African Journal of Inter/Multidisciplinary Studies (AJIMS)
Volume 8, Issue 1, pp. 1-15
e-ISSN: 2663-4589

A Binary LSTM Self-Attention Mechanism Model for Mitigating Academic Ingratiation in the Selection Interview

Junior Vela Vela and Prabhakar Rontala Subramaniam

[1]Discipline of Information Systems and Technology, University of KwaZulu-Natal, Westville, South Africa

Abstract: Organisations often face challenges in detecting and preventing ingratiation, a type of intentional self-misrepresentation during selection interviews that impairs the fairness and accuracy of recruitment decisions. Ingratiation poses substantial risks by allowing the selection of unsuitable personnel, with long-term ramifications for organisational performance and culture. The authors of this paper applied and adapted a deep learning method, specifically a binary LSTM Self-attention model, to mitigate ingratiation during academic selection interviews. As ingratiation manifests itself through strategic linguistic cues and impression-management behaviours comparable to those observed in social media communication, Twitter data serves as an appropriate proxy for modelling such behavioural patterns in an authentic and unconstrained context. Leveraging behavioural data from a Twitter dataset, the model employs deep learning techniques to analyse communication patterns and predict candidate suitability. Combining Long Short-Term Memory (LSTM) networks and self-attention mechanisms enables the model to effectively incorporate complex context-dependent features, thereby enhancing prediction accuracy. This approach not only addresses the limitations of subjective assessment but also aligns with the growing trend of integrating artificial intelligence into human resource procedures. The indicated method was tested on a Twitter dataset, and the findings show that BINSAMLSTM achieved 96% prediction accuracy and 96% F1 score. These results point out to the practical benefits of using this approach, namely: minimising subjective bias and improving consistency in candidate evaluation. While the approach is promising, it also necessitates critical thinking regarding data governance, fairness, and privacy. Compliance with regulatory frameworks such as the Protection of Personal Information Act (POPIA) and the General Data Protection Regulation (GDPR) is critical for maintaining ethical integrity in AI-assisted hiring. This study advances an evidence-based paradigm for mitigating ingratiation, which adds to improving fairness and decision-making in staff selection procedures, particularly in the academic recruiting context.

Keywords: academic; deep learning; ingratiation; machine learning; self-attention mechanism

CORRESPONDENCE

Email: juniorvela06@gmail.com

EDITORIAL DATES

Received: 04 April 2025
Revised: 26 November 2025
Accepted: 30 November 2025
Published: 02 March 2026

DOI: https://doi.org/10.51415/ajims.v8i1.2672

Introduction

Data has always been present, but its interpretation, significance, and applications vary depending on how it is understood and utilised (Nasution et al., 2023). Big data describes the enormous and intricate datasets produced by various sources, such as social media, sensors, transactions, and more (Tahir & Khan, 2023). However, this massive amount of data will be beneficial only if scientists can figure out how to make it speak. Social media has become more pervasive and integrated into many people's daily lives than before. Information searching, self-presentation, amusement, and the development and maintenance of social relationships are just a few of the well-documented advantages of using and participating in social media (Ellison et al., 2007; Quan-Haase & Young, 2010; Whiting & Williams, 2013). Furthermore, the ability of social media big data to yield insightful information is one of its most prominent benefits. Organisations may discover hidden patterns in these enormous datasets, make data-driven choices, and obtain a deeper understanding using advanced analytics and machine learning approaches. These insights have transformed the marketing, banking, e-commerce, and human resources sectors, thereby increasing innovation and operational efficiency (Tahir & Khan, 2023; Perera & Iqbal, 2021).

Companies rapidly use data-driven solutions to streamline the human recruitment and selection process, making it simpler and more effective (Hunkenschroer & Luetge, 2022). Employee recruitment, selection, and evaluation have long been and continue to be one of the most significant and essential fields of research and practice in both work and organisational psychology and human resource management (Nikolaou, 2021). Recruiting fresh talent is a labour-intensive procedure that takes time and money. Posting job advertisements, gathering and reviewing applications, interviewing and assessing applicants, coordinating all correspondence, exchanging calendars, and overseeing the reference-offer procedure are just a few of the many tasks that recruitment professionals must oversee. An applicant's curriculum vitae (CV) is a major requirement in the traditional recruitment process, which apart from being obsolete, it is a pointless endeavour that results in the employment of applicants who ultimately prove to be "misfits". Recently, recruiters have been using resumes only for gathering applicants' data. Interviewing candidates in person is regarded as the best way to find qualified applicants. Nonetheless, there is evidence that job success and interview results are not correlated (Uma et al., 2023).

According to Ganiyu et al. (2010), university administrators are often surprised by the academics' casual approach to their work. The latter suggests that interview processes could have flaws that lead to hiring the wrong people for a job. Impression management (IM), a goal-directed action that might be conscious or unconscious, dishonest or authentic (Gardner & Martinko, 1988; Bozeman & Kacmar, 1997; Bolino et al., 2016), is used frequently during job interviews (Peck & Levashina, 2017). Consequently, organisations face difficulty when it comes to detecting and preventing fraud during selection interviews. In this context, Ingratiation refers to applicants' deliberate attempts to misrepresent themselves throughout the personnel selection (Roulin et al., 2016). Faking potential hire insights harms organisations and individuals (Melchers et al., 2020). Thus, in this paper, the authors address this challenge by introducing a binary LSTM Self-attention model to mitigate ingratiation in academics' selection interviews. It is a deep learning approach that trains a Twitter dataset using LSTM enhanced by a self-attention layer to predict the suitability of academics. The decision to choose this approach was motivated by evidence from literature, which has shown that behavioural insights can be obtained by analysing interactions between individuals on social media platforms (Young et al., 2009; Myslín et al., 2013; Ballantine et al., 2015; Andalibi & Buss, 2020).

This research paper is structured as follows: In section 2, ingratiation-related studies, academic recruitment, and social media behavioural studies are discussed. Section 3 provides the description of the procedures and techniques involved in analysing the Twitter data employed in this study, whereas section 4, provides a summary of the academic suitability model experiments of this study. Section 5 provides the discussion of findings, and section 6 concludes this study.

