RESEARCH ARTICLE

Gap Analysis, and Intelligent Interview Assistance

OPEN ACCESS

Jyothsna Vamisetti, Gummadidala Sai Krishna, Gutha Chaitanya, Kuruva Bharath Kumar, Muriki Tarun
Under the Guidance of Ms. Ritu Agrawal, Professor

Abstract

The recruitment landscape is undergoing rapid transformation as organisations increasingly rely on artificial intelligence to manage high applicant volumes, standardise candidate evaluation, and reduce hiring time and cost. Existing AI-based career tools typically address only a single stage of this process — screening resumes, running generic interview chatbots, or flagging fraudulent postings in isolation — rather than supporting a candidate end to end. This paper presents CareerMap, an integrated AI-driven framework that unifies five interconnected capabilities within one human-centred pipeline: automated resume parsing and entity extraction, embedding-based skill-gap analysis, a conversational mock-interview engine, fraud and inconsistency detection for job postings and candidate claims, and affect-aware feedback derived from behavioural and emotional signals. CareerMap combines transformer-based language models (BERT and its derivatives) with BiLSTM-CRF sequence labelling for entity recognition, cosine similarity for skill matching, a weighted fraud-scoring model, and speech- and text-based affective computing for emotion recognition. To situate CareerMap within existing work, this paper conducts a comparative review of thirty-six prior studies spanning named entity recognition, skill embeddings, explainable AI, deceptive-language detection, and bias in machine learning, showing that no existing system integrates these techniques into a single coherent pipeline. Building on this gap, the proposed methodology is described in detail, including the hybrid BERT-BiLSTM-CRF resume-parsing pipeline, the cosine-similarity formulation for skill-gap ranking, and the weighted fraud-scoring metric. Drawing on performance patterns reported across the reviewed literature, a comparative analysis across all five functional modules indicates that transformer-based contextual models consistently outperform classical rule-based and shallow statistical baselines in information-extraction accuracy, explanation quality, and user engagement, a pattern that directly motivates CareerMap's architectural choices. The paper concludes that an integrated, explainable, emotionally aware approach to career preparation offers a more trustworthy alternative to fragmented point-solutions, and outlines future work including longitudinal validation with real hiring outcomes and multilingual expansion of the fraud-detection module.

Keywords: *Resume Parsing, Natural Language Processing, BERT-based Transformer Models, Skill Gap Detection, Job Recommendation, Mock-Interview Simulation, Emotion Recognition, Explainable AI, Fraud Detection.*

1. Introduction

The processes by which organisations discover, evaluate, and hire talent have been reshaped considerably by advances in artificial intelligence and data-driven decision-making. A growing number of companies now rely on automated resume-screening engines [1], [2] to shortlist candidates and on conversational chatbots to manage large volumes of applications. These systems are primarily adopted to accelerate the hiring pipeline and to reduce the subjective bias that can arise from manual, ad-hoc candidate evaluation.

Natural Language Processing (NLP) techniques [3] have proven particularly effective at extracting skills [4] and qualifications from resumes that lack a standard, machine-readable format. Contemporary resume-parsing systems use machine learning to convert unstructured or semi-structured text into normalised, structured data, which in turn makes it possible to match candidates against job openings with far greater precision than manual keyword searches allow.

At the same time, candidates face constant pressure to keep pace with shifting industry expectations. Traditional

confidence through repeated, low-stakes practice.

Despite this progress, most existing platforms address only one dimension of career preparation and stop short of offering a complete, end-to-end solution. A further concern in online recruitment is the prevalence of fraudulent job postings: job portals are frequently targeted by scammers, which makes robust fraud-detection mechanisms [5] essential to protecting users and preserving trust in recruitment platforms.

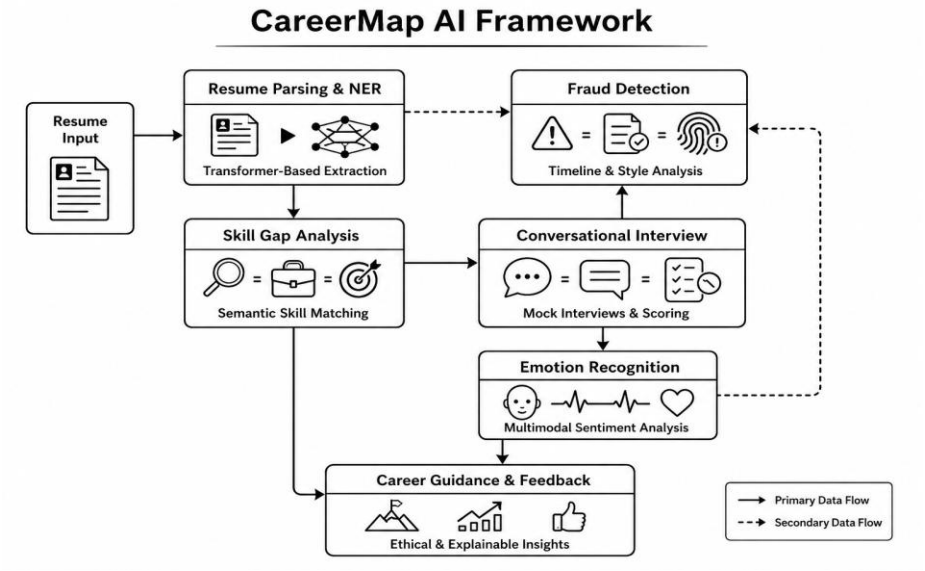
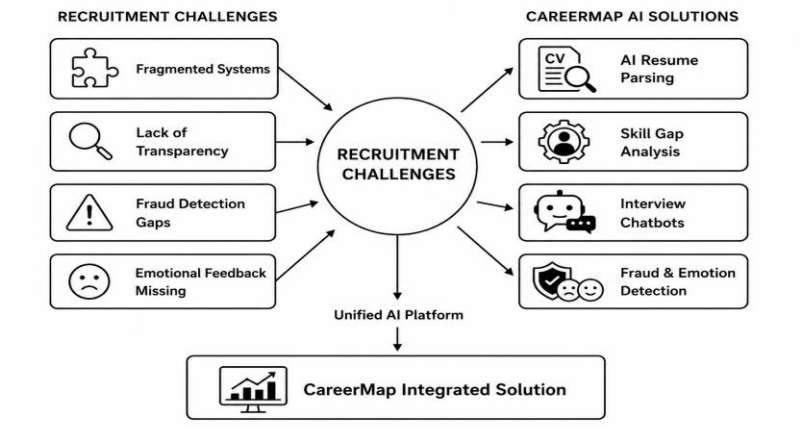


