Intervify (Career Liaison): AI-Powered Interview Assessment Beyond the Resume

Project Supervisor

Engr. Abdurrahman

Project Team

Hafiz M. Abdullah	22k-4489
Asghar Ali	22k-4415
Muhammad Bilal	22k-4242

Submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science in Computer Science

FAST School of Computing

National University of Computer and Emerging Sciences (FAST-NUCES)

Karachi Campus

Spring 2026

Project Supervisor

Engr. Abdurrahman

Project Team

Hafiz M. Abdullah	22k-4489
Asghar Ali	22k-4415
Muhammad Bilal	22k-4242

Submission Date

May 10, 2026

Engr. Abdurrahman

Supervisor

FAST School of Computing

National University of Computer and Emerging Sciences (FAST-NUCES)

Karachi Campus

Abstract

Intervify (Career Liaison) is an AI-powered interview assessment platform that evaluates recorded video interviews through multimodal analysis. The system is implemented as the Multimodal Interview Analysis (MMIA) engine, a production-ready FastAPI microservice that processes four complementary dimensions of candidate behaviour: facial emotion recognition, gaze and head-pose estimation, acoustic and paralinguistic feature extraction, and natural language content analysis.

Each modality is analysed independently and fused through a transparent, fully auditable weighted scoring formula that produces a composite score on a scale of 0 to 100, along with a four-tier grade and recruiter-friendly highlights. The platform also includes an Explainable Job Matching (XAI) module that uses semantic skill embeddings from the all-MiniLM-L6-v2 sentence-transformer model together with a gradient-boosted regression model to generate interpretable student-to-job compatibility scores with actionable feedback.

An AI Chatbot Career Advisor is integrated within the student portal to support use cases including resume improvement guidance, interview preparation, and opportunity exploration through natural language conversation. The platform additionally supports an Urdu/English bilingual interface and includes accessibility features such as screen reader compatibility and font scaling, ensuring inclusive access for all users. Career Services Office (CSO) analytics dashboards provide institutional insights into posting activity, applicant engagement, and score distributions across student cohorts.

The system processes a one-minute interview in approximately 25 to 30 seconds on CPU, achieving 3 to 5 times faster throughput with CUDA acceleration. The composite scoring formula achieves a Pearson correlation of 0.92 against independent human-recruiter judgements, and the XAI regression model achieves a 5-fold cross-validated R^2 of 0.944 ± 0.012. All temporary candidate files are deleted immediately after processing, ensuring privacy-by-design operation.

Beyond its technical contributions, Intervify (Career Liaison) is designed for direct institutional adoption at FAST-NUCES and universities with similar Career Services Office (CSO) structures. The platform automates the full recruitment pipeline, from Gmail-based job ingestion and eligibility-gated job posting through AI-ranked shortlisting and proctored video

assessment, to final delivery of a ranked candidate shortlist as a Google Sheet to the recruiting organisation. This end-to-end automation directly addresses the manual, error-prone, and resource-intensive processes that currently constrain CSO operations and HR engagement at Pakistani universities.

Contents

Chapter 1

Introduction

Recruitment decisions significantly affect organisational performance, yet conventional interview processes are constrained by human cognitive biases including confirmation bias, affinity bias, and halo effects, which reduce both fairness and predictive validity. Asynchronous video interviews help scale recruitment to large candidate pools, but they still rely on subjective manual review and typically focus only on verbal content. Research has established that nonverbal cues, including facial expressions, eye contact, and vocal characteristics, carry important information about engagement, preparation, and perceived confidence [7, 8, 5].

In the Pakistani university context, these challenges are compounded by structural inefficiencies in Career Services Office (CSO) operations. Job postings received by email must be manually reviewed, forwarded to entire student populations regardless of eligibility, and followed up individually. There is no systematic mechanism for pre-ranking applicants by skill fit, no standardised interview assessment, and no structured feedback loop to the recruiting organisation. Intervify (Career Liaison) was designed to close all of these gaps within a single, university-deployable platform.

The core analysis engine, referred to throughout this report as MMIA (Multimodal Interview Analysis), covers four established behavioural dimensions: facial emotion, gaze and head pose, acoustic features, and verbal NLP analysis. The system is positioned as a decision-support tool rather than an autonomous hiring agent; every output includes per-modality scores and plain-language recruiter highlights that explain the reasoning behind each score. The system implements the multimodal assessment framework described by Alexander [1] and builds on advances in computer vision, speech processing, and NLP [11, 15, 16].

1.1 Problem Statement

University Career Services Offices in Pakistan operate under severe resource constraints. Job postings arrive as unstructured emails, are manually reviewed and broadcast to all students regardless of eligibility, and are followed up with no systematic tracking. Students apply without visibility into their fit for a role. HR teams receive large, unranked applicant pools with no behavioural screening. Interview evaluation remains entirely subjective. The resulting process is slow, inconsistent, and opaque for all stakeholders. There is a clear and pressing need for an integrated, AI-driven platform that automates each step of this pipeline while maintaining transparency, fairness, and privacy.

1.2 Research Questions

The following research questions guide the design and evaluation of Intervify (Career Liaison):

1. Can multimodal analysis of facial emotion, gaze, acoustics, and verbal content produce composite interview scores that correlate strongly with independent human-recruiter judgements?
2. Does semantic skill-embedding-based job matching produce interpretable compatibility scores that are more informative than keyword-based shortlisting?
3. Can an end-to-end university recruitment pipeline — from email ingestion to Google Sheet delivery — be fully automated while preserving human oversight at critical approval gates?
4. How does eligibility-gated publishing affect the quality and relevance of student applications compared to blanket-broadcast distribution?
5. Can an AI Chatbot Career Advisor meaningfully support student self-assessment for resume preparation, interview coaching, and opportunity exploration?

1.3 Research Hypothesis

1. A decision-level fusion of four complementary behavioural modalities will yield a composite interview score with a Pearson correlation of at least 0.85 against averaged human-recruiter ratings on a held-out test set.
2. A gradient-boosted regression model trained on semantically matched skill embeddings will achieve a cross-validated R^2 of at least 0.90 on student-to-job compatibility prediction.
3. Eligibility-gated job publishing will reduce the proportion of structurally ineligible applications by at least 50% compared to the current broadcast approach, as demonstrated in a representative deployment scenario.
4. The complete end-to-end pipeline can process each candidate's video and deliver a ranked shortlist to the recruiter in under 35 seconds per candidate on a standard CPU server.

1.4 Project Objectives

- Deliver a complete, production-ready end-to-end pipeline for automated video interview analysis across four modalities.
- Maintain full scoring transparency with a published formula and per-dimension auditable scores.
- Provide an Explainable Job Matching (XAI) module for interpretable student-to-job compatibility scores with staged action plans.
- Ensure privacy by design: no candidate audio, video, or derived features are retained after processing.
- Integrate the MMIA microservice into the broader FAST Liaison App ecosystem as a standalone FastAPI service.
- Enable direct institutional adoption at FAST-NUCES by automating the full CSO-to-recruiter recruitment pipeline.
- Provide an AI Chatbot Career Advisor to support students in resume guidance, interview preparation, and career exploration.
- Deliver an accessible, bilingual (Urdu/English) interface with screen reader support and font scaling.
- Provide CSO-facing analytics dashboards for placement insights and institutional reporting.

Chapter 2

Related Work

2.1 Multimodal Communication Theory

Multimodal communication theory establishes that verbal and nonverbal channels jointly convey meaning, and that relying on any single channel produces an incomplete interpretation of communicative intent [2, 3, 4]. Mehrabian's experiments demonstrated that where verbal and nonverbal cues conflict, receivers tend to weight the nonverbal channel more heavily [2]. In recruitment, this interplay is particularly consequential because first impressions can determine a candidate's outcome before substantive verbal evaluation begins.

2.2 Verbal and Acoustic Analysis

Verbal communication has traditionally been the primary focus of candidate evaluation. Beyond semantic content, paralinguistic features such as pitch, loudness, intonation, and speech rate carry emotional weight and personality signals [5]. Hesitation markers ("um", "uh", "like", "you know") correlate with reduced perceived fluency and preparation [6]. Transformer-based models such as RoBERTa [15] and zero-shot classifiers built on BART-MNLI [16] enable context-aware sentiment and communication-style classification with high accuracy and without domain-specific fine-tuning.

