



## OPEN ACCESS

## \*CORRESPONDENCE

Huixiao Hong,  
✉ huixiao.hong@fda.hhs.gov

RECEIVED 28 February 2025

ACCEPTED 09 May 2025

PUBLISHED 27 May 2025

## CITATION

Dong F, Guo W, Liu J, Patterson TA and Hong H (2025) Pharmacovigilance in the digital age: gaining insight from social media data. *Exp. Biol. Med.* 250:10555. doi: 10.3389/ebm.2025.10555

## COPYRIGHT

© 2025 Dong, Guo, Liu, Patterson and Hong. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Pharmacovigilance in the digital age: gaining insight from social media data

Fan Dong, Wenjing Guo, Jie Liu, Tucker A. Patterson and Huixiao Hong\*

National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, United States

## Abstract

Pharmacovigilance is essential for protecting patient health by monitoring and managing medication-related risks. Traditional methods like spontaneous reporting systems and clinical trials are valuable for identifying adverse drug events, but face delays in data access. Social media platforms, with their real-time data, offer a novel avenue for pharmacovigilance by providing a wealth of user-generated content on medication usage, adverse drug events, and public sentiment. However, the unstructured nature of social media content presents challenges in data analysis, including variability and potential biases. Advanced techniques like natural language processing and machine learning are increasingly being employed to extract meaningful information from social media data, aiding in early adverse drug event detection and real-time medication safety monitoring. Ensuring data reliability and addressing ethical considerations are crucial in this context. This review examines the existing literature on the use of social media data for drug safety analysis, highlighting the platforms involved, methodologies applied, and research questions explored. It also discusses the challenges, limitations, and future directions of this emerging field, emphasizing the need for ethical principles, transparency, and interdisciplinary collaboration to maximize the potential of social media in enhancing pharmacovigilance efforts.

## KEYWORDS

drug safety, artificial intelligence, machine learning, natural language processing, social media, post-market surveillance

## Impact statement

Pharmacovigilance explores the transformative potential of social media in enhancing drug safety monitoring. Traditional methods, while foundational, are limited by delayed data collection and analysis, creating gaps in timely adverse drug event detection. This review advances the field by examining the latest methodologies, including natural language processing and machine learning that enable the extraction of meaningful information from unstructured social media data. These advanced techniques provide

tools to overcome challenges such as data variability and bias, making social media a viable complement to established pharmacovigilance practices. The insight from this review impacts the field by demonstrating how social media can fill critical gaps in real-time adverse drug event detection and provide a broader understanding of public sentiment and patient experiences. By emphasizing the need for interdisciplinary collaboration, ethical principles, and transparency, this review lays the foundation for a more agile, inclusive, and effective pharmacovigilance system, ultimately enhancing public health outcomes.

## Introduction

Pharmacovigilance is an essential component of healthcare, focused on safeguarding patients by vigilantly monitoring and managing the risks associated with medications. Traditional pharmacovigilance methods, such as spontaneous reporting systems and clinical trial data analysis, have played pivotal roles in detecting and preventing adverse drug events. However, these methods have limitations, including delayed access to critical information due to the lag between data collection and availability. This lag impedes the timely monitoring of medication safety, thereby posing challenges to real-time surveillance.

The advent of social media has introduced a new paradigm in pharmacovigilance, providing a platform for individuals to share their experiences and opinions about medications. Platforms like X (former name Twitter) and Facebook serve as repositories of user-generated content, which offer discernment into medication usage patterns, adverse drug events, and public sentiment surrounding pharmaceuticals. Such a reservoir of data can complement traditional data sources by providing near real-time information on medication safety concerns.

Despite its potential, relying solely on social media data for pharmacovigilance has challenges. A significant issue arises from the unstructured nature of social media content, which can introduce variability and noise into the data, thereby complicating the identification of drug safety signals. Furthermore, social media users may not represent the entire population, leading to potential biases in demographic representation and geographic coverage.

To overcome these challenges, advanced analytical techniques such as natural language processing and machine learning are being employed to analyze social media data more effectively. These methods enable the identification of patterns and trends, thereby assisting in the early detection of potential safety concerns and facilitating real-time monitoring of medication safety. It is crucial to ensure the reliability of social media data by validating findings with conventional data sources and maintaining strict privacy and ethical standards in data usage.

The use of social media data in pharmacovigilance represents a promising strategy for enhancing medication safety monitoring. By

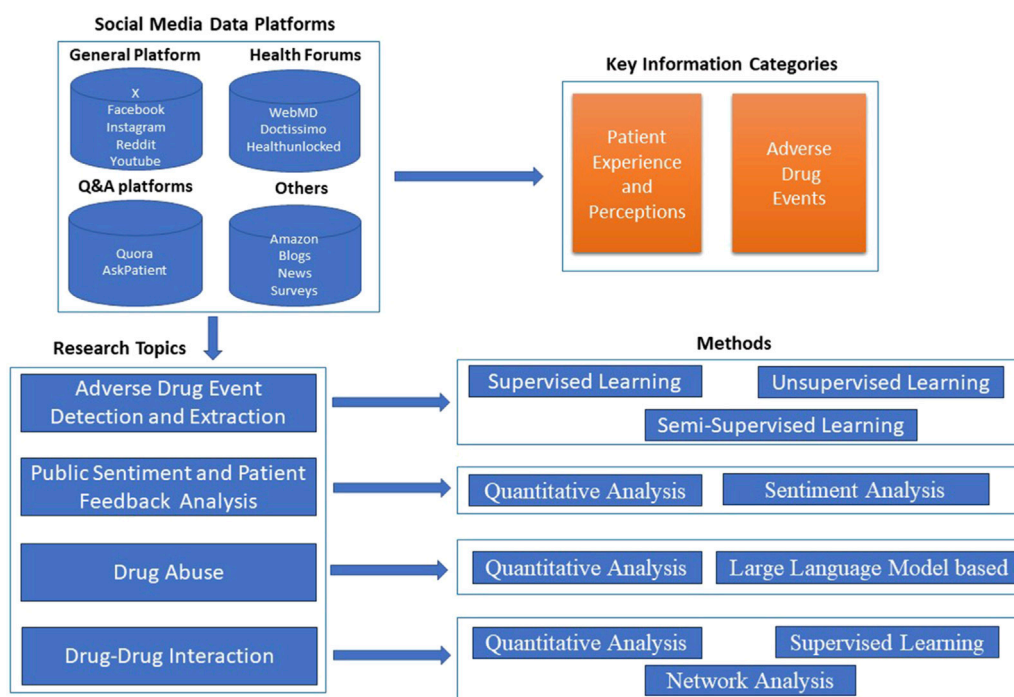
addressing the accompanying challenges and employing rigorous methodologies, social media data can serve as a valuable complement to conventional surveillance data. In the future, prioritizing ethical principles, promoting transparency, and fostering interdisciplinary collaboration will prove indispensable in fully harnessing the potential of social media data in safeguarding public health.

