There are **40+ job roles** supported in the system.
- You will receive **5 carefully curated questions** based on the **job role you selected**.
- These questions come from our internal, cleaned question bank—built from real industry interviews.

Resume based Round

- The system scans your resume for **skills, projects, tools, and achievements**.
- You will receive **3–4 highly personalized questions** derived from your actual resume.
- Follow-up questions are generated **only if your answer is incomplete**, mimicking a rearl interviewer.

The interview is time-bounded, with a maximum duration of 30 minutes.

PART 1 : AUDIO ANALYSIS

Objectively Measuring Candidate Confidence and Communication

DUAL-PATH PROCESSING

The audio stream is analysed in parallel to separate content from delivery:

Content Path (Whisper): High-fidelity transcription (Speech-to-Text) for semantic analysis by the LLM. This evaluates *what* the candidate said.

Signal Path (Acoustic): The raw waveform is analyzed for prosodic features. This evaluates *how* the candidate said it.

Acoustic Feature Extraction

Pacing & Fluency Analysis

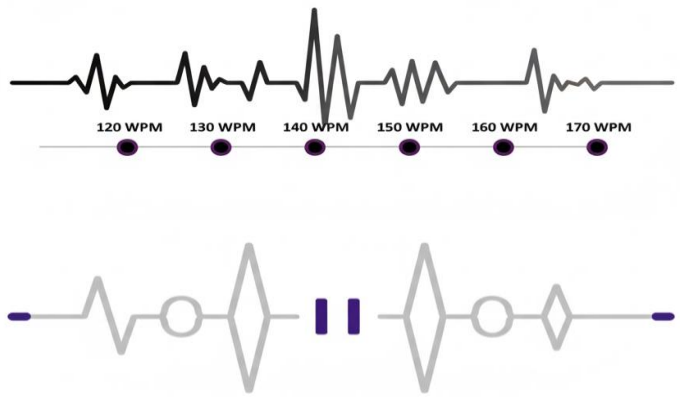
This category measures the *flow* of speech. The goal is to determine if the candidate's speech pattern is confident and natural, or if it's symptomatic of cognitive stress.

Speech Rate (WPM): This is the primary metric for fluency.

Technical Detail: It is calculated as (Total Words/Speaking Time)×60. "Speaking Time" is crucial—it's the total duration *minus* significant silent pauses.

Implication: There is a "confidence band" for WPM (e.g., 140-160 WPM).

- **Too Slow (Low WPM):** Implies high **cognitive load**. The candidate is struggling to formulate the thought, recall facts, or "invent" an answer.
- **Too Fast (High WPM):** Often a sign of **nervousness** or **rote memorization**. The candidate may be rushing through a pre-rehearsed answer, which often sounds unnatural and lacks genuine comprehension.



HESITATION & DISFLUENCY ANALYSIS

This category directly measures cognitive load by tracking disruptions in the speech flow. These disruptions are the strongest indicators that a candidate is stuck.

Pause Frequency & Duration: We analyse silence in the audio.

- **Technical Detail:** We differentiate between *natural pauses* (e.g., at a comma, <0.5s) and *long pauses*(>1.5s).
- **Implication:** A high frequency of long pauses indicates difficulty. The candidate is stopping to think, which, for a technical question, can mean they don't know the answer immediately

Filler Word Count: These are non-lexical "disfluencies."

- **Technical Detail:** The system is trained to detect and count common filler sounds (e.g., "um," "uh," "ahh") and filler words ("like," "so," "you know").
- **Implication:** Fillers are a classic symptom of the brain "buffering"—it's trying to hold its place in the conversation while searching for the next piece of information. A high count directly correlates with higher cognitive load and lower preparation.

PROSODY & ENGAGEMENT ANALYSIS

Prosody is the "music" of speech—the rhythm, stress, and intonation. It is a powerful signal for the candidate's *emotional state* (confidence, engagement, or boredom).

Pitch Variance (Fundamental Frequency, F0):

- **Technical Detail:** F0 is the base frequency of the speaker's voice. We measure its standard deviation (variance) over time.
- **Implication:**
  - **Low Variance (Monotone):** A flat, robotic voice. This signals **disengagement**, **boredom**, or a complete lack of confidence.
  - **High Variance (Dynamic):** A voice with varied pitch. This signals **engagement**, **enthusiasm**, and **confidence**. The speaker is using intonation to emphasize key points, which is a natural behaviour when you are comfortable with the material.

HEURISTIC SCORING MODEL: MAPPING FEATURES TO A SCORE

Our model synthesizes all acoustic features into one final **Communication Score (Scomm)**.
This score is a weighted function of three key areas:

Fluency Score (Sflu)

Purpose: To reward a natural, confident pace and penalize extremes.

Model: A **Gaussian (bell curve) function** based on **Words Per Minute (WPM)**

Rationale: We penalize speaking too slowly (implies hesitation) and too quickly (implies nervousness or a rehearsed script). The model rewards the "sweet spot" of professional speech.

Clarity Score (Sclr)

Purpose: To quantify speech disruptions caused by cognitive load.