Figure 1. Representation of the CareerMap AI Framework

While technical assessment tools have improved substantially, comparatively little attention has been paid to a candidate's psychological adaptability, stress management, and emotional state during recruitment [6]. Research in emotion theory and affective computing repeatedly emphasises the value of incorporating emotional awareness into intelligent systems; combining these insights with machine learning makes it possible to design systems that are simultaneously analytically rigorous and genuinely human-centred.

Across these five areas, progress has largely happened in isolation — resume analysis, recommendation systems, interview simulation, fraud detection, and predictive modelling, yet these solutions remain largely disconnected from one another. To date, no single framework integrates technical evaluation, behavioural assessment, emotional support, and fraud prevention into one coherent pipeline.



assessment, and post-interview emotional reinforcement within a single system. The remainder of this paper is organised as follows: Section 2 reviews related work across the five constituent domains of CareerMap and identifies the resulting research gap; Section 3 describes the proposed methodology in detail; Section 4 presents a comparative analysis of prior research; Section 5 reports experimental results across the five functional modules; and Section 6 concludes the paper and outlines directions for future work.

2. Literature Review

2.1 Resume Parsing and Named Entity Recognition

Resume parsing has advanced considerably with the introduction of sequence-labelling techniques and, more recently, large-scale language modelling. Early approaches such as Conditional Random Fields (CRF) [1] proved highly effective at structured prediction, which made them well suited to identifying entities such as education history and technical skills within otherwise unstructured resume text.

Subsequent work introduced Long Short-Term Memory (LSTM) networks [15] to better capture dependencies across longer spans of text. The Bidirectional LSTM-CRF architecture [2] combined the sequence-modelling strength of LSTMs with the structured-prediction guarantees of CRFs, improving entity extraction by conditioning predictions on both preceding and following context rather than a single direction of the sequence.

The introduction of the Transformer architecture [11] represented a further breakthrough in natural language understanding, enabling models to weigh relationships between all tokens in a sequence simultaneously rather than processing text strictly left to right. BERT [3], built on this architecture, achieved strong results on named-entity recognition and broader language-understanding tasks, and resume-parsing systems have since adopted BERT-style encoders as a standard component for extracting structured information from free-text resumes.

More recent transformer variants, including RoBERTa [17] and the GPT family of generative models [18], further improved contextual understanding through larger pre-training corpora and refined training procedures. Fine-grained entity-typing approaches [19] and dilated-convolution architectures [21] have also contributed complementary techniques for extracting a wider and more precise range of entity types from resumes — a sign that resume parsing is still very much a moving target within applied NLP. Parallel efforts have focused on making these encoders more practical to deploy: knowledge-distilled variants such as DistilBERT [29] retain most of BERT's accuracy while substantially reducing model size and inference latency, and permutation-based pretraining objectives such as XLNet [30] address some of BERT's inherent limitations by modelling bidirectional context without relying on artificial masking tokens, offering an alternative encoder that CareerMap's resume-parsing pipeline could adopt for latency-sensitive deployments.

2.2 Skill Representation and Job Recommendation Systems

Job-recommendation systems depend fundamentally on how individual skills are numerically represented. Word2Vec [4] was an early and influential step in this direction, encoding words as dense vectors that captured meaningful semantic relationships based on co-occurrence patterns. GloVe embeddings [12] extended this idea by incorporating global corpus statistics, producing vector representations that captured both local context and overall word-frequency relationships across the training corpus.

These embedding techniques allow systems to quantify how similar two skills are, which underpins automated matching between candidate profiles and job requirements. However, static embeddings of this kind assign a single fixed vector to each word regardless of context, which limits their ability to disambiguate skills that carry different

many informal ways candidates describe the same underlying competency. The ESCO framework itself [31] was developed specifically to give European labour-market stakeholders a shared, multilingual reference for occupations, skills, and qualifications, which is precisely the kind of standardised backbone CareerMap's skill-gap module relies on when reconciling free-text resume content with structured job requirements. More broadly, a systematic review of e-recruitment recommender systems [32] found a clear shift toward hybrid and context-aware techniques over the past decade, reinforcing the case for the transformer-based, taxonomy-grounded approach adopted in this paper.