Social media platforms have emerged as significant sources of real-time data, where users freely share their experiences and opinions on various topics, including healthcare and medication usage. These platforms enable the collection of patient-reported outcomes, capture discussions on medication usage and safety, and identify the dissemination of misinformation related to drugs and vaccines. Leveraging the capabilities of these platforms empowers researchers and healthcare professionals to delve into the experiences, attitudes, and concerns of the public. This wealth of user-generated content provides a unique perspective on drug safety and efficacy, offering insight that complements traditional pharmacovigilance methods and enhancing our understanding of medication impact in real-world settings.

The use of these platforms addresses a multitude of research questions, spanning from understanding the reasons behind medication changes to evaluating public sentiments towards pharmacovigilance endeavors, and even combating the spread of health-related misinformation. This review examines the current literature on the use of social media data for drug safety analysis, with a focus on the platforms utilized, the key information extracted, the research questions addressed, and the methodologies applied. Additionally, the review discusses the challenges, limitations, and future directions of utilizing social media data for drug safety analysis.

Figure 1 provides an overview of the process for analyzing social media data to support pharmacovigilance. The process starts with data collection from various sources, including general social media platforms (e.g., X, Facebook, Instagram, Reddit, and YouTube), health-specific forums (such as WebMD and other medical forums), Q&A sites (such as Quora and Ask a Patient), and other online content (like Amazon reviews, blogs, news sites, and surveys). The information extracted from these sources is categorized primarily into Patient Experience and Perceptions and Adverse Drug Events. The extracted data undergoes further analysis in four areas:

1. Machine Learning methods for Adverse Drug Event Detection: This involves identifying and categorizing adverse drug events through methods such as supervised, semi-supervised, and unsupervised learning.
2. Public Sentiment and Patient Feedback Analysis: Techniques like quantitative analysis and sentiment analysis are applied to understand public opinion and patient feedback on drug safety.
3. Drug Abuse Monitoring: Quantitative analysis and advanced language models (large language model-based methods) are employed to monitor drug abuse trends based on user discussions.



**FIGURE 1**  
Overview of drug safety analysis with social media data.

4. Drug-Drug Interaction Monitoring: Network analysis and supervised learning are used to identify and assess potential interactions between different drugs.

Through these stages the process integrates various machine learning and analysis techniques to draw meaningful information that can enhance pharmacovigilance practices by capturing real-world patient experiences, monitoring safety signals, and improving our understanding of drug-related issues.

## Social media platforms utilized for pharmacovigilance

Social media platforms provide diverse online environments for sharing user-generated content which contains data useful for drug safety research. General platforms like X,<sup>1</sup> Facebook,<sup>2</sup> Instagram,<sup>3</sup> Reddit,<sup>4</sup> and YouTube<sup>5</sup> offer broad audience

engagement, while health-specific forums such as WebMD<sup>6</sup> as well as Q&A platforms such as Quora<sup>7</sup> and Ask a Patient<sup>8</sup> focus on medical discussions. Additionally, other platforms such as the e-commerce site Amazon<sup>9</sup> and various blogs have user reviews and discussions on drug products, enriching the data pool for drug safety analysis. Collectively, these social media platforms offer data that can facilitate a comprehensive understanding of patient experiences, medication effects, and healthcare trends.

Social media platforms differ in data scale, user demographics, content modality, and real-time accessibility—factors that shape their utilization in pharmacovigilance. X features high-frequency, short-text updates in real time, making it well-suited for detecting emerging safety signals, particularly among its predominantly young U.S. users. Facebook provides large-scale, globally diverse user demographics and facilitates lifestyle sharing, intergenerational interaction, and marketplace activity. Instagram contains visually rich content—photos and videos—primarily around themes like fashion, beauty, and

1 <https://x.com/>

2 <https://www.facebook.com/>

3 <https://www.instagram.com/>

4 <https://www.reddit.com/>

5 <https://www.youtube.com/>

6 <https://www.webmd.com/>

7 <https://www.quora.com/>

8 <https://askapatient.com>

9 <https://www.amazon.com/>

TABLE 1 Data platform Review.

| Category                      | Platform                                                                                       | Start year | Publication                      |
|-------------------------------|------------------------------------------------------------------------------------------------|------------|----------------------------------|
| General social media Platform | X                                                                                              | 2006       | [1–38]                           |
|                               | Facebook                                                                                       | 2004       | [1, 4, 19, 24–26, 28, 29, 38–47] |
|                               | Instagram                                                                                      | 2010       | [1, 4, 41, 48]                   |
|                               | Reddit                                                                                         | 2005       | [4, 16, 17, 41, 49–55]           |
|                               | YouTube                                                                                        | 2005       | [1, 4, 41, 47, 56]               |
| Online Health Forums          | WebMD online health forum                                                                      | 1998       | [40, 57, 58]                     |
|                               | Health forum in French ( <a href="https://www.doctissimo.fr/">https://www.doctissimo.fr/</a> ) | N/A        | [59]                             |
|                               | Health forum ( <a href="https://www.healthunlocked.com">Healthunlocked.com</a> )               | 2010       | [60]                             |
| Q&A platforms                 | Quora                                                                                          | 2009       | [61]                             |
|                               | Ask a Patient                                                                                  | 2000       | [62, 63]                         |
| Others                        | Amazon                                                                                         | 1995       | [64, 65]                         |
|                               | Blogs and News                                                                                 | N/A        | [1, 24, 29, 43, 66–68]           |
|                               | Online Surveys                                                                                 | N/A        | [69–76]                          |

