eISSN 2783-6851

VILNIUS TECH
Vilnius Gediminas Technical University

NEW TRENDS in COMPUTER SCIENCES

2026
Volume 4
Issue 1
Pages 93–110
https://doi.org/10.3846/ntcs.2026.26251

CANDIDATE ASSESSMENT IN ASYNCHRONOUS VIDEO INTERVIEWING: A HYBRID HYPERBOLIC FUZZY INTEGRATED MCDM FRAMEWORK WITH GENERATIVE AI EXPERTISE

Abhishek YADAV [1], Raghunathan KRISHANKUMAR [2]✉

[1]*Department of Mathematics, Amrita School of Physical Sciences, Amrita Vishwa Vidyapeetham, Coimbatore, India*
[2]*Information Technology Systems and Analytics, Indian Institute of Management Bodh Gaya, Bodh Gaya, India*

Article History:
- received 28 February 2026
- accepted 7 July 2026

Abstract. Asynchronous Video Interviewing (AVI) is increasingly used in large-scale recruitment; however, its evaluation remains constrained by opaque predictive models, limited interpretability, and reliance on subjective human judgement. Interview data is often vague and complex. In this study we propose a novel "Multi-Criteria Decision Making (MCDM)" framework designed to navigate this uncertainty by identifying the underlying influence relationships that drive AVI assessment outcomes. The framework integrates "Hyperbolic Fuzzy Sets (HYFS)" with a multi-perspective "Decision Making Trial and Evaluation Laboratory (DEMATEL)" approach to estimate criteria weights and identify directional influence structures while employing state-of-the-art generative AI models as scalable AI-based evaluative agents. To address variability and potential bias in human ratings, a multi-attitudinal variance-based method is used to compute panel participant weights, and interviewees are subsequently ranked using a HYFS-integrated "Weighted Aggregated Sum Product Assessment (WASPAS)" technique. The framework is demonstrated on a real-world AVI dataset comprising 138 interview recordings evaluated across 18 behavioural and non-verbal criteria. The results show that Eye Contact, Smiled, Speaking Rate, Friendly, Paused, and Calm act as dominant drivers, while micro-level speech disfluencies such as fillers exhibit comparatively lower influence once broader behavioural cues are considered. Overall, the proposed framework yields rankings aligned with benchmark scores while offering enhanced interpretability and directional influence insight, supporting fair and trustworthy automated hiring systems.

Keywords: Hyperbolic Fuzzy Sets, multi-perspective DEMATEL, generative AI evaluative agents, influence-based MCDM, WASPAS, asynchronous video interviewing.

✉Corresponding author. E-mail: *raghunathan.k@iimbg.ac.in*

1. Introduction

Asynchronous Video Interviewing (AVI), often referred to as on-demand or digital interviewing, has emerged as a transformative technology in modern human resource management (Lukacik et al., 2022). An AVI system allows candidates to record video responses to predetermined questions at their convenience, which are subsequently reviewed by recruiters or automated systems (Lukacik & Bourdage, 2025). It aims to make hiring easier by letting interviewers and candidates participate at different times and locations (İlhan et al., 2025). Administratively, it offers scalability, consistency in questioning, and significant reductions in time-to-hire (Mejia & Torres, 2018). Strategically, it enables organisations to screen larger talent pools while attempting to maintain a standardised assessment framework (Biswas et al., 2024).

The recruitment landscape is currently facing several significant inefficiencies (Dalgleish, 2005; Maree et al., 2019). One of the most pressing is the sheer volume of applications in a digital-first job market, which places immense pressure on human resource departments (Aly et al., 2025). Traditional synchronous interviews are resource-intensive, often leading to scheduling bottlenecks and extended hiring timelines (Grabara et al., 2016). Furthermore, human evaluators are susceptible to cognitive fatigue and unconscious biases, such as the "halo effect" or affinity bias, which can compromise the fairness of the selection process (Neira & Aracena, 2020; Pesak & Hlavacs, 2026). While AVIs address the logistical constraints, the assessment of these video records remains a challenge (İlhan et al., 2025). Relying solely on manual review reintroduces the very biases and scalability issues the technology aims to solve, while existing automated scoring systems often lack the nuance to capture complex behavioural cues in a fair and effective manner.

In accordance with growing adoption of AVI systems, in this study we pose a few key research questions and tackle them:

(RQ1) What are the influence relationships between different criteria that can be used by an automated system to evaluate asynchronous video interview performance (e.g., engagement, structure, non-verbal cues)?

(RQ2) How can these criteria be effectively weighted and prioritised under environments of high uncertainty and vagueness?

The literature review reveals that while various "Multi-Criteria Decision Making (MCDM)" techniques have been employed to address assessment challenges, significant limitations persist in existing approaches. Most fuzzy-based MCDM studies rely on traditional fuzzy sets, such as Intuitionistic, Pythagorean, or Fermatean fuzzy sets, which face decision space limitations due to their sum-based constraints. Furthermore, traditional studies often depend on manual input from a small pool of human experts to determine criteria weights. These methodologies are not only time-consuming and resource-intensive but also prone to subjective inconsistencies.

In response to these challenges, this study proposes a novel framework and showcases the framework on a case study. The study has the following features:

- Adoption of "Hyperbolic Fuzzy Sets (HYFS)": Instead of relying on traditional fuzzy sets with sum-based constraints, this study utilises Hyperbolic Fuzzy Sets (Dutta & Borah, 2023) which employ product-based constraints.
- Data collection via generative AI: Instead of relying on a limited number of human experts, the data for criteria evaluation is solicited from state-of-the-art generative AI models. This ensures a consistent, scalable, and fatigue-free evaluative perspective.
- Influence Analysis and Prioritization: Finally, the criteria are analysed using a Multi-perspective "HYFS-Decision Making Trial and Evaluation Laboratory (DEMATEL)" approach to map influence relationships and determine criteria weights
- Finally, the criteria weights obtained are used on a case study involving rating of interviewees, and ranking them using multi-attitudinal variance and "Weighted Aggregated Sum Product Assessment (WASPAS)".

The literature review, mathematical formulations, and explanations of the individual technical components discussed above, along with the case study, are detailed in the subsequent sections.

2. Literature review

In this section, a literature review of the use of MCDM for AVI and the use of generative AI for MCDM is provided.

2.1. Use of MCDM for AVI

Interviewee selection is inherently a multi-criteria decision problem, involving conflicting objectives such as balancing technical fit of a candidate interviewee with their behavioural fit, or minimizing the cost while maximizing candidate quality. Early studies have relied on classical MCDM approaches such as "Analytic Hierarchy Process (AHP)" (Saaty, 1980) and "Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS)" (Hwang & Yoon, 1981) to aggregate diverse criteria involved in interviewee candidate selection. For instance, Cahigas et al. (2021) utilised a hybrid AHP-TOPSIS MCDM framework for human resource recruitment process while explicitly using criteria such as "Visual Presentation" and "Profile Presentation". Similarly, Silva et al. (2024) extended fuzzy TOPSIS for selecting project managers based on the soft skills such as "Attitude", "Negotiation", and "Empathy" they demonstrated.

