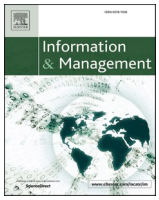
Information & Management 61 (2024) 104010

Contents lists available at ScienceDirect

Information & Management

journal homepage: www.elsevier.com/locate/im



Knowledge-prompted ChatGPT: Enhancing drug trafficking detection on social media

Chuanbo Hu [a], Bin Liu [b], Xin Li [a,*], Yanfang Ye [c], Minglei Yin [d]

[a] Department of Computer Science, University at Albany, 1400 Washington Avenue, Albany, 12222, New York, United States
[b] Department of Management Information Systems, West Virginia University, 83 Beechurst Avenue, Morgantown, 26505, West Virginia, United States
[c] Department of Computer Science and Engineering, University of Notre Dame, 257 Fitzpatrick Hall of Engineering, Notre Dame, 46556, IN, United States
[c] Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506-6109, United States

ARTICLE INFO

Keywords:
Large language models
ChatGPT
Prompt engineering
Drug trafficking
Social media

ABSTRACT

Social media platforms such as Instagram and Twitter have emerged as critical channels for marketing and selling illegal drugs. Detecting and labeling online illicit drug trafficking activities have become an important measure to combat online drug trafficking. Recently, machine learning has been applied to drug trafficking detection. However, the effectiveness of conventional supervised learning methods in detecting drug trafficking heavily relies on access to substantial amounts of labeled data, while data annotation is time-consuming and resource-intensive. Furthermore, these models often face challenges in accurately identifying trafficking activities when drug dealers use deceptive language and euphemisms to avoid detection. To overcome this limitation, we conduct the first systematic study on leveraging large language models (LLMs), such as ChatGPT, to detect illicit drug trafficking activities on social media. We propose an analytical framework to compose *knowledge-informed prompts*, which serve as the interface that humans can interact with and use LLMs to perform the detection task. Additionally, we designed a Monte Carlo dropout-based prompt optimization method to further improve performance and interpretability. Our experimental findings demonstrate that the proposed framework outperforms other baseline language models in terms of drug trafficking detection accuracy, showing a remarkable improvement of nearly 12%. By integrating prior knowledge and the proposed prompts, ChatGPT can effectively identify and label drug trafficking activities on social networks, even in the presence of deceptive language and euphemisms used by drug dealers to evade detection. The implications of our research extend to social networks, emphasizing the importance of incorporating prior knowledge and scenario-based prompts into analytical tools to improve online security and public safety.

1. Introduction

Drug trafficking, the illegal sale or transport of prohibited drugs, is a global issue that has far-reaching impacts on communities, families, and individuals. Illegal drug trade and usage lead to addiction and health problems and have broader social impacts. Drug trafficking organizations are often associated with violence, corruption, and other forms of criminal activity [1–3]. Social media has provided drug dealers with a convenient platform to market and sell their illicit products [4,5]. Social media platforms offer a broad audience reach and provide drug dealers with a level of anonymity that was previously unattainable. These platforms also facilitate communication between dealers and customers, enabling them to coordinate transactions [6]. Detection of drug trafficking on social networks has become critical in tackling drug trafficking. However, detecting drug trafficking activities on social media poses difficulties due to the use of disguised language and euphemisms by drug dealers [7]. Drug dealers employ code words, acronyms, and other disguised language to avoid detection by law enforcement agencies and social media platforms. Consequently, although social media platforms have implemented some mechanisms to combat drug trafficking, they are ineffective due to the challenges above.

Several research studies have developed task-specific machine learning models to detect drug trafficking activities on social media, including detection of drug dealers [8,9], drug trafficking events [7,10,11], and drug-related hashtags [12]. The problem of detection of illicit

* Corresponding author.
E-mail address: xli48@albany.edu (X. Li).

https://doi.org/10.1016/j.im.2024.104010
Received 5 July 2023; Received in revised form 9 July 2024; Accepted 15 July 2024
Available online 24 July 2024

drug trafficking can be formulated as a supervised machine learning problem, where the input is typically the text or/and images on social media, and the output is drug trafficking activities to be detected. As such, different machine learning approaches, such as Convolutional Neural Networks (CNN) [7,8,11], long short-term memory networks (LSTM) [9], and Transformer networks [7,13], can be applied to process the social media data. However, such task-specific models have several limits. First, task-specific models apply supervised learning to train the models, requiring high-quality labeled data to achieve good performance. Unfortunately, annotating a large dataset on social media can be challenging and time-consuming. Second, as platforms improve their detection capabilities, drug dealers constantly adapt their text-based trafficking techniques to avoid detection. This poses a substantial challenge for supervised learning models, as they may struggle to decipher disguised language and euphemisms, potentially resulting in the oversight of crucial information related to drug trafficking activities.

Meanwhile, recent years have observed the emergence of powerful large language models (LLMs) such as GPT [14] and LLaMA [15]. These models, trained on huge datasets, have demonstrated surprising capabilities in various natural language processing (NLP) tasks, such as natural language understanding, generating coherent and contextually relevant responses, and solving complex tasks through text generation. LLMs have shown promise in helping with various real-world tasks, including complex math problems [16], clinical decision support [17–20], public health [21,22], open education [23], and global warming [24]. Unlike the supervised learning paradigm, which requires large labeled data to train a task-specific model, LLMs can perform NLP tasks with just a few or no examples, achieving results similar to those of state-of-the-art supervised models [25]. Therefore, LLMs provide an alternative solution to detect illicit drug trafficking on social media.

Inspired by the advances in large language models (LLMs), in this paper, we propose to apply LLMs to detect illicit drug trafficking activities from text data on social media (e.g., textual information of posts on Instagram). Although images often go with text data on social media, previous studies [7,8] show that text data dominate the detection of illicit drug trafficking activities. Unlike supervised learning that trains a model with input-label pairs, the LLM-based approach performs a downstream task by reformulating the task with an appropriate textual *prompt*, which bridges the task and the LLMs to enable in-context learning in an autoregressive manner [14]. To apply LLMs for prediction tasks, we must modify the original text into a textual prompt [26]. Consequently, the problem of LLM-based drug trafficking detection boils down to the design of appropriate prompts.

To this end, we propose an analytical framework to compose *knowledge-informed prompts*, which serve as the interface that humans can interact with and use LLMs to detect drug trafficking activities. The prompt combines knowledge, questions, and original text to detect. To construct meaningful prompt knowledge, our proposed framework integrates *prior domain knowledge* regarding drug trafficking behaviors, terminologies, and strategies used by drug dealers, and *acquired knowledge* extracted from an LLM, ChatGPT in particular, with a few examples. We further design an optimization method to optimize the prompts. The prompt optimization uses an iterative strategy [27] to elicit more accurate factual knowledge about drug trafficking than manually created prompts on the benchmark. Meanwhile, Monte Carlo dropout [28] is applied to effectively incorporate prior knowledge specific to drug trafficking behaviors rather than supervised relation extraction models.

Finally, we assess our proposed framework through extensive experiments on an illicit drug trafficking dataset we collected from Instagram. Our framework demonstrates superior performance compared to baseline models. Furthermore, thorough evaluations, including ablation experiments and assessments with varying input shots, demonstrate the importance of prompt design, the value of domain knowledge integration, and the optimal threshold of input shots to maximize performance.

In summary, our research makes the following contributions:

- We conduct the first systematic study on leveraging large language models (LLMs), such as ChatGPT, to detect illicit drug trafficking activities on social media. Consistent with the knowledge contribution framework (KCF) [29], our study contributes to deep learning information systems research (DL-ISR) by applying emerging LLMs to novel domain-specific applications in online security and public safety.
- We propose a new analytical framework to compose knowledge-informed prompts, which serve as the interface that humans can interact with and use LLMs to detect drug trafficking activities. We also developed a Monte Carlo dropout-based prompt optimization method to enhance the effectiveness of the prompts further.
- We demonstrate the effectiveness of the proposed framework through extensive experiments on an illicit drug trafficking dataset we collected from Instagram.

Our research contributes to the broader objective of improving online security and public safety. Through the integrated prompts of prior knowledge and ChatGPT knowledge, our approach provides an effective tool to detect drug trafficking activity on social media. Our research offers a practical and applicable tool for law enforcement agencies, social media managers, and other stakeholders concerned with online security and public safety.

The remainder of this paper is organized as follows: Section 2 provides an overview of related work in drug trafficking detection and large language models, Section 3 describes the proposed method in detail, Section 4 presents experimental results and analysis, and Section 5 concludes the article while outlining potential future avenues of research.

