Implementation of the concept of a “swarm of virtual experts” in the formation of semantic networks in the field of cybersecurity based on large language models

Dmitry Lande

National Technical University of Ukraine – Igor Sikorsky Kyiv Polytechnic Institute

Leonard Strashnoy

The University of California, Los Angeles (UCLA)

Abstract

This paper introduces a method for constructing semantic networks in the field of cybersecurity using Large Language Models (LLM) based on the concept of a "swarm of virtual experts." The method involves multiple interactions with LLM through various prompts and roles, where each prompt is treated as a separate "virtual expert." This approach minimizes errors such as hallucinations and redundant links by aggregating results from different models and prompts. Additionally, a validation step involving human experts is introduced to enhance the reliability of the data obtained. The proposed approach demonstrates significant potential in the automated processing of complex data and can be applied to create semantic networks in cybersecurity tasks, data analysis, and information retrieval.

Keywords: Swarm of Virtual Experts, Large Language Models (LLM), Cybersecurity, Semantic Networks, Automated Analysis, LLM Hallucinations, Data Validation.

Introduction

The modern field of cybersecurity is faced with a rapid increase in the volume and complexity of cyber threats, which poses challenges for specialists related to the analysis and processing of huge amounts of data. Attacks are becoming more sophisticated and widespread, including the use of advanced social engineering techniques, phishing and complex multi-layered cyberattacks. Analysis of such threats requires high computational resources and expert knowledge to quickly identify attack patterns and vulnerabilities in systems.

However, even the most experienced professionals face limitations in manual data processing. In this regard, artificial intelligence (AI) and, in particular, large language models (LLMs) such as GPT are growing in importance. These models are capable of efficiently processing and analyzing large volumes of unstructured data, identifying patterns and connections that might go unnoticed with traditional analysis methods. Using LLM allows you to speed up threat detection and automate many processes related to cybersecurity, which is especially important in conditions where every second of delay can lead to significant consequences for the organization.

In the era of rapid development of AI technologies, LLMs have become a tool for extracting entities (concepts) and relationships between them from texts. This process underlies the creation of semantic networks, which can be used in a variety of fields, from big data analysis to knowledge system building. However, several significant problems arise when applying these models, such as hallucinations, redundant or irrelevant connections, and omission of important concepts. This is especially critical in security-related applications, where a missed threat can have serious consequences.

To partially overcome these problems, an approach has been proposed that is based on multiple LLM calls with different prompts, roles and models, as well as on aggregation of results, i.e. The proposed approach is based on repeated interactions with LLMs through different requests, with different roles and on different models. This allows you to increase the likelihood of finding relevant concepts and connections, but the question arises of the optimal number of queries and roles, as well as the need for human verification of the results. This method is called a "swarm of virtual experts", where each launch of a prompt represents an "expert" participating in the process of building a semantic network. This approach allows for the collection and aggregation of various responses, which reduces the likelihood of errors and increases the accuracy of the analysis. This article discusses the steps in this process, its possible limitations, and suggests ways for improvement.

An additional innovation proposed in this paper is the introduction of a validation step involving human experts. This helps ensure that the resulting semantic networks are not only statistically valid but also practically meaningful. This hybrid approach—a combination of machine and human experts—can significantly improve the accuracy and reliability of the results.

"Swarms of virtual experts" in the field of cybersecurity can be used to analyze texts, reports, and data on incidents of cyber threats and attacks, for example, phishing attacks, social engineering and other types of cyber attacks when solving the following problems:

1. **Threat identification**: The role of LLM in extracting concepts and identifying relationships between cyber attacks, defense techniques, and types of vulnerabilities.
2. **Risk assessments**: LLMs acting as virtual experts help assess the risks of various attacks based on historical data and expert knowledge.
3. **Implementation of the protection strategy**: Discussion of how virtual expert roles can model possible defense strategies based on cyberattack data.

Literature review

In recent years, the cybersecurity field has seen rapid growth in the use of artificial intelligence and machine learning models for threat detection and data analysis. The introduction of large language models (LLMs) such as GPT has revolutionized natural language processing tasks, including extracting concepts and relationships from unstructured data. Several studies have demonstrated the potential of LLM in detecting phishing attacks, analyzing social engineering tactics, and identifying cyberattack patterns [1, 2].

