International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Smart Innovations in Computer Science and Applications
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Security News Intelligence Agent Using Web Crawling and Large Language Models for Multi-Level Cybersecurity Awareness

Jay Sunil Dhanwalkar
Department of Science and Technology, G.H.Raisoni Skill Tech University, Nagpur, India

ABSTRACT

Cybersecurity risks like ransomware attacks, phishing campaigns, data breaches, and zero-day vulnerabilities have significantly increased as a result of the quick expansion of digital infrastructure and internet-based services. For people, organizations, and cybersecurity experts, timely awareness of such threats is essential. However, cybersecurity-related information is frequently presented in complicated technical language and is highly fragmented across multiple online platforms, making it difficult for novices to understand and time-consuming for experts to analyze. Current news aggregation systems lack intelligent analysis, adaptive explanation mechanisms, and career-oriented guidance, instead concentrating on content retrieval.

In order to improve cybersecurity awareness and knowledge sharing, this research paper suggests a Security News Intelligence Agent that combines automated web crawling and Large Language Models (LLMs). The suggested system retrieves news articles about cybersecurity. For people, organizations, and cybersecurity experts, timely awareness of such threats is essential. However, cybersecurity-related information is frequently presented in complicated technical language and is highly fragmented across multiple online platforms, making it difficult for novices to understand and time-consuming for experts to analyze. Current news aggregation systems lack intelligent analysis, adaptive explanation mechanisms, and career -oriented guidance, instead concentrating on content retrieval.

Keywords: Cybersecurity, Web Crawling, Large Language Models, AI Agents, Security News Analysis, Threat Intelligence, Job Recommendation, Artificial Intelligence.

1. Introduction

In recent years, the rapid expansion of digital technologies, cloud computing, and internet-based services has significantly increased the dependency of individuals, organizations, and governments on cyberspace. While this digital transformation has improved efficiency and connectivity, it has also exposed systems to a wide range of cybersecurity threats. Cyberattacks such as ransomware, phishing, malware infections, data breaches, denial-of-service attacks, and zero-day exploits are becoming more frequent and sophisticated. These attacks pose serious risks to sensitive data, financial assets, national security, and personal privacy.

Existing news aggregation systems primarily focus on collecting and displaying content based on keywords or categories. While these systems improve accessibility by centralizing information, they lack intelligent analysis and personalization capabilities. Most platforms present the same information to all users, regardless of their expertise level or interests. This one-size-fits-all

approach fails to address the diverse needs of users ranging from students and beginners to experienced cybersecurity professionals. Recent advancements in **Artificial Intelligence (AI)** and **Natural Language Processing (NLP)** have opened new possibilities for addressing these challenges. Large Language Models (LLMs) have demonstrated remarkable capabilities in understanding, summarizing, and generating human-readable text. These models can analyze large volumes of unstructured data and extract meaningful insights, making them suitable for cybersecurity news analysis. By leveraging LLMs, it is possible to transform raw and technical cybersecurity information into structured, understandable, and actionable knowledge.

This research addresses these challenges by proposing a **Security News Intelligence Agent** that combines automated web crawling with AI-driven content analysis. The proposed system retrieves cybersecurity news from trusted online sources and processes the content using Large Language Models. A key feature of the system is its ability to generate **multi-level explanations**, including executive summaries for decision-makers, technical explanations for professionals, and beginner-friendly explanations for non-technical users. This adaptive approach ensures that cybersecurity information is accessible and meaningful to a broad audience.

1.1. Motivation

The motivation of this research is therefore threefold:

- To reduce the complexity and fragmentation of cybersecurity information
- To enhance accessibility through adaptive, multi-level explanations
- To bridge the gap between cybersecurity knowledge and career guidance

1.2. Contribution

The key contributions of this research can be summarized as follows:

- Development of an automated cybersecurity news retrieval system using web crawling
- Integration of Large Language Models for intelligent content analysis
- Design of a multi-level explanation framework for diverse user expertise
- Implementation of a job recommendation mechanism based on cybersecurity trends and skills
- Creation of a scalable and extensible system architecture suitable for real-world deployment

2. Related work

Cybersecurity awareness, threat intelligence, and automated news analysis have been active areas of research in recent years due to the increasing number of cyber incidents worldwide. Researchers have explored various approaches to collect, analyze and disseminate cybersecurity-related information. This section reviews existing work related to cybersecurity news aggregation, threat intelligence platforms, text analysis using Natural Language Processing, and the application of Large Language Models in security-related domains.