2. Related work

2.1. Drug trafficking detection on social media

As social media platforms have emerged as critical channels for drug marketing and illegal sales, how to effectively and efficiently detect illicit drug trafficking activities becomes important in addressing this issue. There are a few studies that build task-specific machine learning models to detect drug trafficking activities on social media, including detection of drug dealers [8,9], drug trafficking events [7,10,11], and drug-related hashtags [12]. For example, Li et al. [9] applied long short-term memory networks (LSTM) to process textual data of Instagram posts to detect and characterize illicit drug dealers on Instagram. Similarly, Zhao et al. [11] combined SVM and TextCNN to detect illicit drug ads. Since social media posts contain textual and image data, some research is on applying the multimodal approach to drug trafficking detection [7,8,10]. Qian et al. [30] employed a heterogeneous graph to capture multi-modal content and relational structured information from social media to detect illicit drug traffickers. More recently, Hu et al. [12] proposed a framework that combined Bidirectional Encoder Representations from Transformers (BERT) with Graph Convolutional Network (GCN) to classify drug-related hashtags on Instagram.

Existing studies that built task-specific models for drug trafficking detection have several limits. First, they need a large amount of high-quality labeled data to train the models. However, data annotation is time-consuming and resource-intensive. Second, as drug dealers constantly update their deceptive language and euphemisms to avoid detection, a well-trained model might fail to perform the detection task. As large language models advance, the challenges associated with drug trafficking detection in social media are expected to be effectively addressed.

2.2. Large language models

Large language models (LLMs) are sophisticated artificial intelligence models trained on vast amounts of text data and demonstrate

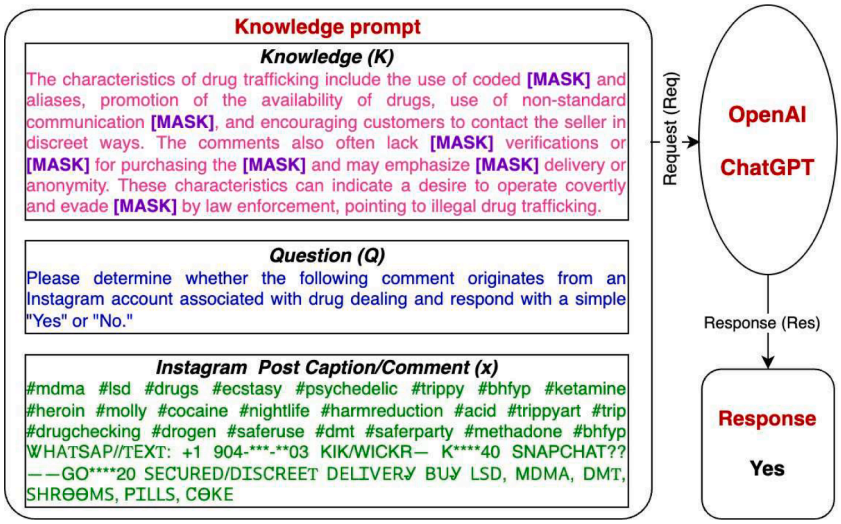


Fig. 1. Illustration of the proposed *knowledge-informed prompt*, which is the interface for humans to interact with and use ChatGPT for drug trafficking detection. The prompt $x' : [K, Q, x]$ is a tuple of knowledge K, question Q, and original text x. The pink font represents knowledge K with masked words in purple, the blue font represents question Q, and the green font represents original text x (e.g., Instagram post captions/comments).

advanced language processing capabilities [31]. Several notable large language models have been developed, including OpenAI's GPT (Generative Pre-trained Transformer) series and Google's BERT (Bidirectional Encoder Representations from Transformers). These large language models are typically pre-trained on massive datasets, often comprising billions of sentences from various sources such as books, websites, and articles. During pre-training, the models learn to predict missing words in sentences, thereby gaining a deep understanding of grammar, context, and semantic relationships. One prominent example of a large language model is ChatGPT [14] developed by OpenAI.

Once pre-training is completed, LLMs can perform various natural language processing (NLP) tasks with just a few or no examples, achieving results similar to those of state-of-the-art supervised models [25]. LLMs also have shown promise in helping with various real-world tasks, including complex math problems [16], clinical decision support [17], public health [21].

2.3. *Prompt engineering based on language models*

Prompts are the interface for humans to interact with and use LLMs. The LLM-based approach performs a downstream task by reformulating the task with an appropriate textual prompt, which bridges the task and the LLMs to enable in-context learning in an autoregressive manner [14]. Prompt engineering [26] has received increasing attention recently. Researchers have explored different methods to design prompts that can improve the performance of language models on specific tasks. For example, Shin et al. [27] proposed a method for designing prompts to improve the performance of language models on various natural language processing tasks. Their approach, AutoPrompt, leverages a small set of labeled examples to automatically generate prompts that guide the language model toward the desired task. They achieved significant performance gains across various benchmarks by fine-tuning the model with task-specific prompts. A proposed method facilitates a chain-of-thought prompting approach, enabling expansive language models to tackle intricate reasoning tasks by generating a sequence of intermediate steps [32]. This methodology sheds light on the emergent property of model scale while expanding the repertoire of reasoning tasks language models can proficiently undertake. An interactive system called PromptChainer has been developed to facilitate the visual programming of LLM chains, empowering individuals without AI expertise to prototype AI-infused applications [33]. However, these methods fail to take advantage of domain knowledge to enhance the performance of the large-language model.

3. Methodology

In this section, we first formulate the problem of drug trafficking detection on social media and then introduce the knowledge-prompted large-language model approach to it.

3.1. *Problem formulation*

Our goal is to design an effective system to detect drug trafficking activities on social media platforms accurately. Specifically, our objective is to classify whether a social media post (e.g., a post on Instagram or a tweet on Twitter) contains information related to the marketing and sales of illicit drugs so that the classification can be applied to monitor drug trafficking activities on social media platforms. In this paper, we only use textual data from the posts. However, images often go with the text on social media, but textual data dominate in detecting illicit drug trafficking activities [7,8]. Let x denote the textual data in a post; our goal is to build a predictive model that takes x as input and predicts a drug trafficking label $y \in \{0,1\}$, where $y = 1$ denotes the post is related to drug trafficking activities and $y = 0$ otherwise. The block highlighted in green in Fig. 1 shows an example of the textual data of an Instagram post annotated as a drug trafficking post. In particular, we propose to leverage large language models (LLMs), such as ChatGPT, so that we can

(1) To leverage the rich knowledge of LLMs acquired through training with large datasets, and
(2) To enable the detection with a few (i.e., few-shot learning where the LLM model is given a few demonstrations of the task) or even no (i.e., zero-shot learning) examples.

Different from supervised learning methods, which train a model with large size of input-label pairs $\{(x_i, y_i)\}_{i\in[N]}$, we need to modify the original text x into a textual *prompt* x' to perform the drug trafficking task on LLMs (e.g., ChatGPT). However, designing appropriate prompts is not a nontrivial task, as the prompts have an important impact on the effectiveness of LLMs. As a result, the problem of LLM-based drug trafficking detection boils down to the design of appropriate prompts, which is the focus of this paper. We will use ChatGPT as the LLM in this paper.

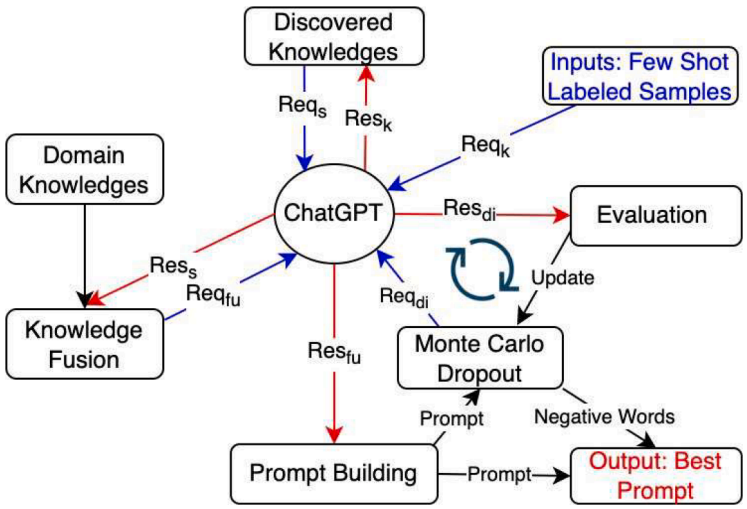


Fig. 2. Framework of our proposed knowledge prompted ChatGPT. Blue arrows represent requests to ChatGPT, while red arrows signify the corresponding responses from ChatGPT. Inputs to the framework are indicated by text in blue, whereas outputs are denoted by text in red.

Table 1
List of symbols used to describe the proposed method.