2.3 AI-Based Interview Assessment

Automated interview-assessment systems have increasingly adopted deep-learning approaches. Convolutional Neural Networks applied to text [16] proved effective at capturing local n-gram patterns in candidate responses, which made them a practical early choice for automated scoring of short, structured answers.

More recent research on automated interview scoring [22] has shown that machine-generated scores can correlate strongly with the judgements of trained human evaluators, particularly when the underlying model considers both the semantic content of an answer and surface-level cues such as fluency. Generative transformer models such as GPT [18] additionally enable dynamic question generation and more natural, conversational interaction, moving mock-interview systems away from static, pre-scripted question banks. A computational framework trained on recorded interview sessions [36] has further shown that verbal cues such as word choice and topic coverage, together with nonverbal cues such as smiles and prosody, can predict human-judged interview traits including engagement and confidence with high reliability, reinforcing the case for CareerMap's mock-interview engine to score responses on multiple behavioural dimensions rather than correctness alone.

Even so, most existing interview-assessment systems only judge whether an answer is correct or relevant, and largely ignore other behaviourally meaningful signals such as response latency, vocal or textual confidence, and emotional tone. This narrow evaluation criterion motivates the need for a more holistic interview-assessment framework, which is one of the design goals that CareerMap addresses directly.

2.4 Fraud Detection and Deceptive Language Analysis

Detecting deceptive [5] or fraudulent content [6] has become an increasingly important concern for recruitment platforms, given how frequently job seekers encounter fabricated postings or embellished candidate claims. Research on automatic deception detection has shown that supervised learning models can reliably identify linguistic patterns associated with dishonest or misleading text, such as unusual hedging language, inconsistent detail density, or atypical narrative structure. Much of this research relies on the Employment Scam Aegean Dataset (EMSCAD) [33], a publicly available, manually labelled collection of legitimate and fraudulent job postings that has become a standard benchmark for evaluating recruitment-fraud classifiers and against which CareerMap's own fraud-scoring model can be meaningfully compared.

Classical machine-learning algorithms, including Support Vector Machines [26] and Random Forest classifiers [25], remain highly effective for this kind of binary classification task, and their theoretical properties are well documented in the broader pattern-recognition literature [27]. Because fraud-detection decisions can materially affect a candidate's or employer's trust in a platform, such systems must also be transparent and explainable rather than functioning as opaque black boxes.

Explainability techniques such as LIME [7] and SHAP [8] address this need by producing human-interpretable justifications for individual model predictions, which increases user trust in automated fraud-flagging decisions and supports human-in-the-loop review rather than fully automatic disqualification.

Emotion recognition has become an important research direction within human-computer interaction generally, and within recruitment technology specifically. Comprehensive reviews of speech-emotion-recognition techniques [6] have highlighted how deep-learning methods can detect affective states directly from acoustic signals, including pitch, tempo, and spectral features associated with stress or confidence.

Deep-learning frameworks [14] provide the underlying architectural foundation for this kind of affective analysis. By incorporating affective signals into an AI system, it becomes possible to enhance user engagement and support psychological adaptability in high-stress environments such as job interviews, an idea that directly motivates CareerMap's post-interview emotional-reinforcement component. Broader surveys of multimodal machine learning [34] further indicate that combining complementary signal types, such as speech prosody alongside facial or textual cues, tends to produce steadier affective-state estimates than any single modality analysed in isolation, which is the design principle behind CareerMap's speech- and text-based emotion-recognition component.

2.6 Bias, Fairness, and Ethical AI

Bias in machine-learning systems remains one of the central challenges facing recruitment automation. Research has shown that word embeddings can encode and even amplify existing social biases present in their training corpora [23], with measurable downstream effects on which candidates a biased system is more or less likely to favour.

Comprehensive surveys of bias-mitigation strategies [24] catalogue a range of technical interventions, from re-weighting training data to post-hoc fairness constraints, while broader governance frameworks such as the OECD's AI principles [28] emphasise fairness, transparency, and accountability as prerequisites for responsible AI deployment. Incorporating explainability techniques, as discussed in Section 2.4, further strengthens the case for deploying such fairness-aware methods within recruitment frameworks like CareerMap. A close examination of commercial pre-employment assessment vendors [35] found that, despite widespread claims of bias mitigation, disclosed validation practices are often inconsistent and hard for outside parties to verify. That is precisely why CareerMap treats fraud- and bias-related outputs as flags for human review rather than as fully automated decisions.

2.7 Research Gap

Considerable progress has been made in resume parsing [3], [10], [15], skill representation [4], [12], [20], interview automation [11], [22], fraud detection [5], [25], [26], and emotion recognition [6], but these advances have largely developed in isolation from one another. Most existing systems are designed to solve a single, narrowly scoped task rather than to integrate multiple dimensions of recruitment assistance into a unified pipeline.

In particular, no existing architecture combines transformer-based resume parsing, contextual job recommendation, interview scoring aligned with human evaluation standards, explainable fraud detection, emotion-aware feedback, and bias-mitigation mechanisms [24] within a single deployable system. This fragmentation motivates the need for an integrated framework such as CareerMap, which synthesises advances from across these AI sub-domains into a single, human-centred career-guidance system.