2.1 Cybersecurity News Aggregation Systems

Traditional news aggregation systems primarily rely on keyword-based search and RSS feeds to collect articles from multiple sources. Platforms such as Google News and general-purpose news aggregators provide access to a wide range of cybersecurity-related articles. While these systems improve accessibility by centralizing content, they are not specifically designed for cybersecurity analysis

2.2 Threat Intelligence Platforms

Threat intelligence platforms (TIPs) are widely used by organizations to collect and analyze security-related data. These platforms integrate information from multiple sources, including

vulnerability databases, malware repositories, and security advisories. Research has shown that threat intelligence platforms can improve incident response and proactive defense strategies. However, most threat intelligence platforms are designed for enterprise environments and require specialized knowledge to operate

2.3 Large Language Models for Text Analysis

Large Language Models represent a significant advancement in text understanding and generation. Models such as GPT-based architectures and LLaMA-based models have demonstrated strong capabilities in summarization, question answering, and contextual reasoning.

2.4 Research Gap Identification

Based on the review of existing literature, several research gaps can be identified:

- Lack of intelligent cybersecurity news analysis systems that go beyond basic aggregation
- Absence of adaptive, multi-level explanations for users with different expertise levels
- Limited integration of Large Language Models in end-to-end cybersecurity awareness platforms
- Lack of systems combining cybersecurity news analysis with career guidance
- Insufficient focus on user-centric design and accessibility

3. Research Methodology

3.1. Problem statement

The rapid evolution of cyber threats and the increasing dependency on digital systems have made cybersecurity awareness a critical necessity. While a vast amount of cybersecurity-related information is available online, effectively accessing, understanding, and utilizing this information remains a significant challenge. The core problem addressed in this research is the inefficient accessibility and usability of cybersecurity news and threat-related information for users with diverse expertise levels.

3.1.1 Fragmentation of Cybersecurity Information

Cybersecurity news and threat intelligence are distributed across a wide range of platforms, including news websites, security blogs, vendor advisories, research forums, and social media channels. This fragmentation forces users to manually browse multiple sources to stay informed about emerging threats, vulnerabilities, and security incidents. Such manual tracking is time-consuming and inefficient, particularly for professionals who require timely insights to make critical security decisions

3.1.2 Complexity and Lack of Accessibility

Another major challenge is the technical complexity of cybersecurity content. Most cybersecurity news articles are written for experienced professionals and assume prior knowledge of security concepts, protocols, and attack mechanisms. Terms such as zero-day vulnerabilities, exploit chains, advanced persistent threats, and malware variants are frequently used without explanation. As a result, beginners, students, and non-technical users struggle to comprehend critical information.

3.1.3 Absence of Adaptive Explanation Mechanisms

Existing cybersecurity news platforms typically provide a single representation of information, regardless of the user's background or expertise. This one-dimensional approach fails to address the diverse needs of users. Beginners require simplified explanations to understand basic concepts, while professionals need detailed technical analysis to support decision-making.

3.1.4 Limited Integration of Artificial Intelligence

While some platforms employ basic Natural Language Processing techniques for summarization or classification, they lack advanced AI-driven interpretation capabilities. Traditional NLP models struggle with context understanding, evolving threat terminology, and nuanced security concepts. As a result, the generated summaries often fail to capture the significance and implications of cybersecurity incidents.

3.1.5 Formal Problem Statement

To design and implement an intelligent system that automatically retrieves cybersecurity news from multiple online sources, analyzes the content using advanced AI techniques, generates adaptive multi-level explanations tailored to different user expertise levels, and provides relevant cybersecurity job recommendations based on user skills.

4. Research Methodology

The proposed **Security News Intelligence Agent** is designed as an intelligent, modular, and scalable system that automates the retrieval, analysis, and presentation of cybersecurity news. The system integrates web crawling techniques with Artificial Intelligence–based content analysis to provide meaningful insights to users with varying levels of expertise. In addition to awareness generation, the system supports career guidance by recommending suitable cybersecurity job roles based on analyzed content and user skills.

4.1 High-Level System Description

At a high level, the system follows a **client–server architecture** consisting of four primary layers:

1. **User Interface Layer (Frontend)**
2. **Application Logic Layer (Backend)**
3. **Data Storage Layer (Database)**
4. **AI Analysis Layer (Large Language Models)**

4.2 User Interface Layer

The User Interface Layer serves as the interaction point between the system and the end user. It is implemented as a web-based interface that allows users to:

- Enter cybersecurity-related keywords (e.g., ransomware, phishing, zero-day)
- View retrieved cybersecurity news articles
- Select articles for detailed analysis
- Provide their skill set for job recommendation
- View AI-generated explanations and career suggestions

4.3 Application Logic Layer

The Application Logic Layer acts as the core processing unit of the system. It is responsible for handling user requests, coordinating data flow between modules, and invoking AI-based analysis services.

