BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

An Explainable AI Framework for Detecting Fraudulent Internship and Job Offers

[1]Priti Anil Zope, [2]Prof. Swarupa Kamble

[1]Student, TE-C (AI & DS), PRN: 72426510K, [2]Faculty Guide
[1]Dept. of Artificial Intelligence and Data Science
[1]Ajeenkya DY Patil School of Engineering, Pune, India

Abstract

Online portals have become commonplace for internships and job opportunities. With the increasing number of fraudulent internship and job offers, students find it difficult to differentiate between legitimate opportunities. Students face various risks such as fake recruiters, phishing, registration fee frauds, unrealistic offers, and other deceptions. This paper presents a study on an explainable artificial intelligence approach to detect fraudulent internship and job offers. It reviews existing artificial intelligence, machine learning, deep learning, natural language processing, and transformer based recruitment fraud detection algorithms and techniques and explainability methods such as SHAP and LIME for providing an interpretable explanation for a model's outcome. It highlights the limitations of previous approaches such as using generic datasets and lacking internship-specific features, limited explainability and no user-centric explanations.

Keywords: Artificial intelligence, cyber security, explainable AI, fraud detection, internship fraud, job fraud, LIME, machine learning, natural language processing, SHAP.

I. Introduction

The Despite the rapid growth of online recruitment platforms, enabling students and job seekers to search for online internships and job offers with much more convenience, the rise of such services has also led to a surge in fraudulent job ads and online recruitment scams. In this regard, students may encounter numerous risks and challenges while searching for a job, including scammers, phishing links, registration fees, unreal stipends, fake certificates, and dishonest job descriptions.

The use of machine learning approaches has been considered for detecting fraudulent job posts and various forms of recruitment fraud. Chairatanasopha et al. [1] investigated machine learning methods for detecting fraudulent types of jobs, whereas Akram et al. [2] explored the use of deep learning methods for combating Online Recruitment Fraud (ORF). It is possible to employ natural language processing (NLP) techniques to analyse the text of job descriptions and detect anomalies. Moreover, transformer-based language models such as BERT have been considered for capturing the context of job posting-related text for fraud detection purposes. For instance, Taneja et al. [3] proposed Fraud-BERT, which was designed to perform context-aware online recruitment fraud detection, whereas Jonathan et al. [4] evaluated BERT for fake job posting detection with data augmentation. Finally, Patil et al. [5] proposed a hybrid approach to fake job detection, which involved the combination of NLP, machine learning, and DistilBERT. Overall, the papers mentioned

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

above demonstrate different applications of artificial intelligence and deep learning to recruitment fraud detection.

However, the detection of fake job offers may not be sufficient for those who are interested in understanding why exactly a particular offer was identified as fake. Explainable Artificial Intelligence (XAI) can report such details by providing comprehensive explanations regarding the reasons behind a specific prediction. Using the methods such as SHAP and LIME can help identify the factors influencing a given outcome, thus making the results more transparent and reliable. Patil et al. [5] demonstrated the use of SHAP and LIME for explaining the prediction results of the fake job detector based on the DistilBERT model. Recently, the AI-based recruitment scam detectors [6] and multimodal fake job offer detectors [7] have been introduced. Therefore, this paper aims to present an explainable model for detecting fraudulent internship and job offers with a specific focus on internship-related indications, paying special attention to transparency and interpretability.

II. Problem Statement

The increasing number of false internship and job offers makes finding legitimate offers more challenging than before. The current set of detection techniques is often non-transparent, non-explanatory, and provides little or no reasoning for its decisions.

III. Objectives

The key objectives of this study include:

1) To provide an analysis of fraudulent internship and job offer frauds.

2) To showcase the current AI and XAI-based methods for detecting such scams

3) To demonstrate the explanation of the results using SHAP and LIME values

4) To compare the proposed approach to the existing solutions.

IV. Relevance to Current Technology Trends

The present article builds upon the following four trends:

- Generative artificial intelligence (AI): AI-generated job offers might be even more convincing and thus more deceptive.
- Explainable AI (XAI): XAI might provide more transparent and thus more accurate results.
- Online recruitment growth: Students might use Internet-based job offers more frequently, thus being more exposed to fraud.
- AI-driven fraud detection: More advanced fraudulent job offer detection methods can employ machine learning and natural language processing (NLP).

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

V. Literature Survey

Recent studies have investigated the use of artificial intelligence, machine learning, deep learning, NLP, and transformer-based techniques for detecting fraudulent job advertisements in online recruitment scams. Machine learning procedures can detect patterns in recruitment information connected with fraudulent job advertisements, while deep learning procedures can learn complex patterns from the same data [1], [2].