Model: An **inverse penalty function** based on **Pause Frequency** and **Filler Word Count**

Rationale: The score starts at 100 (perfect clarity) and is algorithmically reduced for every "um," "uh," or long, unnatural pause

Engagement Score (Seng)

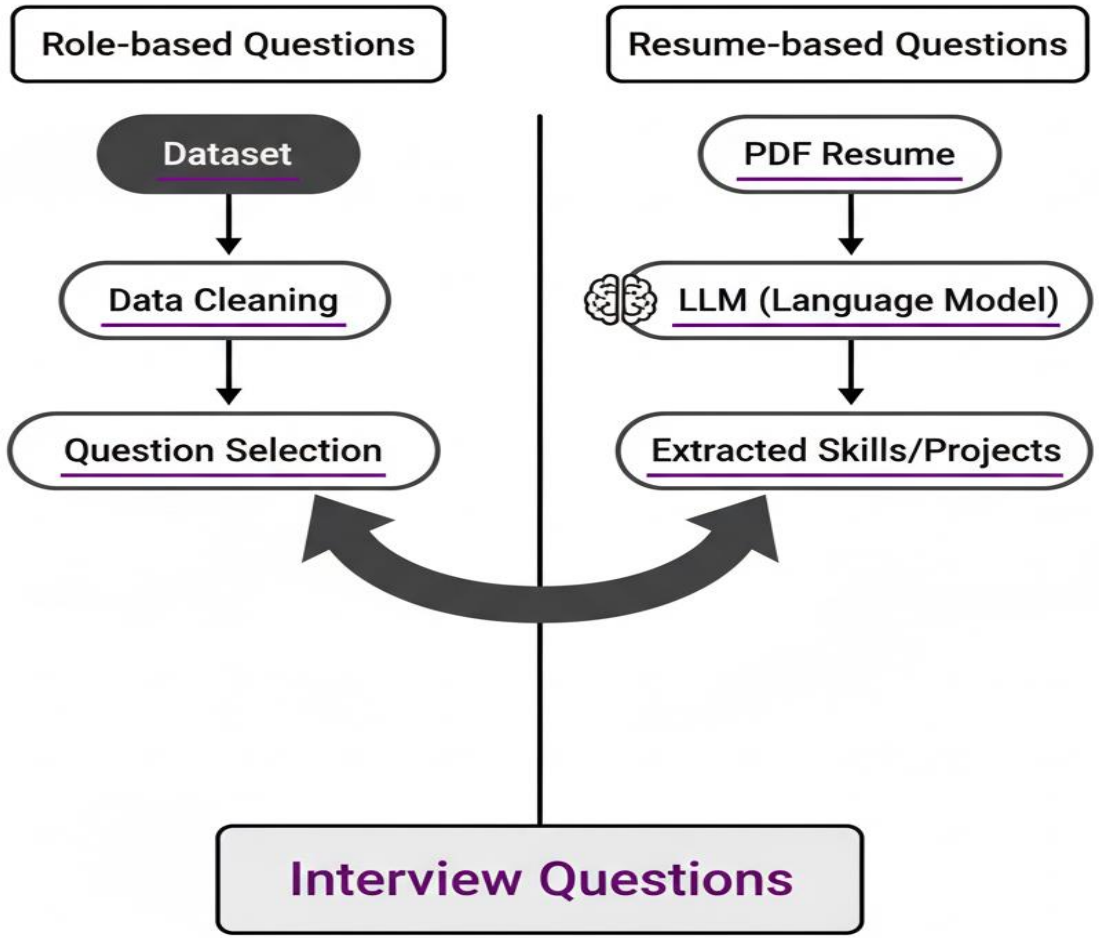
Purpose: To measure vocal confidence and prosody (the "music" of speech).

Model: A **normalized score** based on **Pitch Variance** and **Vocal Energy**.

Rationale: A dynamic, energetic voice (high variance) signals confidence and enthusiasm. A flat, monotone voice (low variance) signals disengagement or low conviction.

PART 2: QUESTION GENERATION

Unified Question Generation Framework



ROLE-BASED QUESTIONS

- Interview questions were systematically extracted from the **AI Recruitment Pipeline Dataset**.
- We cleaned and curated questions **per job role** and stored them in JSON files.
- For every interview, the system selects **5–7 high-quality questions** relevant to the chosen role.
- Uses random sampling to ensure **variation across interviews**.
- Ensures the questions align with **role expectations and required competencies**..
- Retrieval-ready architecture: FAISS index + Sentence-BERT can be activated to retrieve similar questions automatically.
- Each question includes a clearly defined *Canonical Answer*, ensuring consistent evaluation

RESUME-BASED QUESTIONS

- Resume is parsed directly from PDF using **PyPDF2**, extracting skills + projects using an LLM in **JSON-only mode**.
- The **entire resume-based question generation module is LLM-driven**, using the Gemini API for extraction, analysis, question creation, and evaluation.
- The LLM analyzes the candidate's resume to identify key skills, tools, and experience areas.
- It generates around 5 **personalized, resume-driven interview questions** based on these identified themes.
- For each response, the LLM evaluates the answer against an **expected competency** to check completeness.
- If the response is not proper, the system generates up-to 2 **follow-up questions** to probe deeper.
- This creates a structured, dynamic, adaptive interview experience tailored to the candidate's background.