travel, attracting a younger audience but offering limited textual data. Reddit hosts in-depth, community-driven discussions on topics such as technology and health through anonymous forums, though it lacks real-time immediacy. YouTube delivers both long- and short-form video content across educational and entertainment topics to a global, multi-generational audience, but pharmacovigilance efforts on the platform require resource-intensive video analysis. Health forums focus on structured, medically-oriented discussions within smaller, specialized communities. E-commerce platforms like Amazon provide valuable insight into over-the-counter medication safety through consumer reviews. Meanwhile, blogs and online surveys, though useful for exploring historical public opinion, are declining in relevance compared to more dynamic and interactive social media channels. These variations highlight the importance of tailoring pharmacovigilance strategies to each platform—leveraging the strengths of real-time surveillance on X or visual content on Instagram, while accounting for constraints in scalability, content richness, or demographic coverage.

Social media has evolved significantly since its early days in the 1970s, when online communication was limited to basic chat rooms and forums. In the 2000s, platforms such as Facebook, X (formerly Twitter), and Instagram emerged, enabling users to share experiences in real time. Over time, social media content has expanded from text-based posts to multimedia formats, including images, videos, and live interactions. Access for researchers has also changed—while many platforms initially offered open APIs for data analysis, recent restrictions have limited the availability of user-generated content, posing new

challenges for social media-based research. Researchers are employing adaptive strategies such as web scraping, synthetic data generation, and collaborative research agreements with platforms to obtain unrestricted access to social media data for pharmacovigilance research, thereby mitigating the challenges posed by recent API limitations. Table 1 summarizes popular social media platforms and drug safety studies based on data extracted from these platforms.

X has emerged as a pivotal resource in pharmacovigilance, providing insight into public perceptions and experiences regarding medications. Researchers have effectively leveraged this platform to extract valuable information, including adverse drug events, user sentiments, and discussions on drug safety and efficacy. Through the application of analytical methods such as natural language processing, machine learning, and sentiment analysis, they have managed the challenges posed by extensive data volumes and noise in the user-generated content, thereby enhancing drug safety monitoring and public health strategies. For instance, Sule et al. [1] utilized X to combat the dissemination of COVID-19 misinformation from physicians, thereby contributing to the improvement of public health communications. Pathak and Catalan-Matamoros [2] explored X's potential as an early warning system for drug safety signals, with the aim of strengthening pharmacovigilance practices. Hua et al. [8] investigated public perceptions of COVID-19-related medications using X, offering valuable insight for pharmacovigilance and policy-making decisions. Sharma et al. [11] conducted a scoping review utilizing sentiment analysis to comprehend patient experiences with pharmacotherapy. Rezaei et al. [12] applied deep learning techniques to detect adverse drug

events on X, thereby enhancing the efficiency of safety concern identification. Lardon et al. [23] evaluated X as a supplementary data source for pharmacovigilance and compared it with traditional monitoring systems. Khademi et al. [32] developed a model by coupling topic modeling and transformer-based learning for the early detection of vaccine safety signals in X posts, and the detected adverse events were generally aligned with those reported in a spontaneous vaccine safety surveillance system, showcasing the potential of social media data for early detection of emerging vaccine safety issues. These examples underscore the versatility and value of X posts in complementing traditional pharmacovigilance methods, particularly during public health crises like the COVID-19 pandemic, and in understanding off-label medication use.

Facebook offers important user-generated data for pharmacovigilance, covering discussions, comments, and posts on medications, adverse drug reactions, and drug safety feelings. This data helps improve post-marketing drug safety surveillance. For example, Pierce et al. [26] explored the possibility of detecting drug adverse events in Facebook and X earlier than their reports in the FDA Adverse Event Reporting System (FAERS), demonstrating that social media platforms such as Facebook and X can be used for early detection of certain adverse events. Powell et al. [28] investigated the usefulness of Facebook in post-marketing drug safety surveillance by examining its effectiveness in capturing patient experiences and concerns on medications, indicating that Facebook provides data for the detection of early warning signs of potential drug safety. Coloma et al. [29] used Facebook in their evaluation of social media networks' contributions to drug safety surveillance, demonstrating the usefulness of the patient-reported information in pharmacovigilance. These studies collectively highlight the significant role of Facebook in pharmacovigilance, supporting better regulatory decisions and patient care.

Instagram, a popular social media platform owned by Facebook, Meta, allows users to share photos, videos, and engage with others through likes, comments, and direct messaging. For instance, Li et al. [48] focused on developing a machine learning approach for identifying and profiling illicit drug dealers on Instagram by analyzing posts and comments for hashtags and language patterns associated with illegal drug dealings. Through an examination of posts and comments, this study identified key information like hashtags and language patterns indicative of illegal drug activities. This research addresses concerns about drug sales on social media and suggests a mechanism that may help counteract such occurrences.

Reddit's diverse communities offer a wealth of user-generated health data, making it a valuable platform for pharmacovigilance and public health research. For instance, Godinez et al. [51] analyzed Reddit discussions to gain an understanding of the experiences and concerns of individuals transitioning between HIV pre-exposure prophylaxis medications,

thereby informing healthcare policies. Guo et al. [50] utilized the Reddit platform for real-time tracking and analysis of COVID-19 symptoms, facilitating early detection and response to the pandemic. Furthermore, Szczypka et al. [55] explored Reddit discussions on Delta-8-tetrahydrocannabinol, shedding light on public perceptions and potential health risks associated with its use. Lastly, Sharp et al. [53] conducted a comprehensive analysis of 11 years of Reddit posts related to dietary supplements among military personnel, providing valuable data on safety, efficacy, and usage patterns within this specific population. These examples collectively underscore the significant role of Reddit in enhancing our understanding of drug safety and informing public health strategies.

YouTube, being a widely utilized video-sharing platform, provides a rich source of information on medication usage and health-related topics. For example, Hansen et al. [56] assessed the accuracy of safety information conveyed in YouTube videos about medication usage during pregnancy. Their aim was to ensure that pregnant women have access to reliable and evidence-based guidance to facilitate informed decision-making for their well-being.