However, a significant problem that remains is LLM's tendency to "hallucinate," where the model generates irrelevant concepts or fails to detect key relationships between important cybersecurity entities [3].

The virtual expert swarm method aims to solve these problems by using multiple iterations of queries against one or more LLMs to extract concepts and relationships. There are a number of studies devoted to the extraction of concepts and relationships using LLM and other artificial intelligence methods. For example, in [4] methods for automatically constructing semantic networks from text data using LLM were studied.

Recent studies show that similar methods can be used to improve accuracy in other areas such as medical diagnostics and financial analysis [5-7].

The technique of using role models to increase the diversity of concepts is also an important topic. Research [8] proposes techniques similar to the "swarm of experts", but without the use of repeated execution of prompts with different roles. Also, an important area is the inclusion of the human factor at the verification stage.

The approach based on the use of a "swarm of virtual experts" echoes the concepts of a swarm of intelligent agents, which are discussed in [9].

While previous work has focused on creating specialized agents for specific tasks, the swarm approach offers extensive capabilities for flexible adaptation to complex scenarios.

The issues of generating correct and relevant entities when prompts are repeatedly executed are also raised in a number of articles, such as [10], where the impact of cognitive biases and prompt reuse on the quality of extracted data is studied.

The concept of a "swarm of virtual experts"

A “swarm of virtual experts” refers to a set of queries (prompts) executed in LLM. Each such request (or series of requests) is considered to be the opinion of one “expert”.

In the context of the term “swarm of virtual experts,” an important aspect is the consistency of their actions. Although each "expert" represents a separate run of the prompt, there are several mechanisms that ensure consistency of results.

1. **Human factor (launching prompts)**. Consistency is largely ensured by the person formulating and setting the prompts. A person, as a system operator, performs several key functions:
   - Formulates the same or similar queries for different launches, which creates the necessary unity of the context.
   - Monitors results, discarding irrelevant connections and clarifying further prompts based on previous responses.
   - Ultimately, it aggregates the results taking into account the frequency of occurrence of connections, thereby maintaining the coherence of the final semantic network.
2. **Internal mechanisms of LLM (memory and learning)**.
   - **Memory of previous prompts**. Many LLMs, such as GPT, do not have "long-term" memory between individual query sessions, meaning they do not learn in real time. However, during a single session (within a single conversation or a series of related requests), the model may retain the context of previous prompts. This allows the LLM to take previously provided information into account, providing some coherence in responses. If the prompts are formulated sequentially, the model can build connections based on the concepts already provided.
   - **Model training**. Although in most scenarios LLMs do not learn on the fly from the provided data, the internal response generation mechanisms are still based on huge amounts of pre-trained information. This means that the connections between concepts are not random: they are built on previously learned patterns. Thus, each prompt run by an “expert” is based on an already agreed upon knowledge base, which contributes to the consistency of results.

These two aspects—the formulation of prompts by humans and the maintenance of internal coherence of the model—ensure the coordinated functioning of the “swarm of virtual experts.” Human factors direct and aggregate efforts, and LLM, although operating independently in each run, generates responses based on a common pre-trained context.

Methodology

In order to increase the variety and accuracy of the extracted connections, several techniques have been proposed:

1. Extracting entities and relationships using LLM. The basis of the method is the consistent extraction of concepts and relationships between them from the text. This involves using the same prompt, which is reformulated and run multiple times to minimize the impact of repeated responses. Repeated execution of the prompt helps to “reboot” the system, which increases the likelihood of finding more relevant connections. Frequently rebooting the system allows you to generate more varied

results, avoiding repetition and increasing the likelihood of extracting unique and relevant relationships.

2. Using roles. To diversify responses and minimize errors, a role changing technique was added. Each query to an LLM is formulated with a role in mind, such as that of an expert in a particular field, which facilitates the extraction of new, contextually relevant connections.
3. Application of various LLM models. The variety of models allows you to reduce the influence of the features of a particular model on the result and increase the reliability of the extracted data. Each model generates its own unique entities and relationships, which increases the completeness of the resulting semantic networks.
4. Aggregation of results. After collecting data from different models and in different roles, it is aggregated. Links are assigned weights corresponding to the frequency of their occurrence in different networks. Low frequency associations can be eliminated to eliminate spurious or insignificant data.
5. Human verification. At the final stage, it is proposed to include a human expert to verify the networks. This step is necessary to minimize errors made by LLM, especially in cases where important relationships are omitted and non-existent ones are added.