NLP Proves to be Particularly Effective in the Context of Analyzing the Language of Job Descriptions

The transformer-based design allows for the contextual analysis of the interdependencies between words and phrases that can be leveraged to detect deception (e.g., [3] and [4]). Patil et al. [5] proposed a hybrid solution that uses NLP, machine learning, and DistilBERT to detect fake job posts with the interpretation relying on the SHAP and LIME methods, whereas Raghav et al. [8] combined a BiLSTM model with LIME. The latest solutions include FakeJobFinder that implements an AI-powered and multimodal approach to extract and analyze various forms of information, including text, images, PDFs, and URLs. The tool was introduced by Banik et al. [6] who also suggested an AI-powered method to combat online job scams. In another study, Raam Vishaal et al. [7] introduced a similar solution that combines DistilBERT, optical character recognition, and website verification.

In addition to detection, XAI is critical to ensuring that the predictions are easy to understand by the users. SHAP and LIME can highlight the variables that contribute to a certain prediction and offer more transparent explanations. The reviewed articles can offer a basis for considering recruitment fraud detection, while the explainability, internship-related indicators, and the user-oriented explanations can offer further considerations. Fig. 1 below groups the eight reviewed articles according to the method, while Table 1 summarises the reviewed articles.

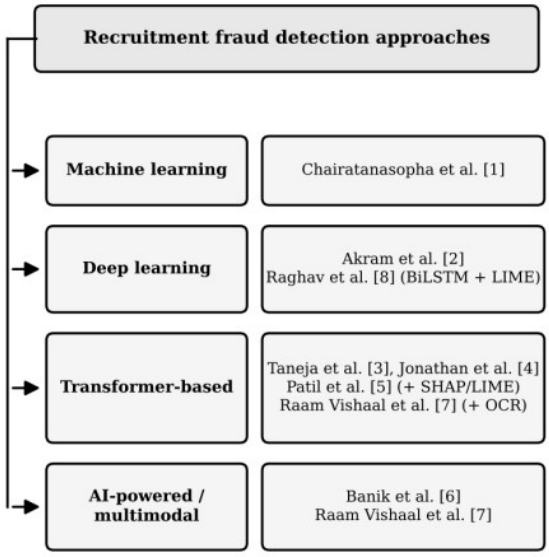


Fig. 1. Reviewed studies grouped by approach.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Table 1: Literature Survey Matrix

Study	Method	Key finding and limitation
Raghav et al. (2026) [8]	BiLSTM + LIME	Detects fraudulent recruitment postings and provides human-interpretable explanations using LIME. Limited use of transformer-based explainability and internship-specific indicators.
Raam Vishaal et al. (2026) [7]	DistilBERT + OCR + website verification	Uses text, images, PDFs, and URLs for better fraud detection. Does not use explainable AI and is not internship-specific.
Patil et al. (2025) [5]	DistilBERT + NLP + SHAP/LIME	Uses NLP and DistilBERT for fake job detection and applies SHAP and LIME to identify factors influencing predictions. Mainly focuses on fake job postings and does not specifically address internship-related fraud indicators.
Taneja et al. (2025) [3]	BERT / transformer	Uses a transformer-based approach for context-aware online recruitment fraud detection. Limited focus on explainability and internship-specific indicators.
Jonathan et al. (2025) [4]	BERT + data augmentation	Studies BERT-based fake job posting detection and the effect of data augmentation on model performance. Does not focus on explainable AI or internship-specific features.
Banik et al. (2025) [6]	AI + NLP	Presents an AI-powered approach for combating online recruitment scams. User-understandable explanations are not the main focus.
Akram et al. (2024) [2]	Deep learning	Explores deep learning approaches for detecting Online Recruitment Fraud. Does not provide clear explanations for the prediction results.
Chairatanasopha et al. (2023) [1]	Machine learning	Detects different types of fraudulent job advertisements using machine learning approaches. Limited focus on explainability and internship-specific fraud.

VI. Research Gaps

The literature survey leads to the identification of five research gaps:

1) The lack of job-focused research: The existing literature focuses on fraudulent job postings with limited consideration of internship-specific fraud.

2) Limited model explanation: Most of the detection models only provide predictions but do not explain the rationale behind the results. Techniques such as SHAP and LIME are primarily used for visualization and are not utilized to their maximum potential in providing justification for the final result.