Request Handler Module: Receives user input from the frontend and validates it

- **Web Crawling Module:** Fetches cybersecurity news articles from online sources
- **Content Processing Module:** Cleans and preprocesses retrieved article content
- **Analysis Controller:** Coordinates AI-based content analysis and explanation generation

- **Job Recommendation Module:** Maps cybersecurity topics and skills to industry roles

4.4 Data Storage Layer

The Data Storage Layer is responsible for persisting cybersecurity news articles and related metadata. A cloud-based NoSQL database is used to store:

- Article titles
- Article URLs
- Cleaned article content
- Timestamps of retrieval

4.5 AI Analysis Layer

The AI Analysis Layer forms the intelligence core of the system. This layer leverages **Large Language Models (LLMs)** accessed via the Groq API to perform advanced text analysis tasks. The AI layer is responsible for:

- Understanding the context of cybersecurity news articles
- Extracting key insights from unstructured text
- Generating explanations at multiple levels of complexity
- Producing job recommendations based on analyzed topics and user skills

4.6 System Workflow

The overall workflow of the system can be summarized as follows:

1. The user enters a cybersecurity-related keyword through the frontend interface.
2. The backend receives the request and triggers the web crawling module.
3. Relevant news articles are fetched from online sources.
4. Retrieved content is cleaned and stored in the database.
5. The user selects an article and provides skill information.
6. The backend sends the article content to the AI analysis layer.
7. The AI generates executive, technical, and beginner-level explanations.
8. The job recommendation module suggests relevant cybersecurity roles.
9. The processed results are returned to the frontend for display.

5. Proposed System Architecture

The proposed **Security News Intelligence Agent** is designed using a modular and layered architecture that enables efficient data collection, intelligent analysis, and user-centric presentation of cybersecurity news. The architecture follows a **client–server model**, integrating web crawling mechanisms with Artificial Intelligence–based processing modules. This design ensures scalability, flexibility, and ease of maintenance while supporting future enhancements.

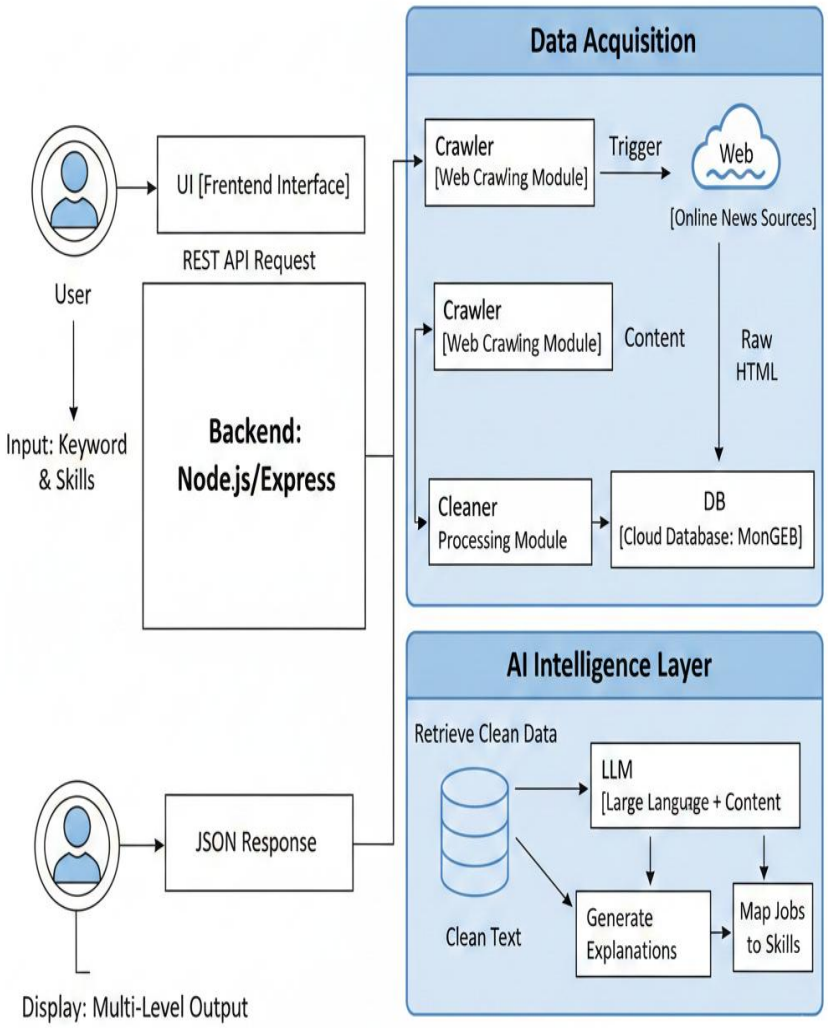


Fig. No. 1. Proposed System Architecture

5.1 Architectural Layers

The proposed architecture consists of the following major layers:

1. **Presentation Layer (Frontend)**
2. **Application Layer (Backend Services)**
3. **Data Layer (Database)**
4. **Intelligence Layer (AI and LLM-Based Analysis)**

5.2 Presentation Layer

The Presentation Layer represents the frontend component of the system and serves as the primary interaction point for users. This layer is responsible for collecting user inputs and displaying processed outputs generated by the backend and AI modules.

Key functionalities of the Presentation Layer include:

- Accepting cybersecurity-related keywords from users
- Displaying fetched news articles in a structured format
- Allowing users to select specific articles for analysis

- Collecting user skill information for job recommendations
- Presenting AI-generated explanations in multiple formats

5.3 Application Layer

The Application Layer forms the core of the system and manages communication between the frontend, database, and AI services. This layer is implemented using server-side technologies and exposes RESTful APIs to handle client requests. The Application Layer is composed of the following modules:

5.3.1 Request Management Module

This module handles incoming requests from the frontend. It validates user inputs, manages request routing, and ensures secure communication between system components.

5.3.2 Web Crawling Module

The Web Crawling Module is responsible for retrieving cybersecurity news articles from online sources. It uses keyword-based search mechanisms to identify relevant articles and fetch associated metadata such as title and URL.

5.3.3 Content Processing Module

Once articles are retrieved, the content processing module cleans and preprocesses the data. This includes removing advertisements, navigation elements, and irrelevant text. The cleaned content is prepared for storage and AI-based analysis.

5.3.4 Job Recommendation Module

This module analyzes cybersecurity topics and user-provided skills to recommend suitable roles in the cybersecurity industry. The recommendations are generated using AI-driven reasoning rather than static rule-based mapping.

5.4 Data Layer

The Data Layer provides persistent storage for cybersecurity news articles and associated metadata. A cloud-based NoSQL database is used to store:

- Article title
- Article URL
- Cleaned article content
- Retrieval timestamp

5.5 Intelligence Layer

The Intelligence Layer is the most critical component of the proposed architecture. It leverages **Large Language Models (LLMs)** to perform advanced analysis of cybersecurity content.

Key responsibilities of the Intelligence Layer include:

- Understanding the semantic context of cybersecurity articles
- Identifying key threats, vulnerabilities, and attack trends
- Generating explanations at different levels of complexity
- Producing job recommendations aligned with cybersecurity domains

5.6 Data Flow in the Proposed Architecture

The data flow within the proposed architecture follows a sequential and controlled process:

1. The user submits a keyword through the frontend interface.
2. The request is forwarded to the backend via a REST API.
3. The Web Crawling Module retrieves relevant cybersecurity news articles.
4. Retrieved content is cleaned and stored in the database.
5. The user selects an article and submits skill information.
6. The backend sends the article content to the AI Analysis Layer.
7. The AI module generates multi-level explanations and job recommendations.
8. The processed output is returned to the backend.
9. The frontend displays the results to the user.

This structured data flow ensures efficient communication and minimizes latency between components.

5.7 Architectural Advantages

The proposed architecture offers several advantages:

- **Modularity:** Individual modules can be modified or extended independently
- **Scalability:** Cloud-based components support increasing data volume and users
- **Flexibility:** AI prompts and models can be updated without redesigning the system
- **Maintainability:** Clear separation of concerns simplifies debugging and updates

5.8 Security and Reliability Considerations

Given the cybersecurity-focused nature of the system, security and reliability are important design considerations. Input validation, controlled API access, and error handling mechanisms are incorporated to ensure stable operation. Sensitive configuration details such as API keys are managed using environment variables to prevent unauthorized access.