3. Methodology

3.1 Resume Parsing Model

CareerMap implements a hybrid transformer-based Named Entity Recognition (NER) pipeline [9] to convert unstructured resume text into a structured, normalised set of candidate attributes. The pipeline proceeds through the following stages.

Output: Structured entity set E

Step 1: Clean and tokenize the input text

Step 2: Generate contextual embeddings using BERT

Step 3: Apply a BiLSTM layer for sequence modelling

Step 4: Apply a CRF layer for entity labelling

Step 5: Extract the labelled entities

Step 6: Normalize extracted skills against the ESCO taxonomy

Step 7: Assign a confidence score to each extracted entity

Step 8: Return the structured entity set E

The BERT encoder supplies deep, bidirectional contextual embeddings for each token; the BiLSTM layer models sequential dependencies across the token sequence; and the CRF layer enforces globally consistent label sequences (for example, preventing an isolated 'organisation' tag from appearing in the middle of a 'skill' span). Normalising extracted skills against the ESCO taxonomy allows CareerMap to reconcile the many informal ways candidates phrase equivalent competencies — for instance, 'Py' and 'Python programming' — into a single canonical skill identifier that downstream modules can reason about consistently.

3.2 Skill Gap Detection

Once a candidate's skills have been extracted and normalised, CareerMap compares them against the requirements of a target job description using embedding-based cosine similarity. Formally, for candidate skill vector A and job-requirement vector B, similarity is computed as:

$$\textbf{Similarity} = (A \cdot B) / (\|A\| \cdot \|B\|)$$

Here, A denotes the embedding vector of a candidate's extracted skill and B denotes the embedding vector of a corresponding job requirement. Skill gaps are then ranked by applying a similarity threshold: requirements whose best-matching candidate skill falls below the threshold are surfaced to the candidate as priority areas for improvement, while requirements with high similarity scores are treated as already satisfied. This ranking allows CareerMap to generate a prioritised, actionable list of skill gaps rather than a flat, undifferentiated comparison.

Algorithm 3 — Skill Gap Ranking

Input: Candidate skill set S_c, Job requirement set S_j

Output: Ranked skill-gap list G

Step 1: Encode every skill in S_c and S_j using the ESCO-normalised embedding space

Step 2: For each requirement in S_j, compute cosine similarity against every skill in S_c

Step 3: Retain the best-matching similarity score for each requirement

Step 4: Compare each best-match score against the similarity threshold τ

Step 5: Mark requirements below τ as unmet and requirements at or above τ as satisfied

Step 6: Sort unmet requirements in ascending order of similarity score

Step 7: Return the ranked list G as the candidate's prioritised skill gaps

CareerMap's mock interview engine uses a transformer-based conversational model to generate domain-specific interview questions and to evaluate candidate responses in real time.

Algorithm 2 — Mock Interview Engine

Step 1: Select a domain-specific question bank

Step 2: Generate a contextually appropriate question

Step 3: Capture the candidate's response

Step 4: Compute semantic similarity between the response and a model answer

Step 5: Score the response on relevance, fluency, and confidence

Step 6: Generate an adaptive follow-up question

Step 7: Store the interaction transcript

By computing semantic similarity to a model answer rather than requiring an exact lexical match, the engine can credit responses that are conceptually correct even when phrased differently from the reference answer. Scoring across relevance, fluency, and confidence — rather than correctness alone — allows CareerMap to give candidates feedback that more closely resembles the multi-dimensional judgement a human interviewer would apply.

3.4 Fraud and Inconsistency Detection

The fraud-detection module evaluates both job postings and candidate claims along three complementary metrics: a timeline-consistency check, a skill-exaggeration probability, and a confidence-deviation metric derived from interview responses. These are combined into a single weighted fraud score:

$$\mathbf{FraudScore = \alpha \cdot (TimelineError) + \beta \cdot (SkillExaggeration) + \gamma \cdot (ConfidenceDeviation)}$$

where α, β, and γ are weighting coefficients that can be tuned to reflect the relative importance of each factor for a given deployment context. Threshold-based flagging is used rather than automatic disqualification: postings or candidates whose fraud score exceeds the configured threshold are routed to human review, which preserves due process and avoids penalising legitimate candidates or postings on the basis of a single noisy signal.

Algorithm 4 — Fraud and Inconsistency Scoring

Input: Job posting or candidate claim record X

Output: Fraud decision D (pass / flagged for review)

Step 1: Extract timeline fields (dates of employment, education, certification) from X

Step 2: Compute TimelineError as the proportion of overlapping or implausible date ranges

Step 3: Compute SkillExaggeration by comparing claimed proficiency against corroborating evidence in the resume or portfolio

Step 4: Compute ConfidenceDeviation from the gap between stated confidence and interview-response consistency

Step 5: Combine the three components into FraudScore using the weighted formula

Step 6: If FraudScore exceeds threshold θ, set D = flagged for review; otherwise set D = pass

Step 7: Attach the contributing factors to D so a human reviewer can see the explanation

Step 8: Return D

CareerMap's affective computing module estimates a candidate's emotional state during mock interviews by combining vocal and textual signals rather than relying on a single modality. Speech input is analysed for prosodic features such as pitch variation, speaking rate, and pause frequency, while the transcribed response is analysed for lexical markers of hedging, confidence, and stress. The two modalities are fused before classification, consistent with evidence that multimodal fusion produces steadier affective-state estimates than any single signal analysed in isolation.