An important part of the "swarm of virtual experts" methodology is the definition of roles from the point of view of which entities and relationships are extracted. These roles can be chosen not only by the individual but also by the LLM themselves. For example, if you are studying the area of cyberattacks such as phishing or social engineering, you can set a separate prompt for LLM asking for a list of roles from which the analysis will be performed. In response, the LLM may offer roles such as: “security engineer”, “system administrator”, “company director”, “accountant” and others. Each role assumes a specific view of the problem, thanks to which it is possible to highlight different aspects of the entities and relationships being studied. This allows us to create a more complete and multifaceted semantic network. Repeated use of such roles helps to minimize possible cognitive biases that could arise when working with one model or one expert.

One of the key issues is determining the optimal number of virtual expert roles. An excessive number of roles can lead to a decrease in network quality due to information noise, while a lack of roles will not allow identifying all possible connections. For this purpose, the approach of statistical analysis of the increase in new connections can be used. In the proposed “virtual expert swarm” model, estimating the number of roles and their human confirmation are important elements to improve the accuracy and reliability of semantic networks generated by LLM. Let's consider these two aspects:

1. **Estimation of the number of virtual expert roles**:

   Expert roles in the model serve to increase the diversity of responses. Each role provides a different context, or perspective, from which the LLM can view concepts and their relationships. This helps minimize the impact of potential errors or omissions that may occur with requests from one role.

   It is important to find the optimal number of roles. Too many roles can generate redundant, irrelevant connections, while too few roles will not produce enough diversity in responses. The optimum can be sought through statistical analysis: after several iterations of prompts with different roles, one can observe at what number of roles the most stable and relevant network of connections appears.

   The role counting method may be based on the concept of convergence. If, when adding a new role, the frequency of appearance of new concepts and relationships begins to decrease, this may signal that most of the relevant concepts have already been taken into account, and further addition of roles is ineffective. This can be formalized by observing the increase in new concepts with each additional role and establishing a threshold at which the addition of new roles becomes statistically insignificant.

2. **Human confirmation of results**:

   The human role in the model is the final test of the aggregated semantic network. After the LLM processes many prompts from different roles, a human expert performs verification by analyzing the relationships between concepts, their weight and relevance.

   A human expert helps eliminate:

   - Hallucinations LLM (non-existent connections);
   - Unimportant or insignificant connections (erroneous conclusions based on irrelevant data);
   - Omissions of important connections that the model may have missed.

For the convenience of validating results, interactive interfaces can be introduced that will allow a person to observe the weights of connections (obtained based on the frequency of their identification by virtual experts) and adjust them if necessary.

Thus, the system gains an additional degree of reliability through the inclusion of a human expert who confirms or corrects the results provided by virtual experts.

Mathematical formalization

Let's denote:

- $S = \{s_1, s_2, \ldots, s_n\}$ — a set of entities extracted from the text;
- $R = \{r_{ij}\}$ — a set of connections between entities, where r_{ij} denotes a relationship between entities s_i and s_j;
- P — a set of prompts for extracting connections;
- $M = \{M_1, M_2, \ldots, M_k\}$ — a set of different LLM models;
- $W(r_{ij})$ — the weight of the relationship between entities s_i and s_j, which depends on the frequency of repetition of this connection.

For each prompt $p \in P$ and each model $M_i \in M$:

- Extracting multiple entities S_p^l and many connections R_p^l.
- Link weight is updated: $W(r_{ij}) = W(r_{ij}) + 1$, If $r_{ij} \in R_p^l$.

Introduction of a threshold θ. Connection r_{ij} is included in the final network if its weight $W\left(r_{ij}\right) \geq \theta$, Where θ is defined as a statistical threshold based on the number of runs of virtual experts and the frequency of connections.

To determine the optimal threshold θ It is suggested to use the following methods:

- Let $f\left(r_{ij}\right)$ — frequency of occurrence of communication r_{ij} in the results of different prompts and models.
- Let TTT be the total number of calls (prompts and models).
- Let's determine the frequency distribution of connections $P(f)$, which shows the probability of connections occurring with different frequencies.