Research on drug safety is increasingly leveraging patient-generated data from health-specific forums and online consultation platforms, such as WebMD. These studies utilize quantitative analysis and natural language processing techniques to extract information directly from patient reviews and discussions, providing a unique perspective on medication adherence, adverse drug reactions, and patient satisfaction. By comparing patient-reported reasons for medication changes against formal adverse event reporting systems and assessing the cognitive levels and intervention preferences of patients with hypertriglyceridemia, these research efforts enhance our understanding of real-world medication experiences and safety concerns, ultimately contributing to the improvement of pharmacovigilance practices.

In addition to WebMD, numerous health forums cater to different languages and focus on various topics. For instance, Abdellaoui et al. [77] used a topic modeling approach to identify instances of noncompliance to drug treatment in patient forum posts. Similarly, Karapetiantz et al. [78] found a discrepancy between personal experiences and negative opinions with the human papillomavirus vaccine in web forums. They also found that descriptions of adverse drug reactions are less detailed in forums compared to the French Pharmacovigilance Database,<sup>10</sup> however, forums provide more unexpected reactions [79]. These examples illustrate the potential of health-specific forums and online consultation platforms in enhancing our understanding of drug safety and improving pharmacovigilance practices. They emphasize the importance of considering patient experiences and

10 <https://ansm.sante.fr/page/enhanced-surveillance-of-medicines>



perspectives in different languages and contexts. Such an approach not only enriches available data for analysis but also ensures a more comprehensive and inclusive understanding of drug safety.

Q&A platforms like Quora and patient feedback websites such as<sup>8</sup> play a crucial role in drug safety research by providing a space for patients to share their experiences, ask questions about medications, and discuss their treatment preferences. For instance, Xu et al. [61] analyzed the discussions about COVID-19 vaccine clinical trials on Quora, providing insight into public sentiment and common questions about the trials, which can inform future communication strategies and patient education efforts. Similarly, Song et al. [62] conducted a social media listening infosurveillance study to evaluate the needs and experiences of patients with hypertriglyceridemia. They identified common concerns and preferences that can be utilized to guide the development of patient-centered care strategies for this population. Moreover, Duh et al. [63] explored whether social media data could aid in the early detection of drug-related adverse events. They found that patient discussions on platforms like<sup>8</sup> can provide early warning signs of potential adverse events, thereby enhancing the timeliness and effectiveness of pharmacovigilance efforts. These examples underscore the potential of Q&A platforms and patient feedback websites in drug safety research, emphasizing the importance of incorporating patient voices in the study to improve drug safety practices.

E-commerce platforms like Amazon serve as valuable repositories of consumer-generated data for drug safety research. By examining user reviews, ratings, and Q&A sections, researchers can gain insight into the real-world performance and safety of over-the-counter medications and health products. For instance, Adams et al. [64] used an automated method to uncover safety and efficacy issues related to joint and muscle pain treatments from Amazon reviews. Similarly, Gartland et al. [65] focused on creating crowdsourced training datasets for pharmacovigilance intelligent automation, illustrating the potential of platforms like Amazon Turk to provide valuable data for training machine learning models in pharmacovigilance studies. These examples highlight the significance of e-commerce platforms in drug safety research, enhancing our understanding of medication safety and efficacy in real-world contexts.

In the past, news articles and blog posts were frequently used as primary data sources for drug safety research. However, their utilization has witnessed a decline in recent years. Despite the shift from blogs to real time social media platforms, traditional news and blog data can still yield valuable insights for drug safety research. For example, de Vries et al. [66] analyzed a series of healthcare provider communications in newspaper articles spanning from 2001 to 2015 in the Netherlands, demonstrating the value of such media in drug safety research. Similarly, Matsuda et al. [68] analyzed patient

narratives sourced from disease-specific blogs, a component of the TOBYO database,<sup>11</sup> revealing valuable real-world medication experiences and emphasizing the potential of such platforms in bolstering drug safety research and pharmacovigilance efforts. These examples emphasize the significant role of news and blog data in understanding public perceptions and experiences related to drug safety.

Online surveys are valuable tools in drug safety research, providing a platform for collecting large-scale patient-generated data. For instance, Grundmann et al. [70] used an online survey to investigate patterns of Kratom use and its health impact in the US. Similarly, Wysota et al. [75] and Nguyen et al. [76] utilized online surveys to understand consumers' knowledge, perceptions, and usage of cannabidiol products. The widespread adoption of online surveys signifies their pivotal role in advancing drug safety research by providing a robust platform for data collection and analysis.

## Information extracted from social media platforms

In the realm of pharmacovigilance, social media has emerged as a pivotal source of information, offering unprecedented access to a wealth of data regarding patient experiences, perceptions, and discussions related to medication use and safety. This section delves into the various ways social media data can be harnessed to enhance our understanding of patient experiences, detect and analyze adverse drug events, assess vaccine efficacy and safety, and combat the spread of misinformation. It highlights the importance of leveraging social media platforms to inform healthcare practices, improve pharmacovigilance efforts, and address the challenges posed by misinformation, ultimately aiming to enhance public health and safety. Table 2 shows the categories of key information extracted from social media.

## Patient experiences and perceptions

Social media platforms have become invaluable for capturing patient experiences and perceptions regarding medication use, benefits, and safety. Golder et al. [40] and Micale et al. [57] examined online discussions and personal narratives, uncovering reasons behind medication changes, such as adverse events experienced by patients on statin therapies and other treatments, along with factors influencing their satisfaction and concerns. Yamaguchi et al. [42] explored the impact of social media information on self-medication choices, focusing on fenbendazole, and highlighting how such information can lead to

11 <https://www.toby.jp/>

TABLE 2 Extracted key information categories from social media data.