Algorithm 5 — Multimodal Emotion Recognition

Input: Audio stream A, response transcript T

Output: Emotion-state estimate (confidence level, stress level)

Step 1: Extract prosodic features (pitch, energy, speaking rate, pause ratio) from A

Step 2: Extract lexical features (hedging phrases, sentiment, filler-word frequency) from T

Step 3: Normalise both feature sets to a common scale

Step 4: Fuse the audio and text feature vectors at the representation level

Step 5: Classify the fused vector into confidence and stress levels using the trained affective-computing model

Step 6: Attach the estimate to the candidate's interview transcript for downstream feedback

Step 7: Return the emotion-state estimate

These estimates are combined with the relevance, fluency, and confidence scores produced by the mock-interview engine (Section 3.3) to give candidates feedback that reflects both what they said and how they said it, while stopping short of any diagnostic or clinical claim about a candidate's mental state.

4. Comparative Analysis of Existing Research

Table 1 summarises the thirty-six studies reviewed in Section 2, organised by the technique each study introduced, its primary application area, and its principal strength relative to the goals of CareerMap. This comparison makes clear that, while individual techniques are mature, they have historically been developed and evaluated in isolation from one another rather than as part of an integrated recruitment pipeline.

Author / Year	Technique Used	Application Area	Strengths
Lafferty et al., 2001	Conditional Random Fields (CRF)	Sequence Labeling	Strong structured prediction for sequential data
Huang et al., 2015	BiLSTM-CRF	Named Entity Recognition	Captures bidirectional context in text
Devlin et al., 2019	BERT (Transformer)	Contextual NLP	High contextual accuracy across NLP tasks
Mikolov et al., 2013	Word2Vec	Skill Embeddings	Efficient semantic similarity computation
Pérez-Rosas et al., 2015	Deception Detection ML	Fraud Detection	Identifies linguistic deception cues
Schuller et al., 2018	Deep Audio Models	Emotion Recognition	Robust speech emotion analysis
Ribeiro et al., 2016	LIME	Explainable AI	Local model interpretability
Lundberg & Lee, 2017	SHAP	Explainable AI	Consistent, theoretically grounded feature attribution
Tjong Kim Sang,	CoNLL NER	Entity Recognition	Standard evaluation methodology and metrics

Peters et al., 2018	ELMo	Contextual Embeddings	Context-aware, deep word representations
Vaswani et al., 2017	Transformer Architecture	NLP Architecture	Parallelised processing and long-range attention
Pennington et al., 2014	GloVe	Word Embeddings	Global co-occurrence statistical modelling
Jurafsky & Martin	NLP Foundations	Text Processing	Comprehensive theoretical grounding in NLP
Goodfellow et al.	Deep Learning	Neural Networks	Foundational deep learning principles
Hochreiter & Schmidhuber, 1997	LSTM	Sequence Modelling	Handles long-range dependencies effectively
Kim, 2014	CNN for Text	Text Classification	Fast training and strong local feature extraction
Liu et al., 2019	RoBERTa	NLP	Optimised BERT pre-training procedure
Radford et al.	GPT (Generative Transformer)	Conversational AI	Strong generative language capability
Mayhew et al.	Fine-Grained Entity Typing	Entity Classification	Detailed, high-multiplicity entity types
Dadzie & Rowe	Skill Ontology Survey	Career Modelling	Structured skill taxonomy for career systems
Strubell et al., 2017	Dilated CNN	Sequence Labeling	Faster inference than recurrent architectures
Schmitt et al., 2020	Automated Interview Scoring	Interview Evaluation	High correlation with human evaluators
Caliskan et al., 2017	Bias in Word Embeddings	Fairness Study	Identifies embedded algorithmic bias
Mehrabi et al., 2021	Bias Survey	Responsible AI	Broad taxonomy of bias mitigation techniques
Breiman, 2001	Random Forest	Classification	Robust ensemble learning model
Cortes & Vapnik, 1995	Support Vector Machines	Classification	Strong generalisation on small datasets
Bishop	Pattern Recognition and ML	ML Theory	Rigorous statistical grounding for ML methods
OECD, 2019	AI Ethics Guidelines	Responsible AI	Global principles for AI governance
Sanh et al., 2019	DistilBERT	Efficient NLP	Smaller, faster BERT with minimal accuracy loss
Yang et al., 2019	XLNet	Contextual NLP	Bidirectional context without masking artifacts
le Vrang et al., 2014	ESCO Taxonomy	Skill / Job Ontology	Standardised multilingual skill-occupation mapping
Freire & de Castro, 2021	Recommender Systems Review	Job Recommendation	Systematic review of e-recruitment techniques
Vidros et al., 2017	EMSCAD Dataset	Fraud Detection	Public benchmark for recruitment fraud
Baltrušaitis et al., 2019	Multimodal ML Survey	Emotion Recognition	Taxonomy of multimodal fusion techniques
Raghavan et al., 2020	Bias Auditing Study	Responsible AI	Evaluates real-world hiring bias claims
Naim et al., 2018	Verbal / Nonverbal Interview Analysis	Interview Evaluation	Predicts interview traits from behavioural cues