In order to choose the optimal threshold θ, you can use a significance criterion based on the number of starts and the dispersion of link frequencies:

$$\theta = \mu f + \alpha \sigma_f,$$

Where:

- μf — average value of the frequency of occurrence of connections;
- σ_f — standard deviation of frequency;
- α — significance coefficient (for example, $\alpha = 1.5$ to weed out less significant connections).

This threshold allows you to filter out connections that are random and do not have significant weight.

The number of virtual expert calls N directly affects the reliability of the network construction. To estimate the optimal number of launches, you can use the method of stabilizing communication frequencies. Let:

- N — Number of starts;
- $f\left(r_{ij}\right)$ — communication frequency r_{ij} after N launches.

If when increasing N increase in the number of new connections $\Delta f\left(r_{ij}\right)$ becomes insignificant (for example, $\Delta f\left(r_{ij}\right) < \varepsilon$, Where ε - small number), this indicates the achievement of stability. Thus, the optimal N can be determined based on an analysis of the growth of connections:

$$N = \min\left\{N' : \frac{1}{|R|}\sum_{r_{ij}\in R}\Delta f\left(r_{ij}\right) < \varepsilon\right\}.$$

This criterion allows you to choose N, in which the model stabilizes and new experts do not add significant connections.

Here is a formalization for estimating the number of roles and confirmation:

1. Let R — the number of roles from which data is requested from the LLM. For every role i, we get a lot of concepts C_i and many connections between them S_i.
2. Aggregated set of connections: $S = \bigcup_{i=1}^{R} S_i$, assigning a weight to each link $w(s)$, which is equal to the frequency of occurrence of a connection among roles:

   $w(s) = ($ number of roles involved s meets $w) / R$

3. Enter a threshold value τ, which determines which connections are considered significant. For example, if a relationship occurs in less than τ share of roles, it is discarded:

   $$S^* = \left\{s \in S \mid w(s) \geq \tau\right\}.$$

   To estimate the required number of roles R^* you can observe the growth of new connections with the addition of new roles. If the increase in the number of new concepts and connections becomes negligible (for example, $\frac{\Delta S_i}{\Delta R} \to 0$), you can stop the increase in the number of roles.

4. After building the network S^*, a human expert carries out the final verification, confirming or rejecting the proposed connections, which ensures the final adjustment of the network.

To determine the optimal threshold for the frequency of connections τ, you can use the following method:

- A series of experiments are carried out with different values τ, starting from minimum values (taking into account all connections) to increased thresholds.
- The connection with weight 1 may be the most frequent when the number of experts is small. To prevent this in a sample of experiments with different numbers of roles and different τ An analysis is carried out that allows us to estimate at what threshold value the increase in new connections becomes negligible.
- This value will be considered optimal.

Let's take an example from the field of cybersecurity. Let's say we need to build a cause-and-effect network corresponding to the concept of “phishing attack”. To do this, we will first build such a network without using a “swarm of virtual experts”, and then apply the proposed concept, based on the role-based approach and repeated setting of prompts.

In both cases, we will use hierarchical clarification of concepts, leading to a phishing attack. As a basic request, consider the prompt:

> List the main reasons (indicated by short keywords) that lead to a phishing attack. Present the results as a list in the format "reason; phishing attack", for example, "human error; phishing attack"

LLM system response:

Human error; phishing attack

Weak passwords; phishing attack

Lack of awareness; phishing attack

Unprotected email systems; phishing attack

No multi-factor authentication; phishing attack

Outdated software; phishing attack

Social engineering; phishing attack

Poor security policies; phishing attack

Suspicious links; phishing attack

Insufficient spam filters; phishing attack

At the second level, similar prompts are formed, but for concepts already obtained using LLM, for example:

> Identify the main reasons (indicated by short keywords) that lead to Human Errors leading to a phishing attack. Present the results as a list in the format "reason; Human error"

The table shows nodes with the highest network degree, obtained by simple hierarchical query input

To implement a small swarm of virtual experts, first, using LLM, we define the main roles of experts, for which we enter the prompt:

> Specify the roles of experts involved in analyzing a phishing attack, such as "organization manager" or "system administrator"

We get the following set of roles:

- Organization Manager
- System Administrator
- Security Analyst
- Incident Response Team Member
- Network Administrator
- IT Support Specialist
- Forensic Analyst
- Compliance Officer
- User Awareness Trainer
- Legal Advisor
- Threat Intelligence Analyst