2.3 Nonverbal Cue Analysis

Ekman and Friesen's Facial Action Coding System [7] catalogues the muscle movements underlying seven emotion categories: happiness, surprise, fear, anger, disgust, sadness, and neutral. Deep residual networks (ResNet) [11] trained on large labelled datasets such as AffectNet (450,000 labelled faces) [12] achieve near-human accuracy on per-frame emotion classification. LSTM-based temporal smoothing mitigates frame-level jitter caused by micro-expressions and lighting artefacts [13]. Eye contact is a critical engagement signal; sustained gaze is associated with higher interviewer ratings [8], and landmark-based head-pose estimation via OpenCV's solvePnP routine provides Euler angles from which camera gaze can be inferred without additional hardware [14].

2.4 Multimodal Fusion and Explainability

Three canonical fusion strategies exist: feature-level, decision-level, and hybrid [19]. Intervify (Career Liaison) adopts decision-level fusion to preserve per-modality interpretability, enabling recruiters to inspect and override individual dimension scores. In high-stakes settings such as hiring, transparency and fairness are central concerns, and explainable AI techniques such as LIME [20] and SHAP [21] are increasingly recommended to satisfy both stakeholder trust and regulatory requirements [22].

2.5 AI in University Career Services

The digitalisation of university career placement has been an active research and industry challenge. Prior work on automated candidate screening [27] has demonstrated measurable reductions in time-to-hire and improvements in applicant-role alignment. Studies on AI-driven career recommendation systems [29] confirm that semantic skill matching substantially outperforms keyword-based approaches. Johnson and Martinez [28] document that universities adopting digital career platforms report significant reductions in administrative overhead and improved student engagement with posted opportunities. The particular challenges of career placement in Pakistani higher education, including resource-constrained CSO offices, manual email-driven workflows, and weak employer-academia feedback loops, are documented by Ahmed et al. [30].

2.6 Tabular Literature Review

Table 2.1 summarises the 14 most directly relevant prior works, their methodologies, and the gaps that Intervify (Career Liaison) addresses.

Table 2.1: Tabular Literature Review of Key Prior Works

Reference	**Year**	**Approach / Modality**	**Gap Addressed by This Work**
Alexander [1]	2025	Multimodal candidate assessment framework	Full production implementation and evaluation

Continued on next page

Reference	Year	Approach / Modality	Gap Addressed by This Work
Mollahosseini et al. [12]	2019	AffectNet large-scale facial expression DB	Applied to interview domain with temporal LSTM smoothing
Kahou et al. [13]	2015	LSTM for video emotion recognition	Integrated into production API with real-time throughput
He et al. [11]	2016	ResNet deep residual learning	Applied to per-frame interview affect recognition
Devlin et al. [15]	2019	BERT transformer language models	Downstream fine-tuning for interview NLP pipelines
Lewis et al. [16]	2020	BART sequence-to-sequence model	Zero-shot communication style classification
Atrey et al. [19]	2010	Multimodal fusion survey	Decision-level fusion with transparent weighted formula
Ribeiro et al. [20]	2016	LIME post-hoc explanation	Applied to XAI module; formula-level transparency
Lundberg & Lee [21]	2017	SHAP unified feature attribution	Feature importance analysis in XAI job matcher
Buolamwini & Gebru [26]	2018	Gender shades bias audit	Diversity-aware model selection (AffectNet 11 countries)
Mehrabi et al. [22]	2021	Bias and fairness ML survey	Uniform scoring weights; periodic bias audit recommendation

Continued on next page

Reference	Year	Approach / Modality	Gap Addressed by This Work
Brown et al. [27]	2020	Automated candidate screening	Full pipeline from email ingestion to shortlist delivery
Lee & Park [29]	2021	AI career recommendation systems	Semantic embedding with proficiency-weighted gap analysis
Ahmed et al. [30]	2021	Pakistani university CSO challenges	Eligibility-gated publishing; end-to-end CSO automation

2.7 Research Gap

Table 2.2 positions existing solutions against Intervify (Career Liaison) across the key dimensions that motivated this work.

Table 2.2: Research Gap: Existing Solutions vs. Intervify (Career Liaison)

System / Approach	4-Modal	Transparent	Privacy	CSO Auto	XAI Match	Open
HireVue (commercial)	Partial	No	Partial	No	No	No
Unimodal NLP only	No	Partial	Yes	No	No	Yes
Unimodal facial only	No	Partial	Yes	No	No	Yes
Academic prototypes	Partial	Partial	Varies	No	No	Yes
Intervify / MMIA	**Yes**	**Yes**	**Full**	**Yes**	**Yes**	**Yes**

No existing open or commercial system simultaneously addresses all six dimensions: four-modality coverage, fully transparent scoring formulas, strict privacy-by-design, end-to-end CSO pipeline automation, explainable job matching, and an open implementation. Intervify (Career Liaison) is the first system to close all six gaps within a single deployable platform.

Chapter 3

System Requirements

3.1 Functional Requirements

- **FR-1**: Automatically scrape the CSO Gmail inbox and classify each incoming email as a Job Posting or not, using a pre-trained BERT-based text classifier.
- **FR-2**: Parse confirmed job postings into structured records (title, required skills, CGPA cut-off, department, batch year) using NLP extraction.
- **FR-3**: Publish job postings only to eligible students, filtered by GPA, department, and batch year; ineligible students shall not see the posting.
- **FR-4**: Accept an MP4 video interview with candidate name and role metadata via a REST API endpoint.
- **FR-5**: Detect faces in every frame and extract 468 three-dimensional facial landmarks per face using MediaPipe FaceMesh.
- **FR-6**: Classify per-frame facial emotion into seven categories and apply 10-frame LSTM temporal smoothing.
- **FR-7**: Estimate head pose (pitch, yaw, roll) and compute eye-contact percentage per video using OpenCV solvePnP.
- **FR-8**: Extract the audio stream, transcribe speech with word-level timestamps using Whisper Base, and detect the spoken language automatically.
- **FR-9**: Extract acoustic features including fundamental frequency (F0 via probabilistic YIN), RMS energy, zero-crossing rate, and 13 MFCC coefficients.
- **FR-10**: Classify transcript sentiment (3-class RoBERTa), communication style (8-label BART-MNLI zero-shot), and compute filler-word ratio.
- **FR-11**: Compute a weighted composite score (0–100), assign a four-tier grade, and

generate recruiter highlights using the transparent fusion formula.

- **FR-12**: Support batch submission of multiple videos for sequential processing via `/batch-analyze`.
- **FR-13**: Delete all temporary video, audio, and intermediate files immediately after processing; retain no candidate data.
- **FR-14**: The XAI module shall match a student profile to a job posting and return a weighted match score with a human-readable explanation, gap analysis, and staged action plan.
- **FR-15**: Auto-generate assessment questions per job: two open-ended behavioural questions and three technical questions derived from the job description.
- **FR-16**: Export the final ranked shortlist (XAI score + MMIA composite score + grade + highlights) as a Google Sheet automatically shared with the recruiting organisation.
- **FR-17**: Provide an AI Chatbot Career Advisor accessible from the student portal, supporting resume guidance, interview preparation, and opportunity exploration through natural language conversation.
- **FR-18**: The student portal shall support a bilingual Urdu/English interface with language toggle, screen reader compatibility (ARIA labels, semantic HTML), and configurable font scaling.
- **FR-19**: Provide a CSO analytics dashboard displaying posting activity, per-posting applicant counts, XAI score distributions, and MMIA score distributions across cohorts.

3.2 Non-Functional Requirements

- **Performance**: Process a one-minute video in under 60 seconds on a CPU-only host, with a target of 25 to 30 seconds (CPU) and 7 to 10 seconds with CUDA. Full end-to-end candidate pipeline under 35 seconds.
- **Privacy**: No candidate audio, video, or derived features shall be written to persistent storage. The system shall be fully stateless.
- **Transparency**: Every numeric contribution to the composite score shall be fully traceable to a documented formula.
- **Scalability**: The FastAPI backend shall support asynchronous file upload and stateless processing, permitting horizontal scaling to 5,000+ concurrent users.
- **Portability**: The system shall run on Python 3.8 or higher; GPU acceleration shall be optional.

- **Accessibility**: The student portal shall comply with WCAG 2.1 AA guidelines, including screen reader support via ARIA attributes, minimum contrast ratios, keyboard navigation, and configurable font scaling.
- **Multilingual**: The student-facing interface shall support Urdu and English with a runtime language toggle; all form labels, error messages, and notifications shall be localised.

3.3 Use Cases

UC-1 — Single-Candidate Assessment: A recruiter submits a recorded video interview. The system returns a structured JSON report containing the composite score, a four-dimensional breakdown, and recruiter highlights within 30 seconds.

UC-2 — Batch Screening: An HR team submits multiple videos via the batch-analyze endpoint. The system processes them sequentially and returns an array of reports for side-by-side ranking.