However, transitioning to evaluation of interviewees through AVI adds an extra layer of complexity as it requires evaluation of non-verbal cues from a pre-recorded video. A recent study extends MCDM to scenarios where candidates are evaluated based on non-verbal cues. Alshammari et al. (2023) proposed a presentation skill scoring system that integrates multi-model computer vision analysis with fuzzy Delphi and AHP to score human presentation performance while considering facial expressions, eye contact, hand gestures.

However, most existing AVI systems primarily rely on machine learning (ML) or deep learning (DL) models that directly predict hiring outcomes or candidate suitability scores, often treating the decision process as a black box (Diefenhardt, 2025; Germanier et al., 2025). These approaches rarely incorporate explicit multi-criteria structures or decision-maker preferences, limiting transparency, interpretability, and managerial control over the evaluation process.

2.2. Use of generative AI in decision-making

Recent advances in generative AI, particularly "Large Language Models (LLMs)" have led to growing interest in their applications to decision-making and decision support systems. For instance Shan and Li, (2025) propose a decision support system for emergency response in case of disasters, that integrates generative AI. Singh et al. (2025) discuss various applications of generative AI in supply chain management as a decision support tool.

Unlike traditional AI models trained on narrowly defined predictive tasks, generative AI models are capable of synthesizing knowledge, reasoning across domains, and generating structured evaluations (Dhamani & Engler, 2026). These qualities of generative AI make them suitable as support tools in MCDM studies. Hackl et al. (2023) evaluated whether generative AI are reliable and consistent raters, and found that the ratings provided by the selected generative AI had high interrater reliability. This cements the idea that generative AI could act as an alternate expert in MCDM studies. Yadav et al. (2025b) extend this idea and explores the

usage of generative AI models as an alternate expert in a fuzzy MCDM framework case study alongside human experts and found that "divergence of opinions between the AI experts and human experts in the extremes and alignment of opinions in the non-extremes". Similarly, Yadav (2025) proposed a purely generative AI driven MCDM framework for prioritization of sustainable city factors. Despite these findings, the existing literature remains limited in both scope and methodological depth. Moreover, limited attention has been paid to influence relationship modelling when using generative AI as experts in MCDM studies.

3. Methods

3.1. Mathematical preliminaries (Hyperbolic Fuzzy Sets)

Traditional fuzzy sets such as Intuitionistic fuzzy sets (IFS), Pythagorean fuzzy sets (PFS), Fermatean fuzzy sets (FFS) face significant limitation in terms of their decision space due to their sum-based constraints. Dutta and Borah (2023) proposed Hyperbolic fuzzy sets (HYFS) that use product-based constraints therefore overcoming the decision space limitations of traditional fuzzy sets. Specifically, sum-based constraints in IFS ($\mu + \gamma \leq 1$) and PFS ($\mu^2 + \gamma^2 \leq 1$) restrict the feasible region of (membership, non-membership) pairs to a subset of the unit square, whereas the product constraint $\mu \cdot \gamma \leq 1$ is satisfied by all $(\mu, \gamma) \in [0,1]^2$, effectively permitting the full unit square as the decision space. This expanded feasibility allows membership and non-membership to be expressed independently, without one necessarily diminishing the other, a property particularly valuable when modelling asymmetric expert judgements about influence. It is worth emphasising that, because both the membership and non-membership degrees are by definition restricted to the interval [0,1], their product can never exceed unity; the hyperbolic (product) constraint is therefore automatically satisfied for every admissible (membership, non-membership) pair, which is precisely what makes the entire unit square available as the decision space.

Definition 1 (Dutta & Borah, 2023). For a reference set R, a HYFS HF_R can be defined according to equation (1). In the context of this study, the reference set R represents the set of evaluation criteria used to assess performance in asynchronous video interviews. Formally, $R = \{C_1, C_2, \ldots, C_n\}$, where each element $r \in R$ corresponds to a distinct interview evaluation criterion such as engagement, eye contact, or response structure.

$$HF_R = \{\mu_{hf}(r),\ \gamma_{hf}(r) \mid r \in R\}; \tag{1}$$

$$0 \leq \mu_{hf}(r),\ \gamma_{hf}(r) \leq 1; \tag{2}$$

$$\mu_{hf}(r) \cdot \gamma_{hf}(r) \leq 1, \tag{3}$$

where $\mu_{hf}(r)$ represents the degree of membership/preference, whereas $\gamma_{hf}(r)$ represents the degree of non-membership/non-preference for a given element $r \in R$.

Definition 2 (Banik & Dutta, 2025; Dutta & Borah, 2023). For reference sets A and B, the unary and binary operations on their respective HYFS HF_A and HF_B are given as Eqs. (4)–(9):

$$Score: S(HF_A) = 2\mu_{hf}(a) - \mu_{hf}(a) \cdot \gamma_{hf}(a), a \in A; \tag{4}$$

$$Power: HF_A^p = \left\{ \mu_{hf}(a)^p, \left(1-\left(1-\gamma_{hf}(a)\right)^p \right) \middle| p>0,\ a \in A \right\}; \tag{5}$$

$$Scalar\ multiplication: HF_A \cdot s = \{ \left(1-\left(1-\mu_{hf}(a)^s \right) \right), \gamma_{hf}(a)^s \middle| s>0,\ a \in A \}; \tag{6}$$

$$Complement: HF_A^{'} = \left\{ 1-\gamma_{hf}(a), 1-\mu_{hf}(a) \mid a \in A \right\}; \tag{7}$$

$$Addition: HF_A \oplus HF_B = \left\{ \begin{matrix} \mu_{hf}(a)+\mu_{hf}(b) \\ - \\ \mu_{hf}(a)\cdot\mu_{hf}(b), \\ \gamma_{hf}(a)\cdot\gamma_{hf}(b) \end{matrix} \middle| a \in A,\ b \in B \right\}; \tag{8}$$

$$Multiplication: HF_A \otimes HF_B = \left\{ \begin{matrix} \mu_{hf}(a)\cdot\mu_{hf}(b), \\ \gamma_{hf}(a)+\gamma_{hf}(b) \\ -\gamma_{hf}(a)\cdot\gamma_{hf}(b) \end{matrix} \middle| a \in A,\ b \in B \right\}. \tag{9}$$

The Eqs. (1)–(9) would aid in the integration of HYFS into the framework proposed in the study. Few terms to note for the sections that follow are: i) "Experts" refers to entities that give their inputs on criteria and their relationships. ii) "Participant/ Panel Participant" refers to entities that rates an "Interviewee" on various criteria based on their video interview. iii) "Interviewee" refers to the alternative of this study's MCDM case study.