6. Methodology

The methodology of the proposed **Security News Intelligence Agent** describes the systematic approach followed to design, develop, and evaluate the system. The methodology integrates automated data collection, data preprocessing, Artificial Intelligence–based analysis, and user-centric information presentation

6.1 Overview of the Methodological Framework

The methodological framework of the proposed system consists of the following major stages:

1. User Input Acquisition
2. Web Crawling and News Retrieval
3. Content Cleaning and Preprocessing
4. Data Storage and Management
5. AI-Based Content Analysis
6. Multi-Level Explanation Generation
7. Job Recommendation Generation
8. Result Presentation

6.2 User Input Acquisition

The methodology begins with user interaction through the web-based interface. Users provide cybersecurity-related keywords such as “ransomware,” “phishing,” or “zero-day vulnerability.” These keywords act as search parameters for retrieving relevant cybersecurity news articles.

6.3 Web Crawling and News Retrieval

Once the user submits a keyword, the backend initiates the web crawling process. The web crawler searches trusted online sources, including cybersecurity news platforms and aggregated feeds, to retrieve articles related to the specified keyword:

- Article title
- Article URL
- Source information

6.4 Content Cleaning and Preprocessing

Raw web content often contains irrelevant elements such as advertisements, navigation menus, headers, footers, and embedded scripts. These elements can negatively impact AI-based analysis if not removed.

The preprocessing steps include:

- Removal of HTML tags and scripts
- Elimination of advertisements and navigation text
- Extraction of the main article body
- Normalization of text format

6.5 Data Storage and Management

After preprocessing, the cleaned article content is stored in a cloud-based NoSQL database. The database stores both the article metadata and processed content, enabling persistent storage and future reuse.

The data storage stage ensures that:

- Articles can be analyzed multiple times without re-crawling
- Historical cybersecurity data is preserved
- System performance is improved by avoiding redundant processing

Using a cloud database enhances scalability and accessibility while simplifying deployment.

6.6 AI-Based Content Analysis

The core intelligence of the proposed system lies in the AI-based content analysis stage. This stage leverages **Large Language Models (LLMs)** to understand and interpret cybersecurity news articles.

The AI model processes the content to identify:

- Key cybersecurity threats and vulnerabilities
- Attack methods and exploitation techniques
- Impact and severity of reported incidents
- Recommended mitigation or defensive strategies

This semantic understanding enables the system to generate meaningful insights rather than simply summarizing text.

6.7 Multi-Level Explanation Generation

One of the key methodological contributions of this research is the generation of **multi-level explanations**. Based on the same analyzed content, the system generates three different formats of explanation.

Multi – Level Explanation Model

Explanation Level	Target Audience	Content Focus
Executive	Managers, Decision Makers	High-level summary, impact on business, significance of the event
Technical	Security Professionals	Attack vectors, specific vulnerabilities (CVEs), exploitation techniques, technical implications
Beginner	Students, Laypeople	Simplified terminology, analogies, non-technical definitions of threats

6.8 Result Presentation

The final stage of the methodology involves presenting the processed results to the user through the frontend interface. The system displays:

- Retrieved cybersecurity news articles
- AI-generated executive, technical, and beginner explanations
- Suggested cybersecurity job roles

7. Algorithm and Mathematical Model

This section presents the formal algorithmic design and mathematical representation of the proposed **Security News Intelligence Agent**. The objective of this section is to describe the internal working of the system in a structured and precise manner, enabling reproducibility and analytical understanding of the proposed approach.

7.1 Mathematical Model

Let the system be represented as a tuple:

S = {U, K, A, D, M, E, J}

where:

- U represents the set of users
- K represents the set of user-provided keywords
- A represents the set of retrieved news articles
- D represents the cleaned and processed article data
- M represents the Large Language Model
- E represents the set of generated explanations
- J represents the set of job recommendations

7.1.1 News Retrieval Function

Let the user input keyword be represented as:

$$k \in K$$

The news retrieval function fetches a set of relevant cybersecurity articles:

$$A = f(k)$$

where:

- f is the web crawling function
- $A = \{a1, a2, a3, ..., an\}$

Each article a_i consists of metadata and raw content.

7.1.2 Content Cleaning and Processing

Raw article content often contains irrelevant data. A preprocessing function is applied:

$$D = g(A)$$

where:

- g represents the content cleaning and preprocessing function
- $D = \{d1, d2, d3, ..., dn\}$
- Each d_i is a cleaned textual representation of article a_i.