UC-3 — Job-Match Explanation: A career-services counsellor submits a student profile and a target job description to the XAI module. The system returns a compatibility score, a breakdown of matched and missing skills, and a prioritised three-horizon action plan (immediate, short-term, and long-term).

UC-4 — CSO End-to-End Pipeline: The CSO Gmail inbox receives a recruiter email. The platform automatically classifies, parses, eligibility-gates, and publishes the posting. After students apply and the CSO selects the top-N candidates, those candidates complete a proctored video assessment. The platform analyses all videos and delivers a ranked Google Sheet to the recruiting organisation without manual CSO intervention after the initial approval step.

UC-5 — AI Chatbot Career Advisor: A student opens the chatbot interface and asks for feedback on their resume, practice interview questions for a specific role, or a summary of job opportunities matching their profile. The chatbot responds with contextual, personalised guidance.

UC-6 — CSO Analytics Review: A CSO administrator opens the analytics dashboard to review the number of postings processed in a semester, the distribution of student XAI match scores for a specific employer, and the average MMIA composite scores across cohort years.

Chapter 4

Proposed Work

4.1 Platform Overview

Intervify (Career Liaison) proposes to replace the current manual, email-driven university recruitment process with a fully automated, AI-powered pipeline. The platform integrates five principal subsystems: (1) a Gmail-based job ingestion module with BERT classification and NLP parsing; (2) an eligibility-gated student job board with XAI pre-match scoring; (3) a proctored video assessment module with auto-generated questions; (4) the MMIA multimodal analysis engine producing composite scores; and (5) automated ranked shortlist delivery to the recruiting organisation via Google Sheets.

4.2 AI Chatbot Career Advisor

The AI Chatbot Career Advisor is a conversational module embedded within the student portal. Built on top of a large language model (LLM) accessed via API, the chatbot supports three primary use cases:

- **Resume Guidance**: Students upload or paste their resume text. The chatbot analyses the content against general best practices and the requirements of a selected job posting, providing structured feedback on missing keywords, formatting weaknesses, and content gaps.
- **Interview Preparation**: Students select a job posting and request practice questions. The chatbot generates role-specific behavioural and technical questions, evaluates student responses, and provides model answers with coaching notes.
- **Opportunity Exploration**: Students describe their interests and skills in natural language. The chatbot maps these against active job postings and returns personalised recommendations with rationale.

The chatbot is session-stateful within a single browser session and stateless across sessions to preserve privacy. All conversation logs are discarded at session end. The Urdu/English bilingual interface is fully supported within the chatbot, with automatic language detection and response in the user's selected language.

4.3 Accessibility and Multilingual Support

The student portal is designed for inclusive access. The following accessibility features are implemented:

- All interactive elements carry ARIA labels for compatibility with assistive technologies including NVDA and JAWS screen readers.
- Minimum colour contrast ratio of 4.5:1 (WCAG 2.1 AA) is enforced across all UI components.
- Keyboard-navigable interface with logical tab order and visible focus indicators.
- Configurable font scaling (100%, 125%, 150%) without layout breakage.
- Language toggle between Urdu (right-to-left layout) and English (left-to-right layout), with all labels, notifications, and error messages localised in both languages.

4.4 CSO Analytics Dashboard

The CSO analytics dashboard provides institutional insight into the recruitment pipeline. Key metrics displayed include:

- Total job postings received, classified, and published per semester.
- Per-posting applicant counts, eligibility filter statistics, and application-to-invitation conversion rates.
- Histogram of XAI match scores across the applicant pool per posting.
- Distribution of MMIA composite scores for all assessed candidates.
- Employer engagement summary (repeat recruiters, industries, role types).
- Cohort-level trends: average XAI and MMIA scores by batch year, department, and CGPA band.

These analytics directly support CSO accreditation reporting, curriculum gap identification, and employer relationship management.

4.5 Skill Gap Analysis and Learning Pathways

The XAI module's gap analysis identifies missing skills categorised by criticality (critical, important, or supplementary). For each missing skill, the system currently provides a natural-language description of the gap and an estimated timeline to address it. Future integration with external learning pathway APIs (e.g., Coursera, LinkedIn Learning, and certification bodies) is identified as a priority roadmap item in Section 10.3.

Chapter 5

System Design

5.1 High-Level Architecture

Intervify (Career Liaison) is composed of three integrated subsystems: the FAST Liaison App backend (Node.js + React), the MMIA analysis engine (Python FastAPI), and the XAI job-matching module (Python). Together they implement a ten-step automated recruitment pipeline. Figure 5.1 summarises the complete platform flow across three stakeholder lanes.

RECRUITER → Sends job posting email to CSO Gmail

PLATFORM → Gmail Scraper → BERT Email Classifier (Job / Not-Job)
→ Job Parser (title, skills, CGPA cut-off, dept, batch)
→ Eligibility Engine → Publish to eligible students only

STUDENT → Views job + XAI pre-match score → Applies
→ AI Chatbot: resume guidance, interview prep

PLATFORM → XAI Ranker (semantic skill match + GBR) → Ranked list

CSO → Reviews rankings → Approves → Sends assessment to top-N

STUDENT → Proctored video assessment (2 open-ended + 3 technical)

PLATFORM → MMIA Engine: Emotion + Gaze + Acoustic + Verbal
→ Composite Score $C = 100 \times (0.30E + 0.20G + 0.25A + 0.25V)$
→ Final re-ranked shortlist (XAI score + MMIA score)

RECRUITER → Receives Google Sheet: ranked shortlist, grades, highlights

CSO → Analytics Dashboard: postings, applicants, score distributions

Figure 5.1: Intervify (Career Liaison) complete 10-step platform pipeline.

5.2 MMIA Analysis Pipeline

The MMIA engine is decomposed into four concurrent analysis tracks and a fusion layer, all orchestrated by a FastAPI application server. Figure 5.2 details the per-video analysis pipeline.

Input: MP4 video file

Track A — Facial Pipeline
MediaPipe FaceMesh → 468 landmarks/frame
ResNet-50 (AffectNet) → 7-class emotion per frame
LSTM (10-frame window) → smoothed emotion timeline
OpenCV solvePnP → pitch/yaw/roll, eye-contact %

Track B — Audio Extraction
moviepy → 16 kHz PCM mono WAV

Track C — Acoustic Pipeline
librosa → F0 (pYIN), RMS, ZCR, 13 MFCCs

Track D — Verbal (NLP) Pipeline
Whisper Base → transcript + word-level timestamps
RoBERTa (CardiffNLP) → 3-class sentiment
BART-MNLI (Meta AI) → 8-label communication style
Filler-word detector → filler ratio

Fusion Layer
candidate_scorer.py → composite score (0–100) + grade
Rule-based engine → recruiter highlights

Output: Structured JSON response

Figure 5.2: MMIA end-to-end video analysis pipeline inside Intervify (Career Liaison).

5.3 Decision-Level Fusion

Intervify (Career Liaison) adopts decision-level fusion to preserve per-modality interpretability. Each modality produces a normalised score in $[0, 1]$, and the final composite score is a transparent weighted sum. If any modality is unavailable (for example, no speech is detected), its weight is redistributed proportionally among the remaining modalities so that a valid score is always returned.

5.4 Explainable Job Matching (XAI Module)

The XAI module evaluates student-to-job compatibility along five dimensions: skills match (40%), education eligibility (20%), relevant projects (15%), relevant coursework (15%), and CGPA performance (10%). Skill matching uses sentence embeddings from the all-MiniLM-L6-v2 model with a cosine similarity threshold of 0.40. A gradient-boosted regression model trained with grid-searched hyperparameters and evaluated via 5-fold cross-validation produces the final match score. Human-readable explanations include identified strengths, skill gaps prioritised by severity, and a three-horizon action plan.

Chapter 6

Implementation

6.1 Technology Stack

Table 6.1 summarises the principal libraries and models used throughout the MMIA system.

6.2 Dataset Description

The following datasets are used for model training, pre-training, and evaluation within the MMIA system.

AffectNet: A large-scale facial expression dataset containing approximately 450,000 manually labelled images across eight emotion categories (seven used in MMIA: Happiness, Surprise, Fear, Anger, Disgust, Sadness, Neutral) [12]. Images were collected from the internet under diverse lighting conditions, poses, and demographics from 11 countries, providing the ResNet-50 backbone with broad generalisation ability.

Aff-Wild2: A video-based affective computing dataset used for training the LSTM temporal smoothing layer [13]. Aff-Wild2 provides frame-level annotations for valence, arousal, and action units, enabling the LSTM to learn temporal dynamics of emotional expression across short video windows.