- Data Protection Officer
- Software Developer
- Chief Information Security Officer (CISO)

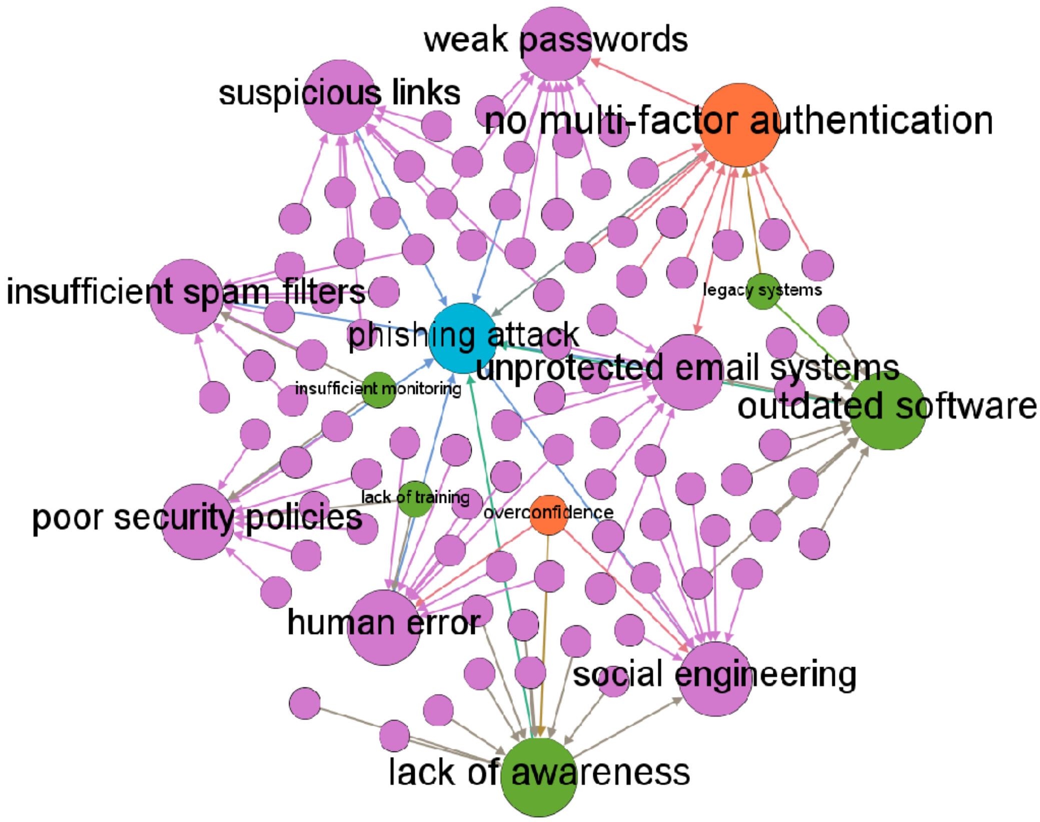

Figure 1 – Network obtained by simple hierarchical query input

From this, using an expert solution, we leave only three roles:

- System Administrator
- Security Analyst
- Incident Response Team Member

On behalf of these roles, enter each of the above requests twice.

Example request:

List the main reasons (indicated by short keywords) that lead to a Phishing Attack from the perspective of a System Administrator. Present the results as a list in the format 'reason; Phishing Attack', for example, 'human error; phishing attack'

In the process of aggregating results, we do not consider connections whose frequency in answers is less than 2. In addition, we involve a human expert in obtaining the final result. In Fig. Figure 2 shows a network obtained using a swarm of virtual experts