Notation	Meaning
Req_k	Request knowledge discovery based on a few-shot Instagram comments.
Res_k	Response extracted knowledge from a few-shot Instagram comments.
Req_s	Request the combination and summarization of the extracted knowledge.
Res_s	Response that combines and summarizes the extracted knowledge.
Req_{fu}	Request the fusion of extracted knowledge and domain knowledge.
Res_{fu}	Response the fusion of extracted knowledge and domain knowledge.
Req_{di}	Request the detection of Instagram comments using the prompt after randomly dropping out the ith iteration.
Res_{di}	Response for the prediction of each input Instagram comment data.

3.2. Overview of proposed framework

As shown in Fig. 1, our proposed *knowledge-informed prompt* x' : $[K, Q, x]$ is composed of knowledge K, question Q, and original text x. Then, the knowledge-informed prompt x' is the interface that humans can use to interact with ChatGPT and detect drug trafficking. ChatGPT utilizes this prompt to generate responses that help detect drug trafficking activities. The response component represents the system output, providing valuable insights and detection results.

To design effective *knowledge-informed prompts* to leverage the capabilities of large language models (LLMs) for the detection of drug trafficking activities on social media, we propose an advanced analytical framework that integrates *prior domain knowledge* and *acquired knowledge* extracted from ChatGPT. As shown in Fig. 2, the framework builds upon ChatGPT as its core, combining the power of the large language model with prior knowledge and acquired knowledge extracted from ChatGPT. Different notations in the framework, as shown in Table 1, represent distinct meanings and functions in terms of knowledge request and response. This framework encompasses several steps to design effective prompts to improve the model's effectiveness in detecting drug trafficking activities.

1. **Knowledge extraction from ChatGPT.** In this step, we leverage the capabilities of ChatGPT to extract knowledge with a few shots of labeled data $\{(x_i, y_i)\}_{i\in[N]}$, where N is a small number. We input relevant text passages or queries into ChatGPT and utilize its language comprehension and generation capabilities to extract key facts, insights, and relationships related to drug trafficking activities. Various sources, such as news articles, online forums, and social networks, extract valuable information.
2. **Knowledge fusion.** The knowledge extracted from ChatGPT is then integrated with domain-specific knowledge to improve the drug trafficking detection capabilities of the framework. Domain-specific knowledge includes information from experts, research articles, and curated databases, providing a deep understanding of drug trafficking patterns, terminology, smuggling techniques, and key entities involved. The integration process aligns and reconciles the extracted knowledge with the domain knowledge, capturing nuanced insights from both sources.
3. **Prompt design based on integrated knowledge.** To effectively prompt the ChatGPT model, we design prompts that leverage integrated knowledge and exploit areas of potential confusion. Confusion knowledge refers to specific aspects or concepts that ChatGPT might struggle to understand or disambiguate accurately. By addressing and clarifying these confusion points in the prompts, we guide the model's attention and enhance its understanding of drug trafficking-related text. The prompts may involve specific questions, context framing, or the inclusion of key keywords related to drug trafficking activities.
4. **Prompt optimization.** Prompt optimization aims to fine-tune the designed prompts to maximize the performance of the ChatGPT model on drug trafficking detection tasks. Monte Carlo drop is employed to improve the prompts' effectiveness iteratively. By analyzing the model's responses to different prompts and measuring their impact on performance, we iteratively refine the prompts to align the model's behavior more closely with the desired drug trafficking detection outcomes.

The pipeline of knowledge extraction, knowledge fusion, prompt design based on confusion knowledge, and prompt optimization is crucial in enhancing the model's understanding and performance in identifying drug trafficking activities. Subsequent sections elaborate on the details for each part.

3.3. Incorporating prior domain knowledge

Domain-specific knowledge holds immense importance in improving the detection of drug trafficking activities on social media platforms. Domain experts possess specialized knowledge and insights into drug trafficking behaviors, terminologies, and strategies used by drug dealers. Integrating this knowledge into the detection framework can improve the accuracy and effectiveness of identifying drug trafficking. To leverage domain knowledge, we focus on three key aspects: hashtags, contact information, and special symbols commonly used by drug

Table 2
Types of domain knowledge for drug trafficking detection.

Name	Meaning	Example
Hashtag	Drug sale-related hashtags	#MDMA #Cocaine #LSD
Contact information	Telephone number, email address, and other private	
social media accounts	Txt/WhatsApp.+1,7**.***.9414 Wichr/snapchat james*****52 kik james*****52	
Special symbol	Using Punctuation, special characters, and emojis to evade detection	M.D.M.A, C.O.C.A.I.N.E, L.s.d M.o.l.l.y, SHRΘ ΘMS CΘKE

dealers to avoid detection. Table 2 provides examples and meanings for each aspect.

- **Hashtag:** Drug sale-related hashtags are frequently used on social media platforms to facilitate communication and advertising among drug dealers and potential buyers. By identifying and analyzing these hashtags, we can gain valuable information about drug trafficking activities. For example, hashtags like "#MDMA" or "#Cocaine" often indicate the sale or discussion of specific drugs.
- **Contact Information:**Drug dealers often share contact information, such as phone numbers or alternative social media accounts, to establish communication with potential buyers. By monitoring and analyzing this contact information, we can detect and track drug trafficking activities. Examples of contact information include phone numbers with specific area codes or messaging app usernames like "Telegram" or "Wickr."
- **Special Symbols:**To evade detection, drug dealers frequently use various techniques, including the use of punctuation marks, special characters, or emojis, to mask content related to drug trafficking. When we recognize and interpret these special symbols, we can reveal the underlying drug-related messages. Examples include variations in drug names using special characters such as "M.D.M.A" or "C.O.C.A.I.N.E," or substituting letters with similar-looking symbols like "SHRΘΘMS."

Incorporating domain-specific knowledge related to hashtags, contact information, and special symbols provides valuable contextual insights that help uncover drug-trafficking discussions and improve the detection system's precision and recall. This domain knowledge will be fused with acquired knowledge extracted from ChatGPT to design prompts.

3.4. Knowledge fusion and prompt engineering

Note that given a text data x, our proposed *knowledge-informed prompt* $x' : [K, Q, x]$ is the combination of knowledge K, question Q, and original text x. Fig. 3 shows the workflow to compose the prompt.

3.4.1. Prompt – knowledge K

In particular, we compose the knowledge part K with *prior domain knowledge* and *acquired knowledge* extracted from ChatGPT with a few shots of labeled data.

Knowledge Discovery with ChatGPT. Beyond the *prior domain knowledge* as discussed in Section 3.3, we also extract relevant knowledge with ChatGPT since ChatGPT is trained on very large datasets. Specifically, we query ChatGPT with a few shots of labeled data $\{(x_i, y_i)\}_{i\in[N]}$, where N is a small number, to guide ChatGPT to generate informative and relevant responses. During fine-tuning, the model learns to identify patterns, relationships, and insights within the drug trafficking domain. This allows the model to acquire implicit knowledge from the dataset and generate informed responses for drug-trafficking-related queries.

Knowledge Fusion as Prompt. Knowledge fusion is a critical component of the prompt design process. It involves integrating relevant information from domain knowledge and knowledge discovered by ChatGPT into the prompts provided to the model. By fusing knowledge into the prompts, we provide the ChatGPT model with additional

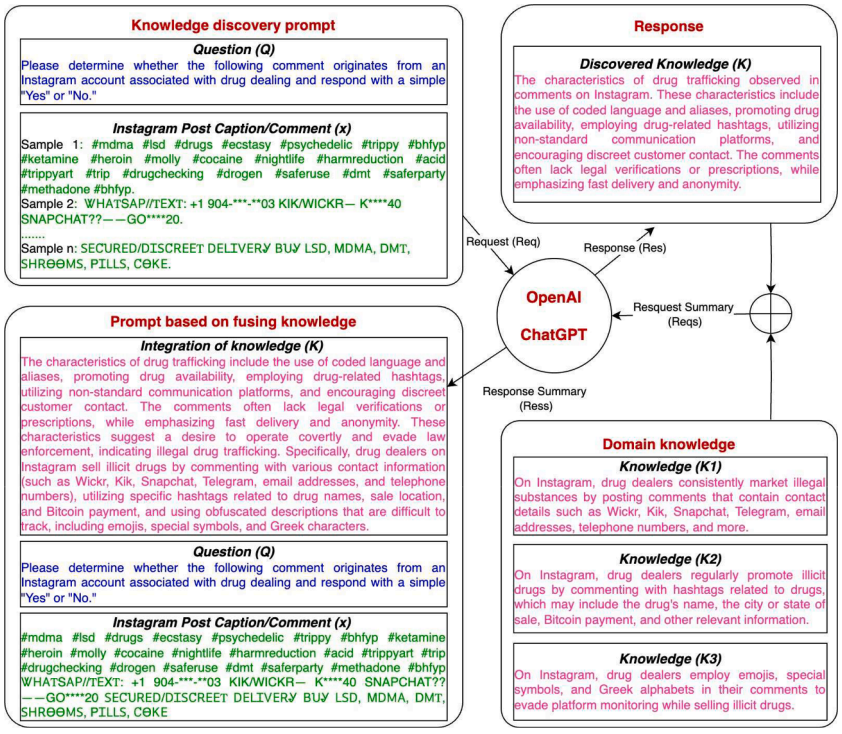


Fig. 3. Illustration of the workflow to compose prompt $x' : [K, Q, x]$, which is a tuple of knowledge K, question Q, and original text x. The pink font represents knowledge K, the blue font represents question Q, and the green font represents original text x (e.g., Instagram post captions/comments).

context and cues related to drug trafficking. This integration helps the model to make more accurate predictions and generate responses that align with the nuances of drug trafficking detection tasks. For example, a prompt could include a snippet of domain knowledge about common drug trafficking routes or the characteristics of illicit drug concealment methods. This prompts the model to focus on these aspects and generate more insightful responses.