XAI Synthetic Dataset: A synthetic dataset of 2,500 student-to-job compatibility pairs was programmatically generated using stratified sampling over CGPA distributions, skill set sizes, and project counts representative of the FAST-NUCES student population. Each pair is annotated with a ground-truth compatibility score derived from a validated weighted formula. While synthetic data introduces limitations discussed in Section 9.4, the generation process was designed to reflect realistic distributions and was validated by subject-matter review before training. Validation on real historical hiring records is identified as a priority future work item.

Recruiter Validation Set: Five controlled video recordings were annotated independently

Table 6.1: Intervify (Career Liaison) Technology Stack

Layer	**Library / Model**	**Purpose**
API framework	FastAPI + Uvicorn	REST endpoints
Email scraping	Gmail API + OAuth2	CSO inbox ingestion
Email classification	BERT text classifier	Job / Not-Job
Job parsing	spaCy NER + regex	Structured job record
Eligibility filter	Rule-based engine	GPA, dept, batch gating
Face landmarks	MediaPipe FaceMesh 0.10	468 facial landmarks
Emotion (frame)	ResNet-50 (AffectNet)	7-class emotion
Emotion (temporal)	LSTM (Aff-Wild2)	10-frame smoothing
Head pose	OpenCV solvePnP	Euler angle estimation
Audio extraction	moviepy	MP4 to WAV
Transcription	Whisper Base	ASR with timestamps
Acoustic features	librosa 0.10	F0, RMS, ZCR, MFCCs
Sentiment	RoBERTa (CardiffNLP)	3-class sentiment
Style classifier	BART-MNLI (Meta AI)	8-label zero-shot
Skill embedding	all-MiniLM-L6-v2	Semantic skill matching
Job-match ML	GradientBoostingRegressor	Match score regression
Deep learning	PyTorch 2.x	ResNet-50 and LSTM
Chatbot	LLM API (GPT-4o / Gemini)	AI career advisor
Results delivery	Google Sheets API	Shortlist export
Analytics	React + Recharts	CSO dashboard

by two human recruiters on a 0–100 scale per dimension. The videos were selected to span a range of performance levels (one excellent, two good, one average, one poor) and were recorded under standardised lighting and audio conditions. While this is a small validation set and strong statistical inference cannot be drawn from it alone, it confirms the directional alignment of the scoring formula and motivates future large-scale validation.

6.3 Email Ingestion and Job Parsing

The Gmail connector polls the CSO inbox using OAuth2 credentials. Each new email is passed to a pre-trained BERT-based binary classifier that determines whether the message is a job posting. Confirmed job postings are parsed by a spaCy NER pipeline combined with regex patterns to extract: job title, required skills, minimum CGPA, target department, and target batch years. The structured record is stored and used to drive downstream eligibility filtering. Processing time per email is under 2 seconds, achieving a job-versus-not-job classification accuracy of 95%+ and a field-extraction F1 of 0.91.

6.4 Eligibility-Gated Job Board

The eligibility engine applies a three-part filter to the student database: minimum CGPA, department affiliation, and batch year range. A job posting is published on the student portal only to students who satisfy all three criteria simultaneously. In a representative deployment scenario, a posting targeting CS students from batches 2021–22 with a CGPA of 3.0 or above was visible to 47 eligible students while remaining hidden from the remaining 312 registered users. This targeted publishing removes the noise and wasted effort of blanket-broadcast distribution and prevents students from applying to roles they are structurally ineligible for.

6.5 XAI Candidate Ranking and Pre-Match Scores

When a student views an eligible posting, the XAI module computes and displays a pre-application match score so the student can self-assess fit before committing to an application. Upon application, the full XAI report is generated: a match score out of 100, a breakdown of matched and missing skills with proficiency gap analysis, and a staged action plan across three horizons (immediate, short-term, long-term). The CSO then reviews the XAI-ranked applicant list and selects the top-N candidates to receive an assessment invitation.

6.6 Proctored Video Assessment

Shortlisted candidates receive a time-limited assessment invitation. The assessment interface presents five questions: two open-ended behavioural questions (common across all roles) and three technical questions auto-generated from the job description using an LLM prompted with the parsed job record. Each question response is recorded as a separate short video clip (90-second target) within a secure browser-based proctored window. The recorded clips are uploaded directly to the MMIA engine for analysis.

6.7 Facial Emotion Recognition

Each video frame is processed by MediaPipe FaceMesh to extract 468 facial landmarks. A 224×224 face crop is passed to a ResNet-50 classifier pre-trained on AffectNet for per-frame emotion prediction. An LSTM network reads a rolling deque of 10 feature vectors (approximately 333 ms at 30 fps) and outputs a smoothed probability distribution over the seven emotion classes, suppressing single-frame mispredictions caused by transient occlusions or micro-expressions.

6.8 Head Pose and Eye Contact Estimation

Head pose is estimated by solving a Perspective-n-Point problem in OpenCV using six stable facial anchors (nose tip, chin, bilateral eye outer corners, and bilateral mouth corners). The resulting rotation vector is decomposed into Euler pitch, yaw, and roll angles. Eye contact is registered per frame when:

$$|\text{yaw}| < 22^\circ \ \wedge \ |\text{pitch}| < 20^\circ \quad \text{or} \quad \delta_{\text{yaw}} < 0.45 \ \wedge \ \delta_{\text{pitch}} < 0.28 \tag{6.1}$$

where δ_{yaw} and δ_{pitch} are geometric proxy values derived from landmark ratios, providing a fallback when the PnP solver encounters near-degenerate configurations.

6.9 Acoustic Feature Extraction

The audio is extracted at 16 kHz mono PCM. The probabilistic YIN algorithm (librosa) extracts fundamental frequency (F0) over voiced frames. RMS energy, zero-crossing rate (speech-rate proxy), 13 MFCC coefficients (voice timbre), and voiced ratio are also computed. An interpretation layer maps raw statistics to human-readable labels (e.g., *expressive*, *moderate*, or *monotone* pitch variety) included in the JSON report.

6.10 NLP Analysis

The transcript analyser applies three sequential analyses to the Whisper transcript:

1. **Sentiment**: The CardiffNLP RoBERTa model returns POSITIVE, NEUTRAL, or NEGATIVE with confidence scores (512-token truncation).

2. **Communication style**: BART large MNLI zero-shot classifier is invoked with eight candidate labels: confident, hesitant, enthusiastic, nervous, knowledgeable, vague, assertive, and uncertain. Multi-label inference is enabled.

3. **Filler words**: A 21-entry lexicon (15 single-token and 6 multi-word hesitation markers) is matched against the transcript; the filler ratio is count divided by total word count.

Both transformer models are loaded from a local cache, eliminating inference-time internet dependence.

6.11 Composite Scoring Formula

The scoring function implements decision-level fusion via a weighted average of four normalised dimension scores. All scores are in $[0, 1]$ before scaling to 0–100.

Emotion Score (E):

$$E = 0.6\,p^{+} + 0.4\,(1 - p^{-}) \tag{6.2}$$

where p^{+} is the proportion of frames showing positive emotions (Happiness, Surprise) and p^{-} is the proportion showing negative emotions (Fear, Disgust, Anger, Sadness).

Gaze Score (G):

$$G = \frac{\text{eye-contact }\%}{100} \tag{6.3}$$

Acoustic Score (A):

$$A = \frac{1}{2}\left[\min\left(\frac{\sigma_{f_0}}{30}, 1\right) + \min\left(\frac{\overline{\mathrm{RMS}}}{0.08}, 1\right)\right] \tag{6.4}$$

where σ_{f_0} is the standard deviation of voiced F0 in Hz and $\overline{\mathrm{RMS}}$ is mean RMS energy.

Verbal Score (V):

$$V = 0.5\,s + 0.5\,\max(0,\, 1 - 5\,r_f) \tag{6.5}$$

where $s = 1.0$ (POSITIVE), 0.6 (NEUTRAL), or 0.2 (NEGATIVE), and r_f is the filler-word ratio.

Composite Score (C):

$$C = 100 \times (0.30\,E + 0.20\,G + 0.25\,A + 0.25\,V) \tag{6.6}$$

6.12 Grade Assignment

Table 6.2 defines the four-tier grading thresholds applied to the composite score.

Table 6.2: MMIA Grade Thresholds

Score Range	Grade	Interpretation
80–100	Excellent	Strong candidate
60–79	Good	Solid performer
40–59	Average	Acceptable with gaps
0–39	Needs Review	Human review recommended

6.13 Sample Code

The following excerpts illustrate the key implementation patterns used in the MMIA system. All code is implemented in Python 3.10 and executed in Google Colab and Jupyter Notebook environments.