PART 3 : MALPRACTICE DETECTION

Preventing Cheating in Online Interviews

- Remote interviews introduce new opportunities but also new risks.
- Candidates may try to gain unfair advantages through **external help, impersonation, or digital cheating tools**.
- To maintain hiring integrity, it becomes essential to actively detect and prevent malpractice in real time.
- This Enables **automated evaluation** to be both **accurate and ethical** and Preserves fairness among all applicants

.

KEY CHALLENGES IN ONLINE HIRING:

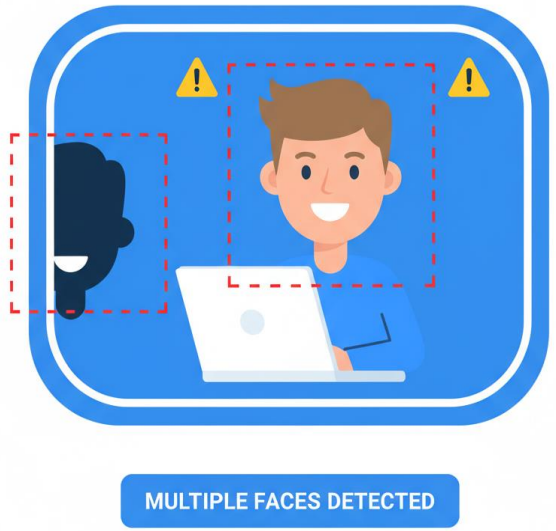
- **Impersonation:** Someone else sits in front of the camera or assists from off-screen.
- **External Help:** Candidates may read from prepared scripts, search answers online, or receive coaching.
- **Multi-Device Cheating:** Using a second screen or phone to look up answers during the interview.
- **Window Switching:** Leaving the interview tab to browse answers.
- **Non-Engagement:** Looking away, turning off camera, or taking unusually long pauses

MULTIPLE FACE / EXTRA PERSON DETECTION:

We detect more than one human face in the webcam feed. On detection, the system marks an incident and generates flag which will be later sent to Evaluation and report genertaion.

Actions / Detection Logic:

- Run a fast face detector (TinyFace) per frame.
- Count faces per frame; if `count ≥ 2` for `N` frames within `M` seconds → **multi-face incident**.
- **Thresholds:** 2+ faces in > 3 frames within 10 seconds → flag.
- Optional: blur secondary face in recorded video and capture snapshot evidence.

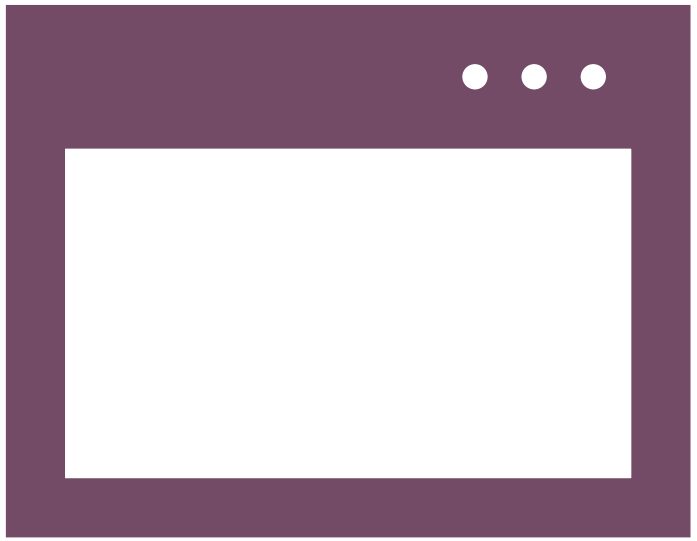


GAZE TRACKING & OFF-SCREEN DETECTION

- We estimate gaze direction and head pose and track the proportion of time the candidate looks away from the screen. We differentiate short, natural glances from suspicious off-screen behavior.
- **Actions / Detection Logic:**
- Compute eye landmarks + head pose (OpenFace).
- Derive gaze vector and map to `on-screen` vs `off-screen` bins.
- **Metrics & thresholds:**
  - Off-screen gaze > 5 seconds continuously → warning.
  - Off-screen gaze proportion > 30% over a question → flag..

WINDOW FOCUS & TAB SWITCHING

- We use browser APIs to detect when the interview tab loses focus, when the window is minimized, or when the user switches to another application. Each focus-loss event is logged and impacts the malpractice score.
- **Actions / Detection Logic:**
- Frontend uses `visibilitychange` and `blur` events to detect focus changes.
- Detect `pagehide`, `visibilityState !== 'visible'`, and `document.hasFocus()` false.
- **Thresholds:**
  - Any focus-loss < 2s → warning (transient).
  - Focus-loss > 2s → flag suspicious.
  - Repeated focus-loss events (> 2 per interview) → increase score.



RESPONSE TIMING & DELAY DETECTION (>10S RULE)

We measure time from end of question audio to the first detected speech segment in the candidate's answer. Delays longer than the configured threshold contribute to malpractice scoring.