Related Works

Although literature on ingratiation, academic recruitment, and social media behavioural analysis is often studied independently, it links in ways that directly influence the current work. Among others, ingratiation studies by Jones and Pittman (1982) and Yan et al. (2020) demonstrate how candidates may intentionally alter their communication styles, via flattery, conformity, or self-promotion, to persuade assessors. Academic recruiting research shows that such impression management compromises merit-based selection by prioritising perceived likability over competence. Meanwhile, social media behavioural analytics by researchers such as Kern et al. (2019) and Menon and Rahulnath (2016) have demonstrated that linguistic and interactional clues on platforms

African Journal of Inter/Multidisciplinary Studies (AJIMS)
Volume 8, Issue 1, pp. 1-15
e-ISSN: 2663-4589

such as Twitter can accurately reflect underlying personality, emotion, and intent. Bringing these threads together, in this study, social media data was used as a behavioural mirror, a real-world setting in which ingratiation-like behaviours can be seen without the performative limitations of formal interviews. Twitter-based language serves as a realistic and ethically manageable surrogate for investigating ingratiation since it models the same types of self-presentational and affective indicators that appear throughout selection processes. This synthesis supports the use of social media data as a supplemental behavioural dataset for training prediction models that detect subtle symptoms of self-misrepresentation, rather than as an alternative to interviews. Empirical studies on ingratiation in interview selection, recruitment of academics, and social media behavioural studies are discussed in the sections that follow.

As the workforce ages, universities are having difficulty finding and luring new faculty members (Larkin & Neumann, 2012; Udjo & Erasmus, 2014; Kissoonduth, 2017; Maslen, 2019). Academics are both teachers and researchers, and their long-term contributions to knowledge production, innovation, and skill development, both individually and nationally, are vital. Universities frequently struggle to balance valuing research output and assessing interpersonal abilities. Research success indicators, such as publications and grant acquisitions, dominate recruiting decisions, possibly overshadowing softer, interpersonal skills critical for teaching and collaborative work. Nacalaban and Jala (2024) observed that academics' interpersonal skills substantially impact their overall performance and the success of their students. Furthermore, Adesina et al. (2016) claim that universities must also choose academics with strong interpersonal skills because their attitudes impact the classroom. Generally, the consensus among scholars and practitioners is that when an individual and their profession are compatible, their work is more likely to be fulfilling and beneficial to society as a whole (Barrick & Mount, 1991; Holland, 1966; Kern et al., 2019).

A significant connection exists between ingratiation in the interview selection process and organisational performance; hence, human resources research has focused much attention on the former. Researchers have used different approaches to detect and tackle ingratiation in selection interviews. Vianello et al. (2024) propose warning messages to discourage ingratiation during the interview process. To ascertain how candidates integrate Impression Management (IM) strategies in job interviews, such as IM profiles, Moon et al. (2024) employed Latent Profile Analysis. They also investigated the construct validity of these strategies by evaluating their relationships with individual differences in applicants and interview results. The best model fit was obtained with a profile solution. The levels of general IM, self-versus other-focus, and honest versus dishonest IM use varied among the profiles. Additionally, the construct validity of the IM profiles was supported by their meaningful relationships with candidate disposition and interview results. Arseneault and Roulin (2021) provide a theoretical model to explain how cultural values influence preferences for and use of impression management methods in job interviews. Relying on prior cross-cultural IM models and the GLOBE cultural framework, they propose that various cultural dimensions are linked to variations in applicants' use of honest and deceptive forms of self-focused, other-focused, and defensive IM tactics in interviews.

A study by Campion et al. (1997), found that structured interviews assist in lessening the impact of IM strategies like ingratiation while simultaneously enhancing the validity and reliability of the interview process. It is more difficult for candidates to shift their focus to likability-based strategies during structured interviews since the questions are pre-planned and centre on relevant behaviours. Standardised criteria for assessing responses also decrease the possibility of interviewers being swayed by ingratiation, as they are instructed to focus solely on objective measurements rather than subjective impressions. This strategy assures consistency in candidate evaluations, reducing the impact of ingratiation on hiring decisions. Suen and Hung (2024) conducted a study to address calls for research into how Artificial Intelligence (AI) and AI interfaces affect candidates' IM in asynchronous video interviews. The study created three Audio-Video interfaces (AVI) and assessed real job seekers' self-reported IMs in four experimental treatments. The study revealed that different AI interfaces increased or decreased candidates' honesty and fraudulent IMs in different ways. An exploratory investigation also showed that an AI interface could reduce candidates' interview anxiety.

Another successful strategy for mitigating ingratiation is to provide interviewers training that raises awareness of ingratiation tactics. Interviewers who understand how ingratiation might influence their perception are better able to focus on key facts while avoiding unconscious bias. Levashina and Campion (2007) discovered that interviewers training, which included activities to raise awareness of various impression management methods, dramatically reduced the impact of ingratiation on recruiting results. The training also covered how to recognise genuine vs carefully prepared responses, helping interviewers to distinguish between ingratiation and authentic candidate attributes. To help interviewers remain objective, training programs can concentrate on making them more aware of various forms of ingratiation, including flattery, adjusting to the interviewer's viewpoints, and other self-promotional strategies. It has been demonstrated that asking behavioural and situational questions during interviews lessens the effect of ingratiation since they call for candidates to give concrete examples rather than generalisations that could be twisted to seem likeable. Tsai et al. (2005) discovered that behavioural inquiries, such as "Tell me about a time when you faced a difficult problem at work", require specific answers that are more difficult for applicants to fake or control through ingratiation. Such discoveries provide motivation for conducting more empirical studies to explore more ways of mitigating ingratiation.

Social media behavioural studies utilising machine learning (ML), and deep learning (DL) have expanded quickly in recent years as academics seek to comprehend complicated social interactions online. As social media usage has expanded, there is growing interest in analysing consumer-generated text to determine personality and behavioural attributes (Han et al., 2020). Social media platforms such as Twitter, Facebook, and Instagram provide much user-generated content that can be evaluated for sentiment, social interactions, viewpoints, mental health, and misinformation. Researchers can quickly process, classify, and anticipate these actions using machine learning and deep learning models, generating previously unachievable insights. Wani et al. (2022) presented an effective deep learning (DL), and artificial intelligence (AI) based model for detecting depressed people on social media. The model trains a convolutional neural network (CNN) and long-short-term memory (LSTM) models utilising hybrid feature-based behavioural-biometric signals that were recorded using Word2Vec and term frequency-inverse document frequency (TF-IDF) models. Experiments have demonstrated that both CNN and LSTM DL models and the hybrid (CNN + LSTM) models produced encouraging outcomes on both single and combined datasets.