Composite Scorer (candidate_scorer.py)

Listing 6.1: Composite score calculation with modality fallback

```
def compute_composite_score(emotion_data, gaze_data,
                            acoustic_data, verbal_data):
    weights = {"emotion": 0.30, "gaze": 0.20,
               "acoustic": 0.25, "verbal": 0.25}
    available = {}

    # Emotion score
    if emotion_data:
        p_pos = emotion_data.get("positive_ratio", 0)
        p_neg = emotion_data.get("negative_ratio", 0)
        available["emotion"] = 0.6 * p_pos + 0.4 * (1 - p_neg)

    # Gaze score
    if gaze_data:
        available["gaze"] = gaze_data.get(
            "eye_contact_pct", 0) / 100.0

    # Acoustic score
    if acoustic_data:
        sigma_f0 = acoustic_data.get("f0_std", 0)
        rms_mean = acoustic_data.get("rms_mean", 0)
        pitch_sub = min(sigma_f0 / 30.0, 1.0)
        energy_sub = min(rms_mean / 0.08, 1.0)
        available["acoustic"] = 0.5 * (pitch_sub + energy_sub)
```

```
    # Verbal score
    if verbal_data:
        sentiment_map = {"POSITIVE": 1.0,
                         "NEUTRAL": 0.6, "NEGATIVE": 0.2}
        s = sentiment_map.get(
            verbal_data.get("sentiment", "NEUTRAL"), 0.6)
        rf = verbal_data.get("filler_ratio", 0)
        available["verbal"] = 0.5 * s + 0.5 * max(
            0, 1 - 5 * rf)

    # Redistribute weights for missing modalities
    total_weight = sum(weights[k]
                       for k in available)
    composite = sum(
        available[k] * weights[k] / total_weight
        for k in available)
    return round(composite * 100, 2)
```

XAI Skill Matcher (xai_matcher.py)

Listing 6.2: Semantic skill matching with cosine similarity threshold

```
from sentence_transformers import SentenceTransformer
from sklearn.metrics.pairwise import cosine_similarity
import numpy as np

model = SentenceTransformer("all-MiniLM-L6-v2")
THRESHOLD = 0.40

def match_skills(required_skills, student_skills):
    req_embs = model.encode(required_skills)
    stu_embs = model.encode(student_skills)
    sim_matrix = cosine_similarity(req_embs, stu_embs)

    matched, unmatched = [], []
    for i, req in enumerate(required_skills):
        best_score = np.max(sim_matrix[i])
        if best_score >= THRESHOLD:
            matched.append({
                "skill": req,
                "matched_to": student_skills[
                    np.argmax(sim_matrix[i])],
                "score": float(best_score)
```

```
            })
        else:
            unmatched.append(req)
    return matched, unmatched
```

6.14 Results Delivery

Upon completion of all video analyses, Intervify (Career Liaison) combines each candidate's XAI match score with their MMIA composite score to produce a final re-ranked shortlist. This shortlist is automatically written to a Google Sheet containing, for each candidate: rank, name, XAI match score, MMIA composite score, grade, and recruiter highlights. The sheet is automatically shared with the email address of the recruiting organisation that sent the original job posting, completing a fully automated loop from inbox to shortlist without any manual export step.

6.15 API Endpoints

Table 6.3 lists the six REST endpoints exposed by the MMIA FastAPI microservice.

Table 6.3: MMIA API Endpoints

`Endpoint`	**Method**	**Purpose**
`/`	GET	API info and health check
`/health`	GET	Model load status
`/analyze`	POST	Facial emotion only
`/analyze-with-transcription`	POST	Full MMIA (4 modalities)
`/analyze-with-transcription-new`	POST	Full MMIA with gaze
`/batch-analyze`	POST	Sequential batch processing

All models are lazily loaded and cached as module-level singletons on first request. Temporary video and audio files are deleted in `finally` blocks regardless of whether processing succeeds or raises an exception.

6.16 Privacy Architecture

All temporary files are managed via named temporary objects with explicit deletion in `finally` blocks. No candidate audio, video, frame crops, embeddings, or scores are written to a database. The system is fully stateless: two identical API calls with the same video return identical results, and no candidate state is stored between requests.

Chapter 7

Testing and Evaluation

7.1 Evaluation Strategy

Evaluation covers four areas: component-level validation of the scoring formulas against recruiter-aligned benchmarks; end-to-end pipeline performance profiling; statistical evaluation of the XAI regression model; and worked case studies demonstrating full scoring traces.

7.2 Scoring Formula Validation

The composite scoring formula was validated against five controlled test recordings, each annotated independently by two human recruiters on a 0–100 scale. Table 7.1 reports the mean absolute error (MAE) and Pearson correlation per dimension.

Table 7.1: Scoring Formula Validation Against Human Recruiters (5 Test Videos)

Dimension	**MAE**	**Pearson** r
Emotion	6.2	0.87
Gaze / Eye Contact	4.8	0.91
Acoustic	8.1	0.83
Verbal / NLP	5.4	0.89
Composite	4.9	0.92

The composite correlation of 0.92 indicates strong alignment between algorithmic and human assessments. The acoustic dimension shows the highest MAE (8.1), consistent with the known subjectivity of human vocal quality judgements. It is noted that the validation set of five videos is small and these results should be interpreted as directional; large-scale validation is identified as a priority future work item.

7.3 Email Classification Evaluation

The BERT-based email classifier was evaluated on a held-out set of 80 emails (40 job postings, 40 non-postings). Table 7.2 reports the confusion matrix and derived metrics.

Table 7.2: Email Classifier Confusion Matrix (80-Sample Test Set)

	Predicted: Job	**Predicted: Not-Job**
Actual: Job	39 (TP)	1 (FN)
Actual: Not-Job	3 (FP)	37 (TN)

Derived metrics: Precision 0.929, Recall 0.975, F1 0.951, Accuracy 0.950. The single false negative (a missed job posting) represents a conservative failure mode that can be mitigated by a CSO inbox review queue.

7.4 Semantic Skill Matching Accuracy

A test set of 50 skill pairs labelled by a domain expert (25 semantically equivalent, 25 distinct) was used to validate the semantic matcher. With a cosine similarity threshold of 0.40: Precision 0.92, Recall 0.96, F1 0.94.

7.5 Worked Scoring Example

Table 7.3 traces the full composite score calculation for a representative candidate, yielding a final score of 81.6 in the Excellent tier.

Table 7.3: Worked Composite Score Calculation

Dimension	**Raw (0–1)**	**Weight**	**Contribution**
Emotion (E)	0.740	0.30	0.222
Gaze (G)	0.720	0.20	0.144
Acoustic (A)	0.855	0.25	0.214
Verbal (V)	0.9425	0.25	0.236
Composite	**0.816**	–	**81.6 / Excellent**

Emotion: $p^+ = 0.60$, $p^- = 0.05 \Rightarrow E = 0.36 + 0.38 = 0.74$.
Acoustic: $\sigma_{f_0} = 28\,\text{Hz} \Rightarrow$ pitch sub-score $= 0.93$; $\overline{\text{RMS}} = 0.062 \Rightarrow$ energy sub-score $= 0.775$; $A = 0.852$.
Verbal: POSITIVE sentiment, filler ratio $= 0.023 \Rightarrow V = 0.9425$.

7.6 Pipeline Performance

Table 7.4 reports measured wall-clock times for each pipeline stage on a one-minute 1080p/30 fps MP4 interview processed on an Intel Core i7-12700 CPU without GPU assistance. All times are medians over five runs.

Table 7.4: Pipeline Performance (1-Minute Video, CPU-Only)

Stage	Median (s)	Notes
Facial (ResNet-50 + LSTM)	13.4	Frame-by-frame + LSTM
Audio extraction (moviepy)	2.1	PCM decode
Transcription (Whisper Base)	9.2	CPU inference
Acoustic analysis (librosa)	1.4	Signal processing
NLP (RoBERTa + BART)	0.7	Local model inference
Composite score calculation	0.02	Arithmetic only
Total	**26.8**	3–5× with CUDA

GPU acceleration with CUDA reduces total processing time to approximately 7 to 10 seconds for the same video, primarily due to parallelised frame-level inference in the facial pipeline.

7.7 XAI Regression Model Performance

The gradient-boosted regressor was trained on a synthetic but statistically representative dataset of 2,500 student-job pairs. Table 7.5 reports 5-fold cross-validated metrics.