Actions / Detection Logic:

- Mark `question_end_time` (when question TTS completes) or `question_display_time`.
- Use VAD on incoming audio; find `first_voice_timestamp`.
- Compute `response_delay = first_voice_timestamp - question_end_time`.

Thresholds:

- `response_delay > 10s` ⇒ **suspicious delay** flag for that question.
- Multiple delayed responses escalate risk.

AGGREGATED MALPRACTICE SCORING & LOGGING

- All incidents are weighted and combined into a malpractice score. The final score contributes to candidate evaluation but does not automatically disqualify

Scoring Model :

- Multiple faces detected: **weight 3.**
- Identity mismatch: **weight 3.**
- Tab-switch (each): **weight 3.**
- Response delay >10s (each): **weight 3.**
- Gaze off-screen for >30%: **weight 3.**

Each unique flag has weight and if repeated multiple time and weight of the duplicates is 1

PART 4 : EVALUATION

Dual-manifold Framework for Technical evaluation

Observation :
Technical evaluation routinely conflates:

- **factual knowledge**
- **the structure & soundness of reasoning**

Objective

Design a principled metric that:
(i) isolates these axes,
(ii) yields reproducible numeric signals,
(iii) admits concise qualitative assessment.

"An important observation is that when reasoning loses its invariants, even small distortions accumulate into destructive steps"

Figure: speaking of destructive steps

DUAL MANIFOLD REPRESENTATION

Core Idea:

Every candidate answer A is evaluated simultaneously on:

- $\mathcal{M}_k$ (Knowledge Manifold),

  $\mathcal{M}_r$ (Reasoning Manifold).

For every answer A we define two embeddings $O_k(A) \in \mathcal{M}_k,$ $O_r(A) \in \mathcal{M}_r,$

Let A* be the canonical solution

Similarity functional: $$\mathcal{F}(A) = \lambda_k \langle O_k(A), O_k(A^*) \rangle + \lambda_r \langle O_r(A), O_r(A^*) \rangle.$$

Interpretation: knowledge alignment measures factual correctness and reasoning alignment checks conceptual structure.

LLM ADJUDICATION

To recover subtleties inaccessible to pure geometry, we condition a state-of-the-art LLM with a highly constrained persona:

"evaluate as a panel of IMO judges, scores in [8,9] require genuine substance & rigorous development of the core ideas; scores >9 demand rare conceptual elegance; scores in [4,7] captures clear flow of thought and a logical approach even if the final result or execution is flawed or incomplete; scores <4 indicates little to no relevant progress or a completely confused/disjointed approach to the problem"

The model produces four sub-scores;

[Conceptual correctness, Precision grounding, Clarity, Creativity] & a synthesis describing trajectory, strengths, risks & recommendations

These qualitative outputs act as “derivatives” of the numerical score, revealing curvature rather than mere magnitude.

CAREER TRAJECTORY EVALUATION

The CV is decomposed into 5 evaluation axes : (using the same parrelel rubric)

- Target Alignment (Relevance)
- Impact Quantifier (Depth & Quality)
- Trajectory Coherence
- Technical Resolution (Specificity)
- Career Velocity & Momentum

Each is scored in [0, 10] under the same strict persona.

Scoring Functional:

$$C(CV) = \frac{1}{5}\sum_{i=1}^{5} s_i, \qquad s_i \in [0, 10].$$

COMPOSITE COMPETENCE INDEX (CCI)

Principle : Demonstrated ability outweighs stated experience.

Define the composite index:

$$\mathrm{CCI}_{final} = 0.75 \cdot \mathcal{F}(A) + 0.25 \cdot \mathcal{C}(CV).$$

The interview defines the main direction in the competency manifold, while the CV acts as a prior

We issue 3 types of **Badges:**

- Hire-Ready: CCI > 8 (representing strong mastery and stability)
- Good Potential: 6 <= CCI <= 8 (promising with some gaps.)
- Weak: CCI < 6 (weak alignment or fragmented reasoning.)

Note: All the scores were later scaled to 100 to acquire better precision

STRUCTURED REPORT GENERATION

we design a report which in itself is a "Structured, mathematical portrait of ability" not merely just achieved scores

- The final PDF cleanly decomposes a candidate into orthogonal dimensions: {Conceptual, Rigor, Clarity, Insight}; revealing strengths & blind spots that raw scores conceal
- The raw score after deduction of malpractice score will be reported as Final
- The Radar geometry makes asymmetries visually legible

We produce Narrative insights, not numerical monotony

- The LLM synthesizes a trajectory, risk and recommendation, transforming dispersed signals into a coherent evaluative narrative.
- The prose highlights why the candidate performs as they do, not merely how well.

Most importantly, it reveals shape not size: Two candidates may both score **7**,
but one may be *"stable but shallow," while the other is "volatile yet brilliant in spikes."*

The narrative disambiguates identical aggregates.

A MORE RELIABLE ALTERNATIVE TO HUMAN GRADING

Another aspect is it creates continuity across assessments. Instead of snapshot scoring, a narrative yields:

- what improved
- what regressed
- what stayed invariant

This is essential for long-term talent evaluation.