3.2. Multi-perspective HYFS-DEMATEL for criteria weight estimation and computation of influence relationships between criteria

In this section a HYFS variant of DEMATEL that considers preference and non-preference perspectives is explained. These two perspectives correspond to the membership dimension and non-membership dimension of HYFS respectively, enabling a two-sided representation of expert opinion that goes beyond single-valued influence assessments.

Definition 1 (Wu & Lee, 2007). The technique DEMATEL extends the concept of digraphs to MCDM while discriminating involved criteria into groups "causes" and "effects". A matrix D of dimension $n \times n$ represents a decision digraph in DEMATEL, where the element of matrix d_{ij} captures the influence of criterion i on criterion j.

Definition 2 (Yadav et al., 2025a). When fuzzy sets, particularly orthopair fuzzy sets, are integrated into DEMATEL, the core characteristics of the orthopair fuzzy set: representing uncertainty using two distinct dimensions: "the preference grade" and "the non-preference grade", can be leveraged to interpret the resulting decision matrix from a multi-perspective view.

The steps of the aforementioned variant of DEMATEL for computing criteria weights and the influence relationships between them are as follows:

Step 1: For each expert *e*, consider the following matrices: $U^{(e)} = \left[u_{ij}^{(e)}\right]_{n\times n}$ and $V^{(e)} = \left[v_{ij}^{(e)}\right]_{n\times n}$, where $U^{(e)}$ represents the preference perspective of an expert *e* on influence of criteria *i* on criteria *j*, and similarly, $V^{(e)}$ represents the non-preference perspective of an expert *e* on influence of criteria *i* on criteria *j*. It is to be noted that $\left(u_{ij}^{(e)}, v_{ij}^{(e)}\right)$ is interpreted as a HYFS possessing the properties mentioned in Eqs. (2)–(3).

Step 2: Aggregate the matrices $U^{(e)} = \left[u_{ij}^{(e)}\right]_{n\times n}$ and $V^{(e)} = \left[v_{ij}^{(e)}\right]_{n\times n}$ to obtain $U = \left[u_{ij}^{\theta}\right]_{n\times n}$ and $V = \left[v_{ij}^{\theta}\right]_{n\times n}$ using the Eqs. (10)–(11) respectively.

$$u_{ij}^{\theta} = \sqrt[K]{\prod_{e=1}^{K} u_{ij}^{(e)}}; \tag{10}$$

$$v_{ij}^{\theta} = \sqrt[K]{\prod_{e=1}^{K} v_{ij}^{(e)}}, \tag{11}$$

where u_{ij}^{θ} represent the aggregated preference perspective value of influence of criteria *i* on criteria *j*, v_{ij}^{θ} represent the aggregated non-preference perspective value of influence of criteria *i* on criteria *j*, *K* represents the total number of experts.

Step 3: Normalize the matrices *U* and *V* to obtain the preference based direct relation matrix N_U and non-preference based direct relation matrix N_V using the Eqs. (12)–(13) respectively.

$$N_U = \frac{U}{max\left(max_i\left(\sum_{i=1}^{n} u_{ij}^{\theta}\right), \quad max_j\left(\sum_{j=1}^{n} u_{ij}^{\theta}\right)\right)}; \tag{12}$$

$$N_V = \frac{V}{max\left(max_i\left(\sum_{i=1}^{n} v_{ij}^{\theta}\right), \quad max_j\left(\sum_{j=1}^{n} v_{ij}^{\theta}\right)\right)}. \tag{13}$$

Step 4: Calculate the preference total relation matrix T_U and non-preference total relation matrix T_V using the Eqs. (14)–(15).

$$T_U = N_U * (I - N_U)^{-1}; \tag{14}$$

$$T_V = N_V * (I - N_V)^{-1}, \tag{15}$$

where *I* represents an identity matrix of the same dimension as N_U and N_V that is $n \times n$. The operator ' * ' in Eqs. (14)–(15) represents the matrix multiplication operation.

Step 5: The row-sums and column-sums are to be calculated for both the preference and non-preference total relation matrix using the Eqs. (16)–(19).

$$R_{U_i} = \sum_{j=1}^{n} T_{U_{ij}}; \tag{16}$$

$$C_{U_j} = \sum_{i=1}^{n} T_{U_{ij}}; \tag{17}$$

$$R_{V_i} = \sum_{j=1}^{n} T_{V_{ij}}; \tag{18}$$

$$C_{V_j} = \sum_{i=1}^{n} T_{V_{ij}}, \tag{19}$$

where R_{u_i} and R_{V_i} represents the sums of i^{th} row of T_U and T_V respectively, and similarly C_{u_i} and C_{V_i} represents the sums of j^{th} column of T_U and T_V respectively.

Step 6: Compute the prominence and impact values for the criteria using the Eqs. (20)–(23).

$$\text{Prominence } P_{U_i} = R_{U_i} + C_{U_i}; \tag{20}$$

$$\text{Prominence } P_{V_i} = R_{V_i} + C_{V_i}; \tag{21}$$

$$\text{Impact } I_{U_i} = R_{U_i} - C_{U_i}; \tag{22}$$

$$\text{Impact } I_{V_i} = R_{V_i} - C_{V_i}, \tag{23}$$

where P_{U_i} and P_{V_i} are the prominence values computed for criteria i from the preference and non-preference perspective respectively, similarly I_{U_i} and I_{V_i} are the impact values computed for criteria i from the preference and non-preference perspective respectively.

Step 7: Compute the criteria weights using Eqs. (24)–(27).

$$ICW_{U_i} = \frac{P_{U_i}}{\sum^{n} P_{U_i}}; \tag{24}$$

$$ICW_{V_i} = \frac{P_{V_i}}{\sum^{n} P_{V_i}}; \tag{25}$$

$$CW_{HF_i} = 2 * ICW_{U_i} - ICW_{U_i} * ICW_{V_i}; \tag{26}$$

$$CW_i = \frac{CW_{HF_i}}{\sum^{n} CW_{HF_i}}, \tag{27}$$

where CW_i represents the criteria weight estimated for criterion i, CW_{HF_i} is computed in a manner inspired from the score function of HYFS mentioned in Eq. (4).

