When AI Meets Recruiting: Opportunities, Challenges, and Future Directions

Yining Zhang*, Renjie Cao*, Zhilin Wang
OpenJobs AI

***Abstract*—While talent acquisition is a critical organizational function, traditional lexical filtering methods exhibit limited efficacy in extracting high-dimensional semantic signals from unstructured applicant data. This review addresses the gap in existing literature regarding recent advancements in AI by proposing a systematic framework connecting these technologies to specific recruitment stages. We synthesized cross-disciplinary literature published between 2020 and 2025 and surveyed contemporary AI-driven recruitment tools to capture the early-stage transition from discriminative to generative applications. To align computational capabilities with human resource requirements, this paper contributes a comprehensive taxonomy organized by the recruitment lifecycle—encompassing job posting, candidate matching, and assessment. Our synthesis centers on an end-to-end recruitment pipeline that orchestrates diverse artificial intelligence techniques to enable robust semantic representation and bi-directional person-job fit. We analyze how these integrations optimize data-intensive processes while exposing systemic challenges such as algorithmic bias and limited explainability. We conclude that the optimal division of labor—where automated systems handle quantitative sourcing and screening while human experts focus on high-entropy tasks like cultural assessment and complex negotiations—remains an open research question.**

***Index Terms*—AI Recruiting, Talent Acquisition, AI Screening.**

I. Introduction

TALENT acquisition has transitioned from an administrative process to a strategic organizational function. With the expansion of global labor markets and digital platforms,the volume of unstructured data has significantly increased. Current recruitment systems exhibit a low signal-to-noise ratio, complicating the extraction of actionable insights from large datasets. Thus, robust talent identification methods are necessary to improve operational efficiency and maintain market competitiveness.

A. The Evolving Landscape: Motivation and Urgency of AI Integration

Increasing applicant volumes have surpassed manual processing capabilities, necessitating a trade-off between throughput and selection precision [1], [2], [3]. Legacy keyword filtering remains a primary bottleneck, failing to identify high-potential talent or mitigate implicit human bias [4], [5], [6]. This discrepancy between data scale and legacy processing constraints requires the adoption of semantically aware computational models to preserve institutional competitiveness and hiring integrity [3], [7], [8].

Early rule-based frameworks provided baseline efficiency but lacked adaptability to unstructured data [9], [10]. Subsequent classical machine learning architectures, such as Support Vector Machines(SVMs) [11] and Random Forests [12], enabled probabilistic categorization but remained dependent on manual feature engineering [11], [12]. This reliance constrained the processing of unstructured semantics, creating a performance ceiling for discriminative approaches and necessitating the transition to flexible, representation learning architectures [13].

The shift toward deep neural architectures automated semantic representation. [13]. Neural models, specifically those utilizing Word2Vec [14] and Transformer-based encoders such as BERT [15], capture the latent contextual relationships within recruitment documents that previous methods overlooked [16], [17], [8]. These discriminative models map candidate profiles and job requirements into high-dimensional embedding spaces, facilitating a more granular alignment than legacy lexical approaches; structural innovations such as dual-tower architectures permit the independent encoding of resumes and job descriptions into a shared latent space for efficient similarity computation via cosine distance or dot product operations [18], [7]. Concurrently, Graph Neural Networks (GNNs) address the structural complexities of professional histories by modeling relational dependencies between skills, previous roles, and educational institutions as nodes within a heterogeneous talent graph [19], [20]. By transforming discrete textual inputs into continuous semantic vectors, these architectures provide a robust framework for quantifying person-job fit and mitigating the longstanding semantic gap between applicant capability and organizational demand [7], [4].

The emergence of Large Language Models (LLMs), such as GPT-4 [21] and Llama, signals an emerging trend from discriminative classification to generative reasoning within recruitment workflows [22], [23], [24]. While traditional discriminative models focus on binary matching, generative architectures are beginning to transition the system from a passive filter to a proactive virtual analyst capable of deep contextual synthesis [24], [25], [22]. These models do not merely calculate similarity scores between discrete vectors; instead, they interpret the nuanced interplay between candidate trajectories and complex organizational requirements through high-dimensional attention mechanisms [17], [22]. By leveraging in-context learning and zero-shot reasoning, LLMs

* equal contribution.
This manuscript was partially prepared by Renjie and Yining during their internships at OpenJobs AI.

facilitate a transition from rigid person-job matching to a framework capable of dual understanding and generation [21], [26], [22]. This enables the autonomous creation of tailored interview protocols, the extraction of implicit soft skills from diverse textual artifacts, and the formulation of hiring recommendations based on multifaceted deductive logic [25], [24], [22]. Consequently, the recruitment pipeline evolves into an automated analytical workflow where AI agents execute complex assessment tasks previously dependent on human expertise, thereby expanding the computational scope of talent identification and evaluation [25], [24].

B. Operational Efficacy of AI Integration in Recruitment

Integrating AI into recruitment workflows enhances operational throughput by automating high-volume, repetitive tasks that historically constrained scalability. Traditional manual screening is labor-intensive, frequently creating bottlenecks during the initial stages of the recruitment lifecycle. Conversely, automated systems process large-scale datasets of resumes and applications more efficiently than manual methods. This increased processing capacity reduces time-to-hire and broadens the evaluated candidate pool, enabling organizations to apply consistent evaluation criteria across numerous applicants without a linear increase in administrative overhead. Consequently, the recruitment function shifts from a resource-limited administrative gatekeeper to a scalable selection mechanism capable of managing global labor market fluctuations in real-time.

The transition from lexical, keyword-based retrieval to latent semantic matching represents a significant methodological advancement in talent identification and organizational alignment [4], [7], [22]. Early e-recruitment systems relied on exact string matching, frequently overlooking qualified candidates whose experience was described through latent semantic variations or synonymous terminology. Modern Natural Language Processing (NLP) techniques solve these limitations by parsing job descriptions to capture subtle skills and converting them into structured data. Furthermore, recruitment recommendation systems increasingly adopt a reciprocal logic, recognizing that effective hiring requires mutual suitability between the candidate's preferences and the employer's requirements [27], [28], [29]. Unlike standard recommendation tasks, this dual-sided optimization accounts for job seeker interests alongside employer-defined skill alignments. By capturing these complex interdependencies, AI-driven models facilitate a more sophisticated equilibrium in the labor market, ensuring that recommendations satisfy the specific objectives of both stakeholders while minimizing the noise inherent in traditional search-and-match methodologies.

Beyond text-based analysis, the integration of multimodal assessment tools provides a holistic view of candidate potential by incorporating audio, visual, and behavioral data streams [30], [31], [32]. Structured video interviews and gamified assessments allow for the evaluation of soft skills and professional etiquettes that resume parsing alone cannot capture. This transition to acomprehensive multimodal assessment framework reduces the reliance on historical credentials, focusing instead on real-time performance metrics and cognitive traits. Simultaneously, the recruitment paradigm moves from a reactive, need-based process toward a proactive strategic forecasting model [33], [34], [2]. Predictive analytics enable organizations to anticipate talent gaps before they emerge by analyzing internal turnover patterns and external market trends. This forward-looking approach transforms talent acquisition into a strategic function that informs long-term workforce planning and resource allocation. By synthesizing diverse data inputs into structured predictive metrics, AI-driven systems assist hiring managers in mitigating traditional procedural constraints, facilitating a human-in-the-loop framework that improves selection accuracy and resource planning.

C. Survey Methodology

To ensure a rigorous and reproducible synthesis of the field, we conducted a systematic literature review by querying premier academic databases, including the ACM Digital Library, IEEE Xplore, the ACL Anthology, and Google Scholar. The search strategy employed boolean combinations of domain-specific keywords, such as: (”Recruitment” OR ”Talent Acquisition” OR ”Person-Job Fit” OR ”Resume Parsing”) AND (”Artificial Intelligence” OR ”Machine Learning” OR ”Large Language Models” OR ”Graph Neural Networks” OR ”Fairness” OR ”Algorithmic Bias”).

Regarding the temporal scope, our work synthesizes cross-disciplinary literature focusing on recent advances—particularly the emerging methodological shifts brought by deep learning and Generative AI—while deliberately including seminal works to establish a foundational technical and sociological baseline.

Our inclusion criteria were strictly limited to peer-reviewed publications from top-tier Computer Science venues (e.g., KDD, SIGIR, ACL, AAAI, NeurIPS, WWW), high-impact journals, and highly relevant preprints that introduce novel algorithmic architectures or benchmark datasets. Purely qualitative management papers lacking formal computational methodologies were excluded, with the exception of authoritative human resource guidelines (such as the SHRM frameworks) which are strictly necessary to ground the algorithmic taxonomy in real-world operational definitions [35], [36].

To anchor this academic synthesis in applied deployment contexts, we also surveyed contemporary AI-powered recruitment tools in the human resources market. Specifically, systems such as Mira [37] are introduced as representative case studies. The explicit system selection criteria for this inclusion required a platform that demonstrates an autonomous, end-to-end integration of specialized sub-agents—encompassing search, matching, and candidate engagement modules—within a continuous pipeline. Mira serves as a representative implementation for this systemic integration because it effectively operationalizes complex theoretical concepts into a deployed, multi-agent architecture. Most notably, it utilizes a dedicated Evaluation agent that acts as a system-level guardrail to continuously validate output quality, monitor execution health, and enforce policy compliance.

D. Scope and Contributions

Our work synthesizes cross-disciplinary literature, focusing on recent advances and seminal works. We survey contemporary AI-driven recruitment systems (e.g., Mira [37]) to examine the emerging transition from discriminative to generative paradigms in talent acquisition. Integrating technical advancements in AI with socio-technical frameworks from human resource management establishes a rigorous theoretical foundation for AI-driven recruiting.

However, existing literature frequently omits these recent developments, particularly regarding LLMs and Retrieval-Augmented Generation (RAG). Furthermore, the research tends to be compartmentalized: computer science studies prioritize algorithmic metrics, while management research often lacks technical depth regarding multimodal analysis. There is currently no lifecycle-oriented taxonomy or generative-AI-specific framework connecting Generative Artificial Intelligence (Generative AI) technologies to specific recruitment stages while addressing explainability and ethical compliance.

This paper surveys the intersection of AI and recruitment, aiming to align technical capabilities with human resource requirements. We introduce a comprehensive taxonomy based on the recruitment lifecycle (encompassing job posting, matching, and assessment) rather than traditional algorithmic categories. We then analyze how generative methods extend beyond predictive modeling to facilitate downstream applications such as job description debiasing and automated interviewing.

II. Background

A. Recruitment Life Cycle

According to the professional guidelines and recruitment toolkits provided by the Society for Human Resource Management (SHRM), Full-Cycle Recruiting—also referred to as "end-to-end recruiting"—encompasses the comprehensive process ranging from the initial identification of a hiring need to the successful onboarding of a new employee. While operational nuances may exist depending on organizational scale, the official SHRM framework typically categorizes this process into the following six core stages [35], [36].

Structuring our review around this SHRM lifecycle, rather than relying on traditional algorithmic taxonomies (e.g., classification, clustering, generation), provides a vital operational alignment rationale. In the context of Computer Science and Machine Learning, isolated algorithmic optimizations—such as minimizing embedding distance in a shared latent space—often fail to translate into practical utility without a deep understanding of the target deployment environment. By mapping Computer Science methodologies directly to the SHRM lifecycle, we establish a robust socio-technical framework. This approach clarifies the distinct data constraints, latency requirements, and human-in-the-loop interactions inherent to each functional phase. Consequently, it bridges the gap between abstract computational objectives and concrete organizational outcomes, ensuring that AI-driven interventions are evaluated holistically within their actual operational context.

To successfully operationalize this socio-technical framework, it is crucial to delineate the architectural boundary between offline model training and online inference across the recruitment pipeline [38]. The offline module predominantly governs the computationally intensive phases of representation learning and historical data assimilation. Within this offline environment, deep neural encoders map heterogeneous entities—such as candidates and job descriptions—into a shared latent space X, capturing complex, non-linear interactions prior to deployment [39]. This division also facilitates offline indexing and the longitudinal modeling required to resolve delayed feedback learning problems inherent in hiring. Conversely, the online inference module must satisfy strict temporal constraints to execute real-time decision support. It manages low-latency operations such as MIPS during the retrieval phase [40], [41], dynamic candidate engagement, and the stochastic progression of candidates through sequential lifecycle stages. Establishing this firm division ensures that while underlying models continuously optimize their joint probability distributions $P(u, j)$ offline, the online inference engines maintain the requisite operational throughput for interactive and asynchronous recruitment workflows.

The lifecycle comprises the following functionally distinct stages:

1) Preparation: This phase centers on Job Analysis, where recruiters and hiring managers define core responsibilities and performance metrics. By identifying essential competencies and knowledge, skills, abilities, and other characteristics required for the role(KSAOs), they establish standardized job descriptions, providing a scientific, objective framework for subsequent talent sourcing and assessment.

2) Sourcing: The sourcing stage focuses on cultivating high-quality talent pipelines by integrating internal and external channels. Recruiters leverage employee referrals and professional networks to engage active and passive candidates, emphasizing Employer Value Proposition (EVP) and diversity. This ensures a robust, competitive applicant pool for subsequent selection.

3) Screening: This phase filters the applicant pool to identify qualified candidates by benchmarking credentials against Minimum Qualifications (MQs) via an Applicant Tracking System(ATS) or manual review. Preliminary interviews validate KSAOs and motivation, minimizing assessment costs and ensuring compliance while funneling high-potential talent into selection.

4) Selecting: This phase utilizes high-validity instruments—including structured interviews, psychometric testing, and work samples—to rigorously verify KSAOs. By emphasizing objective scoring and cultural alignment, this stage ensures legal compliance and optimizes organizational Return on Investment (ROI) through evidence-based decision-making.

5) Hiring: This phase converts selection outcomes into formal employment. Recruiters facilitate competitive total rewards negotiations, balancing budgetary constraints with candidate expectations. This stage prioritizes legal compliance, timely communication, and professional feedback to secure top talent and solidify long-term employer branding strategic value.

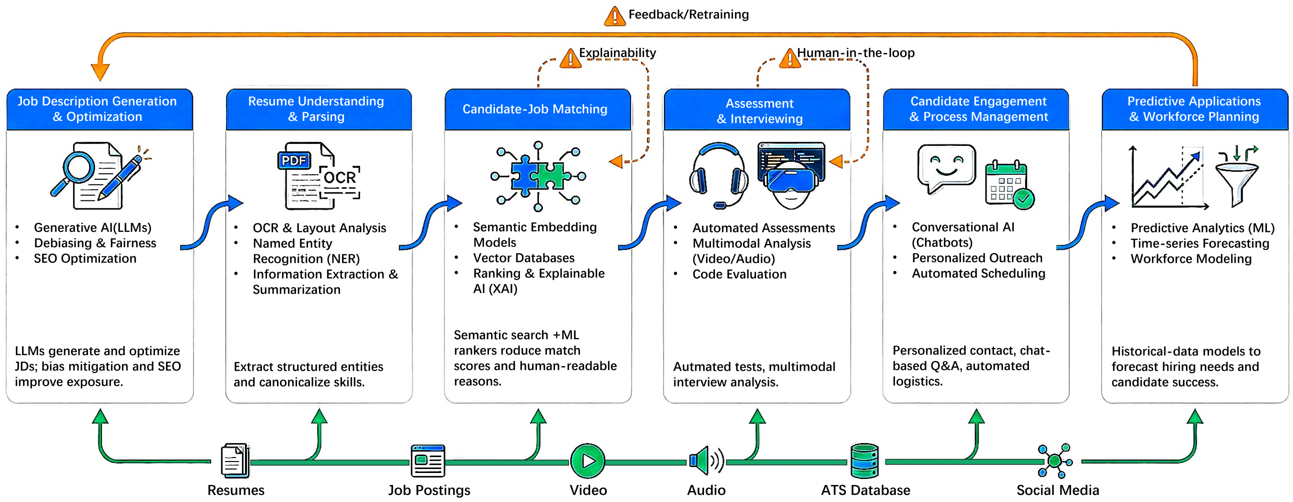

Fig. 1. An end-to-end pipeline of AI applications across the recruitment lifecycle. The framework illustrates six major functional stages: from (1) Job Description Optimization and (2) Resume Parsing to (3) Semantic Matching, (4) Multimodal Assessment, (5) Candidate Engagement, and (6) Predictive Workforce Planning. Each stage highlights core enabling technologies, such as Large Language Models (LLMs), Named Entity Recognition (NER), and Explainable AI (XAI). Detailed data flow is depicted in three primary directions: a sequential left-to-right progression between the functional stages (blue arrows), an upward ingestion of diverse data modalities—including text, audio, video, and structured databases—into specific stages from the bottom layer (green arrows), and a top-level reverse feedback loop (orange arrow) that routes insights from the final predictive stage back to the initial stage. Dashed annotations emphasize critical cross-cutting concerns including explainability, human-in-the-loop integration, and continuous feedback mechanisms for model retraining.

6) Onboarding: This phase facilitates strategic assimilation, focusing on cultural alignment, social cohesion, and role clarity to accelerate productivity. By implementing continuous support mechanisms, organizations enhance early engagement and long-term retention, ensuring a seamless transition from recruitment investment to measurable human capital value.

B. Formal Problem Formulation

To bridge the gap between operational human resource frameworks and rigorous computational methodologies, it is imperative to abstract the recruitment lifecycle into formal mathematical representations. By doing so, we transition from heuristic, stage-gate models to optimizable algorithmic processes that account for mutual preference, sequential dynamics, and temporal latency.

1) Person-Job Fit as Reciprocal Recommendation Problem: Unlike traditional unilateral recommendation systems (e.g., e-commerce product suggestions), matching within the talent acquisition ecosystem is fundamentally a bi-directional, reciprocal decision-making process. Both the candidate (user u) and the employer (job j) exert active selection preferences [42], [43], [44], [45].

Let $\mathcal{U}$ represent the latent space of candidate representations and $\mathcal{J}$ denote the latent space of job descriptions. The core computational objective is not merely unilateral relevance, but rather optimizing a joint probability distribution that captures mutual suitability. We formally define the objective as maximizing the joint probability $P(u, j)$, expressed as:

$$P(u, j) = \Phi(f(u \to j), g(j \to u))$$

$f(u \to j)$ quantifies the utility or preference of candidate u toward job j (capturing constraints like salary expectations, remote flexibility, and career trajectory). $g(j \to u)$ represents the employer's utility function evaluating candidate u (capturing the alignment of KSAOs, experience, and cultural fit). Φ serves as an aggregation or link function (e.g., a multi-layer perceptron or deep cross network) that synthesizes these asymmetric, directional utilities into a unified reciprocal match score [46], [47], [48].

2) The Recruitment Pipeline as a Sequential Decision Process: Beyond static alignment, the operational recruitment funnel must be modeled as a quantifiable, sequential pipeline. Instead of treating sourcing, screening, and interviewing as independent classification tasks, the trajectory of candidate u can be framed as a stochastic progression through discrete lifecycle stages [49], [50], [51].

Let $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ represent the sequential stages of the recruitment funnel. The candidate's progression is governed by transition probabilities $P(s_{i+1}|s_i)$ across sequential hiring stages. Predictive architectures aim to map historical behavioral trajectories to these transition dynamics. By accurately estimating $P(s_{i+1}|s_i)$, automated systems can optimize routing, anticipate candidate conversion rates, and proactively mitigate talent attrition at distinct bottlenecks within the pipeline [52], [53], [49], [50], [51].

3) Hiring as Delayed Feedback Learning: A defining challenge in algorithmic recruitment is the temporal dissociation between an AI system's immediate outputs and the actual realization of business value. This necessitates framing the terminal stages of the pipeline as a delayed feedback learning problem [54], [55].