Human interviewers drift: fatigue, bias, recency noise.

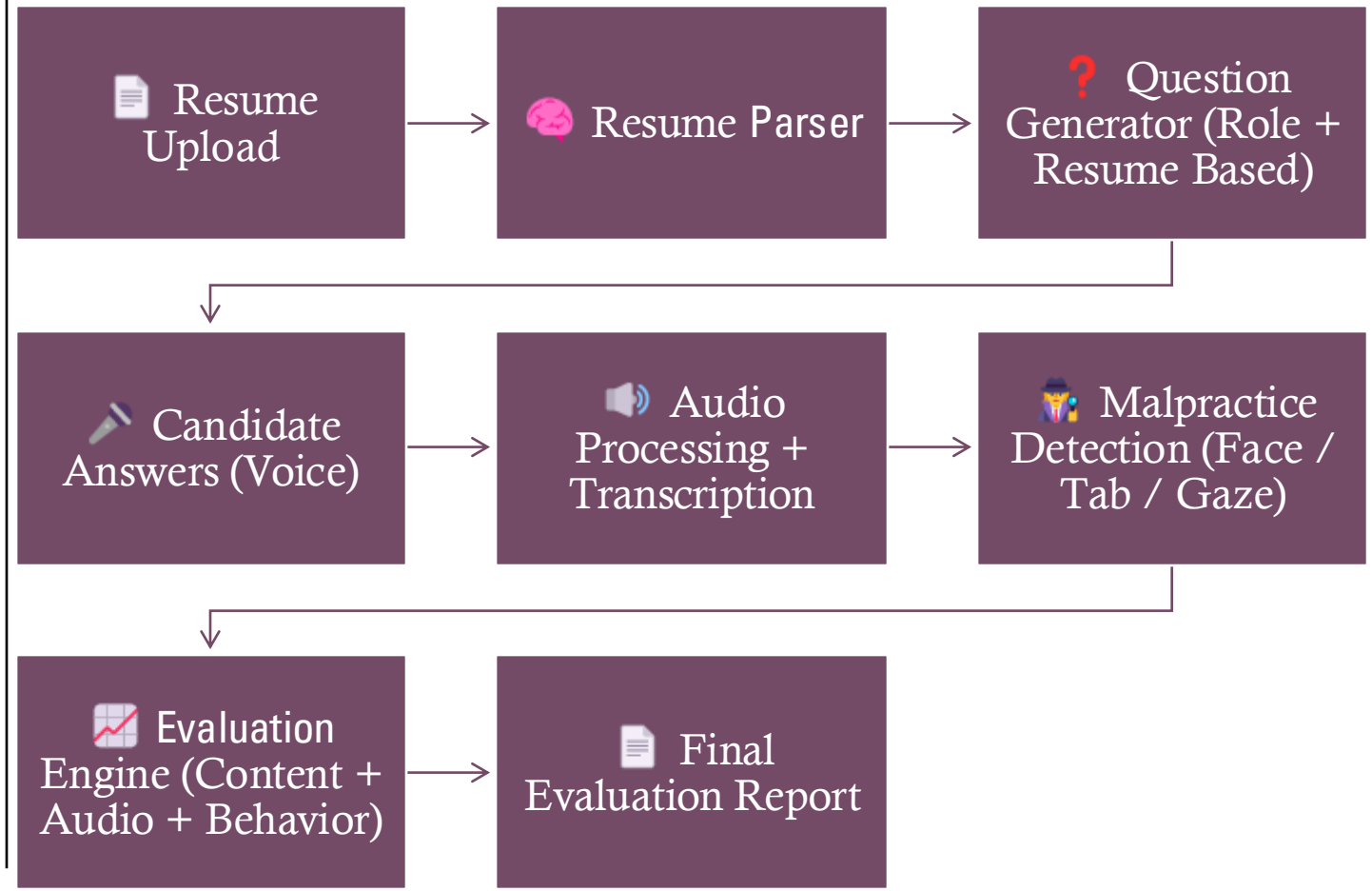
Our system standardizes the evaluation landscape: every candidate solves the same manifold; every score flows through the same computation.

- The result is a report that is reproducible, noise-resistant, and diagnostically rich

The report offers a measured, interpretable alternative to human grading ,one that preserves nuance while ensuring consistency, transparency, and repeatability.