3.4.2. Prompt – question Q

To further enhance the design of knowledge prompts, we identify specific types of questions that generate informative responses related to the detection of drug trafficking. These questions target key aspects of drug trafficking activities, methods, or indicators. For instance, questions like "What are the signs of drug trafficking in online communications?" or "How can drug smuggling be detected at border checkpoints?" guide the model in providing relevant information in its responses. By incorporating carefully crafted questions into the prompts, we guide the ChatGPT model to generate more focused and informative outputs for drug trafficking detection tasks.

3.5. Prompt optimization with Monte Carlo Dropout

We also develop a Monte Carlo dropout based prompt optimization method to further enhance the effectiveness of the prompts. The prompt optimization process aims to fine-tune the designed prompts to maximize the performance of the ChatGPT model on drug trafficking detection tasks. We can identify keywords that significantly influence the model's decisions by systematically analyzing the impact of individual words in the prompt. The prompt optimization is shown in Algorithm 1.

The algorithm, "Prompt Optimization with Monte Carlo Dropout," takes the designed prompt, the number of Monte Carlo dropout iterations, and the dropout probability as input. The algorithm iteratively performs the following steps:

- Create a copy of the designed prompt, optimized_prompt, to preserve the original prompt (Step 1).
- Split optimized_prompt into a list of words.
- For each word in the prompt, perform the Monte Carlo dropout by replacing the word with a "MASK" token (Steps 4-6).
- Perform Monte Carlo simulation on the modified prompt to obtain performance metrics.
- Calculate the change in F1 score, score_change, based on the original and simulated F1 scores.
- Update the importance score of each word in the prompt by incrementing it with score_change.
- Repeat Steps 2-6 for the desired number of iterations.

By optimizing the prompt using Monte Carlo dropout, we can identify and refine the words that significantly impact the model's performance in drug trafficking detection tasks.

The resulting optimized prompt provides valuable insight into the specific words and their contributions to the model's decision-making process. This enhanced understanding enables us to fine-tune the prompt and align the model's behavior more closely with the desired drug trafficking detection outcomes.

Algorithm 1 Word Importance Score Calculation with Monte Carlo Dropout for Prompt Optimization.

```
procedure Calculate Importance Scores(paragraph, num_iterations, dropout_prob)
    Initialize an empty dictionary word_scores
    Split the paragraph into a list of words
    for iter in 1 to num_iterations do
        Create a copy of the list of words called masked_words
        Initialize an empty list called dropout_list
        for word_index in 0 to length(masked_words) - 1 do
            if random.random() < dropout_prob then
                Append word_index to dropout_list
                Set masked_words[word_index] as "MASK"
            end if
        end for
        Create a prompt_knowledge by joining masked_words into a string
        Perform Monte Carlo on prompt_knowledge to get performance metrics
        Compute the change in score_change
        for word_index in dropout_list do
            if word_index not in word_scores then
                Set word_scores[word_index] as 0.0
            end if
            Increment word_scores[word_index] by score_change
        end for
    end for
    Sort word_scores by value in descending order, resulting in sorted_scores
    return sorted_scores
end procedure
```

4. Experiments

4.1. Experimental settings

In this subsection, we describe the experimental setup used to

Table 3
Performance Metrics of Different Models.

Model	Accuracy	Precision	Recall	F1 Score
BERT [34]	76.23%	75.39%	85.91%	79.78%
XLNet [35]	76.13%	75.44%	86.76%	79.92%
ALBERT [36]	77.62%	78.06%	83.14%	80.13%
DistilBERT [37]	80.30%	80.73%	86.03%	82.68%
RoBERTa [38]	82.78%	83.29%	88.29%	85.02%
Ours	**94.58%**	**96.60%**	**93.42%**	**94.98%**

evaluate the performance of the knowledge-prompted ChatGPT framework for drug trafficking detection. The experimental setup encompasses the dataset, model configuration, evaluation metrics, and baselines.

- **Dataset.** To evaluate the framework's effectiveness, we carefully selected a comprehensive dataset called IDDIG that contains text samples related to drug trafficking on Instagram. Similar samples were removed from the IDDIG data, resulting in a final set of 886 samples for the evaluation experiments. There are 486 positive samples (i.e., drug trafficking) and 400 negative samples (i.e., non-drug trafficking). We propose a framework for identifying positive samples. The dataset consists of labeled data for validation and testing, facilitating the computation of performance metrics. It is thoughtfully curated to encompass various drug trafficking scenarios, regions, and languages, ensuring the framework's generalizability. Specifically, a random set of 40 samples was selected as input for the design of knowledge prompts in each experiment. In contrast, the remaining samples were set aside as a test set to assess the performance of the proposed method.
- **Model Configuration.** Our framework is based on the ChatGPT API, specifically using the GPT Turbo 3.5 model. This model undergoes pre-training on a significant corpus of text data to establish a robust language understanding. To accurately detect drug trafficking, we optimize the ChatGPT request prompt by incorporating the knowledge prompt design and optimization techniques mentioned earlier. The model configuration involves several parameters, such as the number of iterations for word dropout, the number of input samples, and the number of words to be dropped out in each iteration. Parameter tuning can be performed to enhance the model's performance.
- **Evaluation Metrics.** To evaluate the performance of the knowledge-prompted ChatGPT framework, various metrics have been used. The chosen metrics encompass precision, recall, F1 score, and accuracy. Precision evaluates the accuracy of identifying drug trafficking instances, while recall measures the effectiveness of correctly identifying actual drug trafficking instances. The F1 score combines precision and recall, offering a balanced measure of the model's performance.
- **Baselines.** To compare the performance of the knowledge-prompted ChatGPT framework, it is important to establish appropriate baselines. Baselines include BERT [34], XLNet [35], ALBERT [36], DistilBERT [37], and RoBERTa [38] for the detection of drug trafficking. These baselines provide reference points to evaluate the effectiveness and superiority of the proposed framework. Baseline models are trained and evaluated on the same dataset as the knowledge-prompted ChatGPT framework, using the same evaluation metrics to ensure a fair comparison.

4.2. Evaluation of knowledge-prompted ChatGPT for drug trafficking detection

Table 3 shows the evaluation results of our knowledge-prompted ChatGPT framework and the baseline approaches for drug trafficking detection. We compare the performance of our framework with several baseline models, including BERT, XLNet, ALBERT, DistilBERT, and RoBERTa. After comparing the outcomes with the baseline models, we noted a significant enhancement across all evaluation metrics. Our framework demonstrated superior performance in drug trafficking detection tasks by achieving higher accuracy, precision, recall, and F1 scores than all other models. Notably, the proposed framework exhibited exceptional improvements, with nearly a 12% increase in drug trafficking detection accuracy and an approximately 10% improvement in F1 score, surpassing other baseline language models.

The significant performance gains can be attributed to incorporating domain-specific knowledge, effective prompt design, and knowledge optimization techniques employed in our framework. The integration of relevant knowledge and the guidance provided through carefully designed prompts contribute to the framework's ability to detect drug trafficking activities accurately.

4.3. Assessing the performance of knowledge prompt ChatGPT with varying input shots

To evaluate the impact of varying input shots on the performance of the knowledge-driven ChatGPT framework for drug trafficking detection, we conducted experiments using different numbers of input shots. As shown in Table 4, the framework's performance consistently improved with an increasing number of input shots. Using 40 shots yielded the best results, achieving an accuracy of 89.84% and an F1 score of 90.24%, while using five shots resulted in an accuracy of 86.23% and an F1 score of 86.94%. These results indicate that the knowledge-driven ChatGPT framework benefits from increasing input shots up to a certain point, improving performance in drug trafficking detection tasks. However, after reaching the optimal point (in this case, around 40 shots), further increases in input shots may not necessarily yield significant performance gains.