Step 8: Perform influence relationship analysis using I_{U_i} and I_{V_i}. In influence relationship analysis mentioned, the preference and non-preference perspectives are to be considered separately. In case of preference perspective if the I_{U_i} value is positive for a criterion i then the criterion is to be classified as that of the "cause" type, if I_{U_i} value is negative for a criterion i then the criterion is to be classified as that of the "effect" type. In case of the non-preference perspective if I_{V_i} value is positive for a criterion i then the criterion is to be classified as that of the "effect" type, if I_{V_i} value is negative for a criterion i then the criterion is to be classified as that of the "cause" type. To map the "effects" to the "causes", a manual threshold is to be applied on the total relationship matrices. These manual thresholds are to be selected in such manner that every "cause" has at least one "effect", while minimizing the number of "effects" mapped to a "cause" to ensure prevention of unnecessary influence mappings. The high-pass and the low-pass threshold algorithms suggested by Yadav et al. (2025a) are

to be utilised for manual selection of threshold for the preference and non-preference total relation matrices to find the preference and non-preference influence mappings respectively.

3.3. Multi-attitudinal variance-based panel participant weight estimation

When participants are elicited to provide their opinions through a structured survey or a questionnaire, it is often the case that their inputs may not necessarily be reliable even if the participants themselves are highly motivated and sincere in their responses. This is often due to two causes: cognitive fatigue, and response bias. This study therefore proposes a multi-attitudinal variance method to estimate how much weight should be given to each participant based on two assumptions: higher variance in participant response suggests lower response bias, and lower variance in participant response suggests focused evaluations therefore lower cognitive fatigue. It is acknowledged that these assumptions are simplifications. High variance may also reflect inconsistency or noise rather than genuine differentiation, and low variance may arise from genuine agreement rather than fatigue. The weighting scheme is therefore best understood as a heuristic that balances two competing risks, and its validity depends on the quality and sincerity of participants' responses. Future work should empirically validate these assumptions across different rater panels. The method proposed by Yadav et al. (2025b) combines both the assumptions to estimate relative weights that should be given to each participant's response. The steps of the method are as follows:

Step 1: For each participant p, a decision matrix $X^{(p)} = \left[x_{ij}^{(p)} \right]_{m \times n}$ is to be constructed, such that x_{ij}^{p} represents value given by participant p to the alternative i on criterion j.

Step 2: The matrix $X^{(p)}$ is to be flattened to a vector of dimension $1 \times m \cdot n$.

Step 3: The flattened vector's variance is to be computed using Eq. (28).

$$var^{(p)} = \frac{1}{m \cdot n} \sum_{ij}^{m \cdot n} \left(x_{ij}^{(p)} - \widehat{x^{(p)}} \right)^2, \tag{28}$$

where $\widehat{x^{(p)}}$ denotes the arithmetic mean of vector $X^{(p)}$.

Step 4: The weights of assumption 1 and assumption 2 as proposed by Yadav et al. (2025b) are to be computed using Eq. (29) and Eq. (30) respectively.

$$wp^{(p)} = \frac{var^{(p)}}{\sum_{p} var^{(p)}}; \tag{29}$$

$$iwp^{(p)} = \frac{1 - wp^{(p)}}{\sum_{p} \left(1 - wp^{(p)} \right)}. \tag{30}$$

Step 5: The assumption 1 and assumption 2 weights are to be combined to get the panel participant weight estimate using Eq. (31).

$$wk^{(p)} = \beta \cdot wp^{(p)} + \left(1 - \beta \right) \cdot iwp^{(p)}, \tag{31}$$

where $wk^{(p)}$ represents the estimated panel participant p's weight, and β is a hyperparameter whose value is set to 0.5 for this study.

3.4. Multi-Attitudinal HYFS-DEMATEL-WASPAS for interviewee rank estimation

Various methods have been prescribed for rank estimation in extant literature. One of the most elegant, straightforward yet comprehensive methods for ranking alternatives of the extant literature is WASPAS proposed by Chakraborty and Zavadskas (2014). WASPAS combines both weighted sum and product sum for ranking the given alternatives. The steps for using WASPAS to rank interviewees are as follows:

Step 1: For each participant *p*, consider a decision matrix $S^{(p)} = \left[s_{ij}^{(p)} \right]_{m \times n}$, where $s_{ij}^{(p)}$ is the score given by participant *p* to the interviewee *i* on criterion *j*.

Step 2: Aggregate the matrices $S^{(p)}$ to a single decision matrix $S = \left[s_{ij}^{\theta} \right]_{m \times n}$ using the Eq. (32).

$$s_{ij}^{\theta} = \prod_{p} \left(s_{ij}^{(p)} \right)^{wk^{(p)}}, \tag{32}$$

where $wk^{(p)}$ represents the relative weight estimated for the participant *p*.

Step 3: Normalize the decision matrix *S* using the Eq. (33) to obtain the normalized decision matrix $N_S = \left[ns_{ij} \right]_{m \times n}$.

$$ns_{ij} = \frac{s_{ij}^{\theta}}{\max\limits_{i} s_{ij}^{\theta}}. \tag{33}$$

Step 4: Compute the weighted sum and weighted product vectors using the Eqs. (34)–(35).

$$wsm_i = \sum_{j} ns_{ij} \cdot CW_j; \tag{34}$$

$$wpm_i = \prod_{j} ns_{ij}^{CW_j}, \tag{35}$$

where wsm_i represents the weighted sum value for interviewee *i*, wpm_i represents the weighted product value for interviewee *i*, and CW_j represents the criteria weight for criterion *j*.

Step 5: Combine the values wsm_i and wpm_i for each interviewee to obtain their WASPAS score value Q_i using Eq. (36).

$$Q_i = \alpha \cdot wsm_i + \left(1 - \alpha\right) \cdot wpm_i, \tag{36}$$

where α is a hyperparameter whose value is set to 0.5 for this study. The WASPAS score value Q_i can further be used to rank the interviewee such that the interviewee with highest score value is assigned rank 1.

In the Figure 1, the complete framework of this study is visualized in an easy-to-understand manner. The notational elements present in figure are to be understood from the Sections 3.2, 3.3, and 3.4. The framework starts with eliciting preference and non-preference-based digraph decision matrices on influence of different criteria on each other from different generative AI models under the assumption of them being the experts, these matrices are used to find the criteria weights. Participants rate the interviewees on the criterion; these ratings are then used as the respective participant's decision matrix and are further used to find the participant's relative weights. The criterion weights, participant's weights, and