PART- 5 : INTEGRATION

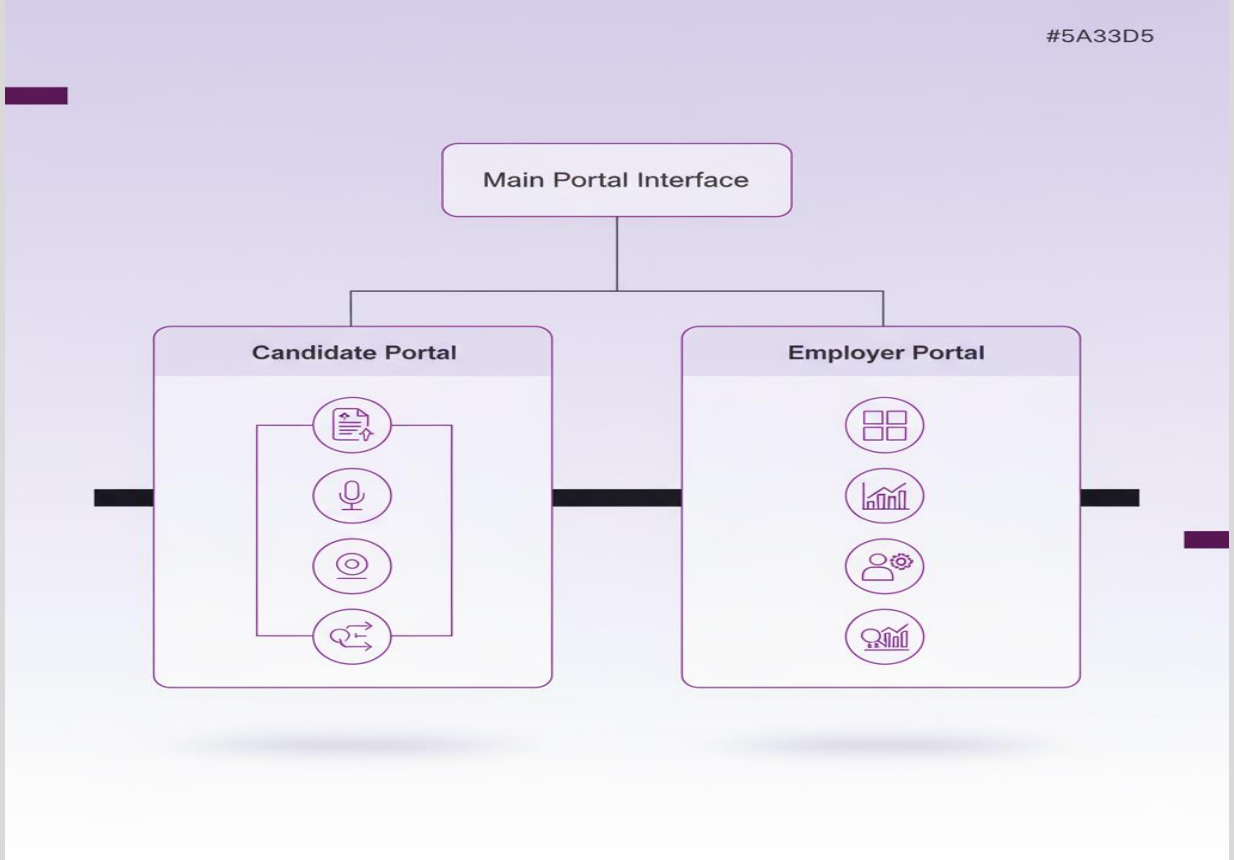
End-to-End AI Interview System



PART 6 : FRONTEND & UI/UX

Unified Landing Interface (Main Portal)

- **A Clean, role-based gateway that directs users to the correct workflow, Candidate interviews or Employer dashboards ensuring clarity, security, and seamless navigation.**



MAIN PORTAL UI

Candidate UI – Gateway for Interview

- **Resume Upload** — Users upload a PDF resume directly through the UI
- **Role Selection** — Dropdown or search-enabled selection of the job role
- **Camera and mic Access** — Asks User for permission to enable camera and mic
- Interview starts only when Resume upload and Permission Grants are completed

Interview Structure:

- Reading aloud questions (**Browser's speech synthesis API**),Live subtitles of questions,
- Audio recording stop and start buttons
- **Real-Time Feedback** — UI updates questions dynamically as backend returns next prompts
- **Session Persistence** — Uses React +Vite session state to maintain continuous interview flow
- **Proctoring UI Hooks** —
    - Shows "Camera Active" indicator and camera preview
    - Shows Microphone Status
    - Alerts when Malpractice flags are detected

UI INTERFACE FOR CANDIDATE

EMPLOYER UI- INTERVIEW INSIGHTS AT A GLANCE

A centralized dashboard for recruiters to review performance, reports & flags

What the Employer UI Enables:

- View **all candidate interview sessions** conducted for their posted roles
- Track **scores, Malpractice Logs, and overall performance**
- Access **evaluation reports, and Per Question Analysis**

Quickly shortlist or reject candidates based on *data-driven insights*

Designed for hiring teams who want clarity, speed, and transparency in the evaluation process.

UI INTERFACE FOR EMPLOYER

HOW FRONTEND COMMUNICATES WITH BACKEND?