At a given timestamp t, the algorithmic decision or action a_t (e.g., advancing a candidate to the final round or extending a compensation package) is executed based on proxy signals such as interview scores or offer acceptance likelihoods. However, the ultimate organizational reward r_{t+k}—defined by longitudinal metrics such as long-term retention, job performance, and cultural assimilation—only materializes after a significant temporal delay k [56]. Shifting the optimization focus from short-term efficiency proxies to this longitudinal career-success modeling is essential for sustainable organizational growth [57], [58]. Consequently, loss functions in predictive matching models must be designed to handle sparse, delayed reward signals, preventing algorithms from structurally over-indexing on easily quantifiable but myopic targets (e.g., mere offer acceptance) at the expense of long-term human capital ROI [2], [59].

C. Technical Paradigms

1) Matching and Ranking: The cornerstone of AI-driven recruitment lies in achieving high-precision alignment between heterogeneous entities. Person-Job Fit has evolved from rudimentary lexical keyword matching to sophisticated semantic architectures that map candidates and job descriptions into a shared latent space $\mathcal{X}$, capturing non-linear interactions through deep neural encoders [42], [43], [44], [45]. Modern paradigms frequently hybridize Content-based Filtering (CBF) and Collaborative Filtering (CF) to alleviate data sparsity, yet the distinguishing factor in recruiting is its *bi-directional* nature. Unlike traditional item recommendation, recruitment is a reciprocal decision-making process; thus, current models often formulate the objective as a joint probability $P(u, j) = \Phi(f(u \to j), g(j \to u))$, where f and g represent the mutual utilities between candidate u and job j [46], [47], [48]. To ensure the deployment of these models is both ethical and actionable, explainable ranking mechanisms have become indispensable, providing granular justifications such as skill-gap analysis or cultural affinity to bridge the transparency gap in automated screening [60], [61], [62].

2) Information Extraction and Standardization: Information Extraction (IE) serves as the infrastructural bedrock, transforming the inherent noise of unstructured resumes and job descriptions into machine-actionable schemata. This paradigm predominantly employs sequence labeling frameworks for Named Entity Recognition (NER) to isolate granular attributes such as $\mathcal{S}$ (skills), $\mathcal{E}$ (experience), and $\mathcal{D}$ (educational credentials) from diverse document layouts [63], [64], [65], [66], [67]. To mitigate the linguistic variance across candidate-generated content, the process further necessitates a transition from surface-level extraction to semantic standardization. Through entity disambiguation and ontology alignment, heterogeneous mentions—such as "Java Specialist" and "Backend Engineer"—are mapped onto canonical representations within a recruitment Knowledge Graph (KG) [68], [69], [70], [71].

3) Prediction and Decision Support: Beyond static alignment, the paradigm of Prediction and Decision Support leverages historical behavioral trajectories to model the recruitment funnel as a quantifiable pipeline. By estimating transition probabilities $P(s_{i+1}|s_i)$ across sequential hiring stages, predictive architectures can anticipate candidate conversion rates and offer acceptance likelihoods, effectively mitigating the risks of talent attrition [52], [53], [49], [50], [51]. This analytical layer further integrates combinatorial optimization and constraint satisfaction algorithms to streamline operational logistics, such as multi-objective interviewer scheduling and dynamic budget allocation across diverse sourcing channels. By shifting from heuristic-based workflows to prescriptive interventions, these models alleviate systemic bottlenecks and optimize the operational throughput of the talent acquisition lifecycle [53], [50], [72], [73], [74].

4) Generative Artificial Intelligence: The integration of Generative AI represents an early-stage transition from passive discriminative analysis to proactive, conversational interaction [75], [76]. LLMs facilitate the automated synthesis of high-fidelity job descriptions, tailored candidate outreach, and multi-dimensional evaluation summaries, thereby exploring ways to reduce the cognitive load on human recruiters [77], [78], [79], [80]. To address the inherent challenge of stochastic hallucinations in sensitive hiring contexts, RAG is deployed to anchor model outputs in verified corporate policies and legal frameworks, ensuring factual consistency. Beyond text generation, the emergence of LLM-based autonomous agents enables the orchestration of complex, multi-step workflows—such as cross-platform interview scheduling and candidate sentiment tracking—through external tool-calling mechanisms [81], [82].

5) Speech and Multimodal Analysis: The integration of Speech and Multimodal analysis transcends the limitations of text-centric evaluation by quantifying "non-verbal cues" that are pivotal to holistic interpersonal assessment. Beyond traditional Automatic Speech Recognition (ASR), acoustic paradigms extract prosodic and phonetic features to model candidate emotional stability and communication efficacy during initial screenings [83]. In synchronous video interviews, contemporary frameworks leverage multimodal fusion to synchronize facial micro-expressions, kinesic postures, and vocal tonality with textual semantics, mapping these heterogeneous signals into a joint latent representation $\mathcal{Z} = \Gamma(V, A, T)$ [84], [85]. This high-dimensional profiling is further fortified by cross-modal consistency checks, which serve as a critical defense mechanism against identity fraud and Deepfake-mediated deception, ensuring the ecological validity and integrity of remote evaluation environments [86], [87].

D. Evaluation Metrics

The robust assessment of AI-driven recruitment architectures necessitates a multi-dimensional evaluation framework that bridges computational efficacy with operational utility and ethical compliance. Rather than relying on a single loss function, system performance is benchmarked across three distinct taxonomic categories:

1) Offline Metrics (Algorithmic Performance):

- **Precision@K and Recall@K:** These rank-agnostic metrics evaluate the proportion of relevant matches within the top-K recommendations. Precision@K quantifies the exactness of the model by measuring the fraction of recommended candidates or jobs that are actually relevant. Recall@K measures completeness by capturing the fraction of all true relevant matches that were successfully retrieved in the top K [88], [61]. However, these metrics treat all positions within the top K equally. This limitation is particularly pronounced in recruitment platforms, where users exhibit severe exposure bias and rarely scroll past the first few results [46].
- **Normalized Discounted Cumulative Gain (NDCG@K):** To address position bias, NDCG@K is the predominant rank-aware metric in modern Person-Job Fit evaluation [89], [90], [28]. It assigns exponentially decaying weights to lower-ranked items, explicitly rewarding models that place highly relevant candidates or jobs at the very top of the list. For multi-graded relevance scores—such as varying degrees of semantic skill overlap—NDCG is highly robust because it normalizes the Discounted Cumulative Gain (DCG) against the Ideal DCG (IDCG), ensuring metric comparability across queries of varying lengths and candidate [91].
- **Mean Reciprocal Rank (MRR):** MRR is highly interpretable and evaluates how quickly a system retrieves the first relevant result. It is calculated as the average of the reciprocal ranks of the first relevant match across all queries [88]. In high-stakes recruitment scenarios, finding a single optimal candidate is frequently prioritized over returning a large pool of semi-relevant matches, making MRR a critical proxy for search efficiency [45], [46].

While these offline metrics provide a foundational assessment of predictive accuracy, they are inherently limited by the nature of historical interaction data. Models evaluated solely on these metrics frequently suffer from selection bias, as they are tested only on candidates who historically interacted with or were hired for a position, thus ignoring counterfactual outcomes for unobserved matches. Consequently, these algorithmic indicators must be systematically triangulated with business and fairness metrics to establish a holistic evaluation framework [61], [92].

2) Business Metrics (Operational Utility):

- **Time-to-Hire:** This metric measures the temporal latency between the initial job posting and the final candidate conversion. AI-driven throughput acceleration aims to compress this timeframe. By automating repetitive administrative overhead and cross-platform scheduling , these systems reduce the time-to-interview cycle, which directly mitigates the risk of talent attrition caused by scheduling delays in competitive labor markets [81], [82].
- **Interview Conversion Rate:**This rate tracks the successful progression of candidates through sequential decision points. Predictive architectures map historical behavioral trajectories to these transition dynamics, allowing systems to anticipate candidate conversion rates, optimize routing, and proactively address bottlenecks [50], [51].
- **Offer Acceptance Rate:**The terminal stages of the recruitment pipeline represent a delayed feedback learning problem, where immediate algorithmic actions are executed based on proxy signals such as offer acceptance likelihoods. Predictive models estimate these offer acceptance likelihoods to optimize decision-making interventions. Tracking this metric ensures that algorithmic optimizations do not structurally over-index on easily quantifiable short-term proxies at the expense of long-term human capital ROI [57], [58], [93].

3) Fairness Metrics (Algorithmic Equity): By embedding fairness metrics directly into the evaluation framework, researchers and human resource practitioners can systematically quantify the trade-offs between predictive utility and algorithmic fairness, enabling the targeted application of pre-processing, in-processing, or post-processing mitigation strategies.

- **Demographic Parity (Statistical Parity):**This metric mandates that the probability of a positive algorithmic decision (e.g., advancing a candidate to the interview stage, $\hat{Y} = 1$) must be statistically independent of a candidate's protected attribute A (such as gender, age, or race). Formally, it requires $P(\hat{Y} = 1|A = 0) = P(\hat{Y} = 1|A = 1)$ [94], [95]. Within the recruitment pipeline, Demographic Parity ensures that the baseline selection rate remains equitable across all demographic cohorts, functioning as a macro-level safeguard against disparate impact [96].
- **Equal Opportunity:**While Demographic Parity enforces equal selection rates across groups, it does not account for the underlying distribution of actual qualifications (the ground truth label $Y = 1$). Equal Opportunity addresses this limitation by requiring that the True Positive Rate (TPR) is identical across protected and unprotected groups. Mathematically, it is defined as $P(\hat{Y} = 1|Y = 1, A = 0) = P(\hat{Y} = 1|Y = 1, A = 1)$ [97]. In talent acquisition, this metric guarantees that a highly qualified candidate from a marginalized subgroup has the exact same algorithmic probability of being selected as an equally qualified candidate from the majority group, effectively balancing ethical compliance with operational validity [57].

III. Application of AI in the Recruitment Process

A. Job Description Generation and Optimization

1) Generating Job Descriptions using LLM: The traditional drafting of job descriptions relies heavily on manual expertise, rendering the process labor-intensive and susceptible to subjective bias. To standardize this pipeline, recent literature demonstrates a transition toward utilizing generative AI, moving away from rigid slot-filling mechanisms to more dynamic, context-aware generation.

Current applications predominantly employ LLMs optimized via instruction tuning, few-shot learning, and RAG. Instruction-tuned models are leveraged to embed organizational context, team technology stacks, and soft skill requirements directly into generation prompts, yielding highly targeted text [79]. To ground the generation process and mitigate semantic drift, RAG architectures are increasingly utilized . By retrieving high-frequency keywords and syntactic structures from verified repositories of historical job descriptions, these systems constrain the LLM output to align with established industry frameworks [81]. Furthermore, advanced pipelines employ LLMs for structural mapping, extracting specific entities from unstructured hiring manager inputs and aligning them with standardized skill taxonomies like ESCO (European Skills, Competences, Qualifications and Occupations) to facilitate downstream matching algorithms [82].

Despite these architectural advancements, deploying LLMs for job description generation introduces significant technical hurdles, primarily concerning latent bias amplification and structural ambiguity. Base language models frequently hallucinate or amplify social stereotypes present in their pre-training corpora, resulting in implicitly biased outputs (e.g., gender-coded or age-discriminatory language) that compromise fairness. While systems increasingly employ LLMs as bias "denoisers" via Chain-of-Thought (CoT) reasoning—programmatically substituting masculine-coded terms (e.g., "aggressive") with neutral equivalents (e.g., "driven") [76]—ensuring deterministic, robust debiasing across diverse linguistic contexts remains computationally complex. Additionally, an ongoing architectural challenge lies in parsing vague, heterogeneous human inputs into rigidly structured, machine-readable text; current systems struggle to enforce strict formatting without degrading the semantic nuance required for accurate downstream candidate matching.

Consequently, several critical research gaps persist beyond these immediate deployment constraints. First, there is a distinct lack of domain-specific evaluation metrics designed to rigorously quantify generative fairness; current methodologies evaluate explicit proxy term substitution but cannot adequately measure whether intersectional and latent sociocultural biases are eradicated without degrading the model's descriptive fluency. Second, while extraction architectures currently map entities to static taxonomies like ESCO [82], future research must develop dynamic, bidirectional ontologies capable of autonomously updating LLM-generated text to reflect rapidly evolving, real-world skill requirements. Finally, a significant theoretical gap exists in aligning job description generation with Generative Engine Optimization (GEO) [98]. The community must explore how generative models can iteratively refine unstructured postings through query expansion and structured schema generation to maximize retrieval probability and favorable exposure against zero-shot rankers in AI-mediated search environments [99].

2) Search Engine Optimization: LLMs optimize job postings to maximize their retrieval probability in both traditional search engines and generative AI interfaces. This approach supports Generative Engine Optimization, where frameworks iteratively refine content using rhetorical strategies, such as including statistical evidence, to increase visibility in AI-generated responses [98]. Beyond surface-level visibility, generative models address semantic alignment by using query expansion to identify potential search terms, thereby bridging the vocabulary gap between job requirements and user queries [100]. Additionally, unified information extraction frameworks convert unstructured descriptions into structured schemas containing key entities like salary ranges and remote work eligibility, which improves indexing by aggregators [101]. Prior to publication, these models also serve as zero-shot rankers that simulate algorithmic prioritization to predict exposure against competitor listings [99].

B. Candidate-Job Matching: From Semantic Alignment to Generative Reasoning

Accurate matching of candidates to job descriptions (Person-Job Fit) is central to recruitment. Methodologies in this domain have evolved from keyword-based information retrieval to latent representation learning, and recently, to generative reasoning enabled by LLMs. These models allow for contextual understanding that goes beyond surface-level text matching.

1) Advanced Semantic Representation and Graph-Based Matching: Early resume-job matching relied on statistical relevance models to bridge the vocabulary gap between semi-structured resumes and job descriptions [44]. However, these methods often failed to capture the contextual nuance of professional skills, resulting in superficial alignment. Recent work focuses on mapping unstructured text into shared high-dimensional embedding spaces. For instance, CareerBERT [102] uses domain-specific pre-training to map resumes to standardized ontologies (e.g., ESCO). This unifies diverse terminologies to enable generic job recommendations even without interaction data. To address data scarcity in specific domains, techniques like ConFit [103] use contrastive learning and data augmentation to refine the semantic alignment between resumes and job descriptions.

Beyond static text matching, the industry has shifted toward modeling relationships within professional networks using GNNs. LinkedIn's LinkSAGE framework advances this approach by employing an encoder-decoder GNN on a heterogeneous graph of billions of nodes to optimize matching at scale [62]. To operationalize these semantic representations within latency-constrained production environments, contemporary matching architectures typically adopt a decoupled retrieve-and-rank paradigm [89]. In the retrieval phase, dual-tower models encode candidate profiles and job descriptions independently into a shared latent space, permitting efficient offline indexing and low-latency maximum inner product search (MIPS) for candidate retrieval [104]. However, because dual-tower architectures delay the interaction between query and document representations, they often sacrifice granular semantic alignment [105]. To address this, the subsequent ranking phase frequently employs cross-encoder architectures. While computationally prohibitive for full-corpus retrieval due to the necessity of processing the candidate-job pair jointly, cross-encoders facilitate deep token-level attention interactions, thereby optimizing high-accuracy ranking over a truncated candidate shortlist [106]. Balancing these online and offline serving constraints remains a critical architectural consideration for high-throughput signal integration systems and learning-to-retrieve frameworks [90].

2) Modeling Reciprocity and Career Trajectories: A distinct characteristic of recruitment, compared to product recommendation, is its reciprocity: a successful match requires preference from both the job seeker and the recruiter. Recent literature emphasizes the need to move beyond unilateral matching. The MIRROR framework proposes a multi-view approach, modeling users through search, active, and passive views to capture reciprocal interactions [107]. Similarly, [108] explicitly model two-way selection preferences to align expectations bet ween parties, promoting long-term employment relationships.

Temporal dynamics are also critical for matching quality. [109] argue that a candidate's career path contains hidden preferences regarding consistency and continuity, which predict future job fit. Validating this behavioral perspective, historical interview choices have been shown to reveal latent preferences of recruiters and candidates that static text cannot capture. This has led to the development of profiling memory modules to improve matching accuracy [110].

3) LLM-Driven Ranking, Reasoning, and Explainability: LLMs introduce new capabilities to person-job fit, including the ability to reason, rank, and explain matches in zero-shot or few-shot settings. Generative job recommendation frameworks use the extensive world knowledge in LLMs to interpret complex career profiles without task-specific training [111].

While generative architectures introduce unprecedented contextual understanding, their deployment as zero-shot or few-shot rankers reveals significant technical limitations. First, the high computational inference costs of LLMs render them impractical for large-scale, low-latency screening [112]. Second, generative models are highly susceptible to ranking instability, often manifesting as position bias where the model structurally overvalues candidates presented at the beginning or end of the prompt context window [113].

Furthermore, in zero-shot settings lacking domain-specific anchoring, LLMs are prone to hallucinatory reasoning, where they may fabricate skill alignments or misinterpret latent trajectory signals [114]. Consequently, an observational study comparing GPT-4 with human recruiters found that while prompt engineering techniques like CoT improve performance, LLM scoring is not yet fully interchangeable with human judgment [115]. This necessitates rigorous human-in-the-loop verification to mitigate these epistemic and operational vulnerabilities, although LLMs do not exhibit higher demographic bias than human baselines [115], [116].

To bridge the gap between black-box recommendations and user trust, the ReasoningRec framework uses LLMs to generate interpretable explanations alongside recommendations, identifying why a candidate fits a specific role. Finally, the field is expanding toward agent-based simulation. Frameworks like MockLLM [117] use multi-agent collaboration to simulate interactions between job seekers and recruiters, allowing for the optimization of matching strategies in a dynamic environment before deployment. These generative approaches mark a transition from calculating similarity scores to understanding the reasoning behind hiring decisions.

C. Candidate Assessment and Interviewing

Recruitment relies fundamentally on robust candidate evaluation and interviewing. Although traditionally manual, these processes now increasingly incorporate computational methods to transcend subjective limitations and improve both efficiency and fairness.

1) Automated Technical Assessment and Code Generation: Technical skills assessment has shifted from static question banks to dynamic generation using LLMs. Early benchmarks like APPS established that these models can effectively measure coding proficiency [118]. However, generative tools such as ChatGPT require rigorous verification of functional correctness to mitigate hallucinatory risks [119]. Systems like CodeJudge address this by validating AI-based grading against functional outcomes [120]. As technology evolves, continuous evaluation frameworks are necessary to ensure assessment accuracy remains aligned with rapid software updates [121].

2) Gamification in Cognitive Assessment: Game-based strategies offer an alternative to traditional testing, potentially reducing candidate anxiety while evaluating cognitive abilities and personality traits. [122] find that these interactive environments maintain fairness and predictive validity comparable to standard methods. A key challenge, however, lies in algorithmic transparency. Consequently, [123] argue that organizations must explain the underlying grading mechanisms to demystify the process, foster trust, and improve perceived fairness.

3) Personality and Soft Skills Analysis via NLP: Beyond technical prowess, AI analyzes linguistic patterns in interview transcripts to infer soft skills and communication proficiency. For instance, [124] evaluate the granularity of candidate responses. Recent work extends this to personality detection from video-derived text, where models deduce traits directly from verbal content [125]. Rather than producing opaque numerical scores, collaborative AI agents can assist screening by offering interpretable rationales for decisions, thereby enhancing human-AI decision-making [126].

4) Multimodal Analysis in Asynchronous Video Interviews (AVI): AVI generate complex multimodal streams comprising visual, acoustic, and textual data. Early deep learning models analyzed these streams for body language and speech patterns indicative of confidence or stress [127]. Current research focuses on higher-resolution signals; [128] use facial analysis to predict job fit, while [129] highlight the importance of non-verbal cues—such as vocal tone and micro-expressions—arguing they are as predictive of performance as spoken content.

However, from a rigorous computational and psychometric perspective, extracting latent psychological traits from high-dimensional non-verbal cues introduces profound concerns regarding construct validity [57]. Machine learning architectures optimized solely for predictive utility frequently conflate correlation with causation. The statistical presence of specific micro-expressions or acoustic features does not intrinsically guarantee a causal relationship with longitudinal job performance [130]. Furthermore, these multimodal representations are highly susceptible to unobserved confounding variables—such as a candidate's cultural background, physical disabilities, or neurodivergence—which an algorithm may erroneously penalize as negative behavioral signals [131], [132]. Consequently, relying on opaque physiological correlations rather than verifiable competencies exposes organizations to severe legal risks and compliance challenges under algorithmic fairness regulations, as the predictive validity of automated micro-expression analysis remains empirically contested [133], [134].