Kholodna's et al. (2021) work sought to develop a machine-learning model for speech-based automatic emotion recognition. The created model was intended to be applied to the public sentiment monitoring system. The machine-learning model produced good results using real-world data, including 90% precision, 89% recall, 0.89% F-measure, 89% accuracy (on validation data), and 88% accuracy on test data. Volkova et al. (2021) initially developed deep learning models for detecting misinformation and disinformation in multilingual and multimodal environments. This work was followed by a psycholinguistic analysis of broad deception categories. They then analysed user behaviour and spread patterns when engaging with false information, identified vulnerable subpopulations and their demographics, and evaluated the speed and scale of diffusion to determine who shares content, how quickly, how much, and how evenly. Ahmad et al. (2021) proposed a hybrid Deep Learning-based model, a Convolutional Neural Network concatenated with Long Short-Term Memory, to demonstrate the effectiveness of the proposed model for detecting significant personality traits, namely: introversion-extroversion, intuition-sensing, thinking-feeling, and judging-perceiving. The evaluations of the model produced improved results, demonstrating that the suggested model can predict the user's personality features when compared to state-of-the-art methodologies. They, therefore, concluded that with the information gained from this study, companies can make more informed decisions on personnel recruiting, and also that they can use the research findings to develop best practices for selecting, managing, and optimising their policies, services, and products.

Machine learning and deep learning applications in social media behavioural studies have made it possible to conduct more comprehensive and accurate evaluations of complex online behaviours. ML and DL techniques offer tools for gaining insights into how people interact, communicate, present themselves, and influence one another on social media, from sentiment and opinion analysis to the detection of hate speech, misinformation, and mental health indicators. Social media research is expected to produce an ever deeper, more complex understanding of human behaviour in digital environments as deep learning models progress. These observations advance scholarly understanding and address social issues, including false information, mental health emergencies, and internet safety. Thus, in this study, the authors propose a data-driven solution to mitigate ingratiation in the interview selection process that uses a binary LSTM Self-attention mechanism model to predict academic suitability.

Materials and methods

In this section, a detailed account of the data and process model utilised to address the challenge of academic ingratiation in selection interviews is presented. The study leverages a Twitter dataset to analyse behavioural patterns and employs an advanced deep learning architecture, specifically, a Long Short-Term Memory (LSTM) network enhanced with a self-attention mechanism to construct a binary classification model for predicting candidate suitability. A Twitter dataset was used for this study. Twitter is a popular microblogging and social networking website that allows users to share, broadcast, and interpret 140-character communications known as tweets. The dataset used for this study was obtained from Vela (2022). The Twitter data consists of tweets from both academic and non-academic users. Also, as depicted in Figure 1, the dataset was categorised into two classes, namely: Suitable [S or 1] with 243070 rows and Not Suitable [NS or 0] with 243070 rows. Vela (2022) used a rule-based technique to label the Twitter dataset, constructing a dictionary of keywords susceptible to be used by academics. This dictionary was used to identify and classify tweets as suitable or not suitable based on the presence of these domain-specific keywords.

	text	Label
0	['today', 'the', 'high', 'court', 'ruled', 'th...	1
1	['we', 'always', 'win', 'elections', 'in', 'zi...	0
2	['recruiting', 'a', 'phd', 'student', 'in', 'b...	1
3	['dear', 'pg', 'students', 'let', 's', 'encour...	1
4	['kuzoti', 'passport', 'hatitaure', 'so', 'pai...	0
5	['is', 'the', 'government', 'aware', 'of', 'th...	1
6	['shoutout', 'to', 'these', 'cadres', 'doing',...	1
7	['nice', 'one', 'vamwe', 'kumuka', 'kunotinets...	0
8	['really', 'why']	0
9	['passport', 'fees', 'anti', 'poor', 'grossly'...	0

Figure 1. Dataset sample

Implementing such recruitment machine learning-based systems, however, poses ethical challenges. Data privacy, informed consent, and fairness or ethical considerations must be prioritised to guarantee that AI-powered recruitment tools work within legal and ethical constraints (Dhiman, 2023). Although Twitter data is public, it may contain personally identifiable information or sensitive opinions that users did not intend to be used for employment-related analysis. Therefore, compliance with the recognised data governance frameworks, particularly the Protection of Personal Information Act (POPIA) in South Africa and the General Data Protection Regulation (GDPR) in the European Union, is critical. These frameworks reaffirm principles such as lawful and transparent processing, purpose limitation, data minimisation, and individual rights protection, all of which are critical when using social media-derived data for recruitment purposes.

LSTM

Long Short-Term Memory (LSTM) is a recurrent neural network (RNN) architecture designed to simulate sequences while addressing typical RNN limitations such as the vanishing gradient problem. LSTMs are extremely effective for tasks that need sequential data, such as time series prediction, natural language processing, and speech recognition. Using the assistance of a memory cell and gating mechanisms that control input flow, LSTMs can capture dependencies over extended periods, in contrast to regular RNNs that have trouble retaining information over lengthy sequences because of diminishing gradients (Zhao et al., 2020; Landi et al., 2021).

Self-attention mechanism

Self-attention operates by projecting each token in the input sequence into three learned representations (Li et al., 2020; Wang et al., 2023): Queries (Q), Keys (K), and Values (V). These are obtained by multiplying the input embeddings $X$with trainable weight matrices:

$Q = XW_Q,\ K = XW_K,\ V = XW_V$

The attention mechanism determines how much each token should focus on each other token (Edelman et al., 2022; Chen et al., 2022). This is done by computing the dot product between a Query and all Keys, scaling the result by $\sqrt{d_k}$ for numerical stability, and applying the SoftMax function. Here, d_k represents the dimensionality of the Key vectors; scaling by $\sqrt{d_k}$ prevents the dot-product values from becoming excessively large as the vector dimension increases, which helps keep the SoftMax distribution well-behaved and stabilises training.

$$Attention(Q, K, V) = SoftMax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Query (Q): Represents what the current token is trying to retrieve from the sequence.
Key (K): Represents the features against which relevance is measured.
Value (V): Contains the information that will be aggregated.

The SoftMax function converts the similarity scores into normalised weights, ensuring they sum to 1. The final output is a weighted sum of the Value vectors, where the weights reflect the attention each token pays to others.

Sequence modelling has been transformed by the self-attention mechanism, enabling models to effectively capture intricate connections in data. Despite its computational difficulties, advancements in effective self-attention are opening the door for its wider use across a range of fields, including the ingratiation research field.

Performance metrics

To address the model evaluation as depicted in Figure 2, different performance metrics were used to predict academic suitability. Experimental evaluation was carried out on factors such as accuracy, precision, Recall, and F1-Score.