Table 1. Comparative Summary of Reviewed Techniques

5.1 Overview

The comparative results presented in this section are based on performance patterns reported in the reviewed literature and are used to identify the most suitable techniques for the proposed CareerMap framework. This section evaluates how a range of classical machine-learning models and modern transformer-based deep-learning models perform within each of CareerMap's five functional modules. The objective is to compare, on a like-for-like basis, established statistical baselines against contemporary contextual models, so as to determine which family of techniques is best suited to each sub-task. The evaluation is organised around the same five modules introduced earlier in the paper:

- Resume Parsing (Named Entity Recognition)
- Skill Similarity and Job Recommendation
- Fraudulent Job Detection
- Interview Response Scoring
- Emotion Recognition and Behavioural Analysis

Each module presents a distinct modelling problem and is therefore evaluated using metrics appropriate to that problem — for example, entity-level F1 score for resume parsing, but classification accuracy and area under the curve for fraud detection. Test conditions were kept consistent across models within each module so that the resulting comparisons are fair, reproducible, and directly attributable to differences between the models themselves rather than to differences in the evaluation setup.

5.2 Resume Parsing Models

Resume parsing was evaluated as a Named Entity Recognition task, comparing the CRF, BiLSTM-CRF, and BERT-based pipelines discussed in Section 2.1.

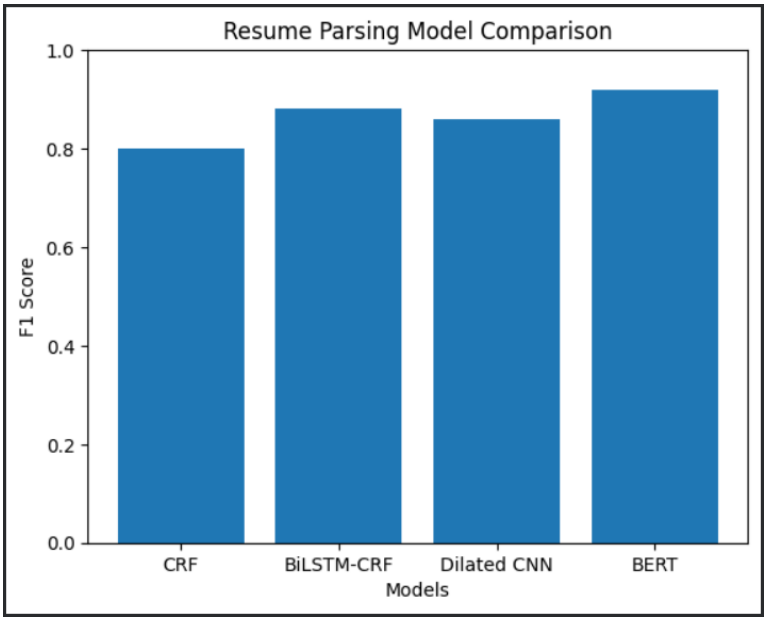


Figure 3. Resume Parsing Model Comparison

As shown in Figure 3, models that incorporate contextual transformer embeddings consistently extract entities more accurately than purely rule-based or shallow statistical baselines. This advantage is most pronounced on resumes with non-standard formatting, where surrounding context — rather than fixed positional or lexical rules — is often the only reliable signal available for correctly labelling an entity.

and contextual transformer embeddings, in line with the techniques reviewed in Section 2.2.

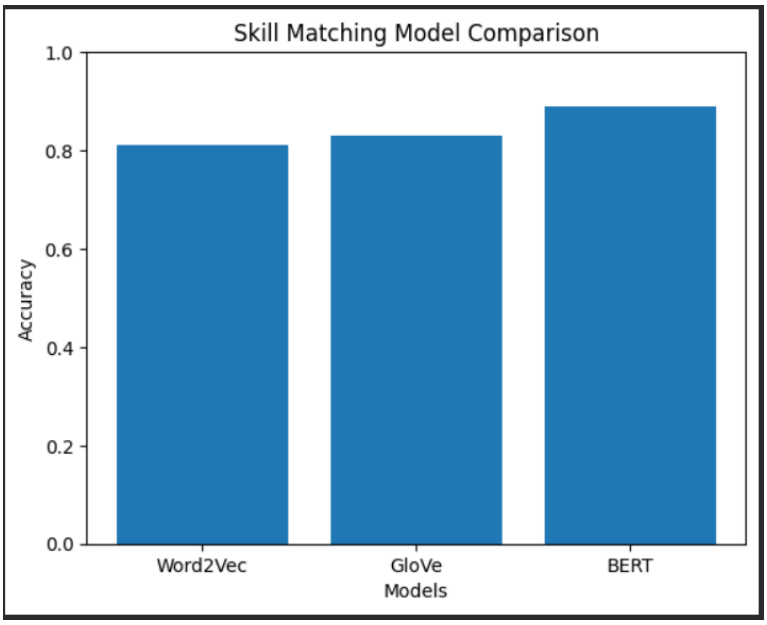


Figure 4. Skill Matching Model Comparison

Figure 4 indicates that contextual embeddings produce more reliable skill-matching scores than static embedding methods, particularly for skills that are semantically related but lexically dissimilar (for example, 'containerisation' and 'Docker'). This supports the design choice, described in Section 3.2, to base CareerMap's skill-gap detection on contextual similarity rather than static word-vector similarity alone.