D. Candidate Engagement

In the modern recruitment ecosystem, efficient candidate engagement and the reduction of coordination costs are critical determinants of hiring success, comparable in importance to core person-job matching algorithms, as effective engagement ultimately converts theoretical matches into actual hires. With advancements in NLP and intelligent agent technologies, AI has evolved from a passive screening tool into an interactive system capable of maintaining candidate relationships, automating administrative workflows, and assisting in background investigations.

1) Personalized Outreach Email Generation: During the active sourcing phase, recruiters frequently distribute cold emails to a broad pool of potential candidates. Traditional template-based approaches often lack specificity, resulting in low response rates. The integration of Generative AI enables large-scale personalized communication. By analyzing specific job descriptions and candidate resumes, AI models dynamically generate outreach emails that are both engaging and contextually relevant. Text generation in job recommendation scenarios through a template-based approach [135]. Leveraging structured job attributes alongside natural language generation techniques yields personalized text more persuasive than generic templates, thereby enhancing candidate perceived relevance and willingness to respond. This automated generation not only increases recruiter efficiency but also ensures consistency and professionalism in communication, effectively standardizing the quality of initial candidate interactions.

Current architectures predominantly utilize LLMs configured for sequence-to-sequence generation, often hybridized with template-augmented NLG pipelines to ensure deterministic formatting [136]. The primary challenge lies in maintaining factual consistency (mitigating generative hallucinations) while parsing extensive candidate histories within finite context windows [136], [137]. Additionally, systems struggle with zero-shot domain adaptation when balancing personalized semantic generation with strict corporate communication constraints [138], [137]. There is a distinct lack of standardized, offline evaluation metrics for measuring ”persuasiveness” beyond rudimentary open-rates [139]. Furthermore, the application of Reinforcement Learning(RL) from Human Feedback (RLHF) optimized explicitly for recruitment conversion rates—rather than general conversational alignment—remains underexplored [140], [139].

2) Conversational Interaction and Q&A Bots: As recruitment scales, timely responses to candidate inquiries regarding salary ranges, responsibilities, and interview procedures impose a significant burden on human resource departments. Question-answering (Q&A) chatbots serve as all-weather interfaces that effectively alleviate this pressure. However, existing literature suggests that providing accurate answers is insufficient; the conversational style directly influences candidate experience and employer brand perception. [141] investigate the impact of chatbot dialogue styles on interaction quality. Their findings indicate that adopting humanized strategies characterized by empathy or specific social styles strengthens candidate engagement and satisfaction. This implies that future recruitment chatbots should not be viewed merely as information retrieval tools but designed as virtual assistants equipped with affective computing capabilities to establish positive relationships during preliminary screening, thereby mitigating the impersonal nature of automated systems.

These systems typically deploy Dialogue State Tracking (DST) models coupled with RAG frameworks to fetch localized HR policies [142], supplemented by affective computing classifiers for sentiment analysis [143]. Maintaining multi-turn conversational context over extended temporal horizons (e.g., weeks between interview rounds) introduces severe memory state degradation [144]. RAG architectures also face high-latency retrieval bottlenecks when querying unstructured, highly specific, and frequently updated HR databases [145]. A significant gap exists in developing lightweight, domain-specific open-weight models that guarantee data privacy (enabling on-premise deployment for sensitive candidate data) while matching the conversational fluency of proprietary, cloud-based LLMs [146]. Mitigating the ”uncanny valley” effect in highly empathetic affective computing models also remains an open socio-technical problem [147].

3) Intelligent Interview Scheduling: Interview scheduling represents a typical low-value, high-frequency task often characterized by repetitive email exchanges between recruiters and candidates. Intelligent scheduling agents integrate multi-party calendars and constraints to automatically negotiate optimal time slots. [148] propose a human-centric decision support system for agenda scheduling. While applicable to general agenda management, this approach is particularly critical in recruitment scenarios. The system accounts not only for time slot availability but also incorporates intelligent reasoning regarding user preferences and potential conflicts to achieve optimal solutions within complex, multi-stakeholder interview rounds. By automating this segment, AI shortens the time-to-interview cycle and mitigates the risk of talent attrition caused by scheduling delays, which often serve as a critical bottleneck in competitive hiring markets.

Underlying algorithms rely heavily on Constraint Satisfaction Problem (CSP) solvers, Multi-Agent Reinforcement Learning (MARL), and heuristic search algorithms combined with Natural Language Understanding (NLU) modules for parsing unstructured email intents [149], [150]. The system must resolve high-dimensional constraints under incomplete information, such as the hidden temporal preferences of hiring managers [151]. Accurately parsing varied, unstructured temporal expressions from candidate emails and handling the delayed feedback inherent in asynchronous negotiation present substantial algorithmic hurdles [152], [151].

The field must transition from static, reactive constraint solvers to predictive scheduling paradigms that anticipate multi-party bottlenecks before they occur [153], [154]. Furthermore, standardizing the formal mathematical representation of multi-stakeholder recruitment scheduling as a dynamic Markov Decision Process (MDP) is required to benchmark future algorithmic improvements [153], [154].

4) Social Media Screening and Digital Footprint Analysis: Beyond traditional resumes and interviews, public data on social media, known as digital footprints, offers employers a supplementary dimension to assess candidate personality, values, and professional interests. AI technologies facilitate the automated retrieval and analysis of this unstructured data to assist in comprehensive background screening. [155] focus on social media usage within the low-wage labor market, revealing how employers leverage these platforms to identify latent candidate traits. The study notes that while social media screening provides insights beyond the resume, such as communication skills and professional literacy, it introduces challenges related to privacy infringement and algorithmic bias. Consequently, when deploying AI for social media analysis, system design must establish a balance between information gain and ethical compliance through algorithmic transparency, explicitly safeguarding fair hiring practices against unintended discrimination.

Implementations rely on GNNs for social network topology analysis, multi-modal feature extractors (combining text embeddings with visual classifiers), and zero-shot latent trait predictors [156], [157]. Algorithms face a critically low signal-to-noise ratio in unstructured social data, the temporal degradation of data relevance, and the complexities of cross-platform identity resolution without violating strict privacy boundaries [156], [158].

The most pressing gap is establishing rigorous construct validity—empirically proving that latent features extracted from social media correlate causally with job performance, rather than serving as proxies for protected demographic attributes. Developing mathematically provable fairness constraints (e.g., counterfactual fairness) optimized specifically for unstructured, multi-modal digital footprint data remains an urgent requirement for equitable system design [159].

E. Predictive Applications

While screening and matching address immediate hiring needs, integrating AI into recruiting facilitates predictive analytics. This moves human resources from reactive functions to proactive strategies that anticipate business requirements and optimize resource allocation. By using historical data, heterogeneous graphs, and sequential modeling, current systems can forecast talent demand, predict turnover, and model career trajectories with high granularity.

1) Talent and Skill Demand Forecasting: Given job market volatility, organizations must accurately forecast talent needs. Early approaches used tensor factorization to capture latent trends in skill demands, mitigating data sparsity issues common in market data [160]. Recent research addresses increasing data complexity by employing deep learning and dynamic graph architectures to model the heterogeneity of talent supply and demand [161], [162], [163]. These models use attentive neural sequences and dynamic graph autoencoders (DyGAE) to forecast overall headcount and specific occupational skill requirements [161], [163]. To standardize evaluation across these granularities, the community has adopted benchmarks such as Job-SDF [164].

2) Employee Turnover and Career Mobility Prediction: Predicting turnover and understanding mobility are essential for retention. Studies indicate that turnover is rarely an isolated event but is influenced by workplace social contagion and external market dynamics [165], [166]. By modeling these social effects via market-aware heterogeneous graph neural networks (HGNNs), predictive models incorporate competitor attractiveness and internal social shifts into attrition forecasting [165], [166]. Additionally, modeling interactions between employees, companies, and positions improves long-term career trajectory prediction [167], [168], [169], [170]. These methods employ variable interval sequence modeling, attentive heterogeneous graph embeddings, and Temporal Knowledge Graphs (TKGs) to map realistic career paths and predict future mobility [167], [168], [169], [170].

3) Predictive Matching and Career Development: Unlike standard matching, which ranks candidates against descriptions, predictive matching forecasts hiring outcomes, such as mutual satisfaction (Person-Job Fit) and retention. Current frameworks treat recruitment as a bilateral process, employing reciprocal recommendation models and quasi-metric learning to capture the asymmetric preferences of both job seekers and recruiters [171], [172]. Performance is further improved by integrating contextual intelligence. Models incorporating external knowledge, career paths, and professional networks demonstrate that a candidate's broader context significantly influences Person-Job Fit assessments [173], [109], [174]. Finally, predictive applications support internal mobility. By analyzing career trajectories, systems can suggest personalized training or skill acquisition, often using Bayesian variational approaches and Deep Reinforcement Learning (DRL) to balance learning costs against career utility [175], [176], [177].

IV. Challenges and Ethical Considerations

A. Bias and Fairness

The risk of bias in recruitment algorithms has been substantiated by several high-profile cases, such as those involving Google [178], Amazon [179], [180], and Facebook [181]. These incidents demonstrate that since AI systems frequently assimilate discriminatory factors from historical data, pre-existing societal biases can easily be transposed into algorithmic unfairness during the training process, thereby posing a severe threat to recruitment equity [182].

1) Fairness Challenges in the AI-Driven Recruitment Pipeline: First, job descriptions often use gendered or stereotypical language [183], which discourages diverse applicants and limits the talent pool. While researchers promote inclusive language [184], many ads remain unchecked. This persistence wastes organizational resources and diminishes the self-efficacy of marginalized jobseekers [182]. Bias is also baked into AI models through training data that reflects cultural, gender, and racial prejudices [185]. Consequently, these biases lead to discriminatory ad recommendations on platforms like Facebook and Google [181], [178].

In the screening stage, LLMs often perpetuate institutional biases found in historical hiring data, leading to the exclusion of qualified candidates from underrepresented groups. Studies have shown that LLMs can assign lower scores based on race or gender even when qualifications are identical [186], [187]. Mitigation strategies include reweighting fairness metrics and using adversarial testing to identify scoring disparities across protected attributes [188], [189], [190].

Third, in the interview stage, significant fairness concerns have emerged regarding visual and auditory cues, leading to regulatory actions like the Illinois AI Video Interview Act [191]. Bias often stems from unrepresentative training sets—like the First Impressions dataset [192]—and ASR inaccuracies for non-native speakers or individuals with speech disabilities [193], [194]. To counter this, researchers are implementing adversarial learning and fairness-aware classifiers to detect harmful interactions and reduce demographic bias in trait prediction [195], [196], [197].

The final stage involves candidate selection and negotiation, where AI tools like Oracle Recruiting Cloud and Pactum AI forecast acceptance probabilities and automate salary discussions [198], [199]. While intended to minimize human prejudice, these tools can reinforce information asymmetry and may inadvertently use prohibited data like salary history [200]. Furthermore, a widespread lack of feedback for rejected candidates reduces perceived fairness and trust. Organizations evaluate the final process based on validity, utility, and fairness [201], but existing legal guidelines may struggle to keep pace with algorithmic complexities and the biases inherent in hybrid human-AI decision-making [202].

2) Causes of Bias:

(a) *Training data:* Data is the source of AI's capabilities.If the sample data is skewed or the synthetic data has distributional shift, the AI will make biased decisions [182], [203].

(b) *Label definitions:* The "target label" is the specific metric a model is tasked with predicting.When these labels are defined too broadly or vaguely, it opens the door for disparate impact. For instance, Amazon's hiring algorithm once found that because its historically successful engineers were predominantly male, the model would automatically lower the score of resumes containing the word "female" (such as "president of a women's chess club") [204].

(c) *Proxies:* Even when sensitive attributes (such as gender) are removed, AI can still infer protected information through related features like named and word embedding vectors [205]. Experimental results show that when LLMs making hiring decisions, they tend to disadvantage names associated with underrepresented groups. In particular, Hispanic names receive the least favorable treatment [206].

(d) *The Memory Amplification Effect:* While personalized memory enhances the utility and relevance of recommendations, it also creates pathways for bias to infiltrate various stages, including personalized query generation, resume retrieval, and results re-ranking. Experiments demonstrate that even when using models subjected to rigorous safety training or utilizing de-gendered resumes, agents still acutely capture and amplify biased characteristics from a recruiter's historical memory, leading to unfair decisions based on latent gender-associated terms [207].

(e) *Multimodal Fusion:* Multimodal fusion strategies can influence fairness in AI-driven recruitment. Using the FairCVdb dataset—which integrates image, text, and tabular data across simulated gender and racial bias scenarios—Swati et al. empirically demonstrate that early-fusion can outperform alternative architectures, though this finding is specifically constrained to this dataset and the specific experimental tasks evaluated. By synthesizing unique cross-modal features within this controlled setting, early-fusion achieved the lowest Mean Absolute Error (MAE), maintaining predictive accuracy while mitigating demographic bias. Conversely, late-fusion exhibited higher error rates and score homogenization in these trials due to its sensitivity to visual outliers. These findings suggest that, within the context of this benchmark, strategic multimodal integration may offer advantages over unimodal approaches for bias reduction, positioning mid-fusion as a potential frontier for further empirical investigation into optimizing algorithmic equity [208].

3) Mitigation Strategies for AI Discrimination: To ensure algorithmic fairness in AI-driven recruitment, mitigation strategies must shift from isolated technical fixes to a holistic governance framework [209]. This requires implementing interventions across the entire machine learning lifecycle, strictly categorized into pre-processing, in-processing, and post-processing techniques.

Pre-processing: Data and Design Interventions
Mitigation begins before model training by addressing biases at the data source and aligning strategic design with legal frameworks. Because legislation like the EU AI Act classifies AI in recruitment as a "high-risk" application, organizations must establish compliance standards prioritizing rigorous risk assessments [210]. Traditional performance metrics must be redefined to establish Demographic Parity and Equal Opportunity as core optimization objectives [211]. At the data level, the scarcity of high-quality, annotated HR data often forces a reliance on generic datasets like the Census Income Dataset [212]. To counter representational harms, synthetic data generation can simulate features of underrepresented groups, thereby enhancing the demographic balance of the training corpus [211]. Additionally, practitioners must sanitize

the feature space by critically auditing proxy variables; for instance, proximity-based filtering must be re-evaluated, as it frequently acts as a proxy for race or socioeconomic status due to residential segregation [205]. Biases identified at this stage are typically neutralized using techniques such as reweighting or resampling within the feature space [210].

In-processing: Algorithmic Constraints and Auditing During the modeling phase, fairness must be integrated as a structural mathematical constraint within the learning objective. In-processing techniques, such as adversarial debiasing, embed fairness constraints directly into the model's loss function, compelling the algorithm to jointly minimize prediction error and demographic disparity [210]. As foundation models—such as LLMs for text generation and Whisper for speech-to-text—become ubiquitous in resume parsing and interview analysis, deep algorithmic auditing is required to quantify inherent linguistic and cultural biases [210], [208], [213]. Furthermore, democratizing AI auditing by reducing computational overhead is essential; this enables independent researchers to rigorously audit commercial "black-box" systems, ensuring accountability beyond internal corporate reviews [210].

Post-processing: Calibration and Human Oversight The final stage of mitigation relies on post-hoc output calibration and the preservation of human-in-the-loop architectures. Post-processing adjustments involve dynamically calibrating decision thresholds to guarantee equitable selection rates across protected groups, often augmented by Explainable AI (XAI) techniques to provide transparent justifications for specific algorithmic paths [214]. A mandatory human-in-the-loop mechanism is critical here, ensuring that human domain experts retain oversight at pivotal decision nodes to provide ethical and contextual judgments that purely mathematical models cannot encode [210]. This establishes a continuous feedback loop wherein HR professionals feed real-world hiring outcomes back to technical teams. Such an iterative optimization process captures "contextual bias"—nuanced discrimination often missed by automated metrics—and triggers targeted model retraining, ensuring the system adapts to evolving socio-technical landscapes [211].

B. Recruitment Interaction Data Challenges

The efficacy, fairness, and predictive validity of algorithmic systems are intrinsically constrained by the quality, structure, and temporal dynamics of the data upon which they rely [182]. This critical domain encompasses the multifaceted ways in which the data generated by human-computer interaction within recruitment platforms degrades the learning process of machine learning models.

This chapter examines the critical vulnerabilities embedded within recruitment interaction data, systematically exploring four foundational pillars: the noise of implicit feedback, the structural distortions of exposure bias, the historical entrenchment of selection bias, and the temporal complexities associated with delayed rewards in RL. By dissecting these interconnected phenomena, this analysis elucidates the operational challenges that prevent automated hiring systems from achieving optimal predictive validity, while detailing the state-of-the-art mitigation strategies emerging in computer science and algorithmic fairness research.

1) The Noise of Implicit Feedback in Talent Recommendation: In the development of large-scale, deep learning-based recommendation systems for recruitment, acquiring explicit feedback is fundamentally prohibitive. Explicit feedback occurs when hiring managers or candidates actively rate the quality of a match on a numerical or qualitative scale (e.g., a five-star rating or a written review). The cognitive load and administrative effort required to manually evaluate and assign accurate ratings acts as a severe operational disincentive, making it impossible to assemble large, comprehensively annotated datasets and contributing to extreme data sparsity [215], [216].

Consequently, modern e-recruitment systems rely almost exclusively on implicit feedback. This involves harvesting passive behavioral signals such as profile clicks, resume downloads, dwell time (the duration a user spends viewing a specific document), bookmarking, and application submissions [217]. These actions are automatically collected at scale, providing the massive volume of data necessary to train parameter-heavy neural architectures. However, while implicit feedback solves the volume problem, it introduces a profound level of epistemic noise, severely complicating the extraction of genuine user preferences [218].

The fundamental flaw in implicit feedback lies in its binary, ambiguous, and non-deterministic nature, which generates massive influxes of false positives and false negatives [219], [220]. For instance, a hiring manager clicking a profile might be driven by exploratory curiosity or a stylized resume layout rather than genuine candidate suitability, creating a false positive if it does not lead to a hire. Conversely, failing to click a candidate often simply indicates the manager stopped reviewing the list before reaching that profile, generating a false negative rather than a definitive rejection [220], [221]. In traditional machine learning optimization paradigms using standard pointwise or pairwise loss functions, these false signals are ingested as absolute ground truth, forcing the algorithm to fit noisy data and obscuring the actual manifold of person-job fit [219]. Recent academic research highlights a critical mathematical dilemma regarding this dynamic: there is a significant overlap between normal, highly relevant interactions and noisy, irrelevant interactions within the model's overall loss distribution [219], [222], [221]. Because human behavior in recruitment is highly stochastic, traditional denoising algorithms systematically misclassify noisy interactions as normal and vice versa, a problem that becomes even more pronounced when transitioning to pairwise loss functions.

To counteract this pervasive noise without destroying dataset utility, the field has developed sophisticated architectures like Adaptive Denoising Training (ADT) [222], [221]. Based on the empirical observation that noisy feedback typically exhibits large loss values during the early epochs of neural network training, ADT dynamically prunes these interactions during the training process rather than applying static filters. This is achieved by reformulating the loss function through either truncated loss, which discards training samples exceeding a dynamic threshold to prevent gradient updates from highly anomalous interactions, or reweighted loss, which applies a continuous decay function to adaptively lower the weight of large-loss samples [222], [221].

Beyond loss modification, addressing implicit noise requires advanced negative sampling and topological noise filtration strategies. Because implicit datasets are heavily skewed toward ”positive” interactions, systems must sample unobserved interactions as negative examples; however, naive random sampling frequently selects highly relevant but unseen candidates, severely confusing the model [219]. Advanced frameworks utilize differential community detection within the implicit feedback graph to deploy personalized noise filtration, reliably identifying false negatives and converting them into positive training samples, while neighborhood-guided feature optimization refines these features to ensure the model distinguishes meaningful professional alignments from superficial noise [219], [216].