| Key information Category           | Key information                                                       | Reference                        |
|------------------------------------|-----------------------------------------------------------------------|----------------------------------|
| Patient Experience and Perceptions | patient experience of adverse events for Discontinuing Statin Therapy | [40]                             |
|                                    | Patient experience of adverse events for medication change            | [57]                             |
|                                    | Patient experience of fenbendazole safety and efficacy                | [42]                             |
|                                    | Negative opinions and personal experience with HPV vaccine            | [78]                             |
|                                    | Patient reported Symptoms                                             | [50]                             |
|                                    | Patient Experiences with Dabigatran                                   | [80]                             |
|                                    | consumer perceptions and attitudes on cannabis products               | [76]                             |
| Adverse Drug Events                | Adverse Drug Events in X posts                                        | [2, 3, 5, 10, 14, 22, 34, 81–83] |
|                                    | Adverse Drug Events in Forums                                         | [59, 72, 84, 85]                 |

adverse events, emphasizing the need for effective communication strategies and vigilant monitoring. In a similar vein, Karapetiantz et al. [78] analyzed HPV vaccine discussions in online forums, identifying a gap between personal experiences and negative public opinions, which contributes to vaccine hesitancy and signals the need for targeted interventions. Guo et al. [50] further illustrated how social media enhances traditional pharmacovigilance by offering a deeper understanding of real-world drug safety through patient-reported symptoms. Vaughan et al. [80] documented patient experiences with dabigatran, offering insight into patient satisfaction, concerns, and adverse effects that stress the importance of understanding patient perspectives. Additionally, Nguyen et al. [76] highlighted consumer perceptions and attitudes toward cannabis products, demonstrating the influence of social media in shaping public opinion and highlighting concerns related to the use and safety of such products. Collectively, these studies showcase social media’s role in strengthening pharmacovigilance by amplifying patient voices and informing healthcare practices that align with patient needs and real-world experiences.

Adverse drug events:

Social media platforms, especially X, are increasingly being used to detect adverse drug events. Pathak et al. [2] and Litvinova et al. [3] demonstrated the potential of X posts as early indicators of safety signals, with Pathak focusing on initial detection and Litvinova refining methods for accurate adverse drug event identification. Yu et al. [5] developed a natural language processing model to analyze X posts for adverse drug events, while Magge et al. [10] used deep learning to extract adverse drug events from the platform. The Web-RADR project, as reported by Gattepaille et al. [14], applied machine learning to identify adverse drug events on X, establishing a benchmark dataset for

adverse event recognition. Masino et al. [22] enhanced automated adverse drug event detection using convolutional neural networks, and Fisher et al. [34] introduced a scalable machine learning framework for identifying drug-related harms on social media. Further advancements in ADE detection on social platforms were achieved by Botsis et al. [81], Dong et al. [82], and Zhang et al. [83] through the application of text mining and BERT-based language models, enhancing post-marketing surveillance and pharmacovigilance practices.

Web forums also serve as valuable source for monitoring adverse drug events. Roche et al. [59] extracted adverse drug events from Doctissimo, a medical forum where users report side effects and allergic reactions. Bulcock et al. [72] highlighted HealthUnlocked, a health discussion forum, as a platform for detecting emerging side effects. Audeh et al. [84] explored French web forums where users share serious adverse drug events, and Karapetiantz et al. [85] analyzed 23 health forums using the V4M Scraper tool to identify adverse drug events, drug interactions, and safety concerns. Together, these studies showcased the potential of social media and web forums as real-world sources of adverse drug event data, enriching drug safety research and enabling faster responses to public safety concerns.

Methods for utilizing social media data

The integration of social media data into drug safety monitoring has significantly advanced the detection of adverse drug events, analysis of public sentiment, monitoring of drug abuse, and identification of drug-drug interactions. Social media platforms, with their vast and diverse user-generated content, offer a unique perspective on real-world patient experiences and public perceptions of medications.

In adverse drug events detection, both supervised and unsupervised learning methods, including models like BERT

TABLE 3 Methods for analyzing adverse drug events, public sentiment, drug abuse, and drug-drug interactions using social media data.

| Topic Category                                            | Method Category            | Details                                                                                                                      | Advantage                                                                               | Limitation                                                              | Reference |
|-----------------------------------------------------------|----------------------------|------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|-------------------------------------------------------------------------|-----------|
| Machine Learning methods for Adverse Drug Event Detection | Supervised Learning        | BERT-based model to extract adverse drug events from X posts                                                                 | High accuracy on X posts                                                                | Requires labeled data for effective training                            | [82]      |
|                                                           |                            | natural language processing methods to detect adverse drug events from clinical text                                         | High accuracy in structured clinical environments                                       | Limited generalizability to non-clinical text and informal language     | [86]      |
|                                                           |                            | Pipeline to extract and normalize adverse drug events                                                                        | Detect adverse drug events in X posts                                                   | Limited to X posts data                                                 | [10]      |
|                                                           |                            | adverse drug event detection across platforms                                                                                | High detection accuracy                                                                 | Limited generalizability to diverse social media platforms              | [12]      |
|                                                           | Unsupervised Learning      | Co-clustering for adverse drug event detection                                                                               | Detects adverse drug event signals by clustering related data without predefined labels | Validation limited to COVID-19 related data                             | [87]      |
|                                                           |                            | Transformer-based adverse drug event detection                                                                               | High sensitivity in capturing emerging adverse drug events                              | Limited to specific drugs and topics in the training set                | [49]      |
|                                                           |                            | CNN model for adverse drug event detection                                                                                   | Detects adverse drug events without predefined labels                                   | Limited in capturing complex adverse drug events                        | [22]      |
|                                                           | Semi-supervised Learning   | Word embeddings-based lexical network for adverse drug events detection                                                      | Severity scoring to assess different adverse drug events                                | Data quality, noise, and bias exist in social media data                | [54]      |
| Public Sentiment and Patient Feedback Analysis            | Quantitative Analysis      | Direct patient feedback analysis through quantitative methods                                                                | Provides insight into patient perspectives                                              | Limited by the accuracy of self-reported data                           | [57]      |
|                                                           | Sentiment Analysis         | WC-CNN model for sentiment analysis                                                                                          | Detects safety signals by examining user behavior over time                             | Limited to the Levothyrox case in France                                | [59]      |
|                                                           |                            | Sentiment analysis of public perception of specific drugs for COVID-19 treatment                                             | Captures public perceptions on certain drugs                                            | Limited to specific drugs and COVID-19 context                          | [39]      |
|                                                           |                            | VADER model for sentiment analysis on cannabidiol use for various conditions                                                 | Provides insight into public perception of cannabidiol use                              | Manual labeling needed for classifier training                          | [9]       |
| Drug Abuse                                                | Large Language Model based | GPT-3 for generating a lexicon for drug abuse detection in social media                                                      | Captures slang and misspellings for drug abuse detection                                | Limited to 98 drugs with limited evaluation                             | [88]      |
|                                                           | Quantitative Analysis      | Frequency analysis of opioid misuse, abuse, addiction, overdose, and death                                                   | Real-time tracking and insight into opioid abuse trends                                 | Limited to opioid-related data                                          | [36]      |
|                                                           |                            | Statistical analysis with data quality metrics and topic modeling for drug abuse content                                     | Structured framework for analyzing drug abuse-related content                           | Limited social media platforms and limited generalizability             | [89]      |
|                                                           |                            | NER for adverse drug event identification; PRR for potential adverse drug event detection; topic modeling to identify themes | Comprehensive analysis and real-time monitoring using social media data                 | Limited to methylphenidate-related posts between 2007 and 2016 only     | [90]      |
|                                                           |                            | Quantitative analysis of annotated posts for drug abuse                                                                      | Supervised classification model to detect drug abuse signals in X posts                 | Requires manually annotated 6,400 X posts containing drug abuse signals | [91]      |
| Drug-Drug Interactions                                    | Quantitative Analysis      | Frequency and rates for posts related to Drug-Drug Interactions are analyzed                                                 | Manually reviewed and evaluated by two blinded investigators                            | Limited data and limited manual validation                              | [92]      |
|                                                           | Supervised Learning        | Drug-Drug Interactions detection in social media data                                                                        | Extracts related Drug-Drug Interactions information from social media platforms         | Requires an annotated Drug-Drug Interactions corpus                     | [93]      |