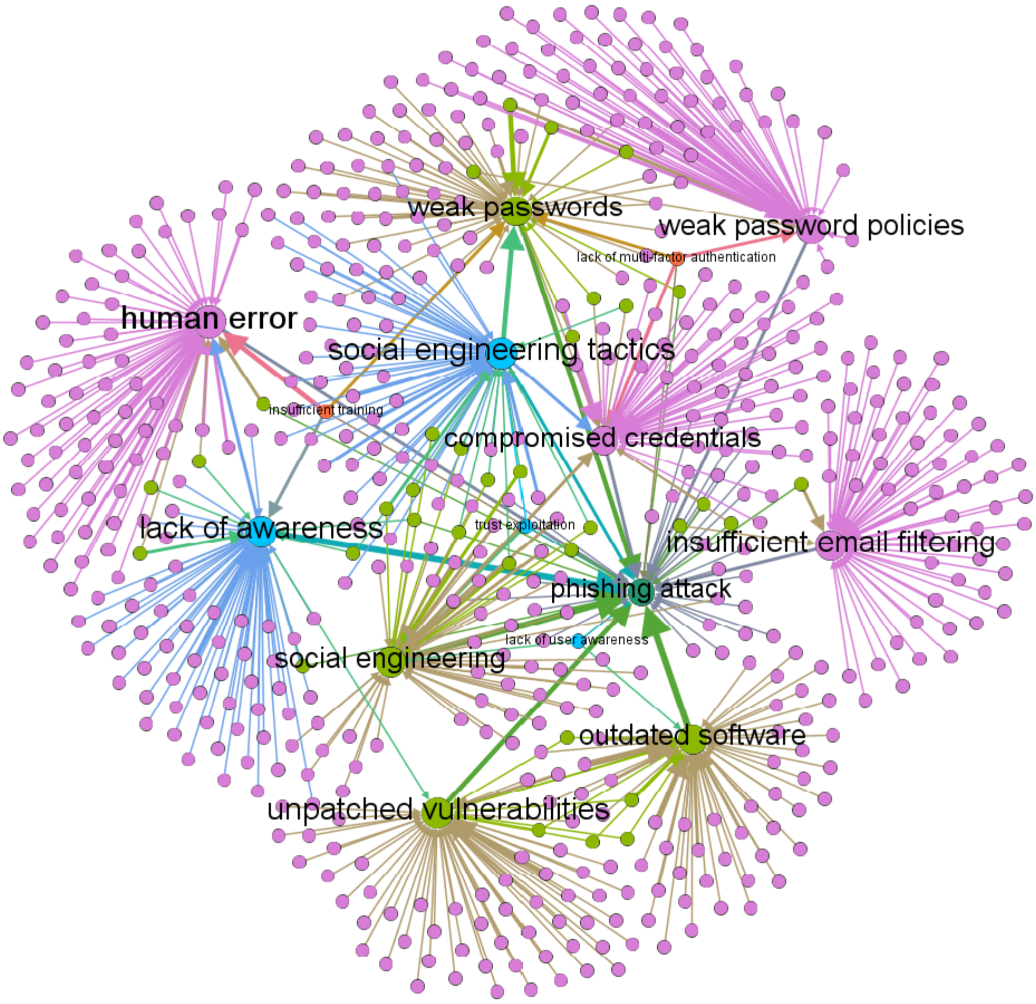

Figure 2 – Network obtained using a swarm of virtual experts

Comparing networks obtained without the participation of a swarm of virtual experts and without its participation using ChatGPT. We will also perform this assessment with the involvement of real and virtual experts. To attract a virtual expert, we will write a prompt that we will set to another LLM - DeepSeek:

> I am interested in the questions of statistical reliability and robustness of the networks. Which one is more reliable and robust for decision-making?

Next, a set of network connections is introduced, obtained by simple hierarchical input of queries (Network 1) and a network obtained using virtual expert swarm technology (Network 2).

After analyzing the structure and content of networks, LLM – DeepSeek gives the answer:

> **Conclusion:**
>
> **Network 2** appears to be more reliable and robust for decision-making as it is more focused on specific aspects such as email security and multi-factor authentication. This reduces the risk of redundancy and repetition in links, increasing the statistical reliability and robustness of the data.
>
> If you need a more detailed and focused network, **Network 2** will be the better choice.

We reach a conclusion that, after reviewing the data, a human expert can agree with.

Conclusions

The proposed "swarm of virtual experts" method demonstrates a new approach to building semantic networks using LLM, which combines multiple queries, roles, different models and aggregation of results, partially solving the problems of hallucinations and redundant connections through aggregation of results. The introduction of a statistical threshold and estimation of the required number of runs of virtual experts increases the accuracy and reliability of the network. Multiple queries with different roles increase the likelihood of finding relevant concepts and their relationships. The importance of the role concept is also reflected in the ability to improve the accuracy and diversity of the network. Statistical analysis of the increase in new connections helps determine the optimal number of roles, which reduces the likelihood of information noise. It is important to note that the consistency of the swarm's actions is achieved not only through multiple launches of prompts but also through human participation, as well as internal LLM mechanisms that maintain the context of requests within a single session.