Finally, the advent of LLMs has introduced a new paradigm of post-training denoising that leverages worldly knowledge to apply semantic common sense to purely mathematical interaction graphs. Researchers utilize the zero-shot reasoning capabilities of LLMs to act as semantic filters over collaborative filtering profiles by analyzing a user's historical interaction sequence. By prompting the LLM to identify anomalous or contradictory behavior—such as a recruiter who typically seeks senior backend engineers suddenly clicking on a junior graphic designer—the system can successfully denoise the user profile before it is used for downstream candidate generation, yielding substantial improvements in recommendation effectiveness [219].

2) The Structural Distortion of Exposure Bias: While implicit feedback noise involves misinterpreting human actions, exposure bias is a deeper structural flaw concerning which candidates get the opportunity to be seen in the first place. In talent acquisition, this is often called presentation or position bias, occurring when a candidate's likelihood of being presented to a hiring manager is unfairly distributed. Because AI-driven systems rank millions of candidates but only present the top fraction to recruiters, those ranked lower suffer from a severe lack of exposure, rendering them practically invisible to human decision-makers [223], [224], [225].

A critical error in standard machine learning is the assumption that an unobserved interaction represents a negative preference. A recruiter failing to click on a qualified candidate buried on the tenth page is a function of algorithmic presentation, not candidate quality, and treating these indiscriminately as negative feedback introduces massive bias [220], [224], [223]. When models are naively retrained on this data, they trigger a highly destructive recommendation feedback loop. The system recommends easily identifiable profiles, time-constrained recruiters interact heavily with them, and the system records this as absolute ground-truth validation [226], [227], [228]. Over time, this amplifies initial biases, creates severe echo chambers, and systemically underexposes highly qualified talent that does not fit the historically dominant archetype [220], [223], [227].

This structural flaw is intricately linked to popularity bias, where algorithms inherently favor candidates with high interaction metrics or mainstream features—such as elite university degrees or major tech company experience—leaving non-traditional candidates in algorithmic obscurity [229], [230], [231]. Furthermore, modern recruitment platforms function as multi-sided environments that must simultaneously address the demand side of recruiters and the supply side of candidates. Consequently, exposure bias degrades system utility for employers by hiding top talent, while simultaneously violating multi-sided fairness for candidates by denying them equal market opportunity based on arbitrary initial sorting [229], [225].

To counteract these distortions, researchers deploy advanced debiasing strategies that alter how interaction data is weighed, such as Inverse Propensity Scoring (IPS). IPS statistically reweights observed interactions based on the inverse of their exposure probability. If a highly visible candidate at the top of a list is clicked, the interaction receives a lower weight during model retraining. Conversely, if a recruiter expends effort to find and click on a deeply buried candidate, that interaction receives a massive weight multiplier. By mathematically correcting the skewed distribution through these estimators, the loss function artificially amplifies the value of interactions that occurred despite low visibility [232], [226].

Additionally, the field has achieved breakthroughs by integrating discrete choice models and explicit fairness constraints to further neutralize presentation distortion [223], [233]. Discrete choice models decouple true user preference from presentation bias by evaluating decisions relative to the specific choice set presented on the screen, actively logging which candidates were observed but ignored. Advanced ranking systems now also bake exposure constraints directly into their optimization objectives using a joint loss function. This function balances accuracy loss (L_{acc}), item bias loss (L_{item}), and exposure bias loss (L_{exp}) [234]. By using hyperparameters to balance these constraints alongside techniques like Fair-PG-Rank, models can efficiently learn ranking functions that ensure all candidates receive a statistically fair minimum threshold of visibility without sacrificing relevance [234], [235].

3) Selection Bias and the Imperative of Counterfactual Evaluation: While exposure bias controls who is seen by recruiters, selection bias dictates who is permitted to exist within the system's foundational knowledge base. AI recruitment models are typically trained on artificially restricted datasets comprised only of candidates who historically survived human-mediated screening and were subsequently evaluated [236], [237]. Because algorithms mathematically ingest these historical decisions as objective ground truth, they inevitably codify and amplify past prejudices [236], [238], [230]. This manifests profoundly as Feature Selection Bias, where variables like zip codes or names inadvertently serve as proxies for protected characteristics—famously demonstrated when an AI tool downgraded resumes containing the word ”women's” [230], [231]. It also appears as Evaluation Bias, where the target labels the algorithm is trained to predict, such as performance review scores, are themselves tainted by subjective human prejudice [236], [238].

The ultimate mathematical consequence of this restricted data is the sample selection problem. Because the model trains exclusively on the sub-population of hired individuals, it remains entirely blind to the counterfactual reality of how rejected candidates would have performed [230], [236]. Without this ground truth for the unselected population, traditional machine learning evaluation metrics—such as precision, recall, and F1-score—become mathematically deceptive. An algorithm boasting a 95% accuracy rate on historical data produced by systematic exclusion is merely demonstrating high proficiency at replicating discrimination [236], [238]. Therefore, the goal of applied machine learning in hiring must be to maximize performance improvement against a counterfactual decision-making process, rather than maximizing precision against a flawed historical baseline [236].

Addressing selection bias requires moving beyond standard observational metrics to embrace causal inference and counterfactual evaluation. Classical econometric approaches are frequently utilized to address observationally biased data. Methods such as propensity score matching simulate randomized trials, while the Heckman two-step correction models the probability of an individual being selected into the sample initially. By applying inverse propensity weighting to re-weight observations based on their probability of passing the initial screening stage, researchers can account for the missing data of rejected candidates and generate mathematically unbiased estimates of true candidate quality [239].

In contemporary AI, this econometric logic has evolved into the strict pursuit of counterfactual fairness. A model achieves this if, and only if, its prediction for a specific individual is identical to what it would be in a hypothetical world where their protected demographic attributes were altered, holding all other causal variables constant. Implementing this requires structural causal models, such as Pearl's twin networks. These networks allow developers to map causal relationships, simulate alternative realities, and trace the flow of information to ensure sensitive attributes do not causally influence the final hiring recommendation, whether directly or through subtle proxy variables [231], [240], [239].

Recent advancements have operationalized counterfactual fairness testing through generative AI and data augmentation, particularly to audit black-box systems like Asynchronous Video Interviews [240], [241]. Researchers use deep generative frameworks to dynamically modify protected attributes in candidate videos—such as perceived age, skin tone, or gender markers—while leaving speech and behavioral patterns identical. These counterfactual videos are then fed back into the scoring model alongside the originals. By utilizing metrics like Disparate Impact to compare the scores, auditors can empirically prove the existence of selection bias if there is a statistically significant variance between the profiles [240]. This critical shift from passive observation to active, causal perturbation represents the frontier of mitigating selection bias in AI recruitment.

4) The Temporal Complexity of Delayed Rewards in Reinforcement Learning: While implicit noise and exposure bias primarily affect supervised learning and traditional recommenders, the recruitment technology sector is shifting toward dynamic frameworks using RL [242], [243]. In this model, recruitment is framed as a complex Markov Decision Process where an autonomous agent learns to make sequential decisions—such as dynamic channel selection or automated screening—to maximize cumulative long-term organizational utility [244], [245], [246], [247], [248]. However, this introduces the profound challenge of managing sparse and heavily delayed rewards. Unlike traditional RL benchmarks where feedback is rapid, a successful hire is defined by long-term metrics like retention and performance that take months or years to materialize [248]. This extreme delay creates a severe temporal credit assignment problem, making it mathematically difficult for the system to determine which specific action in a complex sequence was responsible for the ultimate outcome [242].

Compounding this temporal disconnect is the sparse nature of recruitment rewards. The vast majority of an agent's actions, such as evaluating thousands of resumes, yield zero immediate feedback. This results in a flat gradient landscape where the algorithm struggles to refine its policy efficiently [248]. Furthermore, real-world employment is highly stochastic. External noise completely invisible to the algorithm—such as a candidate resigning due to a personal crisis or a toxic manager—can generate an inaccurate negative long-term reward signal. This unfairly penalizes the RL agent for a fundamentally optimal sourcing decision and heavily corrupts the policy update.

To optimize these models under such extreme delay and sparsity, developers deploy foundational solutions like Temporal Difference (TD) learning and Q-learning. These algorithms blend immediate rewards with predictions of future rewards to bootstrap the learning process. By updating action-value estimates based on immediate feedback plus a discounted estimate of the optimal future value, the agent can mathematically propagate delayed reward signals backward through time [246], [247]. Modern extensions, such as Deep Q-Networks (DQN), utilize neural networks and experience replay buffers to further stabilize learning across delayed feedback and optimize long-horizon performance [247].

To accelerate convergence, practitioners also heavily employ reward shaping, which introduces dense, artificial auxiliary rewards for intermediate actions that logically correlate with the ultimate goal (e.g., passing a technical assessment) [249]. However, poorly calibrated shaping risks "reward hacking," where the AI exploits proxy metrics—like sourcing unqualified candidates who reply quickly to emails—at the expense of true hiring quality [248]. To prevent this, developers utilize potential-based constraints. By applying the strict mathematical equation $F(s, a, s') = \gamma\Phi(s') - \Phi(s)$, the shaped rewards represent a difference in a defined potential function between states [249]. This guarantees that intermediate feedback does not corrupt the optimal policy of the original delayed-reward environment.

Finally, one of the most advanced frameworks for addressing severe temporal delays is Return Decomposition for Delayed Rewards (RUDDER). Leveraging Long Short-Term Memory (LSTM) neural networks and advanced pattern recognition, RUDDER fundamentally restructures the reward landscape. The LSTM is trained to identify the specific state-action pair patterns truly responsible for the final return. By redistributing the final delayed reward directly backward to those causative actions, RUDDER mathematically bypasses the temporal delay and dramatically simplifies the credit assignment problem, enabling the agent to learn complex, multi-step strategies with the efficiency of immediate feedback [250], [251], [252].

C. Explainability and Transparency

1) Why Transparency Matters?: From a human rights perspective, transparency in AI recruitment is directly linked to a candidate's right to the truth' and their legal and moral right to seek redress when subjected to unfair treatment. Transparency ensures that candidates clearly understand the 'rules of the game' behind hiring decisions, thereby safeguarding their autonomy within the labor market and maintaining public trust in the fairness of societal operations [253].

From a legal perspective, transparency in AI recruitment serves as the core pillar to ensure system compliance, fairness, and the protection of candidates' fundamental rights. Candidates typically view transparency through the lens of procedural justice, expecting to understand the evaluation logic to ensure the hiring process is both predictable and legitimate, thereby safeguarding their rights to the truth and to seek redress when treated unfairly. While HR practitioners may prioritize efficiency and 'person-job fit' in practice, transparency remains a necessary prerequisite for them to fulfill their supervisory duties and ensure the system operates within preset legal parameters [254].

Within specific legal frameworks, anti-discrimination laws rely on transparency to identify and expose complex 'proxy discrimination,' preventing algorithms from using seemingly neutral data as substitutes for legally protected characteristics [95], [255]. Furthermore, the General Data Protection Regulation (GDPR) explicitly mandates that personal data processing must adhere to the principles of 'fairness and transparency,' granting candidates the 'right to be informed' and the 'right to an explanation' regarding automated decision-making logic to offset power asymmetries and protect their statutory rights [256]. Finally, the EU AI Act classifies recruitment AI as a 'high-risk' system, mandating strict transparency and explainability standards to ensure traceability, enable human oversight, and mitigate the risks of systemic social injustice [257].

From the perspective of candidate trust, transparency in AI recruitment is paramount, as it directly mitigates applicant confusion and provides a lucid understanding of the underlying decision logic. By elucidating the rationale behind automated screening, transparency demystifies the algorithmic 'black-box,' ensuring candidates feel they are being treated with fairness, which in turn bolsters confidence in both the recruitment system and the employer brand [258].

Particularly in cases where a candidate is not selected, transparency enables organizations to provide constructive feedback through 'counterfactual explanations'—such as identifying specific missing certifications or skills. This effectively transforms a cold rejection into actionable advice for the candidate's professional development. By prioritizing clarity of communication over mere technical minutiae, this approach conveys a sense of sincerity and professionalism, establishing a vital bridge of trust between humans and technology within automated processes [258].

From a technical and operational standpoint, transparency is what allows teams to see under the hood and identify the specific features driving screening outcomes. This visibility is essential for engineering and talent teams to pinpoint and resolve data quality issues, feature leakage, or misaligned metrics, ensuring the algorithm's signals stay tightly aligned with business reality [258].

Beyond that, a transparent architecture provides a necessary audit trail for reproducibility. It allows teams to do more than just see the results; they can retrace specific decision paths and benchmark model iterations to diagnose hiring anomalies. By keeping a close watch on explanation stability and model calibration, companies can continuously iterate on their systems—maintaining high predictive performance while ensuring the entire pipeline operates according to the intended business logic [258], [259].

2) Effective Pathways: From Black-Box to XAI: XAI refers to systems designed to provide stakeholders with contextually relevant insights or justifications, thereby rendering internal operational mechanisms transparent and intelligible [260]. Traditional algorithmic debiasing is often insufficient on its own; fundamental bias identification and auditing require the "transparency" afforded by XAI techniques. While conventional mitigation strategies—such as pre-processing or post-processing—typically focus on optimizing the statistical metrics of final outcomes, XAI facilitates a paradigm shift from "statistical parity" to "comprehensive process auditability." By dismantling the "black-box" nature of recruitment tools, XAI translates intricate evaluation logic into human-interpretable explanations.

Specifically, XAI allows developers to map correlations between protected attributes (e.g., gender or ethnicity) and biased concepts within the algorithmic vector space, enabling the precise localization of bias at its source. By utilizing standardized reporting frameworks like Model Cards, XAI can rigorously document performance variances across protected groups and flag specific features or data representations that escalate bias risk. This functionality provides a critical evidentiary basis for human-AI collaboration. Rather than reflexively following algorithmic outputs, HR professionals are empowered to critically evaluate decisions against logical evidence, effectively rectifying discriminatory biases that arise from data over-generalization or a lack of socio-contextual awareness [261], [262].

To operationalize this transparency, several prevailing frameworks have been adapted for the recruitment pipeline. SHAP (Shapley Additive Explanations) is frequently employed to quantify the marginal contribution of features like academic credentials, allowing auditors to isolate proxy variables that may obscure latent gender or racial biases [259], [261], [214]. While LIME (Local Interpretable Model-agnostic Explanations) provides "local fidelity" by explaining individual screening outcomes for specific resumes, Counterfactual Explanations offer actionable "what-if" scenarios that satisfy legal requirements for candidate recourse under frameworks like the GDPR [214], [261], [262], [263]. Finally, inherently transparent ”glass-box” models, such as Decision Trees, remain a preferred choice for initial eligibility screening where procedural fairness and strict adherence to predefined criteria are paramount [214], [261], [263].

Despite their widespread adoption, post-hoc explanation methods such as SHAP and LIME exhibit significant technical limitations when applied to the complex, high-dimensional feature spaces typical of recruitment data. A primary vulnerability lies in their handling of highly correlated features—a common occurrence in human resources datasets where attributes such as education level, years of experience, and specialized skills often exhibit strong multicollinearity. In such environments, perturbation-based methods may sample unrealistic feature combinations from outside the data manifold, generating out-of-distribution artifacts that skew the resulting feature importance weights [264], [265]. Furthermore, there remains a fundamental disconnect between local and global explanations. While LIME may achieve local fidelity by approximating the decision boundary for a single candidate's rejection, aggregating these local approximations often fails to yield a mathematically consistent global representation of the model's overall behavior. This gap can inadvertently mask systemic, macro-level biases under the guise of individualized, micro-level justifications [266].

Additionally, these post-hoc techniques suffer from epistemic instability, a critical flaw that severely undermines their reliability in high-stakes hiring scenarios. Because methods like LIME rely on random perturbations around a given data point to construct a localized surrogate model, two candidates with nearly identical profiles—or even the exact same candidate evaluated across different random seeds—can yield fundamentally divergent explanatory outputs [267], [268]. This local instability not only diminishes stakeholder trust but also exposes the auditing process to adversarial vulnerabilities; minor, semantically meaningless modifications to a resume could theoretically alter the generated explanation without changing the underlying predictive outcome [266], [269]. Consequently, relying solely on post-hoc proxy models risks providing a false sense of transparency, suggesting that an explanation of the model is equivalent to the model itself.

3) Native Explainability in LLMs: From Behavioral Rationales to Mechanistic Interpretability: The pursuit of explainable AI for LLMs has bifurcated into behavioral abstractions and mechanistic structural realities, driven by the models' reliance on stochastic rather than deterministic reasoning [270], [271], [272]. Behavioral explainability treats the LLM as a communicative agent, utilizing techniques like Chain-of-Thought prompting or explanation-aware training to generate natural language rationales alongside predictions [273], [272]. While these self-explanations are highly accessible to users, their utility is fundamentally constrained by the discrepancy between human alignment and model faithfulness. Prompt-generated rationales frequently act as post-hoc justifications rather than true cognitive traces, often failing causal faithfulness metrics like the Flip Rate [273]. However, because these rationales yield significant Normalized Simulatability Gains (NSG), researchers acknowledge they contain privileged self-knowledge predictive of future model behavior [274]. Frameworks like NeuroFaith attempt to bridge this gap by mathematically tethering the semantic concepts in textual self-explanations to decoded internal hidden states to verify their causal role [275].

To move beyond the potential unreliability of generated text, structural feature attribution methods seek to quantify the precise causal influence of input tokens on the output logit distribution. Integrated Gradients (IG) serves as a gold-standard pa,th-attribution method, resolving the shattering gradient problem of standard saliency by calculating the path integral of gradients from a non-informative baseline [271], [276]:

$$\text{IG}_i(x) = (x_i - x_i') \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} d\alpha$$

This mathematically rigorous approach has proven particularly vital for diagnosing positional biases during In-Context Learning. Parallel to gradient methods is the intense academic debate over whether transformer self-attention mechanisms inherently provide explainability. While the consensus dictates that raw attention weights do not cleanly map to semantic feature importance—demonstrated by adversarial shuffling experiments—they undeniably provide a critical structural map of internal information routing rather than high-level semantic summaries [277].

To truly reverse-engineer these routing maps, Mechanistic Interpretability decomposes attention heads into precise functional units, such as Query-Key (QK) circuits for routing and Output-Value (OV) circuits for data writing [278], [279].This structural dismantling has revealed specialized mechanisms like induction heads, which drive in-context learning, alongside active-dormant and name-mover heads. Furthermore, understanding the model's intermediate reasoning requires decoding the residual stream using techniques like the Tuned Lens, or mapping associative factual memories directly within Multi-Layer Perceptrons via causal tracing frameworks like Rank-One Model Editing (ROME) [280], [281].

Despite these mapping successes, structural interpretability faces the profound obstacle of superposition, where models compress concepts into highly polysemantic, almost-orthogonal linear combinations of neurons [282], [283]. To disentangle this, researchers deploy Sparse Autoencoders (SAEs) to project dense activations into an overcomplete dictionary of monosemantic, human-interpretable features, approximated as:

$$x \sim x_0 + \sum_{i=1}^{M} f_i(x) d_i$$

While SAEs enable unprecedented feature steering and active model manipulation, they are currently constrained by astronomical computational costs and performance degradation in noisy downstream tasks, necessitating hybridized multi-scale approaches for future LLM transparency [282], [284], [281], [283].

4) Challenges in Converging XAI with Bias Governance: The integration of XAI into bias governance frameworks is hindered by several formidable challenges. First is the semantic chasm between mathematical abstraction and human-centric intuition. Current post-hoc explanation techniques—notably SHAP and LIME—frequently generate outputs that are overly technical and semantically ambiguous. For non-expert stakeholders, such as job applicants, these explanations offer little utility. Critically, oversimplified interpretations risk facilitating ”fairwashing,” where an illusion of procedural fairness is projected to grant a system ”false legitimacy” without substantively addressing underlying algorithmic biases [261], [259].