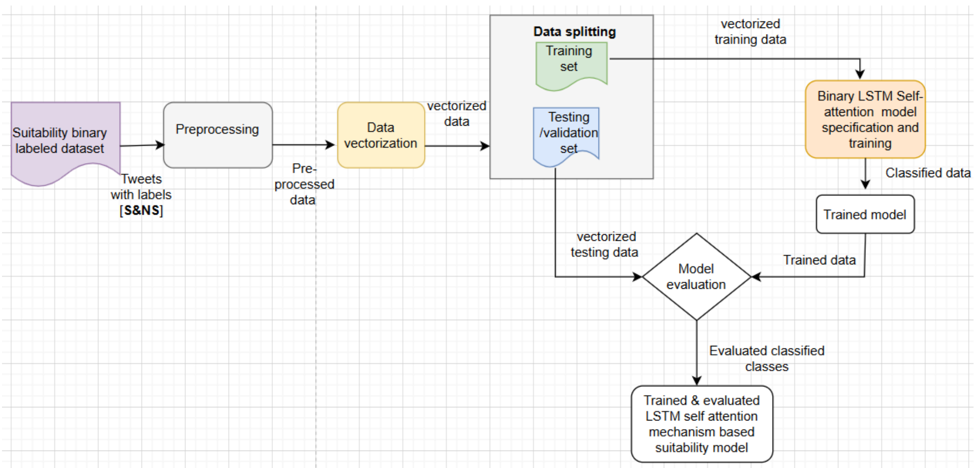

Figure 2. Binary LSTM Self-Attention Mechanism Process model

African Journal of Inter/Multidisciplinary Studies (AJIMS)
Volume 8, Issue 1, pp. 1-15
e-ISSN: 2663-4589

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F1} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Where TN stands for True Negative, TP stands for True Positive, FN stands for False Negative, and FP stands for False Positive.

The methodology used in this study

The purpose of conducting this study was to design and implement a Deep Learning model that detects the suitability of academics from a Twitter dataset. To achieve this study's objective, the authors used a Binary LSTM Self-Attention Suitability process model as a framework. A process model is a condensed depiction of the steps taken to create an artefact (Brown et al., 2021; Kooij et al., 2020). Social media interactions continue to expand at an accelerated rate, as is its use in academia and in business. Sentiment analysis, opinion mining, and social media big data analysis remain topics of research and innovation in these domains. Therefore, the analysis of this data became essential. These data are often analysed using feature extraction, classification, and data pre-processing. Figure 2 gives a clear picture of the fundamental layout as well as the flow and interactions between the various parts of the self-attention mechanism model development of the present study.

Figure 2 displays the steps taken in developing the LSTM self-attention mechanism binary model. The labelled dataset was first pre-processed using rule-based techniques and then vectorised using different word embedding techniques, such as Glove and Keras Tokenizer. The model specification took place before the model was trained. During the model specification, the hyperparameters were optimised. The last step was the model evaluation with the testing/validation set.

Experiments to mitigate academic ingratiation

In this study, five binary classification models were experimented with. All the models in this study adhered to a uniform procedure, transitioning from labelled data to training through pre-processing and data vectorisation, as illustrated in Figure 2. The dataset was trained with various classifiers, including Bi_LSTM, LSTM, Random Forest, Decision Tree, and XGBoost, to make the results more robust. Following training, the models were tested using the test sets. The five experiments conducted and their experimental parameters are depicted in Table 1.

Table 1. Experimental parameters

Name	Description	Parameters
BiLSTM.G.2	Bidirectional Long-Short Memory	Word embedding: GloVe LSTM =32 wt_decay=1e-4 Batch_size =256 Dropout = 0.1 Num_epochs= 30
RF2	Random Forest	Word embedding: Term frequency-inverse document frequency (TF IDF) Classifier: Random Forest n_estimators=100 random_state=42
BINSAMLSTM	Binary Self-attention Long-Short Memory	Word embedding: Tokenizer Classifier: LSTM Layers: input, embedding, LSTM, Self-attention and output-dense epochs=5 batch_size=32

(*Continued*)

Table 1. (Continued)

Name	Description	Parameters
		loss='binary_crossentropy
XGBoost	Extreme Gradient Boosting	Word embedding: TF IDF Classifier: XGBoost objective='binary:logistic', eval_metric='logloss', max_depth=6, learning_rate=0.1, n_estimators=100, n_jobs=-1, random_state=42
Decision Tree	Decision Tree	Word embedding: TF IDF Classifier: Decision Tree random_state=42

Procedure 1 lists the proposed pseudo code and outlines the detailed steps required for implementing the predictive model (Figure 2).

Procedure1: BINSAMLSTM

Input : Binary Labelled data, L_D
Data Variables : TL_D = Tokenised Labelled Data;
VL_D = Vectorised Labelled Data;
X_train = Training Data_Text;
Y_train = Training Data_Label;
X_test = Test Data_Text; Y_test = Test Data_Label
Output: A = Accuracy; *p* = Precision; R = Recall;
L = Test Loss; F1 = F1 score

1. **Data Pre-processing**
   1.1→Removal of special characters (URL, @,#,1)
   1.2→Conversion of text to lowercase
   1.3→TL_D ← Tokenise (L_D)
   1.4→Removal of stop words
   1.5→LL_D ←Lemmatise (TL_D)
2. **Data Vectorisation**
   2.1→VL_D ← Vectorise (LL_D)
3. **Data Splitting (80:20)**
   3.1→X_train ← Y_train ← split (VL_D, 80)
   3.2→X_test ← Y_test ← split (VL_D , 20)
4. **LSTM Model Specification**
   4.1→Input_layer → Output_Shape(X_train, Y_train, 100)
   4.2→Embedding_layer → Output_Shape (X_train, Y_train, 100, 128)
   4.3→LSTM_Layer → Output_Shape(X_train, Y_train, 64)
   4.4→SAM_Layer ← $\frac{\mathbf{Q}_i . K_j}{\sqrt{d_k}}$
   4.5→Dense_layer ← Activate(sigmoid)
   4.6→Model ← combine (input_layer, embedding_layer, LSTM_Layer, Dense_Layer)
   4.7→Compile (Model, Adam, A)
   4.8→Save Model (X_train, Y_train) //Trained Model
5. **LSTM Model Evaluation**
   5.1→Model (X_train, Y_train) ← X_test & Y_test
6. **Return Model Output**
   6.1 *Return Model(A) //Compute Eq (2)*
   6.2 *Return Model(P) //Compute Eq (3)*
   6.3 *Return Model(R) //Compute Eq (4)*
   6.4 *Return Model(L)*
   6.5 *Return Model(F1) //Compute Eq (5)*

African Journal of Inter/Multidisciplinary Studies (AJIMS)
Volume 8, Issue 1, pp. 1-15
e-ISSN: 2663-4589

Classification performance

Table 2 and Figure 3 display the performance metrics of the different classifiers. Following training and evaluation, as depicted in Figure 2, the BINSAMLSTM produced better accuracy, F1 Scores, precision, and recall of 96% with a validation loss of 0.14. The BINSAMLSTM consists of five layers: the input layer, embedding layer, self-attention layer, LSTM layer, and a dense layer. Similarly, the Decision Tree and RF.2 models performed well with 92% accuracy. However, BiLSTM.G.2 and XGBoost have 86% and 84% accuracy, respectively. Figure 4 illustrates the performance of the two classes of the dataset.