4.4. Evaluating model components through ablation experiments

To assess the individual contributions of different model components in the knowledge-driven ChatGPT framework for drug trafficking detection, we conducted ablation experiments. We evaluated the performance when removing or modifying specific prompt sources and knowledge components. The evaluation results are summarized in Table 5.

Removing the prompt knowledge components, resulting in no prompt and no knowledge, led to an accuracy of 87.81%, and an F1 score of 88.70%. When the domain expert provided a drug-related hashtag as the knowledge prompt source, the performance remained consistent with an accuracy of 87.81%. Still, the precision increased to 92.76%, albeit with a slightly lower recall of 84.36% and an F1 score of 88.36%. Using the domain expert's contact information as the prompt source improved precision significantly to 97.46%. However, this change resulted in a decreased recall of 79.01% and a slightly lower accuracy of 87.36%, with an F1 score of 87.27%. Replacing the prompt source with a special symbol guided by the domain expert yielded an accuracy of 90.52%, precision of 89.11%, recall of 94.24%, and an F1 score of 91.60%. These results indicate that the special symbol prompt source effectively captures the model's attention and improves its ability to identify drug trafficking activities. When all knowledge components provided by the domain expert were included, the performance further improved. The framework achieved an accuracy of 90.74% and an F1 score of 90.95%. This highlights the significance of incorporating comprehensive domain knowledge to enhance the framework's performance.

Removing prompt knowledge components resulted in an accuracy of 87.81% and an F1 score of 88.70%. Utilizing a drug-related hashtag as the knowledge prompt source maintained an accuracy of 87.81%, but precision increased to 92.76% with a slightly lower recall of 84.36% and an F1 score of 88.36%. Using the domain expert's contact information as

Table 4
Evaluating the Performance of Knowledge-Prompted ChatGPT Across Different Input Shots.

Shot number	Accuracy	Precision	Recall	F1 Score
5 shots	86.23%	90.63%	83.54%	86.94%
10 shots	88.71%	91.77%	**87.24%**	89.45%
20 shots	87.81%	90.91%	86.42%	88.61%
30 shots	87.13%	95.15%	80.66%	87.31%
40 shots	**89.84%**	**95.41%**	85.60%	**90.24%**

the prompt source significantly improved precision to 97.46%, albeit with decreased recall of 79.01%, accuracy of 87.36%, and an F1 score of 87.27%. Replacing the prompt source with a special symbol guided by the domain expert yielded an accuracy of 90.52%, precision of 89.11%, recall of 94.24%, and an F1 score of 91.60%. Including all knowledge components provided by the domain expert further enhanced performance, resulting in an accuracy of 90.74% and an F1 score of 90.95%. These results demonstrate the effectiveness of different prompt sources and highlight the importance of incorporating comprehensive domain knowledge to enhance the framework's performance in drug trafficking detection.

Next, we examine the effect of Monte Carlo dropout and the number of input shots. Without dropout, our framework with 40 shots achieved an accuracy of 89.84%, precision of 95.41%, recall of 85.60%, and an F1 score of 90.24%. However, the performance significantly improved when incorporating dropout with the same number of shots. The framework achieved an accuracy of 91.87%, precision of 99.52%, recall of 85.60%, and an F1 score of 92.04%. Finally, our complete framework with 40 shots and all knowledge components achieved the best performance, with an accuracy of 94.58%, precision of 96.60%, recall of 93.42%, and an impressive F1 score of 94.98%.

These experiments demonstrate the significance of prompt sources, knowledge components, and the utilization of Monte Carlo dropout in enhancing the accuracy and effectiveness of our framework for drug trafficking detection. The comprehensive integration of domain knowledge and the optimization of model components contribute to the superior performance of our proposed approach.

4.5. Drug trafficking detection case studies

4.5.1. Leveraging Monte Carlo dropout for a knowledge prompt example

In this subsection, we present case studies to demonstrate the effectiveness of our proposed method for drug trafficking detection. The case studies involve the use of positive words and negative words as prompts, highlighting the top 10 words ranked by their calculated importance scores (See Table 6).

The prompts are designed to capture the characteristics of drug trafficking discussions observed in comments on Instagram. The identified positive words, represented in green font, indicate significant features associated with drug trafficking, such as drug names, delivery, non-standard communication platforms, and evasion of law enforcement. For example, "address" and "delivery" are considered strong positive words that express key features of drug trafficking. These words are crucial in improving ChatGPT's performance in detecting drug trafficking. The presence of "address" signifies the use of email addresses as a means of communication, while "delivery" emphasizes fast delivery. These features may have been missing in ChatGPT's understanding of drug trafficking, and their inclusion enhances the model's ability to identify drug trafficking activities accurately. On the contrary, the negative words, represented in red font, signify the words that negatively affect the performance of ChatGPT in detecting drug trafficking activities. These words hinder the model's ability to accurately detect illicit drug-related content on social media platforms. Negative words will be removed from the analysis to enhance the performance of drug trafficking detection.

The calculated importance scores provide information on the relevance and significance of each word in detecting drug trafficking. The top-ranked words offer valuable clues to identify and monitor drug trafficking activities on social media platforms. By utilizing these prompts, our method enables a more targeted and effective approach to detecting drug-related content and identifying potential drug dealers.

Table 6 reveals how "Instagram" emerges as both a significant positive and negative term in our research, highlighting the intricate complexities of language processing within large language models. This phenomenon underscores the context-dependent nature of prompt influence on model performance, illustrating the complex interplay between a single word and the model's response. The dual role of "Instagram" highlights how the surrounding context can significantly sway a word's effect, enhancing the model's accuracy and relevance in certain scenarios while potentially leading to confusion and bias in others. This case exemplifies the critical importance of understanding how specific prompts can affect model behavior, emphasizing the need for careful prompt engineering to optimize performance. In addition, the significance of non-drug related words such as "as," "including," and "and" in our analysis sheds light on a foundational aspect of natural language processing that transcends domain-specific vocabulary. These connective words are essential for the structure and coherence of language, playing a crucial role in shaping the context and meaning of sentences. Their presence in our study demonstrates how strategic use of language, extending beyond domain-specific terms to include general linguistic connectors, can profoundly influence the outcome of model applications. This insight reveals these words' subtle yet powerful impact on the model's interpretation and task performance accuracy, highlighting the complex art of prompt engineering where every element of the prompt, regardless of its direct relevance to the task at hand, can significantly alter the model's output.

The case studies highlight the capabilities of our proposed method in capturing key characteristics of drug trafficking discussions, shedding light on the covert nature of such activities, and providing law enforcement agencies and social media managers with valuable insights for intervention and prevention efforts.

4.5.2. A comparative analysis: proposed framework vs. alternative prompts

In this subsection, we present a comparative analysis of our proposed method for detecting drug trafficking, contrasting it with alternative prompts. There are three alternative prompts: a prompt with no specific topic, a prompt focused on domain knowledge, and a prompt based on extracted knowledge.

- **Prompt a.** No prompt.
- **Prompt b.** Drug dealers on Instagram utilize various methods to sell illicit drugs. They typically comment with contact information (such as wickr, Kik, snapchat, telegram, or telephone number), employ specific hashtags (including names, sale locations, or bitcoin payment), and use obfuscated descriptions with emojis, special symbols and Greek characters to evade tracking, effectively showcasing the drugs they have for sale.
- **Prompt c.** The characteristics of drug trafficking include using coded language and aliases, promoting the availability of drugs, using non-standard platforms, and encouraging customers to contact the seller discreetly. The comments also often lack legal verifications or prescriptions for drug purchasing and may emphasize fast delivery or anonymity. These characteristics can indicate a desire to operate covertly and evade detection by law enforcement, pointing to illegal drug trafficking.
- **Prompt d.** The characteristics of drug trafficking include using coded language and aliases, promoting drug availability, employing drug-related hashtags, utilizing non-standard communication platforms, and encouraging discreet customer contact. The comments often lack legal verifications or prescriptions, emphasizing fast delivery and anonymity. These characteristics suggest a desire to operate covertly

Table 5
Model Components Evaluation via Results from Ablation Studies.

Prompt source	knowledge	Accuracy	Precision	Recall	F1 Score
No prompt	No knowledge	87.81%	90.21%	87.24%	88.70%
Domain expert	Hashtag	87.81%	92.76%	84.36%	88.36%
Domain expert	Contact information	87.36%	97.46%	79.01%	87.27%
Domain expert	Special symbol	90.52%	89.11%	**94.24%**	91.60%
Domain expert	All Knowledge (AK)	90.74%	98.10%	84.77%	90.95%
Ours w/o dropout	40 shots	89.84%	95.41%	85.60%	90.24%
Ours w/o dropout	40 shots + AK	91.87%	**99.52%**	85.60%	92.04%
Ours w/ dropout	40 shots + AK	**94.58%**	96.60%	93.42%	**94.98%**

and evade law enforcement, indicating illegal drug trafficking. Specifically, drug dealers on Instagram sell illicit drugs by commenting with various contact information (such as Wickr, Kik, Snapchat, Telegram, email addresses, and telephone numbers), utilizing specific hashtags related to drug names, sale location, and Bitcoin payment, and using obfuscated descriptions that are difficult to track, including emojis, special symbols, and Greek characters.