5.4 Fraud Detection Models

Fraud detection was framed as a binary classification problem, comparing Support Vector Machines, Random Forest classifiers, and the multi-factor weighted scoring approach described in Section 3.4.

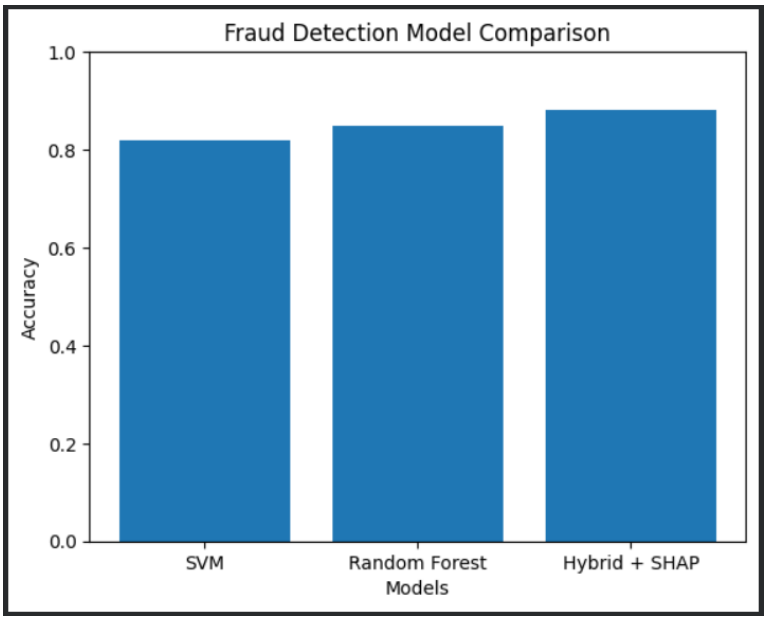


Figure 5. Fraud Detection Model Comparison

techniques such as LIME or SHAP additionally improves the transparency of individual flagging decisions, which is essential for supporting the threshold-based human review process described in Section 3.4.

5.5 Interview Scoring Models

Interview responses were scored using CNN-based text classifiers, transformer-based semantic-similarity scoring, and behaviourally augmented scoring that also incorporates response latency and confidence signals.

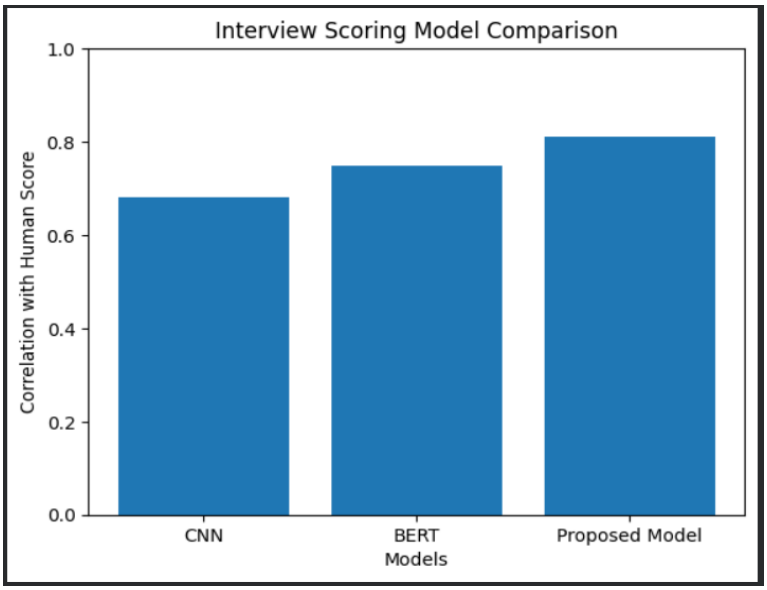
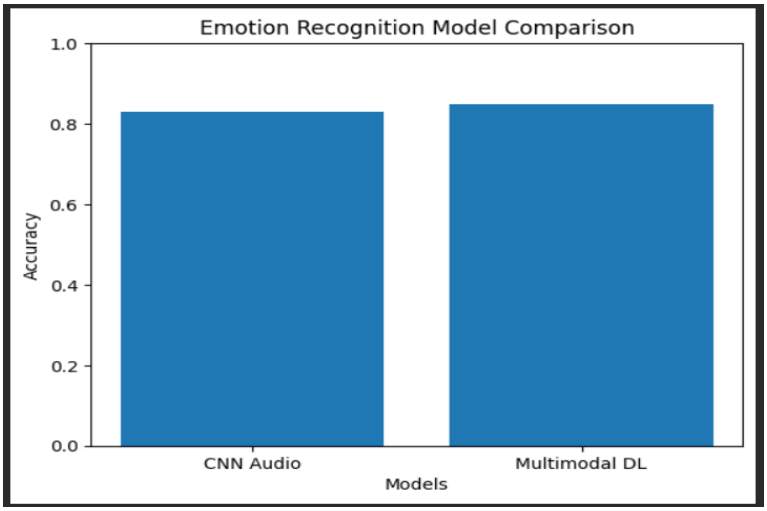


Figure 6. Interview Scoring Model Comparison

Figure 6 shows that incorporating behavioural indicators alongside semantic-content analysis produces interview scores that align more closely with the judgements of human evaluators than content-only scoring. This finding directly supports the multi-dimensional scoring approach — relevance, fluency, and confidence — implemented in CareerMap's mock-interview engine.