Table 2. Classification performance

Classifier	Accuracy	Precision	Recall	F1 score
BiLSTM.G.2	0.86	0.86	0.81	0.86
RF.2	0.92	0.93	0.93	0.93
BINSAMLSTM	0.96	0.96	0.96	0.96
XGBoost	0.84	0.86	0.85	0.85
Decision Tree	0.92	0.92	0.92	0.92

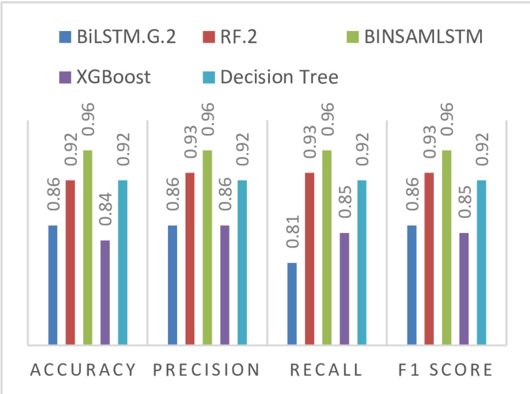

Figure 3. Classification performance using the Twitter dataset

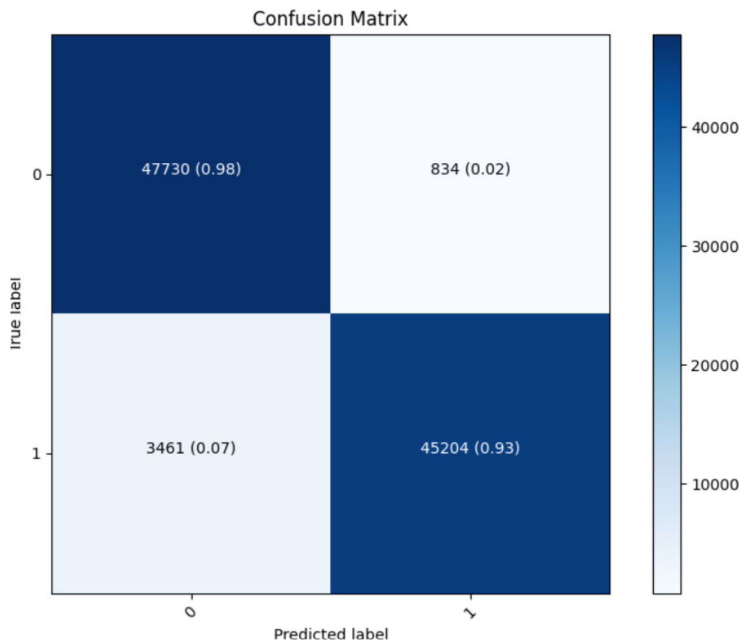

Figure 4. Confusion matrix

Discussion

To contextualise the study's findings, the results were compared with literature by the following authors: Kern et al. (2019), Menon and Rahulnath (2016), as well as Böhmová and Chudán (2018), which employs social media datasets to analyse user behaviour, traits, emotions, attitudes, and sentiments in recruitment processes and/or human resources. Evaluating the performance of the LSTM self-attention binary model is crucial for -understanding its effectiveness in sequence-based tasks such as text classification, sentiment analysis, and anomaly detection. LSTM networks, known for their ability to handle long-term dependencies, work seamlessly with self-attention mechanisms that emphasise key parts of sequences, addressing challenges like vanishing gradients and contextual ambiguity. Metrics like accuracy, precision, recall, and F1-score, along with computational efficiency, provide a comprehensive assessment of the model's real-world applicability. While the binary classification nature simplifies output interpretation, it necessitates careful consideration of class balance and prediction reliability, offering insights into the model's strengths and areas for enhancement.

For instance, Kern et al. (2019) used an XGBoost model to identify personality traits related to various professions in data from 128,279 Twitter users spanning 3,513 vocations, with 74% accuracy. However, their approach was based mostly on surface-level language traits and unchanging word frequencies, limiting its potential to capture contextual nuances in communication. The proposed BINSAMLSTM model solves this restriction with its contextual feature extraction, which uses sequential encoding via LSTM to record the temporal flow of language and a self-attention layer to selectively amplify psychologically and contextually meaningful utterances. This allows for more detailed modelling of ingratiation-related behaviours like strategic flattery or conformity, resulting in much increased prediction precision. Similarly, Menon and Rahulnath (2016) used Twitter data to evaluate emotional intelligence and rate job prospects using meta-attribute extraction and multi-label classification with Naïve Bayes, reaching 87.3% accuracy. While their approach improved affective inference in recruitment analytics, it relied on fixed qualities rather than adaptive context learning. The current model overcomes these limits by autonomously learning representational hierarchies from text, allowing it to distinguish genuine emotional tone from manipulative affective cues, which is a significant gain in ingratiation detection. This deep contextual learning explains why the BINSAMLSTM model outperforms previous techniques in terms of both interpretability and performance, as shown in Figure 5.

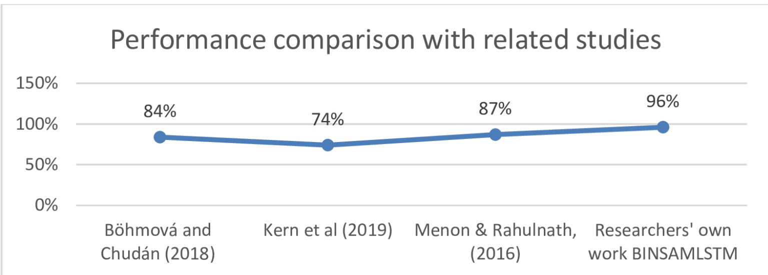

Figure 5. Performance comparison with related studies

Additionally, Böhmová and Chudán (2018) proposed a Facebook-based recruitment model that achieved prediction accuracies ranging from 68%–84%. However, their methodology + was limited in its ability to account for dynamic conversational context, focusing instead on profile metadata and keyword matching. In contrast, the BINSAMLSTM model combines sequence dependency modelling and attention-based weighting, allowing it to detect subtle relational and discourse patterns indicative of authentic versus ingratiating communication, resulting in greater robustness and generalizability.