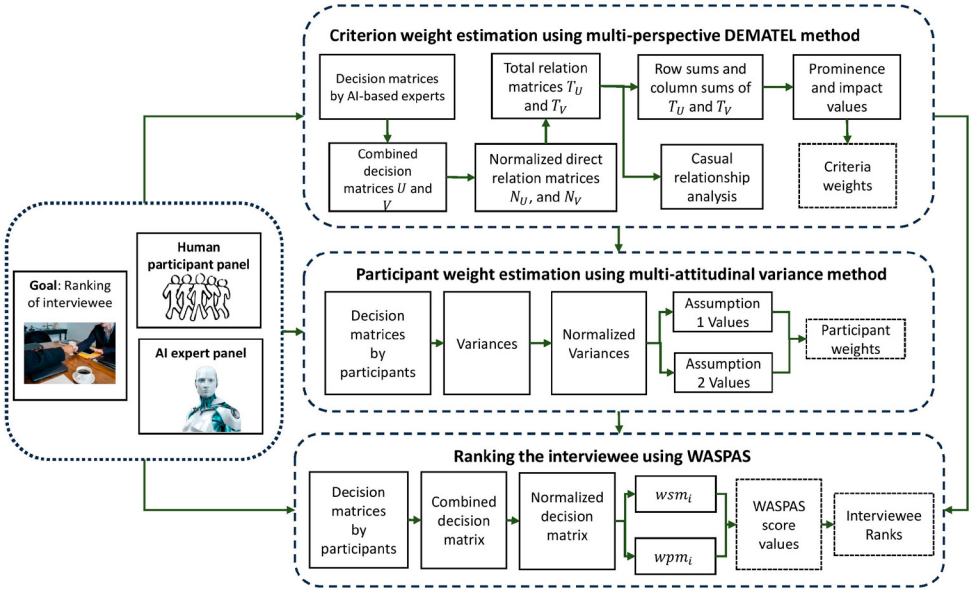

Figure 1. HYFS-DEMATEL-Multi-Attitudinal Variance-WASPAS Interviewee rank estimation framework

decision matrices from participants are then used to rank the interviewees. The framework is further explained in the next section in a form of case study along with the case study's intermediary and final results.

4. A case study

This section presents a case study that focuses on prioritizing interviewees based on their video interviews and analysing the criteria that panel members can use to evaluate candidates through these video interviews. Naim et al. (2015) presents a video interview dataset comprising 138 audio and video recordings and corresponding transcripts of interviews conducted with 69 MIT undergraduate students seeking internships, along with ratings from nine Amazon Mechanical Turk workers across 18 evaluation criteria. The original dataset can be requested from the Rochester HCI Lab (n.d.) following the procedure mentioned in https://roc-hci.com/past-projects/automated-prediction-of-job-interview-performances.

Yadav et al. (2025b) propose the use of generative AI models, particularly those employing reasoning tokens and chain-of-thought reasoning, as an evaluative decision panel in MCDM case studies that have traditionally relied solely on human experts. Building on these ideas, this work extends Section 3 by applying the proposed framework to the dataset presented by Naim et al. (2015).

The generative AI model evaluative panel selected for this case study is as follows: GPT-5.1 by OpenAI (*A*1), and Gemini 2.5 Flash Lite by Google (*A*2). The nine Amazon Mechanical Turk workers of the dataset presented by Naim et al., (2015) form the participant panel members

(as mentioned in Section 3) and for brevity from hereon will be referred *P*1, *P*2, *P*3…*P*9. The criteria taken into consideration are: "Overall" (*C*1), "RecommendHiring" (*C*2), "Colleague" (*C*3), "Engaged" (*C*4), "Excited" (*C*5), "EyeContact" (*C*6), "Smiled" (*C*7), "SpeakingRate" (*C*8), "NoFillers" (*C*9), "Friendly" (*C*10), "Paused" (*C*11), "EngagingTone" (*C*12), "StructuredAnswers" (*C*13), "Calm" (*C*14), "NotStressed" (*C*15), "Focused" (*C*16), "Authentic" (*C*17), "NotAwkward" (*C*18). It is acknowledged that criteria *C*1 and *C*2 represent aggregate outcome assessments rather than independent behavioural cues. Their inclusion in the DEMATEL network is intended to map how lower-level behavioural indicators converge into higher-level hiring judgements, thereby treating the influence structure holistically. Readers should interpret the resulting influence directions from *C*1/*C*2 with this hierarchical distinction in mind.

The steps of applying the decision framework proposed in Section 3 onto the case study along with the results of the case study are as follows:

Step a: Decision matrices of dimension 18 × 18 are constructed using the score values of a criterion's influence on another criterion from both preference perspective and non-preference perspective elicited from the generative AI models for the purpose of criteria weight estimation and criteria influence relationship analysis mentioned in Section 3.2. The context prompt given to the generative AI models is as follows: "You are an expert hiring manager evaluating candidates through video interviews. You are aware of 18 criteria using which a video interview can be evaluated. You are to evaluate the influence of one criterion on another from different perspectives". The prompt given to generative AI models to elicit value $u_{ij}^{(e)}$ (mentioned in Section 3.2) is as follows: "How would you rate the influence criterion <criterion 1> has on <criterion 2> on the scale of 0 to 1? You are to give a decimal value, such that higher value suggests <criterion 1> has a very high influence on <criterion 2>. Think deep to give the decimal value in range 0 to 1 and give nothing else". The prompt to elicit value $v_{ij}^{(e)}$ (mentioned in Section 3.2) is as follows: "How would you rate the non-influence of criterion <criterion 1> on <criterion 2> on the scale of 0 to 1? You are to give a decimal value, such that higher value suggests <criterion 1> has a very little influence on <criterion 2>. Think deep to give the decimal value in range 0 to 1 and give nothing else". The combined decision matrices *U* and *V* formed are given in Supplementary Table 1 and Supplementary Table 2. The prominence values from both the preference and non-preference perspective along with the criteria weights obtained for each criterion are given in Table 1. From the Table 1 it can be seen that the top three criteria as interpreted by the proposed framework and the generative AI models are "Engaged" (C4), "Overall" (C1), and "Not Awkward" (C18), while the three least influential criteria are "Paused" (C11), "StructuredAnswers" (C13), and "NoFillers" (C9).

The influence relationship analysis using the impact values from both the preference and non-preference perspectives is given in Table 2. In the Table 2, the categorization of each criterion to either "cause" or "effect" under both the preference and non-preference perspective is shown. If a criterion happens to be of the "cause" type then it's related "effects" are also mentioned in the Table 2. The "high-pass threshold" computed for preference perspective influence analysis is 0.32437145, and the "low-pass threshold" computed for non-preference influence analysis is 0.12192224.