- Seamless communication through lightweight, scalable APIs

The frontend interacts with backend components through a sequence of API operations:

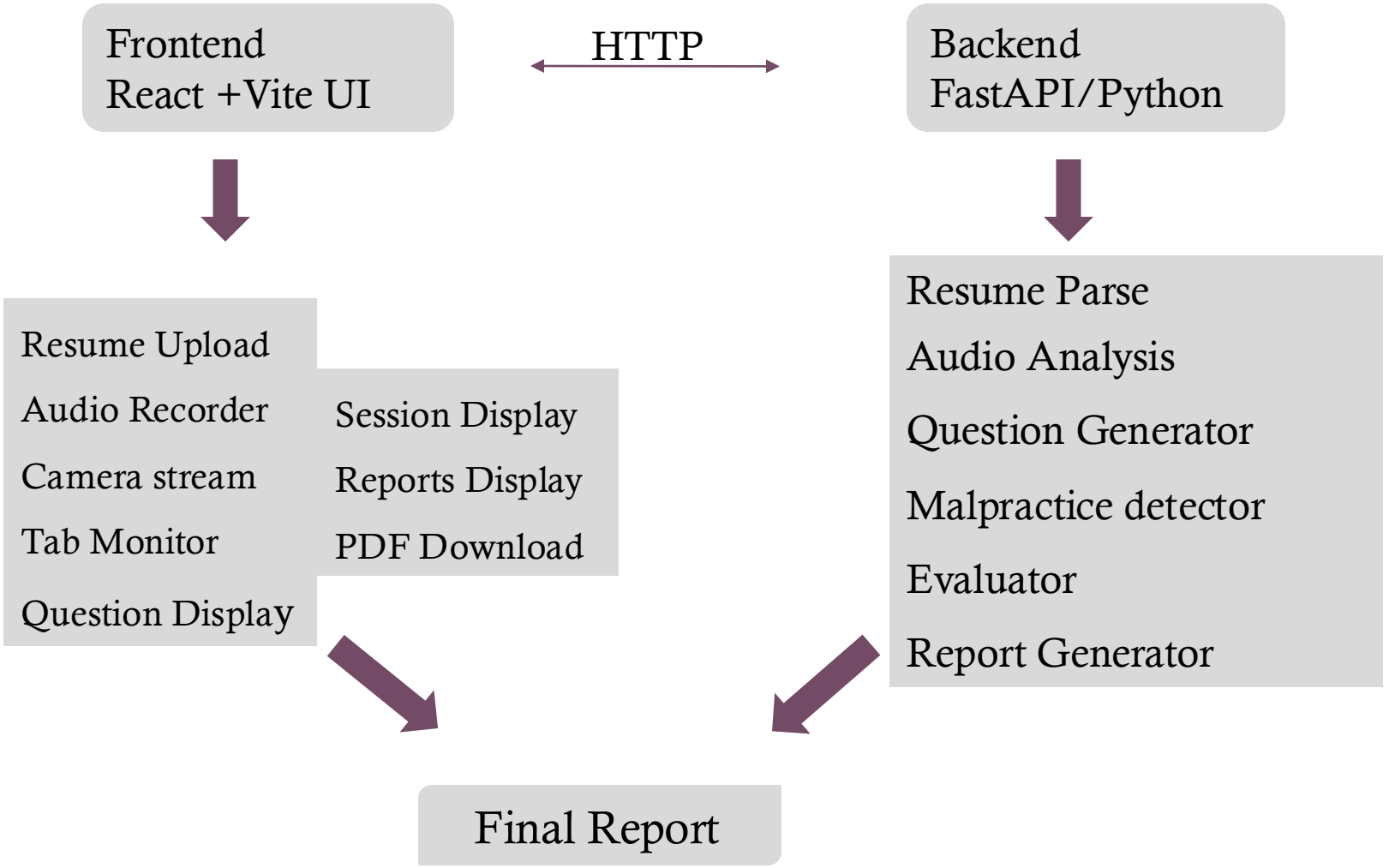
- **Upload Resume API**
- **Start Interview API**
- **Malpractice log API**
- **End Interview API**
- **Get roles API**
- **Process answer API**
- **List sessions API**
- **Get session API**
- **List reports API**
- **Get report API**
- **Get Report Pdf API**

REACT + VITE FRONTEND — WHY WE CHOSE IT

Advantages:

- **Blazing-fast development** using Vite's instant hot-reload (much faster than CRA or Next.js for this use case).
- **Full UI flexibility** — custom components for audio recording, question subtitles, timers, webcam, and alerts.
- **Perfect separation from backend** — communicates with Python APIs cleanly using Axios.
- **Browser-native audio & video access** through the MediaRecorder and getUserMedia APIs.
- **Real-time event handling** for malpractice detection (multiple faces, tab switching, eye-gaze alerts).
- **Reusable component architecture** makes it easy to scale features (candidate UI, employer UI, reports).
- **Production-ready** with Vite build optimizations, code splitting, and static asset handling.
- **Easy deployment** on Netlify / Vercel / Firebase with zero configuration.

FRONTEND–BACKEND INTERACTION (HIGH-LEVEL ARCHITECTURE)



“WHAT CAN BE IMPROVED?”

Advanced UI & Marketplace Integration

- **Improved Employer Portal:**
  A secure, access-controlled interface in which Only companies can access and login , post roles, upload job descriptions.
- **Role Marketplace:**
  All posted roles dynamically sync to the **candidate UI**, users to select real, active openings.
- **Improved Candidate Interface:**
  Cleaner design, real-time subtitles, smoother audio prompts to build a marketplace for the companies for filtered hiring.