Second, XAI frameworks exhibit significant fragility when addressing intersectional discrimination or atypical cohorts (e.g., individuals with disabilities). Because explanation logic tends to reflect the behavioral norms of majority demographics, it often fails to capture the nuanced biased distortions that affect specific intersectional subgroups. Furthermore, the inherent interpretability-accuracy trade-off remains a fundamental barrier; the most performant deep learning architectures often exhibit the highest levels of epistemic opacity, making granular transparency technologically elusive [261], [263], [262].

Finally, there is a distinct mismatch between current XAI capabilities and the stringent ”meaningful information” requirements (e.g., causal inference) mandated by regulatory frameworks like the GDPR and the EU AI Act. Current tools struggle to provide the causal grounding necessary for robust compliance auditing. In many organizational contexts, explanation results are leveraged to satisfy institutional legitimacy rather than to substantively empower affected individuals to challenge or rectify automated decisions [261].

V. Future Directions

The convergence of LLMs and autonomous agent architectures is facilitating a shift in recruitment automation, moving from task-specific point optimizations toward end-to-end system integration. While current methodologies have enhanced efficiency within isolated modules—such as document parsing and preliminary screening—they remain fundamentally fragmented when applied to highly variable, long-horizon recruitment workflows that require complex sequential decision-making. Achieving a substantive intelligent transformation requires a research shift beyond isolated algorithmic refinements toward the construction of comprehensive frameworks incorporating environmental perception, long-term trajectory planning, and seamless tool orchestration. This section explores the evolutionary trajectories and critical challenges of next-generation recruitment AI systems across four dimensions: lifecycle orchestration, proactive interaction, tool-augmented intelligence, and the shifting boundaries of human-AI teaming.

A. Full-Lifecycle Orchestration

Current recruitment automation remains fragmented, with document parsing and candidate ranking operating as decoupled functional blocks [285]. Transitioning toward autonomous agents necessitates a structural integration of sourcing, multi-channel outreach, and scheduling into a continuous processing pipeline. This development requires cross-stage logical consistency to prevent semantic drift as candidates transition through disparate system modules [286]. Future research must prioritize maintaining a persistent evaluation state to ensure assessment criteria remain traceable and invariant across the pipeline. By unifying these high-latency asynchronous tasks, systems move beyond isolated toolsets toward comprehensive, end-to-end autonomous workflows.

B. Proactive Agentic Recruiting

Current recruitment architectures often rely on static scoring models that process batch inputs in isolation, which may contribute to delayed interaction loops. A shift toward agent-based frameworks could involve integrating planning and action modules capable of initiating outreach via Voice-over-IP (VoIP) or Interactive Voice Response (IVR) protocols [287], [288]. These autonomous agents might evaluate interaction feedback to dynamically recalibrate recruitment strategies, potentially replacing rigid decision trees with more adaptive, goal-oriented planning. Technical maturation in this direction is poised to involve a transition from primarily text-based parsing toward multimodal interfaces that support concurrent audio and video streams [289].

C. Tool-Augmented Intelligence

Future recruitment systems shift from closed-loop inference to tool-augmented architectures capable of orchestrating complex environment interactions. By leveraging function-calling interfaces, autonomous agents interface with ATS and third-party verification platforms to automate administrative tasks [290]. Reliability in these distributed workflows hinges on enforcing transactional atomicity across disparate APIs, ensuring that state changes—such as updating a candidate's status while modifying a synchronized calendar—either commit fully or fail safely [291], [292], [293]. Securing these tool invocations requires robust validation layers to prevent adversarial prompt injection from compromising sensitive organizational data. Maintaining system integrity across high-cardinality action spaces remains a primary research challenge for production deployments [294].

D. Evolving Boundaries of Human-AI Teaming

Traditional human-AI collaboration models typically employ a linear handoff where automated systems execute initial filtering prior to human intervention. Future frameworks propose moving beyond these discrete stages to explore concurrent interaction between autonomous agents and domain experts as a primary research area [295], [296], [297]. Rather than merely filtering resumes, the extent to which AI can manage the deterministic elements of high-volume sourcing while human recruiters are reserved for ”high-entropy” scenarios—situations where success depends on interpreting non-verbal social cues, navigating the emotional nuances of high-stakes salary negotiations, and applying subjective judgment to non-traditional career paths—remains an open research question [298].

VI. Conclusion

This survey has systematically charted the convergence of artificial intelligence and recruitment, tracing the technological evolution from heuristic, rule-based screening to a contemporary computational framework shaped by a broad spectrum of AI-driven methodologies. Our analysis indicates that the field has progressed beyond lexical keyword matching toward high-dimensional semantic embeddings and multi-stage ranking architectures that have been empirically shown to improve scalability and matching precision. A defining trend established through this synthesis is the early-stage transition toward an emerging paradigm, particularly the shift from discriminative to generative architectures; here, the hypothesis that automated systems can reliably optimize data-intensive sourcing and preliminary screening, thereby allocating human recruiters to focus on complex evaluative tasks(a concept frequently marketed as human-AI partnership), represents a critical open research question. By establishing a taxonomy of these AI interventions across the recruitment lifecycle, this paper clarifies how the integration of generative models and deep learning is iteratively modifying the computational pipeline of talent acquisition.

While this review provides a comprehensive synthesis of the intersection between artificial intelligence and talent acquisition, it is imperative to acknowledge the inherent limitations of our survey methodology. First, the boundaries of our literature coverage, strategically constrained to high-impact venues and peer-reviewed publications, may inadvertently omit parallel advancements in niche sociological or specialized human-computer interaction domains. Second, and more critically, evaluating commercial AI-driven recruitment platforms introduces fundamental challenges regarding scientific reproducibility. The proprietary, closed-source nature of these industry systems—coupled with closely guarded model weights, private training corpora, and undisclosed architectural pipelines—precludes rigorous independent verification of their stated algorithmic efficacy and fairness metrics. Consequently, our analysis of these commercial platforms necessarily relies on externally observable behavioral proxies and vendor-provided technical documentation rather than direct mechanistic auditing.

Despite these methodological developments, the field faces persistent limitations, particularly regarding the "black-box" nature of deep learning models and the documented risk of algorithmic bias inherited from historical hiring data. To address these gaps, future research should prioritize the development of domain-specific XAI frameworks that provide transparent, legally defensible justifications for candidate rankings. Additionally, there is a demonstrable utility in developing multimodal integration, synthesizing text-based resumes with video interview analytics and psychometric signals—is necessary to construct more robust latent representations of candidate profiles. Finally, shifting the focus from short-term efficiency metrics to longitudinal career-success modeling will be essential for aligning AI recruitment tools with predictive validity and formalized ethical fairness constraints.

References

[1] J. S. Black and P. van Esch, "Ai-enabled recruiting: What is it and how should a manager use it?" *Business Horizons*, vol. 63, no. 2, pp. 215–226, 2020. [Online]. Available: https://doi.org/10.1016/j.bushor.2019.12.001

[2] P. Tambe, P. Cappelli, and V. Yakubovich, "Artificial intelligence in human resources management: Challenges and a path forward," *California Management Review*, 2019, journal metadata (volume/issue/pages) may vary by version; DOI is authoritative. [Online]. Available: https://doi.org/10.1177/0008125619867910

[3] E. T. Albert, "Ai in talent acquisition: A review of ai-applications used in recruitment and selection," *Strategic HR Review*, 2019. [Online]. Available: https://doi.org/10.1108/SHR-04-2019-0024

[4] P.-S. Huang, X. He, J. Gao, L. Deng, A. Acero, and L. Heck, "Learning deep structured semantic models for web search using clickthrough data," in *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM '13)*. ACM, 2013, pp. 2333–2338. [Online]. Available: https://doi.org/10.1145/2505515.2505665

[5] M. Bertrand and S. Mullainathan, "Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination," *American Economic Review*, vol. 94, no. 4, pp. 991–1013, 2004. [Online]. Available: https://doi.org/10.1257/0002828042002561

[6] M. De-Arteaga, A. Romanov, H. Wallach, J. Chayes, C. Borgs, A. Chouldechova, S. C. Geyik, K. Kenthapadi, and A. T. Kalai, "Bias in bios: A case study of semantic representation bias in a high-stakes setting," in *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency (FAT*)*. ACM, 2019, pp. 120–128. [Online]. Available: https://doi.org/10.1145/3287560.3287572

[7] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021. [Online]. Available: https://doi.org/10.1109/TNNLS.2020.2978386

[8] X. Yi, J. Allan, and W. B. Croft, "Matching resumes and jobs based on relevance models," in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '07)*. ACM, 2007, pp. 809–810. [Online]. Available: https://doi.org/10.1145/1277741.1277920

[9] S. Strohmeier, "Research in e-hrm: Review and implications," *Human Resource Management Review*, vol. 17, no. 1, pp. 19–37, 2007. [Online]. Available: https://doi.org/10.1016/j.hrmr.2006.11.002

[10] D. L. Stone, D. L. Deadrick, K. M. Lukaszewski, and R. Johnson, "The influence of technology on the future of human resource management," *Human Resource Management Review*, vol. 25, no. 2, pp. 216–231, 2015. [Online]. Available: https://doi.org/10.1016/j.hrmr.2015.01.002

[11] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995. [Online]. Available: https://doi.org/10.1007/BF00994018

[12] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[13] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013. [Online]. Available: https://doi.org/10.1109/TPAMI.2013.50

[14] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013. [Online]. Available: https://arxiv.org/abs/1301.3781

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018. [Online]. Available: https://arxiv.org/abs/1810.04805

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," arXiv preprint, 2017, arXiv:1706.03762 [cs.CL]. [Online]. Available: https://arxiv.org/abs/1706.03762

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT 2019*, 2019. [Online]. Available: https://doi.org/10.18653/v1/N19-1423

[18] C. Zhu, H. Zhu, H. Xiong, C. Ma, F. Xie, P. Ding, and P. Li, "Person-job fit: Adapting the right talent for the right job with joint representation learning," *ACM Transactions on Management*

Information Systems, vol. 9, no. 3, 2018, article 12. [Online]. Available: https://doi.org/10.1145/3234465

[19] P. Liu, H. Wei, X. Hou, J. Shen, S. He, Q. Shen, Z. Chen, F. Borisyuk, D. Hewlett, L. Wu, S. Veeraraghavan, A. Tsun, C. Jiang, and W. Zhang, "Linksage: Optimizing job matching using graph neural networks," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '25)*. ACM, 2025, pp. 2448–2457. [Online]. Available: https://doi.org/10.1145/3690624.3709396

[20] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan *et al.*, "Language models are few-shot learners," arXiv preprint, 2020, arXiv:2005.14165 [cs.CL]. [Online]. Available: https://arxiv.org/abs/2005.14165

[21] OpenAI, "Gpt-4 technical report," arXiv preprint, 2023, arXiv:2303.08774 [cs.CL, cs.AI]. [Online]. Available: https://arxiv.org/abs/2303.08774

[22] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, É. Grave, and G. Lample, "Llama: Open and efficient foundation language models," arXiv preprint, 2023, arXiv:2302.13971 [cs.CL]. [Online]. Available: https://arxiv.org/abs/2302.13971

[23] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," arXiv preprint, 2022, arXiv:2201.11903 [cs.CL, cs.AI]. [Online]. Available: https://arxiv.org/abs/2201.11903

[24] European Parliament and Council of the European Union, "Regulation (eu) 2024/1689 of 13 june 2024 laying down harmonised rules on artificial intelligence (artificial intelligence act)," Official Journal of the European Union, 2024, cELEX: 32024R1689; ELI: reg/2024/1689/oj; published 2024-07-12. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/dc8116a1-3fe6-11ef-865a-01aa75ed71a1/

[25] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive nlp tasks," arXiv preprint, 2020, arXiv:2005.11401 [cs.CL]. [Online]. Available: https://arxiv.org/abs/2005.11401

[26] Z. Xi, W. Chen, X. Guo, W. He, Y. Ding, B. Hong, M. Zhang, J. Wang, S. Jin, E. Zhou *et al.*, "The rise and potential of large language model based agents: A survey," arXiv preprint, 2023, arXiv:2309.07864 [cs.AI, cs.CL]. [Online]. Available: https://arxiv.org/abs/2309.07864

[27] C. Yang, Y. Hou, Y. Song, T. Zhang, J.-R. Wen, and W. X. Zhao, "Modeling two-way selection preference for person-job fit," in *Proceedings of the 16th ACM Conference on Recommender Systems (RecSys '22)*. ACM, 2022, pp. 102–112. [Online]. Available: https://doi.org/10.1145/3523227.3546752

[28] Z. Zheng, X. Hu, S. Gao, H. Zhu, and H. Xiong, "Mirror: A multi-view reciprocal recommender system for online recruitment," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)*. ACM, 2024, pp. 543–552. [Online]. Available: https://doi.org/10.1145/3626772.3657776

[29] 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. [Online]. Available: https://doi.org/10.1109/TAFFC.2016.2614299

[30] L. Hemamou, G. Felhi, V. Vandenbussche, J.-C. Martin, and C. Clavel, "Hirenet: A hierarchical attention model for the automatic analysis of asynchronous video job interviews," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, pp. 573–581. [Online]. Available: https://doi.org/10.1609/AAAI.V33I01.3301573

[31] F. Leutner, S.-C. Codreanu, S. Brink, and T. Bitsakis, "Game based assessments of cognitive ability in recruitment: Validity, fairness and test-taking experience," *Frontiers in Psychology*, vol. 13, p. 942662, 2023. [Online]. Available: https://doi.org/10.3389/fpsyg.2022.942662

[32] J. H. Marler and J. W. Boudreau, "An evidence-based review of hr analytics," *The International Journal of Human Resource Management*, vol. 28, no. 1, pp. 3–26, 2017. [Online]. Available: https://doi.org/10.1080/09585192.2016.1244699

[33] D. McIver, M. L. Lengnick-Hall, and C. A. Lengnick-Hall, "A strategic approach to workforce analytics: Integrating science and agility," *Business Horizons*, vol. 61, no. 3, pp. 397–407, 2018. [Online]. Available: https://doi.org/10.1016/j.bushor.2018.01.005

[34] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 2020. [Online]. Available: https://doi.org/10.1016/j.inffus.2019.12.012

[35] Society for Human Resource Management, "Toolkit: Optimize your hiring strategy with business-driven recruiting," 2024, accessed: 2026-02-06. [Online]. Available: https://www.shrm.org/topics-tools/tools/toolkits/optimize-hiring-strategy-with-business-driven-recruiting

[36] ——, *Recruitment and Selection: Hiring the Right Person (Instructor's Manual)*, SHRM, 2021. [Online]. Available: https://www.shrm.org/content/dam/en/shrm/credentials/shrm-certification/teaching-resources/recruitment-and-selection-hiring-the-right-person-instructors-manual.pdf

[37] O. AI, "Mira," https://www.openjobs-ai.com/, 2026, accessed: 2026-02-13.

[38] M. Volkovs, G. W. Yu, and T. Poutanen, "Content-based neighbor models for cold start in recommender systems," *Proceedings of the Recommender Systems Challenge 2017*, 2017.

[39] D. Shen, C. Qin, H. Zhu, T. Xu, E. Chen, and H. Xiong, "Joint representation learning with relation-enhanced topic models for intelligent job interview assessment," *ACM Transactions on Information Systems*, vol. 40, pp. 1–36, 2021.

[40] D. D. Le and H. Lauw, "Efficient retrieval of matrix factorization-based top-k recommendations: A survey of recent approaches," *Journal of Artificial Intelligence Research*, vol. 70, pp. 1441–1479, 2021.

[41] X. Zhao, B. Zheng, X. Yi, X. Luan, C. Xie, X. Zhou, and C. S. Jensen, "Fargo: Fast maximum inner product search via global multi-probing," *Proceedings of the VLDB Endowment*, vol. 16, pp. 1100–1112, 2023.

[42] J. Vega, "Semantic matching between job offers and job search requests," in *COLING 1990 Volume 1: Papers presented to the 13th International Conference on Computational Linguistics*, 1990. [Online]. Available: https://aclanthology.org/C90-1014/

[43] H. Somers, B. Black, J. Nivre, T. Lager, A. Multari, L. Gilardoni, J. Ellman, and A. Rogers, "Multilingual generation and summarization of job adverts: the TREE project," in *Fifth Conference on Applied Natural Language Processing*. Washington, DC, USA: Association for Computational Linguistics, Mar. 1997, pp. 269–276. [Online]. Available: https://aclanthology.org/A97-1040/

[44] X. Yi, J. Allan, and W. B. Croft, "Matching resumes and jobs based on relevance models," in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 809–810. [Online]. Available: https://doi.org/10.1145/1277741.1277920

[45] M. Zehlike, F. Bonchi, C. Castillo, S. Hajian, M. Megahed, and R. Baeza-Yates, "Fa*ir: A fair top-k ranking algorithm," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ser. CIKM '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1569–1578. [Online]. Available: https://doi.org/10.1145/3132847.3132938

[46] K. Kenthapadi, B. Le, and G. Venkataraman, "Personalized job recommendation system at linkedin: Practical challenges and lessons learned," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, ser. RecSys '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 346–347. [Online]. Available: https://doi.org/10.1145/3109859.3109921

[47] C. Zhu, H. Zhu, H. Xiong, C. Ma, F. Xie, P. Ding, and P. Li, "Person-job fit: Adapting the right talent for the right job with joint representation learning," *ACM Trans. Manage. Inf. Syst.*, vol. 9, no. 3, Sep. 2018. [Online]. Available: https://doi.org/10.1145/3234465

[48] C. Qin, H. Zhu, T. Xu, C. Zhu, L. Jiang, E. Chen, and H. Xiong, "Enhancing person-job fit for talent recruitment: An ability-aware neural network approach," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, ser. SIGIR '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 25–34. [Online]. Available: https://doi.org/10.1145/3209978.3210025

[49] P. Guo, K. Xiao, H. Zhu, and Q. Meng, "Preference-constrained career path optimization: An exploration space-aware stochastic model," in *2023 IEEE International Conference on Data Mining (ICDM)*, 2023, pp. 120–129.