(Continued on following page)



TABLE 3 (Continued) Methods for analyzing adverse drug events, public sentiment, drug abuse, and drug-drug interactions using social media data.

| Topic Category | Method Category  | Details                                                                   | Advantage                                                                                       | Limitation                                                      | Reference |
|----------------|------------------|---------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|-----------|
|                | Network Analysis | Co-occurrence network for users based on potential Drug-Drug Interactions | Detects relationships and patterns for emerging Drug-Drug Interactions through network analysis | Limited to Instagram platform with privacy and ethical concerns | [94]      |

and co-clustering algorithms, are applied to extract and analyze adverse events from social media posts. For public sentiment analysis, cutting-edge sentiment analysis models are being used to assess patient feedback and identify potential safety signals. Meanwhile, drug abuse monitoring employs large language models and quantitative analysis to detect trends in misuse, addiction, and overdose incidents. The identification of drug-drug interactions is enhanced through network analysis and supervised learning techniques.

These methods are reshaping pharmacovigilance, enabling faster identification of safety concerns, a deeper understanding of patient experiences, and more effective monitoring of public health issues. Table 3 summarizes the various methods and their applications in utilizing social media data to advance drug safety studies.

### Machine learning methods for adverse drug event detection

One of the most active areas of drug safety research is the detection and extraction of adverse drug events from social media posts. Various machine learning approaches—supervised, unsupervised, and semi-supervised—are employed, each with unique strengths and limitations.

Supervised learning approaches, such as those utilizing BERT-based models, achieve high accuracy in extracting adverse drug events but are heavily reliant on large, labeled datasets. This requirement makes them resource-intensive and less adaptable to diverse social media platforms. For example, BERT-based models have demonstrated exceptional performance in extracting adverse drug events from X posts, but their effectiveness hinges on the availability of labeled data [82]. Natural language processing techniques have also been used to detect adverse drug events in clinical documents, achieving good results within structured data but facing challenges when applied to informal or unstructured social media text [86]. Magge et al. [10] used deep learning methods in their DeepADEMiner pipeline to extract and normalize adverse drug event mentions on X posts. Despite its success on specific platforms like X, such models often struggle to generalize across the diverse and evolving landscape of social media environments [12].

Unsupervised learning methods, such as co-clustering, allow adverse drug event detection by grouping related data without

requiring predefined labels. These methods have been particularly useful in monitoring adverse drug events associated with COVID-19, though their validation remains largely confined to this context [87]. Transformer-based [49] and CNN-based [22] models enhance adverse drug event detection sensitivity and capture emerging adverse drug events without relying on labeled data. However, these approaches are often constrained by the drugs and topics present in the training set, making them less adaptable to new or evolving drug discussions.

Semi-supervised learning techniques combine the strengths of supervised and unsupervised techniques, using a smaller labeled dataset alongside a larger pool of unlabeled data. This approach strikes a balance between accuracy and flexibility, although it continues to face challenges with scalability and data quality. For instance, word embeddings-based lexical networks have been used to introduce severity scoring for assessing adverse drug events, addressing the inherent issues with data quality, noise, and bias in social media data [54].

In summary, while advancements in supervised, unsupervised, and semi-supervised learning techniques have significantly improved ADE detection from social media, each method has inherent trade-offs. Future research must focus on enhancing generalizability, scalability, and robustness to fully leverage social media data for pharmacovigilance.

### Public sentiment and patient feedback analysis

Social media platforms provide unique patient perspectives and public sentiment regarding medications, complementing traditional pharmacovigilance methods. Quantitative analysis methods and sentiment analysis techniques contribute to understanding patient experiences and detecting drug safety signals, though challenges with data quality and generalization persist.

Quantitative analysis methods, often involving manual review of patient posts, offer valuable information on patient attitudes and experiences. However, the accuracy and reliability of self-reported data remain significant limitations [57]. Despite these constraints, such methods are instrumental in capturing nuanced patient feedback and identifying trends in medication use and side effects.

Sentiment analysis techniques are frequently used to track public perceptions of medications and detect emerging drug

safety signals. Advanced models like WC-CNN have been employed to monitor sentiment trends linked to specific drugs, such as Levothyrox in France. While effective in capturing localized sentiment, these models often struggle to generalize across different drugs or regions [59].

Studies have also employed sentiment analysis to understand public perceptions of drugs for specific conditions. For instance, sentiment analysis models like VADER have been applied to assess public attitudes toward COVID-19 treatments [39] and the use of cannabidiol for various health issues [9]. While VADER provides valuable public sentiment, its reliance on manual labeling for classifier training presents scalability challenges, particularly when dealing with large, diverse datasets.