Based on the alternative and proposed prompts (i.e., Prompt d.), we utilize sample data and examine the effectiveness of different prompts in identifying drug trafficking activities, as shown in Table 7. It illustrates the comparison by providing four different prompts: prompt a, representing no prompt; prompt b, which includes all knowledge; prompt c, comprising extracted knowledge; and prompt d, a fusion of prompt b and prompt c. Each sample in the table is labeled with either a positive (P) or negative (N) indication based on the presence or absence of drug trafficking elements.

In Case 1, ChatGPT-based methods accurately detect drug trafficking despite the presence of contact information (e.g., WhatsApp, Wickr, Kik, Snapchat) and drug-related hashtags suggest potential drug trafficking activity. This showcases the effectiveness of our approach in identifying drug trafficking activities even when perpetrators employ deceptive techniques. However, there are instances where ChatGPT makes mistakes. Case 2, although unrelated to drug trafficking, shares similarities that may lead to erroneous predictions without prompts. This highlights the importance of prompts in guiding the model and minimizing false positives. Interestingly, Case 3 reveals that ChatGPT can make mistakes even when provided with domain knowledge as a prompt. In some cases, including domain knowledge may introduce noise and negatively impact the model's performance compared to having no prompt at all. This suggests that the prompt design and knowledge extraction must be carefully considered. When Instagram drug trafficking events involve special characters, such as punctuation marks, the automatically extracted knowledge may hurt ChatGPT's predictions. These cases emphasize the need for robust methods to handle different types of input data and ensure accurate detection. Prompt d, which includes prompt b (domain knowledge), may sometimes lead to similar mistakes, as observed in Case 5. This highlights the complexity of balancing the inclusion of domain knowledge and avoiding potential pitfalls associated with prompt-based approaches.

The case studies provide valuable insights into our method's performance, highlighting its ability to distinguish between positive and negative samples based on the prompts used. The results showcase the importance of prompt design and the integration of domain knowledge in enhancing drug trafficking detection mechanisms.

Table 6
Example Prompts by the proposed method. Bold font Represents Positive words, and Italic font Represents Negative words.

Prompt	TopK	P Words	N Words
The characteristics of **drug** trafficking observed in comments on *Instagram*. These characteristics	Top1	addresses	emojis
include the use of coded language and aliases, *promoting* drug availability, *employing* drug-related	Top2	delivery	Instagram
hashtags, utilizing **non-standard** communication platforms, *and* encouraging discreet customer	Top3	enforcement	contact
contact. The comments often lack legal verifications or prescriptions while emphasizing fast	Top4	drug	employing
delivery and *anonymity*. These characteristics suggest a desire to **operate** covertly and **evade** law	Top5	operate	payment
enforcement, indicating illegal drug **trafficking**. Specifically, drug dealers on **Instagram** sell	Top6	as	Wickr
illicit drugs by commenting with various contact information (such **as** *Wickr*, Kik, Snapchat,	Top7	evade	anonymity
email **addresses**, and telephone numbers), utilizing specific hashtags related to drug names, sale	Top8	Instagram	promoting
location, and Bitcoin *payment*, and using obfuscated descriptions that are difficult to track,	Top9	trafficking	including
including emojis, special symbols, and Greek characters.	Top10	on-standard	and

5. Discussions

This section discusses our research's key findings and implications on the knowledge prompted ChatGPT framework for drug trafficking detection. We address the framework's performance, the effectiveness of different model components, and the research's potential limitations and future directions.

5.1. Performance of the framework

The evaluation results demonstrate the superior performance of our knowledge-prompted ChatGPT framework compared to baseline models. Our framework achieved an accuracy of 94.58%, precision of 96.60%, recall of 93.42%, and an F1 score of 94.98%. These results highlight the effectiveness of leveraging domain-specific knowledge, well-designed prompts, and prompt optimization techniques in enhancing drug trafficking detection capabilities.

Comparisons with state-of-the-art models such as BERT, XLNet, ALBERT, DistilBERT, and RoBERTa further validate the superiority of our framework. It consistently outperformed these models in terms of accuracy, precision, recall, and F1 score, indicating its ability to capture drug trafficking activities accurately and make informed predictions.

5.2. Effectiveness of model components

The ablation experiments provided valuable insights into the effectiveness of different model components. Removing the prompt source and knowledge components significantly decreased performance, underscoring the importance of incorporating domain knowledge and well-crafted prompts. Our framework's ability to fuse domain knowledge into the prompts, guided by request-related questions and knowledge optimization techniques, played a crucial role in achieving the observed performance gains.

The analysis of varying input shots showed that the framework's performance improved by increasing shots to an optimal point. Beyond that point, further increases in input shots did not yield significant

Table 7
Contrasting the Proposed Method with Alternative Prompts: A Comparative Analysis. Prompt a. Represents no prompt, Prompt b. Represents Prompt with All Knowledge, Prompt c. Represents Prompt with Extracted Knowledge, Prompt d. Represents Prompt that fusing Prompt b. and Prompt c. "P" represents drug trafficking label or prediction, while "N" represents non-drug trafficking. Italic fonts indicate wrong predictions.

ID	Cases
1	WHαTSαP // TEXT +1 760*****43 ωICKR// KIK //Snαpchαt X****ver SECURED/DISCREET DELIVERY BUY LSD, MDMA, DMT,SHRΘ ΘMS, PILLS, CΘKE & more #lsd#shrooms #shroomforsale #magicshrooms? #tabs
multirow2*2	Label P Prompt a. P Prompt b. P Prompt c. P Prompt d. HOW TO AVOID RELIANCE ON PILLS ? #medication Synovation Medical Group offers a Functional Restoration Program where our physicians help patients overcome chronic pain through an interdisciplinary approach so they can return to normal work and life. This program helps -Increasing the reliance on one's self-Decreasing pain medications #painmedication #healthcare #medicalcare #chronicpain #healthiswealth #opioidcrisis "
3	Label N Prompt a. *P* Prompt b. N Prompt c. N Prompt d. N #psychedelic #psy #psytime #psychedelicrock #magicmushrooms #lsdtabs #shrooms #trippyvibes #mdmaofficials #microdot #purpleacid #blotteracid
4	Label P Prompt a. P Prompt b. *N* Prompt c P Prompt d. P M.D.M.A n N.E.M.B.U.T.A.L, K.U.S.H, C.O.C.A.I.N.E,A.C.I.D,(L.s.d n C.a.r.t) Txt/WhatsApp..+1,*** Telegram:*** Wickr: *** M.Md,L.s.d n M.o.l.l.y,a.d.i.e.s
5	Label P Prompt a. P Prompt b. P Prompt c. *N* Prompt d. P Self Medicated ?#weed #marijuana #cbd #stoner #ganja #indica #sativa #kush #maryjane #dank #medicalmarijuana #bong #hemp #stoned #bhfyp
	Label N Prompt a. N Prompt b. *P* Prompt c. N Prompt d. *P*

performance gains. This suggests a threshold beyond which additional knowledge may not contribute substantially to drug trafficking detection tasks.

5.3. Advancing AI for public safety and law enforcement

The technological innovations brought forth by this research extend the capabilities of large language models, charting new territories in their application for public safety and law enforcement. By pioneering the integration of domain-specific knowledge into the generative capabilities of ChatGPT, our framework exemplifies the potential of AI to tackle sophisticated challenges in digital surveillance and crime prevention. Moreover, developing and implementing a prompt optimization technique tailored for enhancing model interpretability and performance represent significant strides toward the practical deployment of AI in sensitive societal domains.

This body of work not only advances the technical discourse surrounding the use of AI in combating online illicit activities but also sets a foundation for future explorations aimed at refining and expanding the applicability of language models to broader societal challenges. The path forward involves a deepened investigation into multi-modal data analysis and the exploration of next-generation language models, promising to elevate AI-driven detection systems' precision, adaptability, and impact in safeguarding digital communities.

5.4. Limitations and future directions

While our knowledge prompted the ChatGPT framework to demonstrate impressive performance, several limitations should be considered. First, the framework's effectiveness heavily relies on the quality and comprehensiveness of the domain knowledge integrated into the prompts. Further efforts are required to continuously update and refine the knowledge base to adapt to evolving drug trafficking patterns and techniques.