Organisations often grapple with ingratiation and other deceptive behaviours during personnel selection processes. Ingratiation, where candidates strategically manipulate perceptions through flattery or agreement, compromises the integrity of hiring decisions and risks long-term negative impacts on productivity and organisational culture. Traditional detection methods, reliant on human judgment, are prone to biases and inconsistencies, making them unreliable for identifying ingratiation tactics. These challenges highlight the growing need for innovative, data-driven solutions to ensure fairness and accuracy in recruitment. The LSTM self-attention binary model addresses this gap by leveraging deep learning to analyse behavioural data. Unlike conventional techniques, this model incorporates a self-attention mechanism that identifies critical features within input data, providing a nuanced understanding of behavioural patterns. Training the model on a Twitter dataset allows it to glean insights from unfiltered social media interactions, uncovering communication styles, values, and interpersonal dynamics, key indicators of authenticity and professional suitability. The choice of LSTM networks is particularly apt for processing sequential data, capturing long-term dependencies, and ensuring that context-sensitive expressions are not overlooked. Adding a self-attention layer further refines the model's ability to assign weight to relevant features, enhancing prediction accuracy and mitigating bias.

By integrating LSTM networks with self-attention mechanisms, the proposed model enables a context-aware, data-driven, and ethically grounded approach to address ingratiation and improve hiring accuracy. It marks a significant step forward in harnessing social media and deep learning to achieve unbiased, efficient recruitment, transforming how organisations evaluate and select candidates.

Research Implication

This BINSAMLSTM model introduces a significant shift in mitigating ingratiation during hiring by leveraging behavioural insights from real-world data, enabling organisations to move beyond subjective judgment toward evidence-based decision-making. This aligns with broader trends in AI-driven human resource management (Sharma & Sengupta, 2023; Basnet, 2024; Amri, 2024), where predictive analytics and behavioural modelling are increasingly utilised to enhance human resources and recruitment processes. By combining sequence learning with attention-based contextual encoding, the model provides a scalable and data-driven method for improving objectivity and fairness in academic selection decisions.

However, the implementation of such systems presents practical challenges. Addressing the risks of algorithmic bias and representational imbalance is just as crucial. Models trained on non-representative or culturally constrained datasets may unintentionally reproduce language or demographic biases, resulting in unequal ratings of applicants from underrepresented groups. To reduce such risks, organisations should implement techniques such as bias auditing, stratified sampling, and fairness-aware retraining, which ensure that model outputs are systematically examined for discriminatory trends.

Furthermore, the performance of the BINSAMLSTM framework is dependent on the quality and curation of the training dataset, which necessitates rigorous validation to ensure dependability and fairness. While the model provides advanced analytical capabilities, organisations should exercise caution before relying too heavily on algorithmic outputs. Human supervision is essential for contextual interpretation, allowing assessors to account for nuanced social and cultural indicators that AI models might overlook.

By incorporating these ethical safeguards and governance requirements, this study sheds light to the importance of having a framework that combines technological innovation with ethical stewardship, contributing to the developing discussion of responsible and transparent AI in recruitment. The combination of behavioural modelling, regulatory compliance, and fairness-oriented design highlights AI's potential to improve efficiency while still maintaining integrity and equity in academic and organisational recruiting procedures.

Conclusion

Detecting and preventing ingratiation during selection interviews remains a critical challenge for organisations, especially in high-stakes contexts such as academic hiring. The authors of this paper address this issue by introducing a novel solution: a binary LSTM Self-Attention model. Leveraging deep learning, the proposed approach analyses behavioural patterns to predict candidate suitability, with BINSAMLSTM enabling the detection of nuanced, context-dependent signals of ingratiation. In this study, the researchers tapped into the rich behavioural insights available in social media interactions by training the model on a Twitter dataset, offering an innovative method to understand and mitigate fraudulent self-presentation. The study underscores the potential of advanced deep learning techniques to revolutionise personnel selection by integrating behavioural data into decision-making processes. This integration enhances fairness and efficiency in hiring and minimises the impact of ingratiation, fostering improved organisational outcomes.

A key justification for adopting the Self-Attention mechanism LSTM (BINSAMLSTM) lies in its operational characteristics, which balance capturing sequential dependencies and focusing on contextually significant features. Unlike traditional LSTMs, which rely solely on sequential data processing, the self-attention mechanism allows the model to dynamically weigh and prioritise specific inputs. This capability is particularly advantageous in detecting academic suitability, thus, mitigating academic ingratiation, as it enables the identification of subtle behavioural cues that may otherwise be overlooked. Compared to other models, BINSAMLSTM offers computational efficiency while maintaining high accuracy in binary classification tasks, making it a practical choice for resource-constrained environments such as organisational hiring processes. Although the model performed satisfactorily, its binary classification design poses a limitation by failing to provide nuanced levels of academic suitability. Such granularity is important because it could enhance the mitigation of academic ingratiation during selection interviews. Future research could explore multi-class or multi-label models to address academic ingratiation in selection interviews better, paving the way for more nuanced and effective hiring practices.

Declarations

Interdisciplinary Scope: This study brings together insights from computer science, human resource management, behavioural psychology, and computational linguistics to examine ingratiation as a form of self-misrepresentation in academic recruitment. By leveraging social media behavioural data and deep learning techniques, the research integrates concepts from organisational behaviour, natural language processing, and responsible AI. The development and evaluation of a Binary LSTM Self-Attention Model (BINSAMLSTM) demonstrate how data-driven decision-support systems can enhance fairness, mitigate subjective bias, and improve the reliability of academic hiring processes.

Author Contributions: Conceptualisation (VVJ), methodology (VVJ), dataset preparation and modelling (VVJ), analysis and interpretation of results (VVJ), writing- original draft preparation (VVJ), review and editing (RPS). The final version of the manuscript has been read and approved by all authors.

Conflict of Interest: The authors declare that there is no conflict of interest regarding the publication of this article.

Funding: No external funding was received for this research or for the preparation of this manuscript.

Availability of Data: The study includes all essential data used to train and evaluate the model. Additional details are available from the corresponding author upon reasonable request.