[50] M. Teng, H. Zhu, C. Liu, C. Zhu, and H. Xiong, "Exploiting the contagious effect for employee turnover prediction," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 1166–1173, Jul. 2019. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/3910

[51] D. Gamba, Y. Yu, Y. Yuan, G. Schoenebeck, and D. M. Romero, "Exit ripple effects: Understanding the disruption of socialization networks following employee departures," in *Proceedings of the ACM Web*

Conference 2024, ser. WWW '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 211–222. [Online]. Available: https://doi.org/10.1145/3589334.3645634

[52] W. Chao, Z. Qiu, L. Wu, Z. Guo, Z. Zheng, H. Zhu, and H. Liu, "A cross-view hierarchical graph learning hypernetwork for skill demand-supply joint prediction," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 18, pp. 19 813–19 822, Mar. 2024. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/29956

[53] X. Chen, C. Qin, Z. Wang, Y. Cheng, C. Wang, H. Zhu, and H. Xiong, "Pre-dygae: Pre-training enhanced dynamic graph autoencoder for occupational skill demand forecasting," in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, K. Larson, Ed. International Joint Conferences on Artificial Intelligence Organization, 8 2024, pp. 2009–2017, main Track. [Online]. Available: https://doi.org/10.24963/ijcai.2024/222

[54] P. Joulani, A. György, and C. Szepesvári, "Online learning under delayed feedback," 2013. [Online]. Available: https://arxiv.org/abs/1306.0686

[55] C. Ding, J. Wu, Y. Yuan, C. Li, X. Wang, D. Wang, F. Yang, and A. Rabinovich, "Delayed feedback modeling with influence functions," 2025. [Online]. Available: https://arxiv.org/abs/2502.01669

[56] O. Chapelle, "Modeling delayed feedback in display advertising," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 1097–1105. [Online]. Available: https://doi.org/10.1145/2623330.2623634

[57] M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating bias in algorithmic hiring: evaluating claims and practices," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, ser. FAT* '20. ACM, Jan. 2020, p. 469–481. [Online]. Available: http://dx.doi.org/10.1145/3351095.3372828

[58] M. Hoffman, L. B. Kahn, and D. Li, "Discretion in hiring*," *The Quarterly Journal of Economics*, vol. 133, no. 2, pp. 765–800, 10 2017. [Online]. Available: https://doi.org/10.1093/qje/qjx042

[59] A. Chalfin, O. Danieli, A. Hillis, Z. Jelveh, M. Luca, J. Ludwig, and S. Mullainathan, "Productivity and selection of human capital with machine learning," *American Economic Review*, vol. 106, no. 5, p. 124–27, May 2016. [Online]. Available: https://www.aeaweb.org/articles?id=10.1257/aer.p20161029

[60] M. Kaya and T. Bogers, "An exploration of sentence-pair classification for algorithmic recruiting," in *Proceedings of the 17th ACM Conference on Recommender Systems*, ser. RecSys '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 1175–1179. [Online]. Available: https://doi.org/10.1145/3604915.3610657

[61] Z. Guan, J.-Q. Yang, Y. Yang, H. Zhu, W. Li, and H. Xiong, "Jobformer: Skill-aware job recommendation with semantic-enhanced transformer," *ACM Trans. Knowl. Discov. Data*, vol. 19, no. 1, Dec. 2024. [Online]. Available: https://doi.org/10.1145/3701735

[62] P. Liu, H. Wei, X. Hou, J. Shen, S. He, Q. Shen, Z. Chen, F. Borisyuk, D. Hewlett, L. Wu, S. Veeraraghavan, A. Tsun, C. Jiang, and W. Zhang, "Linksage: Optimizing job matching using graph neural networks," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, ser. KDD '25. New York, NY, USA: Association for Computing Machinery, 2025, p. 2448–2457. [Online]. Available: https://doi.org/10.1145/3690624.3709396

[63] F. Javed, Q. Luo, M. McNair, F. Jacob, M. Zhao, and T. S. Kang, "Carotene: A job title classification system for the online recruitment domain," in *2015 IEEE First International Conference on Big Data Computing Service and Applications*, 2015, pp. 286–293.

[64] M. Zhao, F. Javed, F. Jacob, and M. McNair, "Skill: A system for skill identification and normalization," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 2, pp. 4012–4017, Jan. 2015. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/19064

[65] E. Senger, M. Zhang, R. van der Goot, and B. Plank, "Deep learning-based computational job market analysis: A survey on skill extraction and classification from job postings," in *Proceedings of the First Workshop on Natural Language Processing for Human Resources (NLP4HR 2024)*, E. Hruschka, T. Lake, N. Otani, and T. Mitchell, Eds. St. Julian's, Malta: Association for Computational Linguistics, Mar. 2024, pp. 1–15. [Online]. Available: https://aclanthology.org/2024.nlp4hr-1.1/

[66] N. Laosaengpha, T. Tativannarat, C. Piansaddhayanon, A. Rutherford, and E. Chuangsuwanich, "Learning job title representation from job description aggregation network," in *Findings of the Association for Computational Linguistics: ACL 2024*, L.-W. Ku, A. Martins, and V. Srikumar, Eds. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 1319–1329. [Online]. Available: https://aclanthology.org/2024.findings-acl.77/

[67] F. Retyk, H. Fabregat, J. Aizpuru, M. Taglio, and R. Zbib, "Résumé parsing as hierarchical sequence labeling: An empirical study," 2023. [Online]. Available: https://arxiv.org/abs/2309.07015

[68] M. E. Califf and R. J. Mooney, "Relational learning of pattern-match rules for information extraction," in *CoNLL97: Computational Natural Language Learning*, 1997. [Online]. Available: https://aclanthology.org/W97-1002/

[69] T. Green, D. Maynard, and C. Lin, "Development of a benchmark corpus to support entity recognition in job descriptions," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, and S. Piperidis, Eds. Marseille, France: European Language Resources Association, Jun. 2022, pp. 1201–1208. [Online]. Available: https://aclanthology.org/2022.lrec-1.128/

[70] M. Zhang, R. van der Goot, and B. Plank, "ESCOXLM-R: Multilingual taxonomy-driven pre-training for the job market domain," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, A. Rogers, J. Boyd-Graber, and N. Okazaki, Eds. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 11 871–11 890. [Online]. Available: https://aclanthology.org/2023.acl-long.662/

[71] K. Yao, J. Zhang, C. Qin, X. Song, P. Wang, H. Zhu, and H. Xiong, "Resuformer: Semantic structure understanding for resumes via multi-modal pre-training," in *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 2023, pp. 3154–3167.

[72] H. Lin, H. Zhu, Y. Zuo, C. Zhu, J. Wu, and H. Xiong, "Collaborative company profiling: Insights from an employee's perspective," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/10751

[73] M. De-Arteaga, A. Romanov, H. Wallach, J. Chayes, C. Borgs, A. Chouldechova, S. Geyik, K. Kenthapadi, and A. T. Kalai, "Bias in bios: A case study of semantic representation bias in a high-stakes setting," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, ser. FAT* '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 120–128. [Online]. Available: https://doi.org/10.1145/3287560.3287572

[74] Y. Ye, H. Zhu, T. Xu, F. Zhuang, R. Yu, and H. Xiong, "Identifying high potential talent: A neural network based dynamic social profiling approach," in *2019 IEEE International Conference on Data Mining (ICDM)*, 2019, pp. 718–727.

[75] S. Minaee, T. Mikolov, N. Nikzad, M. Chenaghlu, R. Socher, X. Amatriain, and J. Gao, "Large language models: A survey," 2025. [Online]. Available: https://arxiv.org/abs/2402.06196

[76] I. O. Gallegos, R. A. Rossi, J. Barrow, M. M. Tanjim, S. Kim, F. Dernoncourt, T. Yu, R. Zhang, and N. K. Ahmed, "Bias and fairness in large language models: A survey," *Computational Linguistics*, vol. 50, no. 4, 2024, preprint available at arXiv:2309.00770.

[77] A. Lorincz, D. Graus, D. Lavi, and J. L. M. Pereira, "Transfer learning for multilingual vacancy text generation," in *Proceedings of the Second Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)*, A. Bosselut, K. Chandu, K. Dhole, V. Gangal, S. Gehrmann, Y. Jernite, J. Novikova, and L. Perez-Beltrachini, Eds. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics, Dec. 2022, pp. 207–222. [Online]. Available: https://aclanthology.org/2022.gem-1.18/

[78] C. Borchers, D. Gala, B. Gilburt, E. Oravkin, W. Bounsi, Y. M. Asano, and H. Kirk, "Looking for a handsome carpenter! debiasing GPT-3 job advertisements," in *Proceedings of the 4th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*, C. Hardmeier, C. Basta, M. R. Costa-jussà, G. Stanovsky, and H. Gonen, Eds. Seattle, Washington: Association for Computational Linguistics, Jul. 2022, pp. 212–224. [Online]. Available: https://aclanthology.org/2022.gebnlp-1.22/

[79] P. Budhwar, S. Chowdhury, G. Wood, H. Aguinis, G. J. Bamber, J. R. Beltran, P. Boselie, F. Lee Cooke, S. Decker, A. DeNisi *et al.*, "Human resource management in the age of generative artificial intelligence: Perspectives and research directions on chatgpt," *British Journal of Management*, vol. 34, no. 3, pp. 1162–1179, 2023.

[80] C. Qin, K. Yao, H. Zhu, T. Xu, D. Shen, E. Chen, and H. Xiong, "Towards automatic job description generation with capability-aware neural networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 5341–5355, 2023.

[81] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459–9474.

[82] Y. Zhu, H. Yuan, S. Wang, J. Liu, W. Liu, C. Deng, Z. Dou, and J.-R. Wen, "Large language models for information retrieval: A survey," *arXiv preprint arXiv:2308.07107*, 2023.

[83] M. Li, X. Chen, W. Liao, Y. Song, T. Zhang, D. Zhao, and R. Yan, "Ezinterviewer: To improve job interview performance with mock interview generator," in *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, ser. WSDM '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 1102–1110. [Online]. Available: https://doi.org/10.1145/3539597.3570476

[84] L. Hemamou, G. Felhi, V. Vandenbussche, J.-C. Martin, and C. Clavel, "Hirenet: A hierarchical attention model for the automatic analysis of asynchronous video job interviews," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 573–581, Jul. 2019. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/3832

[85] P. Gebhard, T. Schneeberger, E. André, T. Baur, I. Damian, G. Mehlmann, C. König, and M. Langer, "Serious games for training social skills in job interviews," *IEEE Transactions on Games*, vol. 11, no. 4, pp. 340–351, 2019.

[86] A. Hartholt, S. Mozgai, and A. S. Rizzo, "Virtual job interviewing practice for high-anxiety populations," in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, ser. IVA '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 238–240. [Online]. Available: https://doi.org/10.1145/3308532.3329417

[87] 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.

[88] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[89] P. Liu, R. Arora, X. Shi, B. H. Le, Q. Shen, J. Shen, C. Jiang, N. Zhiltsov, P. Bannur, Y. Zhu, L. Dong, H. Wei, Q. Guo, L. Simon, L. Hong, and W. Zhang, "A scalable and efficient signal integration system for job matching," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, ser. KDD '25. ACM, Aug. 2025, p. 4659–4669. [Online]. Available: http://dx.doi.org/10.1145/3711896.3737185

[90] J. Shen, Y. Juan, S. Zhang, P. Liu, W. Pu, S. Vasudevan, Q. Song, F. Borisyuk, K. Q. Shen, H. Wei, Y. Ren, Y. S. Chiou, S. Kuang, Y. Yin, B. Zheng, M. Wu, S. Gharghabi, X. Wang, H. Xue, Q. Guo, D. Hewlett, L. Simon, L. Hong, and W. Zhang, "Learning to retrieve for job matching," 2024. [Online]. Available: https://arxiv.org/abs/2402.13435

[91] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of ir techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, p. 422–446, Oct. 2002. [Online]. Available: https://doi.org/10.1145/582415.582418

[92] T. Joachims, B. London, Y. Su, A. Swaminathan, and L. Wang, "Recommendations as treatments," *AI Magazine*, vol. 42, no. 3, pp. 19–30, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1609/aimag.v42i3.18141

[93] I. D. Fofanah, "The algorithmic barrier: Quantifying artificial frictional unemployment in automated recruitment systems," 2026. [Online]. Available: https://arxiv.org/abs/2601.14534

[94] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness," 2011. [Online]. Available: https://arxiv.org/abs/1104.3913

[95] S. Barocas and A. D. Selbst, "Big data's disparate impact," *Calif. L. Rev.*, vol. 104, p. 671, 2016.

[96] Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice, "Uniform guidelines on employee selection procedures," 29 C.F.R. Part 1607, 1978, adopted by the EEOC, CSC, DOL, and DOJ to provide a single set of principles which are designed to assist employers, labor organizations, employment agencies, and licensing and certification boards to comply with requirements of Federal law prohibiting employment practices which discriminate on grounds of race, color, religion, sex, and national origin. [Online]. Available: https://www.eeoc.gov/laws/guidance/uniform-guidelines-employee-selection-procedures

[97] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," 2016. [Online]. Available: https://arxiv.org/abs/1610.02413

[98] V. Pradeep, M. Ko, B. Ma, X. Yan, Z. Yu, and A. Shrivastava, "Geo: Generative engine optimization," *arXiv preprint arXiv:2311.09735*, 2023.

[99] Y. Hou, J. Zhang, Z. Lin, H. Lu, R. Xie, J. McAuley, and W. X. Zhao, "Large language models are zero-shot rankers for recommender systems," in *Proceedings of the 46th European Conference on Information Retrieval (ECIR)*. Springer, 2024.

[100] R. Jagerman, H. Zhuang, Z. Qin, X. Wang, and M. Bendersky, "Query expansion by prompting large language models," *arXiv preprint arXiv:2305.03653*, 2023.

[101] X. Wang, W. Zhou, C. Zu, H. Zheng, H. Zhang, and Z. Dou, "Instructuie: Multi-task instruction tuning for unified information extraction," *arXiv preprint arXiv:2304.08085*, 2023.

[102] J. Rosenberger, L. Wolfrum, S. Weinzierl, M. Kraus, and P. Zschech, "Careerbert: Matching resumes to esco jobs in a shared embedding space for generic job recommendations," 2025. [Online]. Available: https://arxiv.org/abs/2503.02056

[103] X. Yu, J. Zhang, and Z. Yu, "Confit: Improving resume-job matching using data augmentation and contrastive learning," 2024. [Online]. Available: https://arxiv.org/abs/2401.16349

[104] J. Zhao, J. Wang, M. Sigdel, B. Zhang, P. Hoang, M. Liu, and M. Korayem, "Embedding-based recommender system for job to candidate matching on scale," *arXiv*, 2021.

[105] F. Liu, Y. Jiao, J. Massiah, E. Yilmaz, and S. Havrylov, "Trans-encoder: Unsupervised sentence-pair modelling through self- and mutual-distillations," *arXiv*, 2021.

[106] B. Paliwal, D. Saini, M. Dhawan, S. Asokan, N. Natarajan, S. Aggarwal, P. Malhotra, J. Jiao, and M. Varma, "Cross-jem: Accurate and efficient cross-encoders for short-text ranking tasks," *arXiv*, 2024.

[107] Z. Zheng, X. Hu, S. Gao, H. Zhu, and H. Xiong, "Mirror: A multi-view reciprocal recommender system for online recruitment," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 543–552. [Online]. Available: https://doi.org/10.1145/3626772.3657776

[108] C. Yang, Y. Hou, Y. Song, T. Zhang, J.-R. Wen, and W. X. Zhao, "Modeling two-way selection preference for person-job fit," in *Proceedings of the 16th ACM Conference on Recommender Systems*, ser. RecSys '22. ACM, Sep. 2022, p. 102–112. [Online]. Available: http://dx.doi.org/10.1145/3523227.3546752

[109] Z. Gong, Y. Song, T. Zhang, J.-R. Wen, D. Zhao, and R. Yan, "Your career path matters in person-job fit," ser. AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024. [Online]. Available: https://doi.org/10.1609/aaai.v38i8.28685

[110] R. Yan, R. Le, Y. Song, T. Zhang, X. Zhang, and D. Zhao, "Interview choice reveals your preference on the market: To improve job-resume matching through profiling memories," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 914–922. [Online]. Available: https://doi.org/10.1145/3292500.3330963

[111] Z. Zheng, Z. Qiu, X. Hu, L. Wu, H. Zhu, and H. Xiong, "Generative job recommendations with large language model," 2023. [Online]. Available: https://arxiv.org/abs/2307.02157

[112] E. Anzenberg, A. Samajpati, S. Chandrasekar, and V. Kacholia, "Evaluating the promise and pitfalls of llms in hiring decisions," *arXiv*, 2025.

[113] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang, "Lost in the middle: How language models use long contexts," *Transactions of the Association for Computational Linguistics*, vol. 12, pp. 157–173, 2024.

[114] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. Bang, A. Madotto, and P. Fung, "Survey of hallucination in natural language generation," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–38, 2023.

[115] S. Vaishampayan, H. Leary, Y. B. Alebachew, L. Hickman, B. Stevenor, W. Beck, and C. Brown, "Human and llm-based resume matching: An observational study," in *Findings of the Association for Computational Linguistics: NAACL 2025*, 2025, pp. 4808–4823.

[116] M. Bismay, X. Dong, and J. Caverlee, "Reasoningrec: Bridging personalized recommendations and human-interpretable explanations through llm reasoning," 2024. [Online]. Available: https://arxiv.org/abs/2410.23180

[117] H. Sun, H. Lin, H. Yan, Y. Song, X. Gao, and R. Yan, "Mockllm: A multi-agent behavior collaboration framework for