7.1.3 AI-Based Content Analysis

The cleaned content is analyzed using a Large Language Model M:

$$A' = M(D)$$

where:

- A' represents the semantically analyzed article content
- M performs contextual understanding and information extraction

7.1.4 Multi-Level Explanation Generation

From the analyzed content A′, multiple explanation formats are generated:

$$E = \{Eexec, Etech, Ebeg\}$$

where:

$$Eexec = h1(A')$$

$$Etech = h2(A')$$

$$Ebeg = h3(A')$$

Here:

- $h1$ generates executive-level summaries
- $h2$ generates technical-level explanations
- $h3$ generates beginner-friendly explanations

7.1.5 Job Recommendation Model

Let the user skill set be represented as:

$Sk = \{s1, s2, s3, ..., sm\}$

The job recommendation function is defined as:

$J = r(A', Sk)$

where:

- r maps analyzed cybersecurity topics and user skills to relevant job roles
- J represents the recommended cybersecurity positions

7.2 Algorithmic Representation

The algorithmic steps of the proposed system are described below.

Algorithm 1: Security News Intelligence Agent

Input:

Cybersecurity keyword k, User skill set Sk

Output:

Multi - level explanations E, Job recommendations J

Steps:

1. Accept keyword k from the user
2. Retrieve cybersecurity news articles using web crawler
3. Store retrieved articles in database
4. Clean and preprocess article content
5. Select article for analysis
6. Send cleaned content to Large Language Model
7. Generate executive-level explanation
8. Generate technical-level explanation
9. Generate beginner-level explanation
10. Analyze user skills and article context
11. Generate cybersecurity job recommendations
12. Display results to the user

End Algorithm

7.3 Computational Flow Explanation

The algorithm follows a sequential execution flow, ensuring that each stage depends on the successful completion of the previous stage. This design minimizes redundancy and improves efficiency. The use of cloud-based AI services enables dynamic scaling based on input size and system demand.

7.4 Summary

This section formally defined the mathematical and algorithmic foundation of the proposed Security News Intelligence Agent. By representing system components using mathematical functions and structured algorithms, the design ensures clarity, reproducibility, and extensibility. The algorithm effectively integrates web crawling, AI-based analysis, and job recommendation into a unified framework.

8. Implementation Details

This section describes the practical implementation of the proposed **Security News Intelligence Agent**, including the technologies used, module-wise implementation, and integration of system components. The implementation follows the architectural design and methodology described in previous sections and demonstrates how theoretical concepts are realized in a working system.

8.1 Technology Stack

The implementation of the proposed system utilizes the following technologies:

Technology Stack Used

Component	Technology Used	Functionality
Frontend	Next.js (React)	User interface, collecting keywords/skills, displaying results.
Backend	Node.js with Express.js	API management, routing, request validation.
Database	MongoDB Atlas (Cloud)	Storing article metadata, URLs, and cleaned content.
AI Model	Groq API (LLM)	Semantic analysis, explanation generation, job mapping.
Crawler	Custom / RSS Feeds	Fetching news based on keywords from trusted sources.

8.2 Frontend Implementation

The frontend of the system is developed using **Next.js**, which provides a component-based architecture and efficient rendering capabilities. The frontend is responsible for user interaction and presentation of results.

8.2.1 User Input Interface

The user interface includes:

- A keyword input field for cybersecurity topics
- A list of retrieved cybersecurity news articles
- Article selection dropdown for AI analysis
- Skill input field for job recommendation

8.2.2 Result Display Module

AI-generated outputs are displayed in clearly separated sections:

- Executive Summary
- Technical Explanation
- Beginner Explanation
- Job Recommendations

Each section is visually distinguished to improve readability and comprehension.

8.3 Backend Implementation

The backend is implemented using **Node.js** and **Express.js**, providing RESTful APIs for communication between frontend, database, and AI services.

8.3.1 API Design

The backend exposes the following key APIs:

- **POST/api/news/fetch:** Retrieves and stores cybersecurity news
- **GET/api/news/articles:** Fetches stored articles from the database
- **POST/api/ai/analyze:** Performs AI-based analysis and job recommendation

These APIs ensure modular and controlled access to system functionality.

8.4 Web Crawling Module

The web crawling module automates the retrieval of cybersecurity news articles. It uses keyword-based queries to fetch articles from trusted sources. The crawler extracts:

- Article title
- Article URL

8.6 Database Implementation

The system uses **MongoDB Atlas**, a cloud-based NoSQL database, to store processed articles. The database schema includes:

- Article title
- Article URL
- Cleaned article content
- Timestamp

Using a cloud database eliminates the need for local installation and supports scalability and remote access.

8.7 AI Analysis Module Implementation

The AI analysis module integrates a **Large Language Model** through the Groq API. The backend sends cleaned article content to the AI service using structured prompts.

8.7.1 Prompt Engineering

Different prompts are designed to generate:

- **Executive-level** summaries
- **Technical-level** explanations
- **Beginner-level** explanations

Prompt engineering enables control over the complexity and format of AI-generated content.