Together, quantitative and sentiment analysis methods enrich drug safety assessments, offering a more comprehensive understanding of medication safety and public attitudes. However, ongoing efforts to address issues with data quality, scalability, and model generalization are essential for maximizing their potential in pharmacovigilance.

## Drug abuse

The rise of drug abuse, particularly opioid misuse, has driven significant efforts to monitor drug abuse trends using social media data. By analyzing user-generated content, researchers aim to enable early detection of abuse, addiction, overdose, and other drug-related issues.

Large language models like GPT-3 have been used to generate lexicons for drug abuse detection, capturing slang, misspellings, and colloquialisms frequently used in drug-related discussions. While these lexicons offer a valuable starting point, they typically focus on a narrow range of drugs and often lack comprehensive evaluation, limiting their broader applicability [88].

Quantitative analysis methods, such as frequency analysis, provide real-time tracking of opioid misuse, abuse, addiction, overdose, and related deaths. These methods help identify emerging trends and patterns in drug abuse [36]. More structured approaches, incorporating data quality matrices and topic modeling, provide frameworks for analyzing drug abuse content across social media platforms. However, the diversity and variability of these platforms pose challenges to generalizability and cross-platform applicability [89].

Techniques such as named entity recognition combined with proportional reporting ratios enables more targeted analysis by identifying specific drugs and associated adverse events. For instance, studies have focused on methylphenidate-related social media posts from 2007 to 2016, highlighting the potential of such methods for longitudinal drug abuse research [90].

Supervised learning models further enhance drug abuse monitoring. Sarker et al. [91] developed a supervised classification model trained on 6,400 manually annotated X posts

to identify drug abuse-related content. This approach highlights the importance of labeled data in improving the precision and reliability of drug abuse trend analysis on social media.

By combining advanced natural language processing techniques with structured analytical methods, researchers are gaining a deeper understanding of drug abuse trends. However, challenges related to data quality, platform diversity, and scalability remain areas for future improvement.

## Drug-drug interaction

Monitoring drug-drug interactions using social media data is an emerging trend in drug safety evaluation. Social media provides the advantage of real-time, user-generated data, allowing for quick identification of potential drug-drug interactions across diverse populations. However, this approach faces challenges, including the need for context-rich analysis, issues with data quality, and privacy concerns on certain platforms. Traditionally, drug-drug interactions detection relied on structured data, but social media offers a dynamic and expansive source of information.

Quantitative approaches analyze the frequency and prevalence of drug-drug interactions related posts, which are then manually reviewed by blinded investigators. While informative, these methods are often limited by data availability and require labor-intensive validation processes [92]. Supervised learning models address these limitations by leveraging annotated drug-drug interactions corpora to extract relevant interaction information from user posts, enhancing detection efficiency [93].

Network analysis has also been applied to study potential drug-drug interactions by constructing co-occurrence networks based on user mentions of interacting drugs. This method facilitates the identification of emerging drug-drug interaction patterns and relationships. However, its application is sometimes limited by platform-specific privacy and ethical concerns, especially on platforms like on Instagram [94].

In conclusion, leveraging social media for monitoring drug-drug interactions presents a promising complement to traditional methods, offering real-time information and access to diverse patient populations. While challenges such as data quality, contextual interpretation, and privacy concerns remain, advancements in quantitative analysis, supervised learning models, and network analysis demonstrate the potential of this approach. With continued refinement and ethical considerations, social media could become a vital tool in enhancing drug safety and public health.

## Discussion

Integrating social media data into drug safety research offers valuable insight, but presents several challenges that must be addressed to ensure reliable and actionable findings.

## Data reliability and quality

Social media data is often unstructured and informal, complicating the identification of credible adverse drug event reports amidst speculative or inaccurate posts. The variability and frequent lack of contextual information in social media content further hinder interpretation, reducing the reliability of findings. Addressing these issues requires rigorous screening, validation, and contextual analysis. Moreover, the lack of demographic and medical context from social media users makes it challenging to generalize findings across populations. Comparing and pooling data across social media platforms is challenging due to differences in user demographics, content formats, and engagement patterns. Privacy regulations further restrict access to user profiles, making it difficult to analyze demographic information comprehensively. Additionally, a key limitation of current studies is the inability to fully verify whether social media users are real individuals or automated accounts. While many platforms have implemented stricter user verification measures, restricted access to demographic and account data remains a significant challenge for researchers seeking to ensure the authenticity and representativeness of social media-derived insights.

Social media platforms often restrict access to demographic and account information, posing challenges in verifying user identities. AI tools can help mitigate this by detecting bot-like behavior, such as rapid posting patterns and generic language. However, the rise of advanced language models introduces new concerns, as they can generate large volumes of content that may be difficult to distinguish from genuine user experiences. To enhance data reliability, preprocessing need to include methods for filtering posts from suspicious or abnormal accounts, ensuring that subsequent analysis focuses on real-world health experiences shared by actual users.

Data reliability is a key concern in social media-based pharmacovigilance. For data preprocessing, inclusion criteria focus on posts that explicitly mention drugs, adverse drug events, or health experiences. To ensure quality, irrelevant content, spam, and non-English posts are removed. Additionally, posts from suspicious or abnormal user accounts are filtered out to enhance reliability and prioritize real user experiences. While confirming individual accounts as real people remains a challenge, these preprocessing steps help improve data integrity and the robustness of findings.

## Ethical and privacy concerns

The use of personal social media data in drug safety research introduces significant ethical and privacy challenges. Researchers must comply with privacy laws and adhere to ethical standards to mitigate risks such as discrimination or stigmatization. Implementing robust privacy protections, anonymization

techniques, and transparent data usage practices is essential to maintain public trust and protecting users.