Adding New Features:

- **Mock Interview:** Enabling candidates to try out mock interviews separately with access to report. Feedback helps for better preparation.

Richer Evaluation & Reporting

- **More Polished PDF Reports:**
  Employer-grade visuals, graphs, and highlights.
- **Skill Match Analytics:**
  Highlight how candidate’s resume aligns with job description.

“WHAT CAN BE IMPROVED?”

Better Latency & Scalability

- **Distributed Microservices Pipeline:**
  Split ASR, LLM, evaluation, and video analysis to scale independently.
- **Parallel Processing Architecture:**
  Support **1000+ concurrent interviews** with queue-based orchestration.

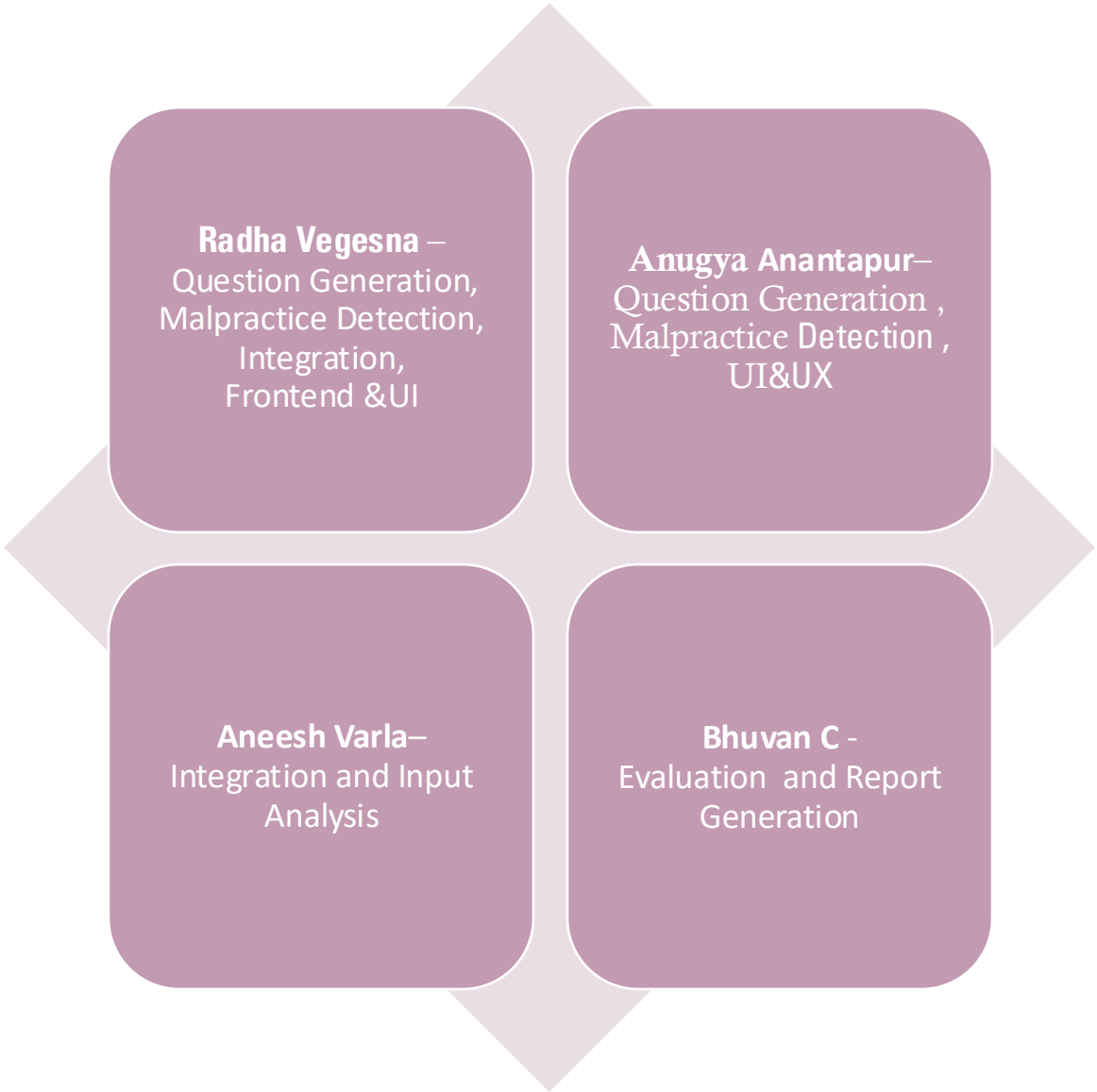
Stronger Malpractice Detection

- **High-Accuracy Face Tracking:**
  Detect cheating behaviors across **entire frame**, not just the center.
- Detect Multiple voices while analyzing theAudio

Bigger Knowledge Base

- **Expand Role Library:**
  Add 100+ roles across tech, HR, marketing, finance, design, operations.
- **Larger Question Bank:**
  Curate thousands of vetted, competency-based questions.
- **Adaptive Questioning:**
  The system learns from previous candidates to improve its query style.

INDIVIDUAL CONTRIBUTIONS



Thank You!