Second, the evaluation focused on a specific dataset and task. Additional evaluations across different datasets and scenarios are needed to validate the framework's generalizability.

Furthermore, the framework's interpretability can be enhanced by exploring techniques such as attention mapping or explainable AI methods. This would enable a better understanding of the model's decision-making process and help address potential biases or limitations.

Future research directions could involve exploring more advanced language models, such as GPT-4 or similar architectures, to improve the framework's performance. Additionally, investigating multi-modal approaches that combine text with other forms of data, such as images or network traffic, could provide a more comprehensive understanding of drug trafficking activities.

5.5. Data security concerns

Utilizing LLMs like ChatGPT in sensitive fields such as law enforcement and public safety naturally raises concerns regarding data security and privacy, especially considering the sensitivity of the data involved and the potential for access by unauthorized parties. The practical implication of these concerns is multifaceted, impacting LLMs' trustworthiness, security, and ethical use in sensitive applications. Key concerns include:

- **Data Privacy** [39]: The risk of exposing sensitive information, whether related to specific individuals, ongoing investigations, or security measures, is a primary concern. Unauthorized access to this data could compromise privacy rights and operational security.
- **Model Integrity** [40]: The integrity of the models themselves can be at risk if adversaries gain the ability to manipulate or reverse-engineer the models in a way that undermines their effectiveness or repurposes them for malicious intent.
- **Adversarial Exploitation** [41]: There is a risk that adversaries could use the same models to develop strategies to evade detection or to gather intelligence on law enforcement techniques and strategies.

To address these concerns, future research may consider the following advanced technological solutions:

- **Homomorphic Encryption** [42–44]: To safeguard data privacy while using LLMs, homomorphic encryption can be employed. This technique allows data to be processed in its encrypted form, ensuring that sensitive information remains secure even during analysis.
- **Federated Learning** [45–48]: Federated learning offers a promising approach to enhance data privacy. Training models on decentralized data sources without needing to centralize sensitive information, significantly reduces the risk of data breaches.
- **Differential Privacy** [49–51]: Implementing differential privacy techniques can protect individual data points within the dataset. Adding noise to the data or the model's outputs makes it substantially more challenging for adversaries to infer specific details about the data subjects.

These strategies are crucial for addressing the dual-use nature of LLMs, ensuring that the benefits of these powerful tools can be harnessed for public safety and law enforcement without compromising data security and privacy.

6. Conclusions

This research presented a knowledge-prompted ChatGPT framework for drug trafficking detection. The framework leverages domain-specific knowledge, well-designed prompts, and prompt optimization techniques to improve the performance of the ChatGPT model in accurately identifying and labeling drug trafficking activities. Our framework outperforms baseline models, showcasing an impressive enhancement in accuracy by 11.8%, precision by 13.31%, recall by 5.13%, and F1-score by 9.96%. By integrating domain knowledge into the prompts, our framework captures nuanced insights, identifies key indicators, and provides actionable information for detecting drug trafficking. The knowledge fusion technique, guided by request-related questions and knowledge optimization, contributes to the interpretability and effectiveness of the framework. We conducted thorough evaluations, including ablation experiments and assessments with varying input shots, to understand the impact of different model components and input variations. The results revealed the importance of prompt design, the value of domain knowledge integration, and the optimal threshold of input shots to maximize performance.

Although our framework exhibits promising results, there are limitations to consider. The framework's effectiveness heavily depends on the quality and comprehensiveness of the domain knowledge integrated into the prompts. The framework's generalizability should be further validated by applying it to diverse datasets and scenarios. Enhancing the framework's interpretability and exploring advanced language models are potential areas for future research. The knowledge-prompted ChatGPT framework has great potential in addressing real-world challenges related to drug trafficking detection. By combining the power of language models with domain expertise, the framework contributes to advancing natural language processing techniques for combating illicit activities. Its performance superiority over baseline models shows its practical applicability and potential for implementation in real-world settings.

Overall, this research contributes to the growing body of knowledge-driven approaches in natural language processing and highlights the importance of incorporating domain-specific knowledge to enhance model performance. The framework of knowledge-prompted ChatGPT is a stepping stone for future advancements in the field, with potential applications in various domains beyond drug trafficking detection, such as community and key player detection.

Author Agreement Statement

AI Disclosure: We confirm that no generative artificial intelligence was used in the preparation of this manuscript.

CRediT authorship contribution statement

Chuanbo Hu: Writing – review & editing, Writing – original draft, Software, Data curation. **Bin Liu:** Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Xin Li:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Conceptualization. **Yanfang Ye:** Supervision, Funding acquisition, Conceptualization. **Minglei Yin:** Data curation, Formal analysis, Visualization.

Declaration of competing interest

We declare there are no financial or personal connections that could be perceived as influencing the research presented in this paper. This declaration is for transparency and to maintain the research's integrity.

Acknowledgements

The NSF partially supports this work under grants CMMI-2146076 and CNS-2125958.

References

[1] S.D. Morris, Corruption, drug trafficking, and violence in mexico, Brown J. World Aff. 18 (2) (2012) 29–43.
[2] S.D. Morris, Drug trafficking, corruption, and violence in mexico: mapping the linkages, Trends Organized Crime 16 (2013) 195–220.
[3] P. Doherty, P. Rudol, A UAV search and rescue scenario with human body detection and geolocalization. Australasian Joint Conference on Artificial Intelligence, Springer, 2007, pp. 1–13.
[4] J. Demant, S.A. Bakken, A. Oksanen, H. Gunnlaugsson, Drug dealing on Facebook, Snapchat and Instagram: a qualitative analysis of novel drug markets in the nordic countries, Drug Alcohol Rev. 38 (4) (2019) 377–385.
[5] B.A. Liang, T.K. Mackey, Prevalence and global health implications of social media in direct-to-consumer drug advertising, J. Med. Internet Res. 13 (3) (2011) e1775.
[6] L. Moyle, A. Childs, R. Coomber, M.J. Barratt, Drugsforsale: an exploration of the use of social media and encrypted messaging apps to supply and access drugs, Int. J. Drug Policy 63 (2019) 101–110.
[7] C. Hu, M. Yin, B. Liu, X. Li, Y. Ye, Detection of illicit drug trafficking events on Instagram: a deep multimodal multilabel learning approach. Proceedings of the 30th ACM International Conference on Information & Knowledge Management, 2021, pp. 3838–3846.
[8] C. Hu, M. Yin, B. Liu, X. Li, Y. Ye, Identifying illicit drug dealers on Instagram with large-scale multimodal data fusion, ACM Trans. Intell. Syst.Technol. (TIST) 12 (5) (2021) 1–23.
[9] J. Li, Q. Xu, N. Shah, T.K. Mackey, A machine learning approach for the detection and characterization of illicit drug dealers on Instagram: model evaluation study, J. Med. Internet Res. 21 (6) (2019) e13803.
[10] X. Yang, J. Luo, Tracking illicit drug dealing and abuse on Instagram using multimodal analysis, ACM Trans. Intell. Syst.Technol. (TIST) 8 (4) (2017) 1–15.
[11] F. Zhao, P. Skums, A. Zelikovsky, E.L. Sevigny, M.H. Swahn, S.M. Strasser, Y. Huang, Y. Wu, Computational approaches to detect illicit drug ads and find vendor communities within social media platforms, IEEE/ACM Trans. Comput. Biol.Bioinf. 19 (1) (2022) 180–191.
[12] C. Hu, B. Liu, Y. Ye, X. Li, Fine-grained classification of drug trafficking based on Instagram hashtags, Decis. Support Syst. 165 (2023) 113896.
[13] T. Zhang, A.M. Schoene, S. Ananiadou, Automatic identification of suicide notes with a transformer-based deep learning model, Internet Interventions 25 (2021) 100422.
[14] T. Brown, B. Mann, N. Ryder, M. Subbiah, J.D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Adv. Neural Inf. Process. Syst. 33 (2020) 1877–1901.
[15] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al., Llama: open and efficient foundation language models, arXiv preprint arXiv:2302.13971 (2023).
[16] I. Drori, S. Zhang, R. Shuttleworth, L. Tang, A. Lu, E. Ke, K. Liu, L. Chen, S. Tran, N. Cheng, et al., A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level, Proc. Natl. Acad. Sci. 119 (32) (2022).e2123433119
[17] S. Liu, A.P. Wright, B.L. Patterson, J.P. Wanderer, R.W. Turer, S.D. Nelson, A. B. McCoy, D.F. Sittig, A. Wright, Using AI-generated suggestions from ChatGPT to optimize clinical decision support, J. Am. Med. Inf.Assoc. 30 (7) (2023) 1237–1245.
[18] T. Zhang, K. Yang, S. Ji, S. Ananiadou, Emotion fusion for mental illness detection from social media: a survey, Inf. Fusion 92 (2023) 231–246.
[19] T. Zhang, J. Leng, Y. Liu, Deep learning for drug–drug interaction extraction from the literature: a review, Briefings Bioinf. 21 (5) (2020) 1609–1627.
[20] T. Zhang, A.M. Schoene, S. Ji, S. Ananiadou, Natural language processing applied to mental illness detection: a narrative review, NPJ Digit. Med. 5 (1) (2022) 46.
[21] S.S. Biswas, Role of Chat GPT in public health, Ann. Biomed. Eng. 51 (5) (2023) 868–869.
[22] K. Yang, S. Ji, T. Zhang, Q. Xie, Z. Kuang, S. Ananiadou, Towards interpretable mental health analysis with large language models. The 2023 Conference on Empirical Methods in Natural Language Processing, 2023.
[23] M. Firat, How Chat GPT can transform autodidactic experiences and open education. Department of Distance Education, Open Education Faculty, Anadolu Unive, 2023.
[24] S.S. Biswas, Potential use of Chat GPT in global warming, Ann. Biomed. Eng. 51 (6) (2023) 1126–1127.
[25] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E.H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, W. Fedus, Emergent abilities of large language models, Trans. Mach. Learn. Res. (2022).Survey Certification
[26] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, G. Neubig, Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing, ACM Comput. Surv. 55 (9) (2023).
[27] T. Shin, Y. Razeghi, R.L. Logan IV, E. Wallace, S. Singh, Autoprompt: eliciting knowledge from language models with automatically generated prompts. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 4222–4235.
[28] P. Goel, L. Chen, On the robustness of monte carlo dropout trained with noisy labels. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2219–2228.