8.8 Job Recommendation Module Implementation

The job recommendation module analyzes cybersecurity topics and user skills. The AI model maps these inputs to relevant cybersecurity roles such as:

- Security Analyst
- Penetration Tester
- SOC Analyst
- Incident Responder

8.9 Error Handling and Validation

Robust error handling mechanisms are implemented across the system:

- Input validation ensures valid user inputs
- API error handling prevents system crashes
- Logging mechanisms assist in debugging

These measures improve system reliability and stability.

8.10 Integration and Testing

All modules are integrated through RESTful APIs. The system is tested using real-time cybersecurity keywords and articles. Testing validates:

- Correct retrieval and storage of articles
- Accurate AI-based analysis
- Proper display of results on the frontend

9. Results and Discussion

This section presents the results obtained from implementing and testing the proposed **Security News Intelligence Agent** and discusses the observed outcomes.

10.1 Results

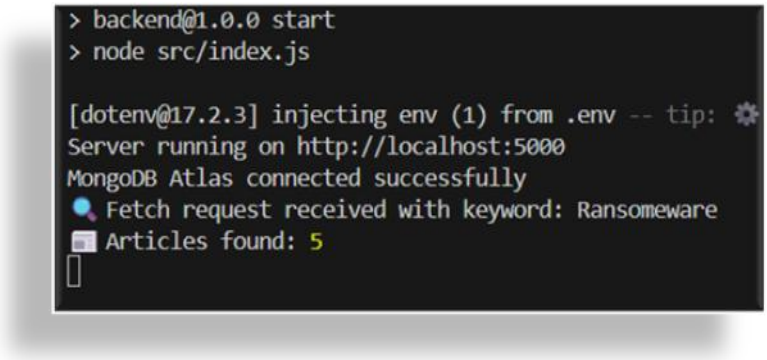
The system successfully retrieved cybersecurity-related news articles based on user-provided keywords. Articles were collected from online sources, cleaned, and stored in the cloud database without manual intervention. This demonstrates the effectiveness of the web crawling and content preprocessing modules.

Phase 1:

- Phase 1: Keyword input

Screenshot No. 1. Keyword Input

- List of fetched articles



Screenshot No. 2. List of Fetched Articles

- Dark cyber theme UI

Screenshot No. 3. Output Displayed On UI

The AI analysis module generated **three distinct levels of explanation** for selected cybersecurity articles:

- **Executive Summary:** Provided concise, high-level insights highlighting the overall impact.
- **Technical Explanation:** Delivered detailed analysis covering attack methods and vulnerabilities.
- **Beginner Explanation:** Simplified complex cybersecurity concepts into easy-to-understand language for non-technical users.

In addition, the job recommendation module successfully suggested relevant cybersecurity roles based on the analyzed content and user-provided skills. These recommendations aligned well with industry-standard cybersecurity job profiles.

10.2 Discussion

The results indicate that the proposed system effectively addresses the challenges identified in earlier sections. By automating news retrieval and integrating AI-based analysis, the system reduces manual effort and information overload. The multi-level explanation framework improves accessibility by adapting content to users with different expertise levels. The integration of job recommendations enhances the practical relevance of thesystem by connecting cybersecurity awareness with career guidance. This feature distinguishes the proposed system from traditional news aggregation platforms.