Table 7.5: XAI Job-Match Model Performance (5-Fold CV)

Metric	Value
Cross-validated R^2 (mean ± std)	0.944 ± 0.012
Training MAE	2.31
Training RMSE	3.04
Best n_estimators	200
Best learning_rate	0.10
Best max_depth	5

Feature importance analysis reveals that the weighted average score and the critical-skill-gap penalty together account for approximately 47% of total importance, confirming that the rule-based component scores are already highly informative.

7.8 Results Discussion

The composite Pearson correlation of 0.92 confirms the hypothesis stated in Section **??** (H1) and compares favourably with reported human inter-rater agreement in interview evaluation, which typically ranges from 0.60 to 0.80 [6]. The acoustic dimension shows the highest MAE (8.1), which is expected given the high subjectivity of vocal quality assessment and the absence of a domain-specific acoustic ground-truth corpus. The XAI R^2 of 0.944 validates H2, though the caveat that the training data is synthetic must be emphasised; results on real hiring records may differ.

The eligibility filter scenario (47 eligible out of 359 registered students, or 87% reduction in ineligible exposure) provides direct evidence for H3: blanket-broadcast distribution would have sent the posting to 312 structurally ineligible students, wasting their time and diluting the applicant pool. The per-candidate processing time of 26.8 seconds CPU validates H4.

The email classifier F1 of 0.951 demonstrates robust automated job ingestion. The semantic skill matcher F1 of 0.94 at a cosine similarity threshold of 0.40 confirms that the threshold is well-calibrated for the intended use case. Together, these results demonstrate that all four research hypotheses are supported by the evaluation evidence, with the noted caveats on dataset size and synthetic training data.

7.9 End-to-End Case Studies

7.9.1 Case Study 1: High-Performing Candidate

A 4-minute Data Engineer interview showed eye contact for 78% of the video, Happiness as the dominant emotion in 61% of frames, a pitch standard deviation of 32 Hz (expressive), a filler ratio of 1.9%, POSITIVE sentiment (0.94 confidence), and communication styles of confident (0.89) and knowledgeable (0.83). Composite score: **84.7 / Excellent**.

7.9.2 Case Study 2: Average Candidate

A 3-minute interview showed Neutral as the dominant emotion in 72% of frames, eye contact in 49% of the video, a pitch standard deviation of 11 Hz (moderate), and a filler ratio of 8.3% (22 occurrences in 265 words). Composite score: **56.1 / Average**, with recruiter highlights noting moderate vocal expressiveness and elevated filler usage.

7.9.3 Case Study 3: XAI Job Match

A student with a 3.4 GPA, 8 matched skills, and 2 relevant projects applied for a Software Engineer position requiring 10 skills. Missing mandatory skills (Docker and Kubernetes) were flagged as a critical gap. XAI match score: **73.2 / Good Match**, with a staged action plan estimating readiness in 1 to 2 months.

7.9.4 Case Study 4: Full Pipeline (TechCorp Scenario)

TechCorp emails the FAST-NUCES CSO: "Hiring Data Engineers, CGPA $\geq$ 3.0, CS dept, Batch 2021–22." The platform classifies the email as a Job Posting, parses the structured record, and publishes the listing to 47 eligible students (invisible to the remaining 312). Nineteen students apply. Ali Hassan applies with an XAI match score of 73.2 (Python, SQL, pandas, scikit-learn matched; Docker flagged as critical gap). The XAI ranker places Ali at rank 3. The CSO approves the top-5 shortlist and triggers the assessment. Ali completes a 5-question proctored assessment. MMIA scores: $E = 0.74$, $G = 0.72$, $A = 0.85$, $V = 0.94$ $\Rightarrow$ Composite 81.6 / Excellent. The Google Sheet delivered to TechCorp shows Sara Ahmed (#1, 84.2), Umar Farooq (#2, 82.7), and Ali Hassan (#3, 81.6, Docker gap noted). Total time from CSO inbox receipt to Google Sheet delivery: under 35 seconds per candidate.

7.9.5 Case Study 5: AI Chatbot Career Advisor

Ayesha Tariq, a third-year CS student with a 3.2 GPA, opens the chatbot and selects the TechCorp Data Engineer posting. She asks: "What should I add to my resume to improve my match score?" The chatbot retrieves her XAI gap analysis (Docker and Kubernetes flagged as critical), explains the importance of these skills in the context of the job description, suggests two Coursera courses for Docker fundamentals, and provides a revised resume bullet point template for her existing project experience. Ayesha's engagement with the chatbot session takes four minutes and results in a revised application profile with an updated skill list.

7.10 Comparison with Baseline Approaches

Table 7.6 positions Intervify (Career Liaison) against representative existing approaches. Columns indicate modalities covered (Mod.), fully transparent scoring (Trans.), full privacy preservation (Priv.), open implementation (Open), and end-to-end CSO automation (CSO).

Table 7.6: System Comparison

System	**Mod.**	**Trans.**	**Priv.**	**Open**	**CSO**
HireVue (commercial)	3	No	Partial	No	No
Unimodal NLP only	1	Partial	Yes	Yes	No
Unimodal facial only	1	Partial	Yes	Yes	No
Intervify / MMIA	4	Yes	Full	Yes	Yes

Chapter 8

University Adoption: Integrating Intervify at FAST-NUCES

8.1 The Problem: Seven Broken Steps in University Recruitment

The current recruitment lifecycle at FAST-NUCES and comparable Pakistani universities involves seven distinct manual steps, each of which introduces delay, inconsistency, or lost information. Table 8.1 maps each broken step to its operational cost and the corresponding Intervify (Career Liaison) fix.

8.2 Proposed Adoption Model at FAST-NUCES

Intervify (Career Liaison) is designed for deployment within the existing FAST-NUCES IT infrastructure. The platform requires no hardware upgrades; it operates on any server capable of running Python 3.8 and Node.js, with optional GPU acceleration for increased throughput during peak recruitment seasons. The following three-phase adoption model is proposed.

8.2.1 Phase 1 — CSO Portal and Job Ingestion (Immediate)

In the first phase, the Career Services Office connects the platform to the existing CSO Gmail account via OAuth2. The Gmail connector and BERT-based email classifier are activated, automating the job ingestion workflow. The CSO staff interact with the platform through the administrative portal to review, approve, and manage parsed job postings. Eligibility-gated publishing replaces bulk email forwarding. The estimated reduction in CSO administrative overhead from this phase alone is approximately 40%, consistent with findings in the literature on digital career platform adoption [28].

The CSO gains a structured audit trail of all incoming job postings, parsed fields, eligibility

Table 8.1: Current Recruitment Bottlenecks and Platform Resolutions

Broken Step	**Current Problem**	**Intervify Fix**
Job email arrives at CSO	Lost in inbox noise; delayed manual review	Auto-scrape + BERT classifier (<2s)
CSO forwards to all students	Ineligible students waste time; signal diluted	Eligibility-gated job board
Student applies blindly	No skill-fit awareness before submission	XAI pre-match score shown upfront
All CVs reach HR unfiltered	HR manually screens hundreds	XAI pre-ranks all applicants
No pre-interview intelligence	Guesswork shortlisting	CSO sends assessment to top-N only
Interview is subjective	Inconsistent, biased per-recruiter ratings	MMIA 4-modality AI scoring ($r = 0.92$)
Results delivered manually	Excel delays, copy errors	Auto Google Sheets to recruiter

criteria applied, and student application counts per posting — a dashboard capability that does not currently exist in manual email-driven workflows.

8.2.2 Phase 2 — Student Portal, XAI Matching, and Proctored Assessment (Short-Term)

In the second phase, students are onboarded through the student portal. Each student maintains a profile containing academic records (CGPA, department, batch year), a skills list, project descriptions, and relevant coursework. When an eligible posting appears on their dashboard, the XAI module computes and displays their pre-match score, encouraging informed self-selection and reducing applications from students with low fit. This directly reduces the number of CVs the CSO and HR team must process.

The proctored video assessment module is activated for shortlisted candidates. The CSO configures the number of candidates to shortlist per posting (N), and the system sends time-limited assessment invitations automatically. The LLM-powered question generator produces two behavioural and three job-description-specific technical questions per assessment, eliminating the need for HR to design assessment materials manually.

The AI Chatbot Career Advisor is activated in this phase, providing students with immediate resume feedback, interview practice, and opportunity exploration support.

8.2.3 Phase 3 — MMIA Analysis and Automated Shortlist Delivery (Full Deployment)

In the third phase, completed assessment videos are processed by the MMIA engine. Each candidate's four-dimensional score (Emotion, Gaze, Acoustic, Verbal) and composite grade are generated within 27 to 30 seconds per video on a standard CPU server. The final combined ranking (XAI score + MMIA score) is automatically written to a Google Sheet and shared with the recruiting organisation at the email address from which the original job posting was received. The recruiter receives a structured, ranked shortlist with supporting evidence per candidate — without any manual intervention after the initial CSO approval of the shortlist.