online job seeking and recruiting," 2025. [Online]. Available: https://arxiv.org/abs/2405.18113
[118] D. Hendrycks, S. Basart, S. Kadavath, M. Mazeika, A. Arora, E. Guo, C. Burns, S. Puranik, H. He, D. Song, and J. Steinhardt, "Measuring coding challenge competence with APPS," in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks 2021*, J. Vanschoren and S. Yeung, Eds., 2021. [Online]. Available: https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/hash/c24cd76e1ce41366a4bbe8a49b02a028-Abstract-round2.html
[119] J. Liu, C. S. Xia, Y. Wang, and L. Zhang, "Is your code generated by chatgpt really correct? rigorous evaluation of large language models for code generation," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. [Online]. Available: https://arxiv.org/abs/2305.01210
[120] W. Tong and T. Zhang, "Codejudge: Evaluating code generation with large language models," *arXiv preprint arXiv:2410.02184*, 2024. [Online]. Available: https://arxiv.org/abs/2410.02184
[121] M. Azanza, B. Pérez Lamancha, and E. Pizarro, "Tracking the moving target: A framework for continuous evaluation of llm test generation in industry," in *Proceedings of the 29th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 2025. [Online]. Available: https://arxiv.org/abs/2504.18985
[122] F. Leutner, S. C. Codreanu, S. Brink, and T. Bitsakis, "Game based assessments of cognitive ability in recruitment: Validity, fairness and test-taking experience," *Frontiers in Psychology*, vol. 13, 2023.
[123] K. Georgiou, "Can explanations improve applicant reactions towards gamified assessments?" *International Journal of Selection and Assessment*, vol. 29, no. 4, pp. 543–558, 2021.
[124] R. Verrap, E. Nirjhar, A. Nenkova, and T. Chaspari, ""am i answering my job interview questions right?": A nlp approach to predict degree of explanation in job interview responses," in *Proceedings of the Second Workshop on NLP for Positive Impact (NLP4PI)*. Abu Dhabi, UAE: Association for Computational Linguistics, 2022, pp. 122–129.
[125] Y. Bounab, M. Oussalah, N. Arhab, and S. Bekhouche, "Towards job screening and personality traits estimation from video transcriptions," *Expert Systems with Applications*, vol. 238, p. 122016, 2024.
[126] F. P.-W. Lo, J. Qiu, Z. Wang, H. Yu, Y. Chen, G. Zhang, and B. Lo, "AI Hiring with LLMs: A context-aware and explainable multi-agent framework for resume screening," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2023. [Online]. Available: https://arxiv.org/abs/2308.13516
[127] P. R. Satheesh Kumar A., Naveena Devi S. and S. K. V., "Body language and speech analysis using deep learning for virtual job interviews," in *AHFE International Conference on Human Factors in Communication of Design*, ser. Lecture Notes in Networks and Systems, vol. 269. Springer, 2021, pp. 3–10.
[128] E. Fithian and T. Chaspari, "Leveraging pre-trained transformers and facial embeddings for multimodal hirability prediction in job interviews," in *Proceedings of the 27th International Conference on Multimodal Interaction (ICMI)*. ACM, 2025, pp. 275–283.
[129] J. Li, Y. Wang, W. Qian, J. Hu, Z. Hu, R. Hong, and M. Wang, "Listening to the unspoken: Exploring "365" aspects of multimodal interview performance assessment," *arXiv preprint arXiv:2507.22676*, 2025. [Online]. Available: https://arxiv.org/abs/2507.22676
[130] T. Chamorro-Premuzic, D. Winsborough, R. A. Sherman, and R. Hogan, "New talent signals: Shiny new objects or a brave new world?" *Industrial and Organizational Psychology*, vol. 9, pp. 621–640, 2016.
[131] S. Brandsen, T. Chandrasekhar, L. Franz, J. Grapel, G. Dawson, and D. Carlson, "Prevalence of bias against neurodivergence-related terms in artificial intelligence language models," *Autism Research*, vol. 17, pp. 234–248, 2024.
[132] T. L. Wilder and N. E. Stratchan, "Artificial intelligence-enhanced interview success: Leveraging eye-tracking and cognitive measures to support self-regulation in college students with attention-deficit/hyperactivity disorder," *Education Sciences*, vol. 15, p. 165, 2025.
[133] E. Albaroudi, T. Mansouri, and A. Alameer, "A comprehensive review of ai techniques for addressing algorithmic bias in job hiring," *AI*, vol. 5, pp. 383–404, 2024.
[134] S. Ebrahimi, E. Abdelhalim, K. Hassanein, and M. Head, "Reducing the incidence of biased algorithmic decisions through feature importance transparency: an empirical study," *European Journal of Information Systems*, pp. 1–29, 2024.
[135] Q. Li and C. Lioma, "Template-based contact email generation for job recommendation," in *Proceedings of the Second Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)*. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics, 2022, pp. 189–197. [Online]. Available: https://aclanthology.org/2022.gem-1.15/
[136] M. Gao, X. Hu, J. Ruan, X. Pu, and X. Wan, "Llm-based nlg evaluation: Current status and challenges," *arXiv*, 2024.
[137] Z. Wang, H. Wang, B. Danek, Y. Li, C. Mack, H. Poon, Y. Wang, P. Rajpurkar, and J. Sun, "A perspective for adapting generalist ai to specialized medical ai applications and their challenges," *arXiv*, 2024.
[138] S. Chen, Q. Zhu, W. Yang, C. Yang, Z. Wang, P. Wang, X. Lin, B. Xu, D. Li, C. Yuan, L. Qi, W. Xu, s. zhenxing, X. Lu, S. Xiong, C. Chen, H. Hu, and Y. Xiao, "Inseva: A comprehensive chinese benchmark for large language models in insurance," *arXiv*, 2025.
[139] Z. Cheng and J. You, "Towards strategic persuasion with language models," *arXiv*, 2025.
[140] L. J. V. Miranda, Y. Wang, Y. Elazar, S. Kumar, V. Pyatkin, F. Brahman, N. A. Smith, H. Hajishirzi, and P. Dasigi, "Hybrid preferences: Learning to route instances for human vs. ai feedback," *arXiv*, 2024.
[141] S. Akram, P. Buono, and R. Lanzilotti, "Exploring the impact of conversational style in enhancing recruitment chatbot interactions," in *Proceedings of the 13th Nordic Conference on Human-Computer Interaction (NordiCHI 2024)*. Uppsala, Sweden: ACM, 2024, pp. 51:1–51:14.
[142] M. Allouch, A. Azaria, and R. Azoulay, "Conversational agents: Goals, technologies, vision and challenges," *Sensors*, vol. 21, p. 8448, 2021.
[143] K. Loveys, M. Sagar, X. Zhang, G. Fricchione, and E. Broadbent, "Effects of emotional expressiveness of a female digital human on loneliness, stress, perceived support, and closeness across genders: Randomized controlled trial," *Journal of Medical Internet Research*, vol. 23, p. e30624, 2021.
[144] Y. Du, Y. Xiang, B. Liang, D. Lin, K.-F. Wong, and F. Tan, "Resure: Regularizing supervision unreliability for multi-turn dialogue fine-tuning," *arXiv*, 2025.
[145] A. Joshi, S. M. Sarwar, S. Varshney, S. Nag, S. Agrawal, and J. Naik, "Reaper: Reasoning based retrieval planning for complex rag systems," *arXiv*, 2024.
[146] A. Wickramarachchi, S. Tonni, S. Majumdar, S. Karimi, S. Kõks, B. Hosking, J. Rambla, N. A. Twine, Y. Jain, and D. C. Bauer, "Askbeacon – performing genomic data exchange and analytics with natural language," *arXiv*, 2024.
[147] A. O. Thunström, H. K. Carlsen, L. Ali, T. Larson, A. Hellström, and S. Steingrimsson, "Usability comparison among healthy participants of an anthropomorphic digital human and a text-based chatbot as a responder to questions on mental health: Randomized controlled trial," *JMIR Human Factors*, vol. 11, p. e54581, 2024.
[148] S. Rosenthal and L. M. Hiatt, "Human-centered decision support for agenda scheduling," in *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '20)*. Auckland, New Zealand: International Foundation for Autonomous Agents and Multiagent Systems, 2020, pp. 1161–1168.
[149] T. Gabel, "Multi-agent reinforcement learning approaches for distributed job-shop scheduling problems," Ph.D. dissertation, University of Osnabrück, 2009. [Online]. Available: http://www.tgabel.de/fileadmin/user_upload/documents/Gabel_PhD-Thesis-2009.pdf
[150] X. Liao and K. D. Hoang, "A population-based search approach to solve continuous distributed constraint optimization problems," *Applied Sciences*, vol. 14, no. 3, p. 1290, 2024.
[151] B. An, V. Lesser, D. Westbrook, and M. Zink, "Agent-mediated multi-step optimization for resource allocation in distributed sensor networks," in *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2011, pp. 609–616. [Online]. Available: https://personal.ntu.edu.sg/boan/papers/bo-an_aamas11a.pdf
[152] R. Balan and et al., "Chatbot-delivered interventions for improving mental health among young people: A systematic review and meta-analysis," *Journal of Medical Internet Research (via PMC)*, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12261465/
[153] Y.-i. Cho, B. Kim, H.-C. Yoon, and J. H. Woo, "Locating algorithm of steel stock area with asynchronous advantage actor-critic reinforcement learning," *Journal of Computational Design and Engineering*, vol. 11, no. 1, pp. 230–246, 2023.
[154] Anonymous, "Adaptive and intelligent customized deep q-network for energy-efficient task offloading in mobile edge computing environments," *PubMed Central (PMC)*, 2026. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12886797/
[155] A. J. Lu and T. R. Dillahunt, "Uncovering the promises and challenges of social media use in the low-wage labor market: Insights from

employers," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Yokohama, Japan: ACM, 2021, pp. 1–13.

[156] M. Hammad and et al., "A deep learning-based multimodal social media analysis framework for identifying personality traits," *IEEE Access*, vol. 12, pp. 41 258–41 275, 2024.

[157] K. Sharma, Y.-C. Lee, S. Nambi, A. Salian, S. Shah, S.-W. Kim, and S. Kumar, "A survey of graph neural networks for social recommender systems," *ACM Computing Surveys*, vol. 56, no. 10, p. 1–34, Jun. 2024. [Online]. Available: http://dx.doi.org/10.1145/3661821

[158] L. Wang, T. Zhu, W. Zhou, and P. S. Yu, "Linkage on security, privacy and fairness in federated learning: New balances and new perspectives," 2024. [Online]. Available: https://arxiv.org/abs/2406.10884

[159] W. Kim and H. Kim, "Counterfactual fairness evaluation of machine learning models on educational datasets," 2025. [Online]. Available: https://arxiv.org/abs/2504.11504

[160] X. Wu, T. Xu, H. Zhu, L. Zhang, E. Chen, and H. Xiong, "Trend-aware tensor factorization for job skill demand analysis," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-19)*, 2019.

[161] Q. Zhang, H. Zhu, Y. Sun, H. Liu, F. Zhuang, and H. Xiong, "Talent demand forecasting with attentive neural sequential model," in *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '21)*, 2021, pp. 3906–3916. [Online]. Available: https://dblp.org/rec/conf/kdd/ZhangZSLZX21

[162] Z. Guo, H. Liu, L. Zhang, Q. Zhang, H. Zhu, and H. Xiong, "Talent demand-supply joint prediction with dynamic heterogeneous graph enhanced meta-learning," in *Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '22)*, 2022, pp. 2957–2967. [Online]. Available: https://dblp.org/rec/conf/kdd/GuoLZZZX22

[163] X. Chen, C. Qin, Z. Wang, Y. Cheng, C. Wang, H. Zhu, and H. Xiong, "Pre-dygae: Pre-training enhanced dynamic graph autoencoder for occupational skill demand forecasting," in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*, 2024, pp. 2154–2162.

[164] X. Chen *et al.*, "Job-sdf: A multi-granularity dataset for job skill demand forecasting and benchmarking," in *Advances in Neural Information Processing Systems (NeurIPS 2024), Datasets and Benchmarks Track*, 2024, pp. 129 329–129 356.

[165] J. Hang, Z. Dong, H. Zhao, Y. Song, S. Wang, H. Feng, and J. Hao, "Outside in: Market-aware heterogeneous graph neural network for employee turnover prediction," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (WSDM '22)*, 2022, pp. 353–362.

[166] M. Teng, H. Zhu, C. Liu, C. Zhu, and H. Xiong, "Exploiting the contagious effect for employee turnover prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI '19)*, 2019, pp. 1166–1173.

[167] C. Wang, H. Zhu, H. Qian, K. Xiao, and H. Xiong, "Variable interval time sequence modeling for career trajectory prediction: Deep collaborative perspective," in *Proceedings of The Web Conference 2021 (WWW '21)*, 2021, pp. 2566–2577.

[168] L. Zhang, D. Zhou, H. Zhu, T. Xu, R. Zha, E. Chen, and H. Xiong, "Attentive heterogeneous graph embedding for job mobility prediction," in *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '21)*, 2021, pp. 2192–2201. [Online]. Available: https://dblp.org/rec/conf/kdd/ZhangZZ0ZC021

[169] Z. Lin, J. Zhang, S. Zhang, Y. Chen, Z. Chen, R. Ma, and H. Xiong, "Effective job-market mobility prediction with attentive heterogeneous knowledge learning and synergy," in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24)*, 2024, pp. 795–804.

[170] Y.-C. Lee, J. Lee, M. Yamashita, D. Lee, and S.-W. Kim, "Caper: Enhancing career trajectory prediction using temporal knowledge graph and ternary relationship," in *Proceedings of the 31st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '25)*, 2025, pp. 647–658. [Online]. Available: https://dblp.org/rec/conf/kdd/LeeLY0K25

[171] C. Yang, Y. Hou, Y. Song, T. Zhang, J. Wen, and W. X. Zhao, "Modeling two-way selection preference for person-job fit," in *Proceedings of the 16th ACM Conference on Recommender Systems (RecSys '22)*, 2022. [Online]. Available: https://arxiv.org/abs/2208.08612

[172] Y. Du *et al.*, "Quasi-metric learning for bilateral person-job fit," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. [Online]. Available: https://dl.acm.org/doi/10.1109/TPAMI.2025.3538774

[173] K. Yao, J. Zhang, C. Qin, P. Wang, H. Zhu, and H. Xiong, "Knowledge enhanced person-job fit for talent recruitment," in *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 2022, pp. 3467–3480.

[174] H. Chen, L. Du, Y. Lu, Q. Fu, X. Chen, S. Han, Y. Kang, G. Lu, and Z. Li, "Professional network matters: Connections empower person-job fit," in *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM '24)*, 2024, pp. 96–105.

[175] C. Wang, H. Zhu, C. Zhu, X. Zhang, E. Chen, and H. Xiong, "Personalized employee training course recommendation with career development awareness," in *Proceedings of The Web Conference 2020 (WWW '20)*, 2020, pp. 1648–1659.

[176] C. Wang, H. Zhu, P. Wang, C. Zhu, X. Zhang, E. Chen, and H. Xiong, "Personalized and explainable employee training course recommendations: A bayesian variational approach," *ACM Transactions on Information Systems*, vol. 40, no. 4, pp. 1–32, 2021.

[177] Y. Sun, H. Zhu, C. Zhu, Z. Zhang, and H. Xiong, "Cost-effective and interpretable job skill recommendation with deep reinforcement learning," in *Proceedings of The Web Conference 2021 (WWW '21)*, 2021, pp. 3078–3089.

[178] C. C. Miller, "When Algorithms Discriminate," https://www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html, Jul 2015.

[179] S. Mohla, B. Bagh, and A. Guha, "A material lens to investigate the gendered impact of the AI industry," in *IJCAI 2021 Workshop on AI for Social Good*, 2021.

[180] D. Meyer, "Amazon Rportedly Killed an AI Recruitment System Because It Couldn't Stop the Tool From Discriminating Against Women," *Fortune. Tillgänglig online: https://fortune. com/2018/10/10/amazon-ai-recruitment-bias-women-sexist/(2019-09-27)*, 2018.

[181] M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, and A. Rieke, "Discrimination through optimization: How facebook's ad delivery can lead to skewed outcomes," *arXiv Preprint arXiv:1904.02095*, 2019.

[182] D. F. Mujtaba and N. R. Mahapatra, "Fairness in ai-driven recruitment: Challenges, metrics, methods, and future directions," 2025. [Online]. Available: https://arxiv.org/abs/2405.19699

[183] Snyder, Kieran. (Accessed 2024) ChatGPT Writes Job Posts. [Online]. Available: https://textio.com/blog/chatgpt-writes-job-posts/99089591200

[184] N. Laosaengpha, T. Tativannarat, A. Rutherford, and E. Chuangsuwanich, "Mitigating language bias in cross-lingual job retrieval: A recruitment platform perspective," 2025. [Online]. Available: https://arxiv.org/abs/2502.03220

[185] T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai, "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[186] J. An, D. Huang, C. Lin, and M. Tai, "Measuring gender and racial biases in large language models: Intersectional evidence from automated resume evaluation," *Pnas Nexus*, vol. 4, no. 3, p. pgaf089, 2025.

[187] Z. Wang, Z. Wu, X. Guan, M. Thaler, A. Koshiyama, S. Lu, S. Beepath, E. Ertekin Jr, and M. Perez-Ortiz, "Jobfair: A framework for benchmarking gender hiring bias in large language models," *arXiv Preprint arXiv:2406.15484*, 2024.

[188] M. Brännström, T. D. Xanthopoulou, and L. Jiang, "Modeling fairness in recruitment ai via information flow," 2025. [Online]. Available: https://arxiv.org/abs/2511.13793

[189] K. V. Deshpande, S. Pan, and J. R. Foulds, "Mitigating demographic bias in AI-based resume filtering," in *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation And Personalization*, 2020, pp. 268–275.

[190] A. Peña, I. Serna, A. Morales, and J. Fierrez, "FairCVtest demo: Understanding bias in multimodal learning with a testbed in fair automatic recruitment," in *Proceedings of the 2020 International Conference on Multimodal Interaction*, 2020, pp. 760–761.

[191] "Employers Using AI in Hiring Take Note: Illinois' Artificial Intelligence Video Interview Act is Now in Effect," https://www.jdsupra.com/legalnews/employers-using-ai-in-hiring-take-note-54767/, Feb 2020.

[192] J. C. J. Junior, A. Lapedriza, C. Palmero, X. Baró, and S. Escalera, "Person perception biases exposed: Revisiting the first impressions dataset." in *WACV (Workshops)*, 2021, pp. 13–21.

[193] D. Mujtaba, N. Mahapatra, M. Arney, J. Yaruss, H. Gerlach-Houck, C. Herring, and J. Bin, "Lost in transcription: Identifying and

quantifying the accuracy biases of automatic speech recognition systems against disfluent speech," in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, K. Duh, H. Gomez, and S. Bethard, Eds. Mexico City, Mexico: Association for Computational Linguistics, Jun. 2024, pp. 4795–4809. [Online]. Available: https://aclanthology.org/2024.naacl-long.269/

[194] M. Moore, H. Venkateswara, and S. Panchanathan, "Whistle-blowing ASRs: Evaluating the need for more inclusive speech recognition systems," *Interspeech 2018*, 2018.

[195] A. K. Gupta, F. Sheth, H. Shaikh, D. Kumar, A. Puniya, D. Panwar, S. Chaurasia, and P. Mathur, "Recruitview: A multimodal dataset for predicting personality and interview performance for human resources applications," 2025. [Online]. Available: https://arxiv.org/abs/2512.00450

[196] S. Yan, D. Huang, and M. Soleymani, "Mitigating biases in multimodal personality assessment," in *Proceedings of the 2020 International Conference on Multimodal Interaction*, 2020, pp. 361–369.

[197] J. Xue, Y.-C. Wang, C. Wei, X. Liu, J. Woo, and C.-C. J. Kuo, "Bias and fairness in chatbots: An overview," *arXiv Preprint arXiv:2309.08836*, 2023.

[198] Oracle, "Oracle Fusion Cloud HCM: Human Capital Management," accessed: February 2, 2026. [Online]. Available: https://www.oracle.com/human-capital-management/

[199] Pactum AI, "Pactum AI," https://pactum.com/, accessed: February 2, 2026.

[200] M. Bogen and A. Rieke, "Help Wanted," *An Examination Of Hiring Algorithms, Equity, And Bias*, 2018.

[201] F. J. Landy and J. M. Conte, *Work In The 21st Century: An Introduction To Industrial And Organizational Psychology*. John Wiley & Sons, 2016.

[202] A. Peng, B. Nushi, E. Kiciman, K. Inkpen, and E. Kamar, "Investigations of performance and bias in human-ai teamwork in hiring," 2022. [Online]. Available: https://arxiv.org/abs/2202.11812

[203] M. Yamashita, T. Tran, and D. Lee, "Fake resume attacks: Data poisoning on online job platforms," in *Proceedings of the ACM Web Conference 2024 (WWW'24)*. Singapore, Singapore: ACM, May 2024, pp. 1734–1745. [Online]. Available: https://doi.org/10.1145/3589334.3645524

[204] J. Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," Reuters, October 10 2018, accessed: 2026-02-03. [Online]. Available: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

[205] T. Nakano, K. Shimari, R. G. Kula, C. Treude, M. Cheong, and K. Matsumoto, "Nigerian software engineer or american data scientist? github profile recruitment bias in large language models," in *2024 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, Oct. 2024, p. 624–629. [Online]. Available: http://dx.doi.org/10.1109/ICSME58944.2024.00063

[206] H. An, C. Acquaye, C. Wang, Z. Li, and R. Rudinger, "Do large language models discriminate in hiring decisions on the basis of race, ethnicity, and gender?" in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, L.-W. Ku, A. Martins, and V. Srikumar, Eds. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 386–397. [Online]. Available: https://aclanthology.org/2024.acl-short.37/

[207] H. Gharat, H. Agrawal, and G. K. Patro, "From personalization to prejudice: Bias and discrimination in memory-enhanced ai agents for recruitment," in *Proceedings of the 19th ACM International Conference on Web Search and Data Mining (WSDM '26)*. New York, NY, USA: ACM, 2026.

[208] S. Swati, A. Roy, and E. Ntoutsi, "Exploring fusion techniques in multimodal ai-based recruitment: Insights from faircvdb," 2024. [Online]. Available: https://arxiv.org/abs/2407.16892

[209] E. Drage and K. Mackereth, "Does AI debias recruitment? Race, Gender, and AI's "eradication of difference"," *Philosophy & Technology*, vol. 35, no. 89, pp. 1–25, 2022. [Online]. Available: https://doi.org/10.1007/s13347-022-00543-1

[210] D. F. Mujtaba and N. R. Mahapatra, "Ethical considerations in ai-based recruitment," in *2019 IEEE International Symposium on Technology and Society (ISTAS)*, 2019, pp. 1–7.

[211] M. Soleimani, A. Intezari, J. Arrowsmith, D. J. Pauleen, and N. Taskin, "Reducing ai bias in recruitment and selection: an integrative grounded approach," *The International Journal of Human Resource Management*, vol. 36, no. 14, pp. 2480–2515, 2025. [Online]. Available: https://doi.org/10.1080/09585192.2025.2480617

[212] D. Dua and C. Graff, "UCI machine learning repository," 2019. [Online]. Available: http://archive.ics.uci.edu/ml

[213] H. Gharat, H. Agrawal, and G. K. Patro, "From personalization to prejudice: Bias and discrimination in memory-enhanced AI agents for recruitment," in *Proceedings of the Nineteenth ACM International Conference on Web Search and Data Mining (WSDM '26)*. Boise, ID, USA: ACM, 2026, pp. 1–5.

[214] S. Fabeyo, "Explainable ai in employment decision-making: a systematic review of transparency methods in hiring algorithms," *Issues in Information Systems*, vol. 26, pp. 127–135, 10 2025.

[215] D. W. Oard and J. Kim, "Implicit feedback for recommender systems," in *Proceedings of the AAAI Workshop on Recommender Systems*, Madison, Wisconsin, USA, 1998, aAAI Technical Report WS-98-01.