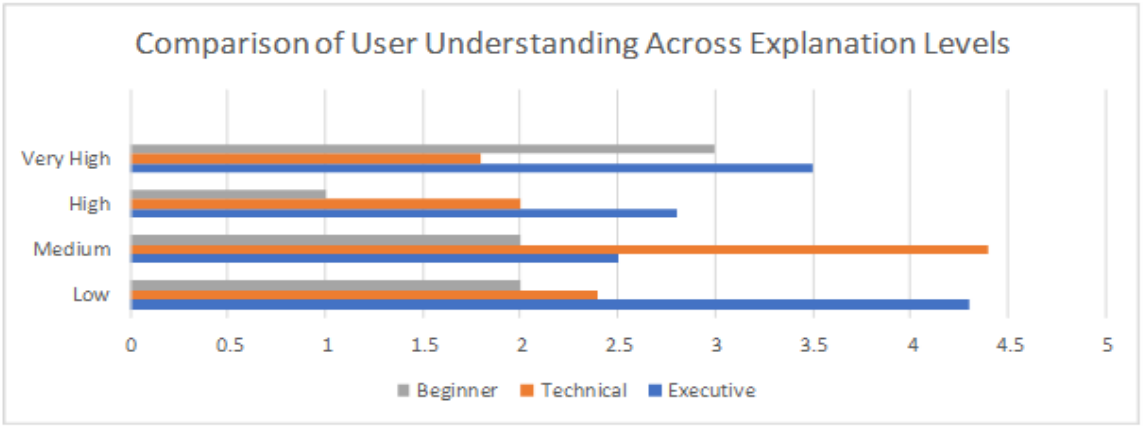


Fig. No. 2. Comparison of User Understanding Across Explanation Levels

10. Conclusion And Future Scope

11.1 Conclusion

This research presented a **Security News Intelligence Agent** that integrates automated web crawling and Artificial Intelligence to enhance cybersecurity awareness. The proposed system successfully retrieves cybersecurity news from online sources, processes the content, and generates adaptive explanations tailored to different user expertise levels. By providing executive, technical, and beginner-level explanations, the system improves accessibility and understanding of complex cybersecurity information.

11.2 Future Scope

Although the proposed system achieves its primary objectives, several enhancements can be considered for future development. The system can be extended by storing AI-generated insights for long-term analytics and trend analysis. Additional news sources and threat intelligence feeds can be integrated to improve coverage and accuracy.

Future work may also include user authentication, personalized dashboards, and improved job recommendation accuracy using skill assessment mechanisms. Deploying the system as a fully cloud-based service and integrating real-time alerting mechanisms can further enhance its applicability in practical cybersecurity environments.

11. References

1. *OWASP Foundation, "OWASP Top 10 – Web Application Security Risks," OWASP, 2023. [Online]. Available: https://owasp.org/www-project-top-ten/*
2. *NIST, "Framework for Improving Critical Infrastructure Cybersecurity," National Institute of Standards and Technology, Version 1.1, 2018.*
3. *S. M. Bellovin, "Security problems in the TCP/IP protocol suite," Computer Communication Review, vol. 19, no. 2, pp. 32–48, 1989.*
4. *MongoDB Inc., "MongoDB Atlas Documentation," MongoDB, 2024. [Online]. Available: https://www.mongodb.com/docs/atlas/*
5. *Groq Inc., "Groq Large Language Model API Documentation," Groq, 2024. [Online]. Available: https://console.groq.com/docs*
6. *Google, "Google News RSS Feeds," Google Developers, 2024. [Online]. Available: https://news.google.com/rss*
7. *DuckDuckGo, "DuckDuckGo Search API Documentation," DuckDuckGo, 2024. [Online]. Available: https://duckduckgo.com/api*
8. *J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," Communications of the ACM, vol. 51, no. 1, pp. 107–113, 2008.*
9. *T. Mikolov et al., "Efficient estimation of word representations in vector space," Proceedings of the International Conference on Learning Representations (ICLR), 2013.*
10. *T. Brown et al., "Language models are few-shot learners," Advances in Neural Information Processing Systems, vol. 33, pp. 1877–1901, 2020.*
11. *Meta AI, "LLaMA: Open and Efficient Foundation Language Models," Meta AI Research, 2023. [Online]. Available: https://ai.meta.com/llama/*
12. *R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," IEEE Symposium on Security and Privacy, pp. 305–316, 2010.*
13. *M. Bishop, Computer Security: Art and Science, Addison-Wesley, 2003.*

14. *J. Andress, The Basics of Information Security: Understanding the Fundamentals of InfoSec, 2nd ed., Elsevier, 2014.*

15. *A. McAfee and E. Brynjolfsson, "Big data: The management revolution," Harvard Business Review, vol. 90, no. 10, pp. 60–68, 2012.*

16. *Next.js Documentation, "Next.js App Router and Client Components," Vercel, 2024. [Online]. Available: https://nextjs.org/docs*

17. *Express.js, "Express – Node.js Web Application Framework," OpenJS Foundation, 2024. [Online]. Available: https://expressjs.com/*

18. *Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.*

19. *A. Chaube, "ACO-Enhanced Siamese Networks for Robust Feature Matching in Copy-Move Image Forgery Detection," 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), Nagpur, India, 2024, pp. 1-6, doi: 10.1109/ICAIQSA64000.2024.10882433.*

20. *Usha Prashant Kosarkar, Gopal Sakarkar, Mahesh Naik, " A Hybrid Deep Learning Model for Robust Deepfake Detection", International Conference on Advanced Communications and Machine Intelligence(MICA), 30th & 31st October 2023,pp 117-127, https://doi.org/10.1007/978-981-97-6222-4_9*