[29] S. Samtani, H. Zhu, B. Padmanabhan, Y. Chai, H. Chen, J.F. Nunamaker Jr, Deep learning for information systems research, J. Manage. Inf. Syst. 40 (1) (2023) 271–301.
[30] Y. Qian, Y. Zhang, Y. Ye, C. Zhang, et al., Distilling meta knowledge on heterogeneous graph for illicit drug trafficker detection on social media, Adv. Neural Inf. Process. Syst. 34 (2021) 26911–26923.
[31] W.X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al., A survey of large language models, arXiv preprint arXiv: 2303.18223 (2023).
[32] J. Wei, X. Wang, D. Schuurmans, M. Bosma, E. Chi, Q. Le, D. Zhou, Chain of thought prompting elicits reasoning in large language models, arXiv preprint arXiv: 2201.11903 (2022).
[33] T. Wu, E. Jiang, A. Donsbach, J. Gray, A. Molina, M. Terry, C.J. Cai, PromptChainer: chaining large language model prompts through visual programming. CHI Conference on Human Factors in Computing Systems Extended Abstracts, 2022, pp. 1–10.
[34] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv: 1810.04805 (2018).
[35] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R.R. Salakhutdinov, Q.V. Le, XLNet: generalized autoregressive pretraining for language understanding, Adv. Neural Inf. Process. Syst. 32 (2019).
[36] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, Albert: a lite bert for self-supervised learning of language representations, arXiv preprint arXiv: 1909.11942 (2019).
[37] V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, arXiv preprint arXiv:1910.01108 (2019).
[38] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: a robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692 (2019).
[39] M. Gupta, C. Akiri, K. Aryal, E. Parker, L. Praharaj, From ChatGPT to ThreatGPT: impact of generative ai in cybersecurity and privacy, IEEE Access (2023).
[40] X. Wu, R. Duan, J. Ni, Unveiling security, privacy, and ethical concerns of ChatGPT, J. Inf. Intell. (2023).
[41] B. Liu, B. Xiao, X. Jiang, S. Cen, X. He, W. Dou, et al., Adversarial attacks on large language model-based system and mitigating strategies: a case study on ChatGPT, Secur. Commun. Netw. 2023 (2023).
[42] A. Acar, H. Aksu, A.S. Uluagac, M. Conti, A survey on homomorphic encryption schemes: theory and implementation, ACM Comput. Surv. (Csur) 51 (4) (2018) 1–35.
[43] X. Yi, R. Paulet, E. Bertino, X. Yi, R. Paulet, E. Bertino, Homomorphic Encryption, Springer, 2014.
[44] M. Naehrig, K. Lauter, V. Vaikuntanathan, Can homomorphic encryption be practical?. Proceedings of the 3rd ACM Workshop on Cloud Computing Security Workshop, 2011, pp. 113–124.
[45] P. Kairouz, H.B. McMahan, B. Avent, A. Bellet, M. Bennis, A.N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, et al., Advances and open problems in federated learning, Found. Trends® Mach. Learn. 14 (1–2) (2021) 1–210.
[46] Q. Yang, Y. Liu, T. Chen, Y. Tong, Federated machine learning: concept and applications, ACM Trans. Intell. Syst.Technol. (TIST) 10 (2) (2019) 1–19.
[47] J. Thrasher, A. Devkota, P. Siwakotai, R. Chivukula, P. Poudel, C. Hu, B. Bhattarai, P. Gyawali, Multimodal federated learning in healthcare: a review, arXiv preprint arXiv:2310.09650 (2023).
[48] T. Che, J. Liu, Y. Zhou, J. Ren, J. Zhou, V. Sheng, H. Dai, D. Dou, Federated learning of large language models with parameter-efficient prompt tuning and adaptive optimization. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023, pp. 7871–7888.Singapore
[49] M. Abadi, A. Chu, I. Goodfellow, H.B. McMahan, I. Mironov, K. Talwar, L. Zhang, Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 308–318.
[50] C. Dwork, Differential privacy: a survey of results. International Conference on Theory and Applications of Models of Computation, Springer, 2008, pp. 1–19.
[51] C. Dwork, Differential privacy. International Colloquium on Automata, Languages, and Programming, Springer, 2006, pp. 1–12.

Dr. Chuanbo Hu received the PhD degree in geographic information science from Wuhan University, Wuhan, China, in 2017. He was a Post-Doctoral Research Fellow with The Chinese University of Hong Kong, Hong Kong, from 2017 to 2018. He is currently a Post-Doctoral Scholar with West Virginia University, Morgantown, WV, USA. His research interests include geospatial artificial intelligence (GeoAI), health geography, and data mining.

Dr. Bin Liu received the PhD degree from Rutgers University. He is an Assistant Professor in the Department of Management Information Systems at West Virginia University. He is interested in data mining and machine learning, and their intersections with healthcare, business analytics, recommender systems, and privacy/security. He has published in premier journals such as the IEEE Transactions on Knowledge and Data Engineering, the ACM Transactions on Knowledge Discovery from Data, the ACM Transactions on Privacy and Security, the ACM Transactions on Intelligent Systems and Technology; and top conferences such as KDD, AAAI, WSDM, and USENIX Security. He currently serves on the editorial board of the Journal of Business Analytics, and has served as a reviewer for many journals, including IEEE TKDE. He has served regularly on program committees of conferences, including KDD, AAAI, CIKM, SDM, and RecSys. He is a member of the ACM and IEEE.

Dr. Xin Li received the BS degree (Hons.) in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 1996, and the PhD degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2000. He was a Member of the Technical Staff with Sharp Laboratories of America, Camas, WA, USA, from August 2000 to December 2002. Since January 2003, he has been a Faculty Member of the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA. His research interests include image/video coding and processing. Dr. Li received various best paper awards at image processing and data mining conferences. He was elevated to an Fellow of IEEE in 2017.

Dr. Yanfang Ye is currently the T. and D. Schroeder Associate Professor with the Department of Computer and Data Sciences, Case Western Reserve University (CWRU), Cleveland, OH, USA. Her research interests include cybersecurity, data mining, machine learning, and health intelligence. Her proposed techniques by advancing AI and data-driven innovations for malware detection have been incorporated into popular commercial cybersecurity products that protect millions of users worldwide. She has expanded her research on health intelligence with focus on combating opioid epidemic and COVID-19 crisis. She was the recipient of the the CSE Research Award in 2019-2020 at CWRU, the NSF Career Award in 2019, the MetroLab Innovation of the Month in May 2020, the IJCAI 2019 Early Career Spotlight, the AICS 2019 Challenge Problem Winner, the SIGKDD 2017 Best Paper Award and Best Student Paper Award (Applied Data Science Track), the IEEE EISIC 2017 Best Paper Award, and the New Researcher of the Year Award in 2016-2017 at WVU.

Mr. Minglei Yin is currently a PhD student in the Lane Department of Computer Science and Electrical Engineering at West Virginia University. His research interests include data mining, multimodal AI, recommendation systems, and cyber-security.