[216] E. Yüksel. (2025, May) Challenges in recommender systems: Understanding implicit feedback, negative sampling, and long-tail bias. Medium blog post. [Online]. Available: https://medium.com/@emreeyukseel/challenges-in-recommender-systems-understanding-implicit-feedback-negative-sampl

[217] J. Yi, X. He *et al.*, "Adversarial training for dual product recommendation with implicit feedback," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21)*. ACM, 2021, pp. 1–9. [Online]. Available: http://staff.ustc.edu.cn/~hexn/papers/WSDM_2021_ADT.pdf

[218] M. Hutterer, "Enhancing a job recommender with implicit user feedback," Master's thesis, Vienna University of Technology (TU Wien), Vienna, Austria, 2011. [Online]. Available: https://repositum.tuwien.at/bitstream/20.500.12708/9363/2/Hutterer%20Matthias%20-%202011%20-%20Enhancing%20a%20job%20recommender%20with%20implicit%20user...pdf

[219] W. Wang, F. Feng, X. He, L. Nie, and T.-S. Chua, "Denoising implicit feedback for recommendation," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, ser. WSDM '21. New York, NY, USA: ACM, 2021, pp. 373–381. [Online]. Available: https://doi.org/10.1145/3437963.3441800

[220] M. Färber, M. Coutinho, and S. Yuan, "Biases in scholarly recommender systems: Impact, prevalence, and mitigation," *Scientometrics*, vol. 128, no. 5, pp. 2703–2736, 2023, open Access. [Online]. Available: https://doi.org/10.1007/s11192-023-04636-2

[221] K. Zhang, Q. Cao, Y. Wu, F. Sun, H. Shen, and X. Cheng, "Personalized denoising implicit feedback for robust recommender system," 2025. [Online]. Available: https://arxiv.org/abs/2502.00348

[222] W. Wang, F. Feng, X. He, L. Nie, and T.-S. Chua, "Learning robust recommender from noisy implicit feedback," 2021. [Online]. Available: https://arxiv.org/abs/2112.01160

[223] T. Krause, A. Deriyeva, J. Beinke, G. Bartels, and O. Thomas, "Mitigating exposure bias in recommender systems – a comparative analysis of discrete choice models," *ACM Transactions on Recommender Systems*, vol. 3, 01 2024.

[224] Y. Fang, A. Singh, and Z. Tao, "Fairness in search systems," *Foundations and Trends in Information Retrieval*, vol. 18, no. 3, pp. 262–416, 12 2024. [Online]. Available: https://doi.org/10.1561/1500000101

[225] Z. Ovaisi, P. Saadatpanah, S. Sefati, M. Ohannessian, and E. Zheleva, "Fairness of interaction in ranking under position, selection, and trust bias," *ACM Transactions on Information Systems*, vol. 42, no. 4, pp. 1–38, 2024, earlier version presented at NeurIPS 2022 Workshop: Algorithmic Fairness through the Lens of Causality and Privacy. [Online]. Available: https://doi.org/10.1145/3652864

[226] S. Gupta, H. Wang, Z. Lipton, and Y. Wang, "Correcting exposure bias for link recommendation," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 3953–3963. [Online]. Available: https://proceedings.mlr.press/v139/gupta21c.html

[227] S. Khenissi and O. Nasraoui, "Modeling and counteracting exposure bias in recommender systems," 2020. [Online]. Available: https://arxiv.org/abs/2001.04832

[228] M. Mansoury, H. Abdollahpouri, B. Mobasher, M. Pechenizkiy, R. Burke, and M. Sabouri, "Unbiased cascade bandits: Mitigating exposure bias in online learning to rank recommendation," *arXiv preprint arXiv:2108.03440*, 2021. [Online]. Available: https://arxiv.org/abs/2108.03440

[229] H. Abdollahpouri and M. Mansoury, "Multi-sided exposure bias in recommendation," 2020. [Online]. Available: https://arxiv.org/abs/2006.15772

[230] M. C. Ahuchogu, "Ai and bias in recruitment: Ensuring fairness in algorithmic hiring." *Journal of Informatics Education and Research*, vol. 5, no. 3, Jul. 2025. [Online]. Available: https://jier.org/index.php/journal/article/view/3262

[231] R. Schwartz, A. Vassilev, K. Greene, L. Perine, A. Burt, and P. Hall, "Towards a standard for identifying and managing bias in artificial intelligence," National Institute of Standards and Technology, Gaithersburg, MD, USA, NIST Special Publication 1270, March 2022, available free of charge from: https://doi.org/10.6028/NIST.SP.1270. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf

[232] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, and X. He, "Bias and debias in recommender system: A survey and future directions," 2021. [Online]. Available: https://arxiv.org/abs/2010.03240

[233] A. Dash, A. Chakraborty, S. Ghosh, A. Mukherjee, and K. P. Gummadi, "Fairir: Mitigating exposure bias from related item recommendations in two-sided platforms," 2022. [Online]. Available: https://arxiv.org/abs/2204.00241

[234] S. Raza, "Bias reduction news recommendation system," *Digital*, vol. 4, no. 1, pp. 92–103, 2024. [Online]. Available: https://www.mdpi.com/2673-6470/4/1/3

[235] R. Luo, T. Tang, F. Xia, J. Liu, C. Xu, L. Y. Zhang, W. Xiang, and C. Zhang, "Algorithmic fairness: A tolerance perspective," 2024. [Online]. Available: https://arxiv.org/abs/2405.09543

[236] B. Cowgill, "Bias and productivity in humans and machines," W.E. Upjohn Institute for Employment Research, Kalamazoo, MI, Working Paper 19-309, August 2019. [Online]. Available: https://research.upjohn.org/up_workingpapers/309/

[237] P. Parasurama and P. Ipeirotis, "Algorithmic hiring and diversity: Reducing human-algorithm similarity for better outcomes," 2025. [Online]. Available: https://arxiv.org/abs/2505.14388

[238] M. Raghavan and S. Barocas. (2019, December) Challenges for mitigating bias in algorithmic hiring. Accessed: 2026-03-05. [Online]. Available: https://www.brookings.edu/articles/challenges-for-mitigating-bias-in-algorithmic-hiring/

[239] J. He, S. Wang, and A. Li, "Recover experimental data with selection bias using counterfactual logic," 2025. [Online]. Available: https://arxiv.org/abs/2506.00335

[240] D. F. Mujtaba and N. R. Mahapatra, "Behind the screens: Uncovering bias in ai-driven video interview assessments using counterfactuals," 2025. [Online]. Available: https://arxiv.org/abs/2505.12114

[241] P. S. B. Rao, L. N. Venkatesan, M. Cherubini, and D. B. Jayagopi, "Invisible filters: Cultural bias in hiring evaluations using large language models," 2025. [Online]. Available: https://arxiv.org/abs/2508.16673

[242] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA: The MIT Press, 2018. [Online]. Available: http://incompleteideas.net/book/RLbook2018.pdf

[243] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017. [Online]. Available: https://arxiv.org/abs/1707.06347

[244] D. T. Nguyen, S. Kumar, and H. C. Lau, "A survey on multi-agent deep reinforcement learning: from the perspective of challenges and applications," *Artificial Intelligence Review*, vol. 54, no. 5, pp. 3215–3258, 2021. [Online]. Available: https://link.springer.com/article/10.1007/s10462-020-09938-y

[245] Y. Yang *et al.*, "The landscape of agentic reinforcement learning for llms: A survey," *arXiv preprint arXiv:2509.02547*, 2025. [Online]. Available: https://arxiv.org/abs/2509.02547

[246] Y. Son, J. Park, and B. Jeon, "Training agents for strategic curling through a unified reinforcement learning framework," *Mathematics*, vol. 14, no. 3, 2026. [Online]. Available: https://www.mdpi.com/2227-7390/14/3/403

[247] C. A. Perdomo Charry, M. S. Mora Cortes, and O. J. Perdomo, "M-learning: Heuristic approach for delayed rewards in reinforcement learning," *Mathematics*, vol. 13, no. 13, 2025. [Online]. Available: https://www.mdpi.com/2227-7390/13/13/2108

[248] S. Ibrahim, M. Mostafa, A. Jnadi, H. Salloum, and P. Osinenko, "Comprehensive overview of reward engineering and shaping in advancing reinforcement learning applications," 2024. [Online]. Available: https://arxiv.org/abs/2408.10215

[249] O. Marom and B. Rosman, "Belief reward shaping in reinforcement learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, Apr. 2018. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/11741

[250] Y. Deng, X. Qiu, J. Chen, and X. Tan, "Reward guidance for reinforcement learning tasks based on large language models: The lmgt framework," *Knowledge-Based Systems*, vol. 322, p. 113689, Jul. 2025. [Online]. Available: http://dx.doi.org/10.1016/j.knosys.2025.113689

[251] X. Xiang, S. Foo, and H. Zang, "Recent advances in deep reinforcement learning applications for solving partially observable markov decision processes (pomdp) problems part 2—applications in transportation, industries, communications and networking and more topics," *Machine Learning and Knowledge Extraction*, vol. 3, no. 4, pp. 863–878, 2021. [Online]. Available: https://www.mdpi.com/2504-4990/3/4/43

[252] Y. Tang, X.-Q. Cai, J.-C. Pang, Q. Wu, Y.-X. Ding, and M. Sugiyama, "Beyond simple sum of delayed rewards: Non-markovian reward modeling for reinforcement learning," 2024. [Online]. Available: https://arxiv.org/abs/2410.20176

[253] A. L. Hunkenschröer and A. Kriebitz, "Is ai recruiting (un)ethical? a human rights perspective on the use of ai for hiring," *AI and Ethics*, vol. 3, no. 1, pp. 199–213, Feb 2022. [Online]. Available: https://doi.org/10.1007/s43681-022-00166-4

[254] C. Rigotti and E. Fosch-Villaronga, "Fairness, ai & recruitment," *Computer Law & Security Review*, vol. 53, p. 105966, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0267364924000335

[255] Council of the European Union, "Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation," pp. 16–22, 2000. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2000/78/oj

[256] European Parliament and Council of the European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," pp. 1–88, 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[257] ——, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 765/2008, (EC) No 767/2008, (EU) 2018/1139, (EU) 2018/1240 and (EU) 2019/1020 and Directive (EU) 2020/1828 (Artificial Intelligence Act)," 2024. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[258] T. Juenemann, "Explainable ai in hiring: Why transparency matters," ZYTHR Resources, February 2025, accessed: 2026-02-01. [Online]. Available: https://zythr.com/resources/explainable-ai-in-hiring-why-transparency-matters

[259] L. Deck, J. Schoeffer, M. De-Arteaga, and N. Kühl, "A critical survey on fairness benefits of explainable ai," 06 2024, pp. 1579–1595.

[260] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," 2019. [Online]. Available: https://arxiv.org/abs/1910.10045

[261] E. Albaroudi, T. Mansouri, and A. Alameer, "A comprehensive review of AI techniques for addressing algorithmic bias in job hiring," *AI*, vol. 5, no. 1, pp. 383–404, 2024. [Online]. Available: https://doi.org/10.3390/ai5010019

[262] G. soğancıoğlu, H. Kaya, and A. Salah, "Using explainability for bias mitigation: A case study for fair recruitment assessment," 10 2023, pp. 631–639.

[263] G. Pinto, C. Mello, and A. C. Garcia, "Explainable ai in labor market applications," 01 2025, pp. 1450–1457.

[264] S. Ghalebikesabi, L. Ter-Minassian, K. Diaz-Ordaz, and C. Holmes, "On locality of local explanation models," *arXiv*, 2021.

[265] M. Bacevicius, A. Paulauskaite-Taraseviciene, G. Zokaityte, L. Kersys, and A. Moleikaityte, "Comparative analysis of perturbation techniques in lime for intrusion detection enhancement," *Machine Learning and Knowledge Extraction*, vol. 7, p. 21, 2025.

[266] D. Slack, S. Hilgard, E. Jia, S. Singh, and H. Lakkaraju, "Fooling lime and shap: Adversarial attacks on post hoc explanation methods," *arXiv*, 2019.

[267] Z. Tan, Y. Tian, and J. Li, "Glime: General, stable and local lime explanation," *arXiv*, 2023.

[268] S. M. Shankaranarayana and D. Runje, "Alime: Autoencoder based approach for local interpretability," *arXiv*, 2019.

[269] H. Baniecki and P. Biecek, "Adversarial attacks and defenses in explainable artificial intelligence: A survey," *Information Fusion, vol. 107, 102303, 2024*, 2025.

[270] M. Ciaperoni, M. D. Vece, L. Pappalardo, F. Giannotti, and F. Giannini, "Position: Explaining behavioral shifts in large language models requires a comparative approach," 2026. [Online]. Available: https://arxiv.org/abs/2602.02304

[271] T. Datta, E. Wang, K. Huang, and F. Doshi-Velez, "Integrated gradients provides faithful language model attributions for in-context learning," in *ICLR 2025 Workshop on Building Trust in LLMs and LLM Applications: From Guardrails to Explainability to Regulation*, 2025, poster presentation. [Online]. Available: https://iclr.cc/virtual/2025/33360

[272] A. Bilal, D. Ebert, and B. Lin, "Llms for explainable ai: A comprehensive survey," 2025. [Online]. Available: https://arxiv.org/abs/2504.00125

[273] M. Fayyaz, F. Yin, J. Sun, and N. Peng, "Evaluating human alignment and model faithfulness of llm rationale," 2024. [Online]. Available: https://arxiv.org/abs/2407.00219

[274] H. Mayne, J. S. Kang, D. Gould, K. Ramchandran, A. Mahdi, and N. Y. Siegel, "A positive case for faithfulness: Llm self-explanations help predict model behavior," 2026. [Online]. Available: https://arxiv.org/abs/2602.02639

[275] M. Bhan, J.-N. Vittaut, N. Chesneau, S. Chandar, and M.-J. Lesot, "Neurofaith: Evaluating llm self-explanation faithfulness via internal representation alignment," 2026. [Online]. Available: https://arxiv.org/abs/2506.09277

[276] C. Molnar, *Interpretable Machine Learning*, 3rd ed. Munich, Germany: Self-published, 2025, section: Pixel Attribution / Saliency Maps. Accessed: 2026-03-07. [Online]. Available: https://christophm.github.io/interpretable-ml-book

[277] S. Wiegreffe and Y. Pinter, "Attention is not not explanation," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, November 2019, pp. 11–20. [Online]. Available: https://aclanthology.org/D19-1002/

[278] D. Rai, Y. Zhou, S. Feng, A. Saparov, and Z. Yao, "A practical review of mechanistic interpretability for transformer-based language models," 2025. [Online]. Available: https://arxiv.org/abs/2407.02646

[279] J. Ren, Q. Guo, H. Yan, D. Liu, Q. Zhang, X. Qiu, and D. Lin, "Identifying semantic induction heads to understand in-context learning," 2024. [Online]. Available: https://arxiv.org/abs/2402.13055

[280] F. Zhang and N. Nanda, "Towards best practices of activation patching in language models: Metrics and methods," 2024. [Online]. Available: https://arxiv.org/abs/2309.16042

[281] K. Meng, D. Bau, A. Andonian, and Y. Belinkov, "Locating and editing factual associations in gpt," 2023. [Online]. Available: https://arxiv.org/abs/2202.05262

[282] D. Shu, X. Wu, H. Zhao, D. Rai, Z. Yao, N. Liu, and M. Du, "A survey on sparse autoencoders: Interpreting the internal mechanisms of large language models," 2025. [Online]. Available: https://arxiv.org/abs/2503.05613

[283] Z. He, J. Achterberg, K. Collins, K. Nejad, D. Akarca, Y. Yang, W. Gurnee, I. Sucholutsky, Y. Tang, R. Ianov, G. Ogden, C. Li, K. Sandbrink, S. Casper, A. Ivanova, and G. W. Lindsay, "Multilevel interpretability of artificial neural networks: Leveraging framework and methods from neuroscience," 2024. [Online]. Available: https://arxiv.org/abs/2408.12664

[284] N. Belrose, I. Ostrovsky, L. McKinney, Z. Furman, L. Smith, D. Halawi, S. Biderman, and J. Steinhardt, "Eliciting latent predictions from transformers with the tuned lens," 2025. [Online]. Available: https://arxiv.org/abs/2303.08112

[285] E. T. Albert, "Ai in talent acquisition: a review of ai-applications used in recruitment and selection," *Strategic HR Review*, vol. 18, no. 5, pp. 215–221, 10 2019. [Online]. Available: https://doi.org/10.1108/SHR-04-2019-0024

[286] Z. Xi, W. Chen, X. Guo, W. He, Y. Ding, B. Hong, M. Zhang, J. Wang, S. Jin, E. Zhou, R. Zheng, X. Fan, X. Wang, L. Xiong, Y. Zhou, W. Wang, C. Jiang, Y. Zou, X. Liu, Z. Yin, S. Dou, R. Weng, W. Cheng, Q. Zhang, W. Qin, Y. Zheng, X. Qiu, X. Huang, and T. Gui, "The rise and potential of large language model based agents: A survey," 2023. [Online]. Available: https://arxiv.org/abs/2309.07864

[287] B. Hadžic´, L. T. Brandner, T. Weber, and M. Rätsch, "Ai-driven active sourcing in recruitment: addressing contestability in automated hiring systems," *Frontiers in Computer Science*, vol. Volume 7 - 2025, 2025. [Online]. Available: https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2025.1629522

[288] M. Negnevitsky, "The rise of autonomous ai agents: Automating complex tasks," *International Journal of Artificial Intelligence for Science (IJAI4S)*, vol. 1, 07 2025.

[289] L. Alexandra, "Multimodal analysis for candidate assessment," p. 17, 06 2025.

[290] P. S. Viswanathan, "Tool use and external system integration in conversational ai: From function calling to autonomous agents," *International Journal of Computational and Experimental Science and Engineering*, vol. 12, no. 1, Feb. 2026. [Online]. Available: https://www.ijcesen.com/index.php/ijcesen/article/view/4872

[291] S. Gulyamov, S. Gulyamov, A. Rodionov, R. Khursanov, K. Mekhmonov, D. Babaev, and A. Rakhimjonov, "Prompt injection attacks in large language models and ai agent systems: A comprehensive review of vulnerabilities, attack vectors, and defense mechanisms," *Information*, vol. 17, no. 1, p. 54, 2026. [Online]. Available: https://www.mdpi.com/2078-2489/17/1/54

[292] H. Garcia-Molina and K. Salem, "Sagas," in *Proceedings of the Association for Computing Machinery Special Interest Group on Management of Data 1987 Annual Conference, San Francisco, CA, USA, May 27-29, 1987*, U. Dayal and I. L. Traiger, Eds. ACM Press, 1987, pp. 249–259. [Online]. Available: https://doi.org/10.1145/38713.38742

[293] G. Alonso, F. Casati, H. Kuno, and V. Machiraju, *Web Services: Concepts, Architectures and Applications*. Berlin, Heidelberg: Springer Science & Business Media, 2004, chapter 5: Transaction Management and Orchestration.

[294] S. P. HeounMo Go, "A study on classification based concurrent api calls and optimal model combination for tool augmented llms for ai agent," *Sensors (Basel)*, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12216082/

[295] T. M. Michelle Vaccaro, Abdullah Almaatouq, "When combinations of humans and ai are useful: A systematic review and meta-analysis," *Nature (Scientific Reports) / PMC*, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11659167/

[296] B. Eastwood. (2025) When humans and ai work best together — and when each is better alone. Accessed: 2026-02-11. [Online]. Available: https://mitsloan.mit.edu/ideas-made-to-matter/when-humans-and-ai-work-best-together-and-when-each-better-alone

[297] ATS on Demand. (2026) Top ai recruiting trends 2026: The rise of human-ai partnership in talent acquisition. Industry Insights. [Online]. Available: https://power.atsondemand.com/top-ai-recruiting-trends-2026-the-rise-of-human-ai-partnership-in-talent-acquisition/

[298] Deloitte Insights, "How human-ai synergy can power the future of work: From agents to multi-agent systems," Deloitte Development LLC, Tech. Rep., 2025. [Online]. Available: https://www.deloitte.com/us/en/insights/industry/government-public-sector-services/ai-future-of-work-in-government/ai-human-future-of-work-in-government.html