However, the key role is played by a human expert who performs the final verification of the aggregated data. The quality of the method is assessed through the involvement of a human expert who acts as a validator of the solutions proposed by the system. This allows you to maintain a balance between an automated and human approach, where the LLM does the main work of creating the network, and a human expert confirms its correctness and helps with optimization. Determining the optimal number of roles and thresholds for connections remains an important direction for future research.

The proposed approach bridges the gap between automated AI-based analysis and human expertise, providing a robust solution for processing complex cybersecurity data. Future research could include expanding the range of roles used in the virtual expert generation, testing the approach on different cybersecurity reporting datasets, and improving the human validation process by integrating active learning methods.

Further research can be aimed at optimizing the method and developing more accurate criteria for aggregation and automation of verification procedures using LLM. Based on the presented results, this method can be useful for building semantic networks in problems of information retrieval, big data analysis and creating cognitive models. The proposed approach can be applied to problems of big data processing, knowledge analytics and the construction of expert systems.

Literature

1. Gallagher, Sean, Ben Gelman, Salma Taoufiq, Tamás Vörös, Younghoo Lee, Adarsh Kyadige, and Sean Bergeron. Phishing and Social Engineering in the Age of LLMs. In *Large Language Models in Cybersecurity: Threats, Exposure and Mitigation*, pp. 81-86. Cham: Springer Nature Switzerland, 2024. DOI: 10.1007/978-3-031-54827-7_8
2. Lande, Dmitry; Strashnoy, Leonard. Hierarchical Formation of Causal Networks Based on ChatGPT. SSRN preprint (May 8, 2023). - 13 p. DOI: 10.2139/ssrn.4440629
3. Gabrijela Perković; Antun Drobnjak; Ivica Botički. Hallucinations in LLMs: Understanding and Addressing Challenges. In 2024 47th MIPRO ICT and Electronics Convention (MIPRO) DOI: 10.1109/MIPRO60963.2024.10569238
4. Dmytro Lande, Leonard Strashnoy. GPT Semantic Networking: A Dream of the Semantic Web - The Time is Now. - Kyiv: Engineering, 2023. - 168 p. ISBN 978-966-2344-94-3
5. Akshay Goel, Almog Gueta, Omry Gilon, et al. LLMs Accelerate Annotation for Medical Information Extraction. Proceedings of the 3rd Machine Learning for Health Symposium, PMLR 225:82-100, 2023. URL: http://dblp.uni-trier.de/db/conf/ml4h/ml4h2023.html#GoelGGLENHJRKSL23
6. Rian Dolphin, Joe Dursun, Jonathan Chow, Jarrett Blankenship, Katie Adams, Quinton Pike. Extracting Structured Insights from Financial News: An Augmented LLM Driven Approach (2024). Preprint arXiv, arXiv:2407.15788. DOI: 10.48550/arXiv.2407.15788

7. Wiest, Isabella C., et al. "LLM-AIx: An open source pipeline for Information Extraction from unstructured medical text based on privacy pre-serving Large Language Models. medRxiv (2024): 2024-09. DOI: 10.1101/2024.09.02.24312917
8. Kommineni, Vamsi Krishna, Birgitta König-Ries, and Sheeba Samuel. From human experts to machines: An LLM supported approach to ontology and knowledge graph construction. arXiv preprint arXiv:2403.08345 (2024).DOI: 10.48550/arXiv.2403.08345
9. Steenstra, I., Nouraei, F., Arjmand, M., & Bickmore, T. W. (2024). Virtual agents for alcohol use counseling: exploring LLM-powered motivational interviewing. arXiv preprint arXiv:2407.08095. DOI: 10.48550/arXiv.2407.08095
10. Lim, S., Kim, Y., Choi, C. H., Sohn, J. Y., & Kim, B. H. (2024). ERD: A Framework for Improving LLM Reasoning for Cognitive Distortion Classification. arXiv preprint arXiv:2403.14255. DOI:10.48550/arXiv.2403.14255