References

Adesina, O. J., Raimi, S. O., Bolaji, O. A., & Adesina, A. E. (2016). Teachers’ attitude, years of teaching experience and self-efficacy as determinants of teachers’ productivity in teachers’ professional development programme in Ibadan Metropolis, Oyo State, Nigeria. *Journal of Emerging Trends in Educational Research and Policy Studies*, *7*(3), 204–211. https://hdl.handle.net/10520/EJC196815

Ahmad, H., Asghar, M. U., Asghar, M. Z., Khan, A., & Mosavi, A. H. (2021). A hybrid deep learning technique for personality trait classification from text. *IEEE Access*, *9*, 146214–146232. https://doi.org/10.1109/ACCESS.2021.3121791

Amri, A. (2024). Trends in human resource management and organisational behaviour. *Economics and Digital Business Review*, *5*(2), 1011–1027. https://doi.org/10.37531/ecotal.v5i2.1374

Andalibi, N., & Buss, J. (2020). *The human in emotion recognition on social media: Attitudes, outcomes, risks*. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–16, New York, NY, USA. https://doi.org/10.1145/3313831.3376680

Arseneault, R., & Roulin, N. (2021). A theoretical model of cross-cultural impression management in employment interviews. *International Journal of Selection and Assessment*, *29*(3–4), 352–366. https://doi.org/10.1111/ijsa.12348

Ballantine, P. W., Lin, Y., & Veer, E. (2015). The influence of user comments on perceptions of Facebook relationship status updates. *Computers in Human Behavior*, *49*, 50–55. https://doi.org/10.1016/j.chb.2015.02.055

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Basnet, S. (2024). Artificial Intelligence and machine learning in human resource management: Prospect and future trends. *International Journal of Research Publication and Reviews*, *5*(1), 281–287. https://doi.org/10.55248/gengpi.5.0124.0107

Böhmová, L., & Chudán, D. (2018). Analysing social media data for recruiting purposes. *Acta Informatica Pragensia*, *7*(1), 4–21. https://doi.org/10.18267/j.aip.111

Bolino, M., Long, D., & Turnley, W. (2016). Impression management in organisations: Critical questions, answers, and areas for future research. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*(1), 377–406. https://doi.org/10.1146/annurev-orgpsych-041015-062337

Bozeman, D. P., & Kacmar, K. M. (1997). A cybernetic model of impression management processes in organisations. *Organizational Behavior and Human Decision Processes*, *69*(1), 9–30. https://doi.org/10.1006/obhd.1996.2669

Brown, P., Von Daniels, C., Bocken, N. M. P., & Balkenende, A. R. (2021). A process model for collaboration in circular oriented innovation. *Journal of Cleaner Production*, *286*, 125499. https://doi.org/10.1016/j.jclepro.2020.125499

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, *50*(3), 655–702. https://doi.org/10.1111/j.1744-6570.1997.tb00709.x

Chen, W., Lu, Y., Ma, H., Chen, Q., Wu, X., & Wu, P. (2022). Self-attention mechanism in person re-identification models. *Multimedia Tools and Applications*, *81*(4), 4649–4667. https://doi.org/10.1007/s11042-020-10494-4

Dhiman, D. B. (2023). Ethical issues and challenges in social media: A current scenario. In *Preprints*. https://doi.org/10.20944/preprints202303.0513.v1

Edelman, B. L., Goel, S., Kakade, S., & Zhang, C. (2022). *Inductive biases and variable creation in self-attention mechanisms*. International Conference on Machine Learning, 5793–5831. https://proceedings.mlr.press/v162/edelman22a.html

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook “friends:” Social capital and college students’ use of online social network sites. *Journal of Computer-Mediated Communication*, *12*(4), 1143–1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x

Ganiyu, G., Okunola, P. O., Oni, A. A., & Adetoro, J. A. (2010). The relationship between vice-chancellors’ leadership behaviour and the work behaviour of lecturers in Nigerian universities: Implications for leadership training for vice-chancellors. Journal of Higher Education in Africa, *8*(1), 123–139. https://doi.org/10.57054/jhea.v8i1.1592

Gardner, W. L., & Martinko, M. J. (1988). Impression management in organizations. *Journal of Management*, *14*(2), 321–338. https://doi.org/10.1177/014920638801400210

Han, S., Huang, H., & Tang, Y. (2020). Knowledge of words: An interpretable approach for personality recognition from social media. *Knowledge-Based Systems*, *194*, 105550. https://doi.org/10.1016/j.knosys.2020.105550

Holland, J. L. (1966). The psychology of vocational choice: A theory of personality types and model environments. *Blaisdell*. https://psycnet.apa.org/record/1966-12652-000

Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics*, *178*(4), 977–1007. https://doi.org/10.1007/s10551-022-05049-6

Jones, E. E., & Pittman, T. S. (1982). Toward a general theory of strategic self-presentation. *Psychological Perspectives on the Self*, *1*(1), 231–262. https://www.researchgate.net/publication/248124372_Toward_a_general_theory_of_strategic_self-presentation

Kern, M. L., McCarthy, P. X., Chakrabarty, D., & Rizoiu, M. A. (2019). Social media-predicted personality traits and values can help match people to their ideal jobs. *Proceedings of the National Academy of Sciences*, *116*(52), 26459–26464. https://doi.org/10.1073/pnas.1917942116

Kholodna, N., Vysotska, V., & Albota, S. (2021). A machine learning model for automatic emotion detection from speech. In *MoMLeT + DS* (pp. 699–713). https://ceur-ws.org/Vol-2917/paper42.pdf

Kissoonduth, K. (2017). *Talent Management: Attracting and retaining academic staff at selected public higher education institutions* University of Unisa. http://hdl.handle.net/10500/24225

Kooij, D. T. A. M., Zacher, H., Wang, M., & Heckhausen, J. (2020). Successful aging at work: A process model to guide future research and practice. *Industrial and Organizational Psychology*, *13*(3), 345–365. https://doi.org/10.1017/iop.2020.1

Landi, F., Baraldi, L., Cornia, M., & Cucchiara, R. (2021). Working memory connections for LSTM. *Neural Networks*, *144*, 334–341. https://doi.org/10.1016/j.neunet.2021.08.030

Larkin, J., & Neumann, R. (2012). Ageing academics: Workforce priorities for universities. *International Journal of Employment Studies*, *20*(1), 3–24. https://search.informit.org/doi/10.3316/informit.891595916444219

Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: Development and validation of an interview faking behavior scale. *The Journal of Applied Psychology*, *92*(6), 1638–1656. https://doi.org/10.1037/0021-9010.92.6.1638

Li, W., Qi, F., Tang, M., & Yu, Z. (2020). Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing*, *387*, 63–77. https://doi.org/10.1016/j.neucom.2020.01.006

Maslen, G. (2019). *Recruitment of top academic leaders is getting harder*. University World News. https://www.universityworldnews.com/post.php?story=20190709153145683