3) Generic datasets: The existing approach uses generic datasets such as the EMSCAD / Fake Job Postings that may not be suitable for detecting internship-specific fraud.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

4) Incomplete set of indicators: The existing literature does not consider registration fees, fake recruiters, unrealistic stipends, phishing links, and fake certificates.

5) No simple explanation: The existing literature does not explain the result in a manner that can be understood by students and job seekers.

These research gaps indicate the need for a transparent AI framework for detecting internship and job fraud.

VII. Scope of the Study

This article focuses on the explainable artificial intelligence (AI) method for detecting fraudulent internship and job offers. The study covers internship and job offer scams, machine learning and deep learning solutions, and natural language processing (NLP) approaches for fraud detection. Besides, the study covers transformer-based language models like BERT, which were applied for fraud detection purposes [3], [4], [2]. Moreover, the article covers internship fraud-specific features like requirements for money payments, unrealistic stipend, and fake certificates. Finally, the study provides an overview of the explainable AI methods, such as SHAP and LIME, for generating explanations for detected frauds [5] and compares different fraud detection solutions. Fig. 2 below represents the general idea of the approach that has been chosen for this study.

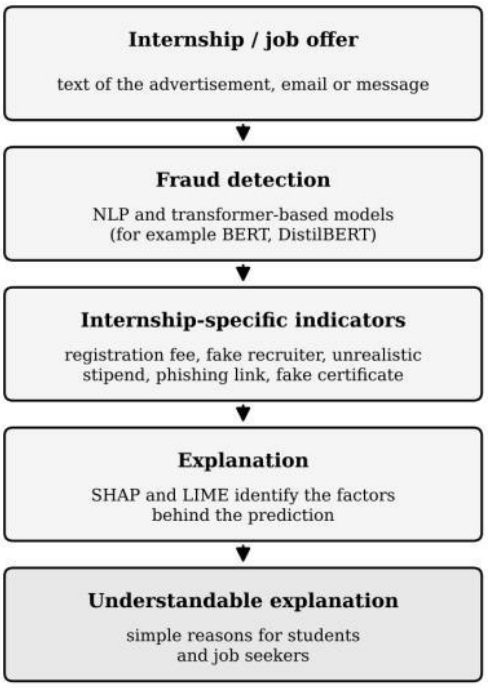


Fig. 2. Overview of the explainable detection approach studied in this paper.

VIII. Conclusion

This paper selected the research topic, defined the problem statement, studied the current state-of-the-art technologies for recruitment fraud detection automation with artificial intelligence (AI), and examined recent IEEE and Springer research papers. The research gap was defined, and the research objectives and

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

scope were finalised. Based on the examined research papers, this study concluded that despite the topic's popularity in scientific research, there is a need for further investigation into the explainability of fraud detection systems, the identification of internship-specific indicators, and user-friendly explanations of fraud detection results. This research aims to fill this gap by proposing a transparent framework for detecting internship and job fraud.

References

[1] P. Chairatanasopha et al., "A machine learning approach to detecting fraudulent job types," *AI & Society*, vol. 38, no. 4, pp. 1925–1941, 2023.
[2] N. Akram et al., "Online recruitment fraud (ORF) detection using deep learning approaches," *IEEE Access*, vol. 12, pp. 109388–109408, 2024.
[3] K. Taneja et al., "Fraud-BERT: Transformer based context aware online recruitment fraud detection," *Discover Computing*, vol. 28, 2025.
[4] J. Jonathan et al., "Optimizing BERT for fake job posting detection: A comparative study for data augmentation integrity and model performance," in *Proc. IEEE ICITCOM*, 2025, pp. 263–268.
[5] K. Patil, A. Shetty, A. Rajagopal, and S. Sonawani, "A hybrid approach to fake job detection using NLP and machine learning," in *Proc. 2025 IEEE 5th Int. Conf. ICT in Business Industry and Government (ICTBIG)*, 2025, doi: 10.1109/ICTBIG68706.2025.11323989.
[6] S. Banik et al., "IntegriHire: Combating online recruitment scams with Groq-powered intelligence," in *Proc. IEEE IEMECON*, 2025.
[7] K. S. Raam Vishaal et al., "AI-powered fake job detection using multimodal hybrid intelligence," in *Proc. IEEE IDCIoT*, 2026, pp. 964–970.
[8] B. K. Raghav, D. Paschal, G. S. Somashekhara, S. Patil, A. Polur, and Praveena, "XAI-Recruit: Explainable AI for job recruitment fraud detection using deep learning," in *Proc. 2026 Int. Conf. Smart Futuristic Technology (ICSFT)*, 2026, doi: 10.1109/ICSFT66733.2026.11507788.