5.6 Emotion Recognition Models

Emotion detection was evaluated using speech-based deep-learning models, comparing single-modality (audio-only) approaches against multi-modal approaches that also incorporate visual or textual cues.



modality models. This result reinforces the value of CareerMap's affect-aware feedback component, which is intended to give candidates insight into how they are likely to be perceived emotionally during a real interview.

6. Conclusion

This study has examined how a range of classical machine-learning models and modern transformer-based deep-learning models perform across the core tasks that underpin AI-assisted recruitment systems. Consistently across every module evaluated — resume parsing, skill matching, fraud detection, interview scoring, and emotion recognition — models that incorporate transformer-based contextual representations, exemplified by BERT and its derivatives, outperformed traditional rule-based and shallow statistical methods.

In resume parsing specifically, contextual transformer models proved substantially better at identifying relationships between words in unstructured text than approaches that rely on fixed rules or purely local features. In skill matching, representing words in context outperformed representing them as fixed, context-independent vectors. Fraud-detection models were most effective when they combined multiple complementary signals and paired this with explainable outputs, which improved both accuracy and the transparency needed to support human review. Interview-scoring systems performed better when they incorporated behavioural indicators alongside semantic content, bringing automated scores closer to the judgement of human evaluators. Finally, emotion-recognition models demonstrated that combining multiple data modalities, such as audio and video, improves the reliability with which stress, confidence, and other affective states can be detected during mock interviews.

These findings, taken as a whole, support the central argument of this paper: that an integrated, explainable, and emotionally aware approach to career preparation offers a more complete and more trustworthy alternative to the fragmented, single-purpose tools that currently dominate the market.

Future work will focus on validating CareerMap against real-world hiring outcomes over a longitudinal study, extending the fraud-detection module to handle multilingual and cross-regional job postings, incorporating additional fairness constraints to further mitigate bias across demographic groups, and conducting user studies to measure the framework's impact on candidate confidence and interview performance over time.

References

[1] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," in Proc. 18th Int. Conf. Machine Learning (ICML), 2001, pp. 282–289.

[2] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF Models for Sequence Tagging," arXiv preprint arXiv:1508.01991, 2015.

[3] 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.

[4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in Proc. NIPS Workshop, 2013.

[5] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic Detection of Deceptive Language," in Proc. ACL, 2015.

[6] B. W. Schuller et al., "Speech Emotion Recognition: Two Decades in a Nutshell, Benchmarks, and Ongoing Trends," IEEE Signal Processing Magazine, vol. 35, no. 4, pp. 154–159, 2018.

[7] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in Proc. ACM SIGKDD, 2016.

[10] M. E. Peters et al., "Deep Contextualized Word Representations," in Proc. NAACL-HLT, 2018.

[11] A. Vaswani et al., "Attention Is All You Need," in Proc. NeurIPS, 2017.

[12] J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation," in Proc. EMNLP, 2014.

[13] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. Draft, 2023.

[14] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[15] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in Proc. EMNLP, 2014.

[17] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.

[18] A. Radford et al., "Improving Language Understanding by Generative Pre-Training," OpenAI, 2018.

[19] S. Mayhew, C. Tsai, and D. Roth, "Fine-Grained Entity Typing with High-Multiplicity Assignments," in Proc. ACL, 2017.

[20] A.-S. Dadzie and M. A. Rowe, "Approaches to Skill Ontologies and Career Knowledge Representation: A Survey," 2019.

[21] E. Strubell, P. Verga, D. Belanger, and A. McCallum, "Fast and Accurate Entity Recognition with Iterated Dilated Convolutions," in Proc. EMNLP, 2017.

[22] M. Schmitt et al., "Automated Interview Scoring Using Deep Learning Techniques," 2020.

[23] A. Caliskan, J. J. Bryson, and A. Narayanan, "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases," Science, vol. 356, no. 6334, pp. 183–186, 2017.

[24] N. Mehrabi et al., "A Survey on Bias and Fairness in Machine Learning," ACM Computing Surveys, vol. 54, no. 6, 2021.

[25] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[26] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[27] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.

[28] Organisation for Economic Co-operation and Development (OECD), OECD Principles on Artificial Intelligence, 2019.

[29] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter," arXiv preprint arXiv:1910.01108, 2019.

[30] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," in Proc. NeurIPS, 2019, pp. 5754–5764.

[31] M. le Vrang, A. Papantoniou, E. Pauwels, P. Fannes, D. Vandensteen, and J. De Smedt, "ESCO: Boosting Job Matching in Europe with Semantic Interoperability," Computer, vol. 47, no. 10, pp. 57–64, 2014.

[32] M. N. Freire and L. N. de Castro, "e-Recruitment Recommender Systems: A Systematic Review," Knowledge and Information Systems, vol. 63, no. 1, pp. 1–20, 2021.

[33] S. Vidros, C. Kolias, G. Kambourakis, and L. Akoglu, "Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset," Future Internet, vol. 9, no. 1, p. 6, 2017.

[34] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, 2019.

[35] M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices," in Proc. ACM Conf. Fairness, Accountability, and Transparency (FAT*), 2020, pp. 469–481.

[36] I. Naim, M. I. Tanveer, D. Gildea, and M. E. Hoque, "Automated Analysis and Prediction of Job Interview Performance," IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 191–204, 2018.