Melchers, K. G., Roulin, N., & Buehl, A. K. (2020). A review of applicant faking in selection interviews. *International Journal of Selection and Assessment*, *28*(2), 123–142. https://doi.org/10.1111/ijsa.12280

Menon, V. M., & Rahulnath, H. A. (2016). *A novel approach to evaluate and rank candidates in a recruitment process by estimating emotional intelligence through social media data*. 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), 1–6. https://doi.org/10.1109/ICNGIS.2016.7854061

Moon, B., Daljeet, K. N., Bourdage, J. S., & Roulin, N. (2024). Impression management profiles in job interviews: Relations with applicant individual differences and interview outcomes. *Journal of Business and Psychology*, *39*(4), 849–870. https://doi.org/10.1007/s10869-023-09918-3

Myslín, M., Zhu, S. H., Chapman, W., & Conway, M. (2013). Using twitter to examine smoking behavior and perceptions of emerging tobacco products. *Journal of Medical Internet Research*, *15*(8), e174. https://doi.org/10.2196/jmir.2534

Nacalaban, E. A., & Jala, J. V. M. (2024). Connecting the gap: Enhancing faculty performance through interpersonal skills. *International Journal of Social Science and Human Research*, *7*(8), 6720–6725. https://doi.org/10.47191/ijsshr/v7-i08-111

Nasution, M. K., Syah, R., & Elveny, M. (2023). What is data science. In A. Patel & N. C. Debnath (Eds.), *Data Science with Semantic Technologies: New Trends and Future Developments* (pp. 1–25). CRC Press. https://doi.org/10.1201/9781003310785

Nikolaou, I. (2021). What is the role of technology in recruitment and selection? *The Spanish Journal of Psychology*, *24*, e2. https://doi.org/10.1017/SJP.2021.6

Peck, J. A., & Levashina, J. (2017). Impression management and interview and job performance ratings: A meta-analysis of research design with tactics in mind. *Frontiers in Psychology*, *8*, 201. https://doi.org/10.3389/fpsyg.2017.00201

Perera, A., & Iqbal, K. (2021). Big data and emerging markets: Transforming economies through data-driven innovation and market dynamics. *Journal of Computational Social Dynamics*, *6*(3), 1–18. https://vectoral.org/index.php/JCSD/article/view/32

Quan-Haase, A., & Young, A. L. (2010). Uses and gratifications of social media: A comparison of Facebook and instant messaging. *Bulletin of Science, Technology & Society*, *30*(5), 350–361. https://doi.org/10.1177/0270467610380009

Roulin, N., Krings, F., & Binggeli, S. (2016). A dynamic model of applicant faking. *Organizational Psychology Review*, *6*(2), 145–170. https://doi.org/10.1177/2041386615580875

Sharma, S., & Sengupta, S. (2023). *AI-driven HR practices: Recent trends in digital HR tools and their perceived benefits*. The International Conference on Recent Innovations in Computing, (pp. 219–231). https://doi.org/10.1007/978-981-97-3442-9_15

Suen, H. Y., & Hung, K. E. (2024). Revealing the influence of AI and its interfaces on job candidates' honest and deceptive impression management in asynchronous video interviews. *Technological Forecasting and Social Change*, *198*, 123011. https://doi.org/10.1016/j.techfore.2023.123011

Tahir, F., & Khan, M. (2023). *Big Data: The Fuel for Machine Learning and AI Advancement* (no.10951)EasyChairhttps://easychair.org/publications/preprint/wLd4

Tsai, W. C., Chen, C. C., & Chiu, S. F. (2005). Exploring boundaries of the effects of applicant impression management tactics in job interviews. *Journal of Management*, *31*(1), 108–125. https://doi.org/10.1177/0149206304271384

Udjo, E. O., & Erasmus, B. (2014). Impact of retirement age policy on the workforce of a higher education institution in South Africa. *Politics & Policy*, *42*(5), 744–768. https://doi.org/10.1111/polp.12092

Uma, V. R., Velchamy, I., & Upadhyay, D. (2023). Recruitment analytics: Hiring in the era of artificial intelligence. In *The adoption and effect of artificial intelligence on human resources management, Part A* (pp. 155–174). Emerald Publishing Limited. https://doi.org/10.1108/978-1-80382-027-920231008

Vela, J. V. (2022). *Developing a predictive model using Twitter dataset for recruiting job-fit candidates in higher education institutions*. University of KwaZulu-Natal (Doctoral dissertation). https://doi.org/10.29086/10413/22277.)

Vianello, M., De Nicolo, F. M., Jonason, P. K., & Palmieri, A. (2024). Warning messages against faking in selection interviews: The role of the Dark Triad traits. In *SSRN*. https://doi.org/10.2139/ssrn.4703704

Volkova, S., Glenski, M., Ayton, E., Saldanha, E., Mendoza, J., Arendt, D., & Greaves, M. (2021). Machine intelligence to detect, characterise, and defend against influence operations in the information environment. *Journal of Information Warfare*, *20*(2), 42–66. https://www.jstor.org/stable/27036529

Wang, Y., Yang, G., Li, S., Li, Y., He, L., & Liu, D. (2023). Arrhythmia classification algorithm based on multi-head self-attention mechanism. *Biomedical Signal Processing and Control*, *79*, 104206. https://doi.org/10.1016/j.bspc.2022.104206

Wani, M. A., ELAffendi, M. A., Shakil, K. A., Shariq Imran, A., & Abd El-Latif, A. A. (2022). Depression screening in humans with AI and deep learning techniques. *IEEE Transactions on Computational Social Systems*, *10*(4), 2074–2089. https://doi.org/10.1109/TCSS.2022.3200213

Whiting, A., & Williams, D. (2013). Why people use social media: A uses and gratifications approach. *Qualitative Market Research*, *16*(4), 362–369. https://doi.org/10.1108/QMR-06-2013-0041

Yan, M., Xie, Y. P., Zhao, J., Zhang, Y. J., Bashir, M., & Liu, Y. (2020). How ingratiation links to counterproductive work behaviours: The roles of emotional exhaustion and power distance orientation. *Frontiers in Psychology*, *11*, 2238. https://doi.org/10.3389/fpsyg.2020.02238

Young, S., Dutta, D., & Dommety, G. (2009). Extrapolating psychological insights from Facebook profiles: A study of religion and relationship status. *CyberPsychology & Behavior*, *12*(3), 347–350. https://doi.org/10.1089/cpb.2008.0165

Zhao, J., Huang, F., Lv, J., Duan, Y., Qin, Z., Li, G., & Tian, G. (2020). *Do RNN and LSTM have long memory?* International Conference on Machine Learning, 11365–11375. https://proceedings.mlr.press/v119/zhao20c.html