Table 1. Prominence values, intermediary values, and criteria weights

	R_{U_i}	R_{V_i}	C_{U_j}	C_{V_j}	P_{U_i}	P_{V_i}	ICW_{U_i}	ICW_{V_i}	CW_{HF_i}	CW_i
C1	6.613	3.011	6.990	2.078	13.603	5.088	0.063	0.051	0.123	0.063
C2	4.531	3.396	6.644	1.982	11.175	5.378	0.052	0.054	0.101	0.052
C3	5.214	3.179	6.219	3.187	11.432	6.366	0.053	0.064	0.102	0.053
C4	6.889	2.586	6.785	1.818	13.674	4.404	0.063	0.044	0.124	0.064
C5	6.074	2.747	5.856	2.668	11.929	5.415	0.055	0.055	0.107	0.055
C6	6.234	2.345	5.602	3.194	11.836	5.539	0.055	0.056	0.106	0.055
C7	5.662	2.802	5.141	3.234	10.803	6.036	0.050	0.061	0.097	0.050
C8	5.679	2.678	5.369	3.313	11.048	5.991	0.051	0.060	0.099	0.051
C9	5.233	2.771	5.528	3.309	10.761	6.080	0.050	0.061	0.096	0.050
C10	6.183	2.771	5.908	3.008	12.091	5.779	0.056	0.058	0.109	0.056
C11	5.028	3.507	4.775	3.685	9.803	7.192	0.045	0.072	0.087	0.045
C12	6.401	2.407	6.621	2.343	13.021	4.749	0.060	0.048	0.118	0.060
C13	5.273	2.565	5.297	2.879	10.570	5.443	0.049	0.055	0.095	0.049
C14	6.695	2.685	5.950	2.953	12.645	5.638	0.058	0.057	0.114	0.058
C15	6.895	2.900	6.361	2.693	13.256	5.593	0.061	0.056	0.119	0.061
C16	6.206	1.883	6.312	2.784	12.518	4.667	0.058	0.047	0.113	0.058
C17	6.671	2.570	5.935	2.349	12.606	4.919	0.058	0.050	0.114	0.058
C18	6.642	2.823	6.828	2.149	13.470	4.972	0.062	0.050	0.121	0.062

Table 2. Influence relationship analysis using Multi-perspective HYFS-DEMATEL

	I_{U_i}	I_{V_i}	Preference perspective category	Non-preference perspective category	Preference perspective impact	Non-preference perspective impact
C1	–0.377	0.933	Effect	Effect	–	–
C2	–2.113	1.414	Effect	Effect	–	–
C3	–1.005	–0.008	Effect	Cause	–	C2, C4
C4	0.104	0.768	Cause	Effect	C1, C2, C3, C4, C5, C6, C9, C12, C13, C16, C18	–
C5	0.218	0.079	Cause	Effect	C1, C2, C3, C4, C12, C16, C18	–
C6	0.632	–0.848	Cause	Cause	C1, C2, C3, C4, C5, C12, C16, C18	C1, C2, C4, C5, C6, C12, C17, C18
C7	0.520	–0.432	Cause	Cause	C1, C2, C3, C4, C10, C12, C18	C4, C12
C8	0.309	–0.635	Cause	Cause	C1, C2, C4, C12, C16, C18	C1, C2, C4, C18
C9	–0.296	–0.537	Effect	Cause	–	C2, C4, C18
C10	0.276	–0.236	Cause	Cause	C1, C2, C3, C4, C5, C6, C12, C14, C15, C16, C18	C1, C2, C4
C11	0.252	–0.179	Cause	Cause	C18	C4
C12	–0.220	0.064	Effect	Effect	–	–
C13	–0.024	–0.314	Effect	Cause	–	C1, C2, C4, C12, C13, C17, C18

End of Table 2

	I_{U_i}	I_{V_i}	Preference perspective category	Non-preference perspective category	Preference perspective impact	Non-preference perspective impact
C14	0.744	–0.268	Cause	Cause	C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C12, C13, C15, C16, C17, C18	C1, C2, C14, C15, C18
C15	0.534	0.207	Cause	Effect	C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C12, C13, C14, C15, C16, C17, C18	–
C16	–0.106	–0.902	Effect	Cause	–	C1, C2, C4, C6, C8, C12, C13, C14, C15, C16, C17, C18
C17	0.737	0.221	Cause	Effect	C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C12, C13, C14, C15, C16, C18	–
C18	–0.185	0.674	Effect	Effect	–	–

Note: The preference high-pass threshold computed is 0.32437145, and the non-preference low-pass threshold computed is 0.12192224.

Step b: Out of the nine workers that form the rating participant panel, only 8 have rated all interviewees. Therefore, ratings given by these 8 workers (i.e. *P*1, *P*2…*P*8) have been considered. A total of 8 decision matrices of dimension 138 × 18 (i.e. 138 interviewees rated on 18 criteria) are taken and weights of the 8 workers that form the rating participant panel are computed using the steps mentioned in Section 3.3. The intermediary results and the weights computed for the 8 workers are presented in Table 3.

Table 3. Panel participants' (Amazon Mechanical Turk Workers) weights computed and intermediary results of multi-attitudinal variance method

	Variance	$wp^{(p)}$	$iwp^{(p)}$	Panel Participant Weight
*P*1	1.778	0.105	0.128	0.116
*P*2	1.335	0.079	0.132	0.105
*P*3	3.336	0.196	0.115	0.156
*P*4	1.437	0.085	0.131	0.108
*P*5	1.713	0.101	0.128	0.115
*P*6	2.503	0.147	0.122	0.135
*P*7	2.033	0.120	0.126	0.123
*P*8	2.854	0.168	0.119	0.143

Step c: The weights of criteria obtained from step a, and the weights of panel participants obtained from step b, and the 8 decision matrices of dimension 138 × 18 used in step b are used to find the scores of interviewees using the steps mentioned in Section 3.4. The top 10 and the bottom 10 interviewees computed as per following steps of Section 3.4 are shown in Table 4 along with intermediary results of computing the Q_i (mentioned in Section 3.4).

It is to be noted that in Table 4 the interviewees are denoted using the symbols used in the dataset itself and the notations used in the Table 4 come from Section 3.4. To compare the results of the proposed framework, the top 10 and the bottom 10 interviewees as per the dataset's own aggregate scores are given in Table 5. The Spearman rank correlation between the results of the proposed framework and the dataset's own aggregate score is 0.9889 ($p = 1.38 * 10^{-114}$), and the Kendall tau correlation is 0.9191 ($p = 1.45 * 10^{-57}$). These correlation metrics indicate strong rank-order agreement with the dataset's own aggregate scores; however, since both rankings are derived from the same underlying participant ratings, this agreement reflects internal structural consistency rather than independent predictive validity.