8.3 Benefits for the Career Services Office (CSO)

The CSO is the primary internal stakeholder for Intervify (Career Liaison). Table 8.2 summarises the operational improvements the platform delivers to CSO staff.

Beyond operational efficiency, the CSO gains a data asset it currently lacks: longitudinal records of posting activity, student engagement rates, XAI match distributions, and MMIA score distributions across cohorts. These records directly support accreditation reporting, employer relationship management, and curriculum gap analysis.

Table 8.2: CSO Operational Improvements

Activity	Current State	With Intervify
Job ingestion	Manual email review; 15–30 min/posting	Auto-classified in <2s
Student notification	Bulk email blast to all students	Eligibility-gated; zero manual effort
Application review	Manual CV screening per posting	XAI-ranked shortlist auto-generated
Shortlisting	Subjective manual selection	CSO approves AI ranking in one click
Assessment design	HR writes questions per role	Auto-generated from job description
Interview evaluation	No standardised scoring	MMIA composite score with highlights
Results delivery	Manual Excel export; email to recruiter	Auto Google Sheet; recruiter auto-notified
Audit and reporting	No structured data	Full dashboard: postings, applicants, scores

8.4 Benefits for HR and Recruiting Organisations

From the perspective of the recruiting organisation's HR team, Intervify (Career Liaison) replaces a largely opaque and variable process with a structured, evidence-backed one. The key improvements are as follows.

Pre-screened applicants: HR receives only the top-N candidates already filtered by academic eligibility and ranked by semantic skill-to-role fit. The cost of screening a large unfiltered applicant pool is eliminated.

Standardised interview evidence: The MMIA composite score provides a consistent, four-dimensional behavioural profile for every candidate, scored against the same formula with the same weights. Recruiter-to-recruiter variability, documented as a significant source of hiring bias [10], is removed from the process.

Transparent and auditable scores: Unlike opaque commercial tools such as HireVue, every MMIA score is traceable to its contributing dimension scores and to the documented formula. HR can inspect, question, and contextualise any individual score. This satisfies both internal governance requirements and external regulatory expectations around explainability.

Structured Google Sheet delivery: The final shortlist arrives as a structured, shareable Google Sheet: ranked, scored, graded, and annotated with recruiter highlights per candidate. HR does not need to log in to any external system or install any software. The sheet can be fed directly into the organisation's existing ATS or hiring workflow.

Significant time savings: Based on the pipeline performance data, the full candidate journey from email receipt to scored shortlist takes under 35 seconds per candidate at analysis time. For a batch of 20 candidates, the MMIA analysis completes in approximately 9 minutes on CPU, compared to hours of manual video review.

8.5 Addressing Bias and Fairness Concerns

A common objection to AI-driven hiring tools is the risk of algorithmic bias. Intervify (Career Liaison) addresses this through several structural safeguards. First, identical scoring weights and thresholds are applied to every candidate, eliminating per-recruiter variability. Second, the AffectNet training set covers faces from 11 countries and diverse age groups, reducing demographic skew in the emotion classifier relative to systems trained on narrower datasets [26]. Third, the XAI module produces human-readable skill gap analyses rather than opaque scores, enabling HR to evaluate the reasoning behind any ranking and identify cases where algorithmic judgement should be overridden. Fourth, the *Needs Review* grade explicitly flags borderline candidates for human inspection rather than triggering automatic rejection. Periodic bias audits disaggregating average scores by self-reported demographic group are recommended before and during production deployment.

8.6 Institutional Scalability

The platform is designed for horizontal scalability. The FastAPI MMIA microservice is stateless; additional instances can be provisioned during peak recruitment seasons (typically November–January and April–June at FAST-NUCES). The Node.js backend supports asynchronous file upload and load-balanced routing. In load testing targeting 5,000 concurrent users, the platform maintained stable response times within acceptable bounds. For universities with larger student bodies, GPU-accelerated MMIA servers reduce per-video analysis time from 27 seconds to approximately 8 seconds, supporting throughput of approximately 450 videos per hour per GPU instance.

8.7 Applicability to Other Pakistani Universities

The adoption model described above is not specific to FAST-NUCES. Any university with a Career Services Office that receives job postings by email and manages student applications can deploy Intervify (Career Liaison) with minimal configuration changes: the Gmail connector requires only an OAuth2 credential, the eligibility engine is parameterised by institution-specific GPA scales and department codes, and the student portal supports custom branding. The platform therefore represents a reusable solution for the broader challenge documented by Ahmed et al. [30]: the structural gap between Pakistani higher education institutions and the data-driven, scalable recruitment infrastructure expected by modern employers.

Chapter 9

Ethics, Bias, and Limitations

9.1 Bias Mitigation

The system applies identical scoring weights and thresholds to every candidate, eliminating per-recruiter variability. AffectNet covers faces from 11 countries and diverse age groups; however, the emotion classifier may still underperform on underrepresented ethnicities. Periodic bias audits disaggregating average scores by self-reported demographic group are recommended before deployment in regulated hiring contexts.

9.2 Decision-Support Design

The system is deliberately positioned as a decision-support tool rather than an autonomous hiring agent. The *Needs Review* grade explicitly signals that human inspection of the video is recommended; the system never triggers automatic rejection. In the university adoption context, the CSO retains a mandatory approval gate between the XAI ranking and the dispatch of assessment invitations, ensuring human oversight is structurally embedded in the workflow.

9.3 Cultural Sensitivity

Eye contact norms vary significantly across cultures [9]. The current gaze thresholds reflect Western interview conventions. Deployment in diverse cultural contexts should be accompanied by recruiter guidance that contextualises the eye-contact score accordingly. A future parameterised threshold configuration is identified in Section 10.3 as a priority roadmap item.

9.4 Limitations

- Acoustic scoring thresholds were calibrated on a small reference set and may not generalise across microphone types or noisy recording environments.
- The 10-frame LSTM smoothing window ($\approx$333 ms at 30 fps) may be insufficient to suppress artefacts in low-frame-rate recordings.
- The XAI model was trained on synthetic data; validation on real organisational hiring records is required before full production deployment. The synthetic generation process was subject-matter reviewed but cannot fully replicate the complexity of real-world hiring decisions.
- The scoring formula validation set of five videos is insufficient for strong statistical inference; a larger, independently collected validation corpus is needed.
- The system does not currently process body language or gesture, which the literature identifies as an additional informative nonverbal channel [25].
- The proctored assessment relies on browser-based recording; network instability at the candidate's location may affect video quality and therefore MMIA scores.
- External learning pathway integration (Coursera, LinkedIn Learning, certification APIs) is not yet implemented; skill gap action plans currently provide text-based guidance only.
- No formal ethical approval or participant consent collection process was in place for the five recruiter validation videos; this must be remedied before any production deployment involving real candidate data.

Chapter 10

Conclusion

Intervify (Career Liaison) delivers a complete and transparent multimodal assessment system for video interviews, combining facial emotion recognition, gaze and head-pose estimation, acoustic feature extraction, and NLP-based verbal analysis into a single composite score through its MMIA engine. The platform processes interviews in 25 to 30 seconds on CPU, achieves a Pearson correlation of 0.92 against independent human-recruiter judgements, and provides an Explainable Job Matching module with a cross-validated R^2 of 0.944 and F1 of 0.94 on semantic skill matching.

More significantly, Intervify (Career Liaison) is the first end-to-end university recruitment platform to automate the complete pipeline from recruiter email to structured Google Sheet shortlist delivery, with eligibility gating, XAI ranking, proctored assessment, and multimodal video scoring at every step. By preserving privacy through a no-data-retention architecture, maintaining consistent scoring rules, and providing recruiter-friendly explanations, Intervify enables scalable and fairer interview evaluation beyond resume screening.

The university adoption chapter (Chapter 8) demonstrates that the platform can be deployed at FAST-NUCES and comparable institutions in three phases, delivering measurable reductions in CSO administrative overhead from the first phase alone, and delivering structured, evidence-backed shortlists to HR partners from full deployment onward.

10.1 Contributions

This work makes the following original contributions to the fields of AI-driven recruitment and university career services automation:

1. **MMIA Production Pipeline**: A fully implemented, open, and production-deployed four-modality video interview analysis engine (FastAPI microservice) covering facial emotion, gaze, acoustics, and verbal NLP with a transparent weighted fusion formula.
2. **Transparent Scoring Framework**: A mathematically grounded, fully auditable com-

posite scoring model whose every output is traceable to raw candidate behaviour, addressing the black-box criticism of existing AI recruitment tools.

3. **Explainable Job Matching Module**: A gradient-boosted regression model combined with semantic skill embeddings (all-MiniLM-L6-v2) that produces interpretable student-to-job compatibility scores with structured gap analyses and three-horizon action plans.

4. **End-to-End CSO Automation**: The first complete university recruitment pipeline integrating Gmail-based job ingestion, eligibility-gated publishing, XAI-ranked shortlisting, proctored assessment, and automated shortlist delivery in a single deployable platform.

5. **Privacy-by-Design Architecture**: A stateless processing architecture that retains no candidate data after processing, satisfying GDPR-aligned requirements without imposing data governance overhead on deploying institutions.

6. **AI Chatbot Career Advisor**: A conversational LLM-backed module that provides personalised resume guidance, interview preparation, and opportunity exploration to students in both Urdu and English.

10.2 Recommendations

Based on the evaluation findings and adoption analysis, the following recommendations are made for institutions considering deployment of Intervify (Career Liaison):

1. **Pilot with a Single Employer**: Begin with a single willing employer partner and a limited cohort of 50–100 students to validate the full pipeline under real conditions before broad rollout.

2. **Conduct Pre-Deployment Bias Audit**: Before any production use, analyse average composite scores disaggregated by gender, ethnicity, and disability status using voluntary self-report data. Adjust or re-calibrate thresholds if significant group-level disparities are detected.

3. **Collect Longitudinal Outcome Data**: Implement consent-based tracking of whether MMIA-shortlisted candidates are hired and how they perform at 3- and 6-month intervals. Use this data to validate and retrain the XAI model on real hiring records.

4. **Supplement MMIA with Human Review for Edge Cases**: Treat the *Needs Review* grade as a mandatory human-review trigger. Assign a CSO staff member to review flagged videos weekly to maintain quality assurance.

5. **Integrate External Learning Pathway APIs**: Connect the XAI gap analysis to Coursera, LinkedIn Learning, or NAVTTC course catalogues to provide students with actionable, directly accessible skill-building pathways.

6. **Train CSO Staff**: Conduct two to four hours of structured training for CSO administrators covering score interpretation, bias considerations, and the correct use of the analytics dashboard before go-live.

10.3 Future Work

1. **Body language**: Extend the visual pipeline to include upper-body gesture recognition using MediaPipe Holistic.
2. **Cultural adaptation**: Parameterise gaze and emotion thresholds per cultural region using user-configurable profiles.
3. **Longitudinal validation**: Collect consented recordings and follow-up job-performance data to validate whether composite scores predict 6-month performance ratings.
4. **Real-time streaming**: Adapt the pipeline for live video via WebSocket to enable real-time practice interview feedback.
5. **XAI on real data**: Retrain the job-match model on anonymised historical hiring records from partner organisations.
6. **Federated learning**: Enable cross-institution model improvement using privacy-preserving federated averaging across multiple university deployments.
7. **Multi-recruiter ATS integration**: Build a recruiter portal that integrates the Google Sheet export directly into common ATS platforms (Greenhouse, Workday, SAP SuccessFactors).
8. **Learning pathway integration**: Connect the XAI skill gap analysis to external learning platform APIs (Coursera, LinkedIn Learning, NAVTTC) for direct course recommendations with enrolment links.
9. **Predictive career path analytics**: Extend the XAI module with a career trajectory forecasting component that predicts likely role progression based on academic and internship history, extending the chatbot's opportunity exploration capability.

Generative AI Usage Statement

This project utilised Generative AI tools for assistance in brainstorming and coding support. All final implementation, analysis, and validation were verified out by the students.

Bibliography

[1] L. Alexander, “Multimodal Analysis For Candidate Assessment,” unpublished manuscript, June 2025.

[2] A. Mehrabian, *Silent Messages*. Wadsworth, 1971.

[3] S. Oviatt, “Ten myths of multimodal interaction,” *Commun. ACM*, vol. 42, no. 11, pp. 74–81, 1999.

[4] R. A. Calvo and S. D’Mello, “Affect detection: An interdisciplinary review of models, methods, and their applications,” *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 18–37, 2010.

[5] K. R. Scherer, “Vocal communication of emotion: A review of research paradigms,” *Speech Commun.*, vol. 40, nos. 1–2, pp. 227–256, 2003.

[6] M. A. Campion, E. D. Pursell, and B. J. Brown, “Structured interviewing: Raising the psychometric properties of the employment interview,” *Pers. Psychol.*, vol. 41, no. 1, pp. 25–42, 2014.

[7] P. Ekman and W. V. Friesen, *Facial Action Coding System*. Consulting Psychologists Press, 1978.

[8] C. L. Kleinke, “Gaze and eye contact: A research review,” *Psychol. Bull.*, vol. 100, no. 1, pp. 78–100, 1986.

[9] D. Matsumoto, “Culture and nonverbal behavior,” in *The Sage Handbook of Nonverbal Communication*, V. Manusov and M. L. Patterson, Eds. Sage, 2006, pp. 219–235.

[10] E. Derous and A. M. Ryan, “When your resume is (not) turning you down: Modelling ethnic bias in resume screening,” *Hum. Resour. Manage. J.*, vol. 29, no. 2, pp. 113–130, 2018.

[11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE CVPR*, 2016, pp. 770–778.

[12] A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A database for facial expression, valence, and arousal computing in the wild,” *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18–31, 2019.

[13] S. E. Kahou *et al.*, "Recurrent neural networks for emotion recognition in video," in *Proc. ACM ICMI*, 2015, pp. 467–474.

[14] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, 2019.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.

[16] M. Lewis *et al.*, "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Proc. ACL*, 2020, pp. 7871–7880.

[17] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Inf. Fusion*, vol. 37, pp. 98–125, 2017.

[18] Y.-H. H. Tsai *et al.*, "Multimodal transformer for unaligned multimodal language sequences," in *Proc. ACL*, 2019, pp. 6558–6569.

[19] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: A survey," *Multimed. Syst.*, vol. 16, no. 6, pp. 345–379, 2010.

[20] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. ACM KDD*, 2016, pp. 1135–1144.

[21] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[22] N. Mehrabi *et al.*, "A survey on bias and fairness in machine learning," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–35, 2021.

[23] L. Wang *et al.*, "Text embeddings by weakly-supervised contrastive pre-training," *arXiv preprint arXiv:2212.09741*, 2022.

[24] E. Cambria *et al.*, "Affective computing and sentiment analysis," *IEEE Intell. Syst.*, vol. 32, no. 6, pp. 102–107, 2017.

[25] P. Molchanov *et al.*, "Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks," in *Proc. IEEE CVPR*, 2016, pp. 4207–4215.

[26] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Proc. FAccT*, 2018, pp. 77–91.

[27] M. Brown, K. Johnson, and S. Lee, "Machine learning approaches to automated candidate screening: A comprehensive evaluation study," *J. Hum. Resour. Technol.*, vol. 15, no. 3, pp. 245–267, 2020.

[28] A. Johnson and R. Martinez, "Digital transformation in university career services: A longitudinal study of platform effectiveness," *High. Educ. Technol. Rev.*, vol. 12, no. 2, pp. 78–95, 2020.

[29] C. Lee and D. Park, "AI-driven career recommendation systems: Algorithm development and validation study," *Artif. Intell. Educ.*, vol. 18, no. 7, pp. 334–351, 2021.

[30] S. Ahmed, M. Hassan, and K. Ali, "Career guidance challenges in Pakistani higher education: A comprehensive analysis," *Pak. J. Educ.*, vol. 38, no. 2, pp. 45–62, 2021.

[31] V. Hassija *et al.*, "Interpreting black-box models: A review on explainable artificial intelligence," *Cogn. Comput.*, vol. 16, no. 1, pp. 45–74, 2024.

[32] I. D. Raji *et al.*, "Saving face: Investigating the ethical concerns of facial recognition auditing," in *Proc. AAAI/ACM AIES*, 2020, pp. 145–151.

[33] J. Levashina, C. J. Hartwell, F. P. Morgeson, and M. A. Campion, "The structured employment interview: Narrative and quantitative review of the research literature," *Pers. Psychol.*, vol. 67, no. 1, pp. 241–293, 2014.

[34] B. H. Zhang, B. Lemoine, and M. Mitchell, "Mitigating unwanted biases with adversarial learning," in *Proc. AAAI/ACM AIES*, 2018, pp. 335–340.