Table 4. Top 10 and the bottom 10 interviewees as per the proposed framework, along with the intermediary results

Top 10					Bottom 10				
Interviewee	wpm_i	wsm_i	Q_i	Rank	Interviewee	wpm_i	wsm_i	Q_i	Rank
p57	0.9440	0.9461	0.9451	1	p30	0.5414	0.5632	0.5523	138
pp57	0.9360	0.9377	0.9369	2	pp69	0.5557	0.5797	0.5677	137
pp81	0.9065	0.9078	0.9071	3	pp47	0.5877	0.6017	0.5947	136
pp33	0.9012	0.9060	0.9036	4	p52	0.5895	0.6174	0.6034	135
pp42	0.8931	0.8982	0.8957	5	pp48	0.5975	0.6147	0.6061	134
pp24	0.8948	0.8961	0.8955	6	pp6	0.6011	0.6197	0.6104	133
p32	0.8865	0.8943	0.8904	7	p69	0.6015	0.6237	0.6126	132
pp14	0.8880	0.8901	0.8891	8	p61	0.6073	0.6233	0.6153	131
p83	0.8857	0.8900	0.8878	9	p47	0.6297	0.6450	0.6374	130
p84	0.8823	0.8866	0.8844	10	p56	0.6415	0.6542	0.6479	129

Table 5. Top 10 and the bottom 10 interviewees as per the dataset's own aggregate scores

Interviewee	Rank	Interviewee	Rank
p57	1	pp69	138
pp57	2	p30	137
pp81	3	pp6	136
pp33	4	p69	135
pp14	5	pp48	134
pp24	6	p52	133
p32	7	pp20	132
p84	8	p61	131
p76	9	pp30	130
pp76	10	p21	129

5. Discussion and conclusions

This study set out to address two gaps in the evaluation of asynchronous video interviews which are absence of explicit influence modelling between the evaluation criteria, and over-reliance on either opaque predictive models or small, fatigue prone human expert evaluators.

The results obtained from the proposed HYFS-DEMATEL-multi-attitudinal variance-WASPAS framework provide several insights that are both theoretically and practically relevant. First, with respect to the research question RQ1, the influence analysis reveals that AVI is not driven by isolated criteria, but by a network of interdependencies. Criteria such as "EyeContact" (C6), "Smiled" (C7), "SpeakingRate" (C8), "Friendly" (C10), "Paused" (C11), and "Calm" (C14) consistently emerge as drivers (i.e. are categorized into the "cause" category by both the preference and non-preference perspectives). Second, the prominence-weight results directly address the research question RQ2 by demonstrating the HYFS-based weighting approach. Theoretically, HYFS permit independent expression of preference and non-preference across the full unit square, unlike sum-constrained orthopair fuzzy sets that restrict this space; however, a direct experimental comparison is beyond the scope of this study and remains a direction for future work.

The relatively high weights assigned to "Engaged" (C4), "Overall" (C1), and "Not Awkward" (C18) reflect the generative AI's ability to synthesize behavioural theory with practical hiring heuristics. At the same time criteria such as "Paused" (C11), "StructuredAnswers" (C13), and "NoFillers" (C9) receiving lower weights indicate that micro-level speech disfluencies are perceived as less important once broader "Engagement" and authenticity cues are accounted for.

The ranking results further support the framework's practical relevance. The high overlap between the top-ranked and bottom-ranked interviewees suggests broad rank-order consistency. It should be noted that since both rankings derive from the same underlying participant ratings, this comparison assesses structural alignment rather than constituting fully independent external validation. Accordingly, the very high level of agreement should not be interpreted as evidence of predictive superiority of the proposed framework over the benchmark, since both rankings originate from the same underlying participant ratings rather than from an independent ground-truth outcome.

Finally, the use of generative AI as evaluative agents proves to be both feasible and informative. This study provides initial evidence that large language models can serve as scalable expert surrogates in MCDM contexts, particularly where human expert elicitation is costly or impractical. Given that only two LLMs were used in a single case study, broader claims about generalisability should be treated with caution pending replication across more models, domains, and datasets. This does not eliminate human judgement from the process, rather it repositions human involvement to validation, governance, and ethical oversight roles.

While the study has several merits, the following limitations should be noted. First, the influence structures derived from DEMATEL reflect expert-like judgements from two generative AI models rather than empirically observed causal mechanisms; the relationships are therefore better understood as plausible directional dependencies rather than causal claims in the formal sense. Second, the case study relies on a single dataset and two LLMs, which constrains generalisability; results may differ across datasets, AI model families, and application domains. Third, the validation comparison is not fully independent as both the proposed ranking and the benchmark aggregate ranking are computed from the same underlying participant ratings, so the observed alignment assesses structural consistency rather than external validity. Fourth, no formal baseline comparison against simpler MCDM approaches is

provided, leaving open the question of whether the added complexity yields materially better results. Fifth, the variance-based participant weighting rests on assumptions about response bias and cognitive fatigue that have not been empirically verified in this study. Finally, while generative AI provides consistency, it may encode latent biases from its training data. Future research should explore hybrid expert panels, conduct sensitivity analysis across AI model families, include formal baseline comparisons, and extend the framework to real-time or multimodal AVI settings.

In summary, this work contributes a practically actionable framework for AVI evaluation. It demonstrates that explainable, influence-aware, and uncertainty-aware decision-modelling is not only possible in automated hiring contexts but also necessary for building fair, trustworthy, and effective recruitment systems.

Author contributions

AY and RK conceived the study and were responsible for its design and data analysis development. Under RK's supervision, AY handled data collection, analysis, and interpretation. AY wrote the first draft of the manuscript, which RK reviewed and revised.

Disclosure statement

The authors declare no conflicts of interest.

References

Alshammari, R. F. N., Rahman, A. H. A., Arshad, H., & Albahri, O. S. (2023). Real-time robotic presentation skill scoring using multi-model analysis and fuzzy delphi–analytic hierarchy process. *Sensors*, *23*(24), Article 9619. https://doi.org/10.3390/s23249619

Aly, A. M., Alawida, M., Aldhaen, G., El Alem, A., & Albloushi, S. (2025). Optimizing recruitment with AI: A smarter approach to candidate evaluation. In *2025 International Conference on Computational Intelligence and Approaches Applications*, *ICCIAA*. IEEE. https://doi.org/10.1109/ICCIAA65327.2025.11013684

Banik, A. K., & Dutta, P. (2025). An efficient decision making method based on hyperbolic fuzzy environment with new score function and its application in determining crime prone zones. *International Journal of Machine Learning and Cybernetics*, *16*, 9807–9833. https://doi.org/10.1007/s13042-024-02404-z

Biswas, S., Jung, J.-Y., Unnam, A., Yadav, K., Gupta, S., & Gadiraju, U. (2024). "Hi. I'm Molly, Your Virtual Interviewer!" Exploring the impact of race and gender in AI-powered virtual interview experiences. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *12*, 12–22. https://doi.org/10.1609/hcomp.v12i1.31596

Cahigas, M. M. L., Robielos, R. A. C., & Gumasing, M. J. J. (2021). Application of multiple criteria decision-making methods in the human resource recruitment process. In *Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management*. IEOM Society International, Singapore. https://doi.org/10.46254/AN11.20210265

Chakraborty, S., & Zavadskas, E. K. (2014). Applications of WASPAS method in manufacturing decision making. *Informatica*, *25*(1), 1–20. https://doi.org/10.15388/Informatica.2014.01

Dalgleish, S. (2005). Recruiting quality. *Quality*, *44*(6), 14.

Dhamani, N., & Engler, M. (2026). *Introduction to generative AI*. Simon and Schuster.

Diefenhardt, F. (2025). Automating the managerial gaze: Critical and genealogical notes on machine learning in personnel assessment. *The International Journal of Human Resource Management*, *36*(14), 2516–2549. https://doi.org/10.1080/09585192.2025.2470302

Dutta, P., & Borah, G. (2023). Construction of hyperbolic fuzzy set and its applications in diverse COVID-19 associated problems. *New Mathematics and Natural Computation*, *19*(1), 217–288. https://doi.org/10.1142/S1793005723500072

Germanier, E., He, M., Rufai, A. M., Garner, P. N., Bangerter, A., Renier, L. A., Schmid Mast, M., & Orji, K. (2025). Identifying storytelling in job interviews using deep learning. *Computers in Human Behavior Reports*, *19*, Article 100688. https://doi.org/10.1016/j.chbr.2025.100688

Grabara, J. K., Kot, S., & Pigoń, Ł. (2016). Recruitment process optimization: Chosen findings from practice in Poland. *Journal of International Studies*, *9*(3), 217–228. https://doi.org/10.14254/2071-8330.2016/9-3/17

Hackl, V., Müller, A. E., Granitzer, M., & Sailer, M. (2023). Is GPT-4 a reliable rater? Evaluating consistency in GPT-4's text ratings. *Frontiers in Education*, *8*, Article 1272229. https://doi.org/10.3389/feduc.2023.1272229

Hwang, C.-L., & Yoon, K. (1981). Methods for multiple attribute decision making. In C.-L. Hwang & K. Yoon (Eds.), *Lecture notes in economics and mathematical systems*: *Vol. 186*. *Multiple attribute decision making* (pp. 58–191). Springer. https://doi.org/10.1007/978-3-642-48318-9_3

İlhan, Ü. D., Güler, B. K., Turgut, D., & Duran, C. (2025). Opportunities and challenges of asynchronous video interviews: Perceptions of human resources professionals from Türkiye. *PLoS ONE*, *20*(6), Article e0325932. https://doi.org/10.1371/journal.pone.0325932

Lukacik, E.-R., & Bourdage, J. S. (2025). Does design matter? The (limited) effects of six asynchronous video interview design features on impression management, reactions, and evaluations. *International Journal of Selection and Assessment*, *33*(1), Article e12511. https://doi.org/10.1111/ijsa.12511

Lukacik, E.-R., Bourdage, J. S., & Roulin, N. (2022). Into the void: A conceptual model and research agenda for the design and use of asynchronous video interviews. *Human Resource Management Review*, *32*(1), Article 100789. https://doi.org/10.1016/j.hrmr.2020.100789

Maree, M., Kmail, A. B., & Belkhatir, M. (2019). Analysis and shortcomings of e-recruitment systems: Towards a semantics-based approach addressing knowledge incompleteness and limited domain coverage. *Journal of Information Science*, *45*(6), 713–735. https://doi.org/10.1177/0165551518811449

Mejia, C., & Torres, E. N. (2018). Implementation and normalization process of asynchronous video interviewing practices in the hospitality industry. *International Journal of Contemporary Hospitality Management*, *30*(2), 685–701. https://doi.org/10.1108/IJCHM-07-2016-0402

Naim, I., Tanveer, M. I., Gildea, D., Mohammed, & Hoque, M. E. (2015). *Automated analysis and prediction of job interview performance*. arXiv. https://doi.org/10.48550/arXiv.1504.03425

Neira, C. C., & Aracena, J. H. (2020). "Hiring someone like myself…". Discourse and discrimination in recruitment and selection processes in large firms in Chile. *Discurso y Sociedad*, *14*(4), 790–822.

Pesak, P., & Hlavacs, H. (2026). A stealth serious game about hiring bias. In A. Thomas, M. Meyer, & M. Zank (Eds.), *Lecture Notes in Computer Science*: *Vol. 16243*. *Serious games*. *JCSG 2025* (pp. 27–33). https://doi.org/10.1007/978-3-032-10518-9_4

Rochester HCI Lab. (n.d.). https://roc-hci.com/past-projects/automated-prediction-of-job-interview-performances

Saaty, T. L. (1980). The analytic hierarchy process. McGraw-Hill. https://doi.org/10.21236/ADA214804

Shan, S., & Li, Y. (2025). Research on the application framework of generative AI in emergency response decision support systems for emergencies. *International Journal of Human–Computer Interaction*, *41*(15), 9191–9208. https://doi.org/10.1080/10447318.2024.2423335

Silva, L. F. da, Oliveira, P. S. G. de, Grander, G., Penha, R., & Bizarrias, F. S. (2024). Soft skills fuzzy TOPSIS ranked multi-criteria to select project manager. *International Journal of Information and Decision Sciences*, *16*(1), 19–45. https://doi.org/10.1504/IJIDS.2024.136280

Singh, V., Hughes, L., Albashrawi, M. A., Jeon, I., & Dwivedi, Y. K. (2025). Generative artificial intelligence (GenAI) in procurement and supply chain management: Applications, opportunities and challenges. *Journal of Systems and Information Technology*, *28*(1), 123–144. https://doi.org/10.1108/JSIT-03-2025-0139

Wu, W.-W., & Lee, Y.-T. (2007). Developing global managers' competencies using the fuzzy DEMATEL method. *Expert Systems with Applications*, *32*(2), 499–507. https://doi.org/10.1016/j.eswa.2005.12.005

Yadav, A. (2025). Prioritizing sustainable city factors: A generative AI-driven fermatean fuzzy prioritization framework. *Management Analytics and Social Insights*, *2*(2), 108–122.

Yadav, A., Krishankumar, R., Ravichandran, K. S., & Ecer, F. (2025a). Analyzing barriers to discrimination mitigation for women's empowerment to achieve SDG 5–gender equality. *Sustainable Development*, *34*(S2), 757–777. https://doi.org/10.1002/sd.70356

Yadav, A., Krishankumar, R., Ravichandran, K. S., Zemlickien, V., & Turskis, Z. (2025b). Prioritising strategies to improve girls' access to quality education: A hybrid hyperbolic fuzzy MCDM framework with human and AI expertise. *International Journal of Computers Communications & Control*, *20*(6). https://doi.org/10.15837/ijccc.2025.6.7240