## Misinformation and bias

Social media is prone to misinformation and biases that can hinder accurate adverse drug event detection. Distinguishing genuine patient experiences from misleading or biased content requires careful fact-checking and verification to ensure data reliability. Misinformation on social media can distort perceptions of drug safety, emphasizing the need for robust methods to counteract its influence. A key research gap is the need for a deeper exploration of strategies to mitigate the impact of misinformation and bias, as these factors influence all reported findings and approaches discussed in this review. Future research could focus on developing and evaluating advanced AI-driven algorithms for detecting and filtering misinformation. Additionally, platform regulations, combined with multi-source verification mechanisms, could play a crucial role in limiting the spread of biased or inaccurate content while still addressing user privacy concerns. A more systematic investigation into these mitigation strategies would strengthen the reliability of social media-based research findings.

## Text mining and modeling challenges

Text mining models face limitations when processing the informal language common on social media, such as slang, abbreviations, and non-standard terms. This issue can reduce the accuracy of automated adverse drug event detection. To address these issues, researchers need to improve text mining models and standardize data validation methods. Combining social media data with traditional pharmacovigilance methods provides a more comprehensive understanding of drug safety.

## Regulatory frameworks

The regulatory landscape for incorporating social media data into pharmacovigilance is still evolving. Clear and standardized guidelines are needed to govern ethical data use and alignment with established pharmacovigilance practices. Such frameworks are crucial for maintaining public trust and facilitating the effective use of social media in drug safety research.

## Conclusion

The integration of social media data into drug safety research offers promising opportunities to enhance pharmacovigilance. Social media's vast, real-time data can complement traditional

methods, enabling early detection of adverse drug events, monitoring public sentiment, and capturing patient experiences on medication risks and benefits. However, unlocking the full potential of social media in drug safety research requires addressing significant methodological, ethical, and regulatory challenges.

This review highlights the need for rigorous methodologies, including advanced data validation, standardized data processing protocols, and specialized text mining models tailored to the unique characteristics of social media content. Ethical considerations, such as privacy protection and transparent practices, are essential to maintaining public trust. Additionally, as social media evolves, developing a strong regulatory framework to govern the use of social media data in pharmacovigilance is important.

Future research should focus on developing real-time monitoring systems that can utilize social media for the early detection of drug safety concerns. Hybrid models that integrate social media data with traditional pharmacovigilance data hold promise for the potential to improve the accuracy, speed, and scope of safety signal detection. Standardizing data collection and analytical methodologies across studies will enhance consistency and reliability, enabling more actionable findings. Addressing cultural and linguistic diversity in social media content will also be essential to adapt these approaches to global pharmacovigilance efforts.

Collaboration among academia, industry, regulatory agencies, and social media platforms will play a key role in advancing these goals. Such partnerships can foster resource sharing, methodological alignment, and innovation in drug safety monitoring. Regulatory agencies can provide guidance on ethical and privacy standards, while social media platforms can facilitate data access and support anonymization efforts. These collaborative efforts will help establish a robust and responsive pharmacovigilance ecosystem, enhancing post-market surveillance and protecting public health worldwide. By integrating social media data with established pharmacovigilance practices, researchers and healthcare providers can improve post-marketing drug safety monitoring,

enhance public health outcomes, and build a collaborative ecosystem that benefits patients and society.

## Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## Author disclaimer

This article reflects the views of the authors and does not necessarily reflect those of the U.S. Food and Drug Administration. Any mention of commercial products is for clarification only and is not intended as approval, endorsement, or recommendation.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was funded by the US Food and Drug Administration (FDA).

## Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. We would like to acknowledge that ChatGPT was used to assist in correcting English grammar errors during the preparation of this manuscript.

## References

1. Sule S, DaCosta MC, DeCou E, Gilson C, Wallace K, Goff SL. Communication of COVID-19 misinformation on social media by physicians in the US. *JAMA Netw Open* (2023) 6(8):e2328928. doi:10.1001/jamanetworkopen.2023.28928
2. Pathak R, Catalan-Matamoros D. Can Twitter posts serve as early indicators for potential safety signals? A retrospective analysis. *Int J Risk & Saf Med* (2023) 34(1): 41–61. doi:10.3233/jrs-210024
3. Litvinova O, Matin FB, Matin M, Zima-Kulisiewicz B, Tomasik C, Siddiquea BN, et al. Patient safety discourse in a pandemic: a Twitter hashtag analysis study on #PatientSafety. *Front Public Health* (2023) 11:1268730. doi:10.3389/fpubh.2023.1268730
4. Arillotta D, Floresta G, Guirguis A, Corkery JM, Catalani V, Martinotti G, et al. GLP-1 receptor agonists and related mental health issues; insights from a range of social media platforms using a mixed-methods approach. *Brain Sciences* (2023) 13(11):1503. doi:10.3390/brainsci13111503
5. Yu D, Vydiswaran VGV. An assessment of mentions of adverse drug events on social media with natural language processing: model development and analysis. *JMIR Med Inform* (2022) 10(9):e38140. doi:10.2196/38140
6. Xiong S, Batra V, Liu L, Xi L, Sun C. Detecting personal medication intake in twitter via domain attention-based RNN with multi-level features. *Comput Intelligence Neurosci* (2022) 2022:1–7. doi:10.1155/2022/5467262
7. Klein AZ, O'Connor K, Levine LD, Gonzalez-Hernandez G. Using twitter data for cohort studies of drug safety in pregnancy: proof-of-concept with  $\beta$ -blockers. *JMIR Form Res* (2022) 6(6):e36771. doi:10.2196/36771
8. Hua Y, Jiang H, Lin S, Yang J, Plasek JM, Bates DW, et al. Using Twitter data to understand public perceptions of approved versus off-label use for COVID-19-related medications. *J Am Med Inform Assoc* (2022) 29(10):1668–78. doi:10.1093/jamia/ocac114

9. Turner J, Kantardzic M, Vickers-Smith R. Infodemiological examination of personal and commercial tweets about cannabidiol: term and sentiment analysis. *J Med Internet Res* (2021) 23(12):e27307. doi:10.2196/27307
10. Magge A, Tutubalina E, Miftahutdinov Z, Alimova I, Dirkson A, Verberne S, et al. DeepADEMiner: a deep learning pharmacovigilance pipeline for extraction and normalization of adverse drug event mentions on Twitter. *J Am Med Inform Assoc* (2021) 28(10):2184–92. doi:10.1093/jamia/ocab114
11. Sharma C, Whittle S, Haghighi PD, Burstein F, Keen H. Sentiment analysis of social media posts on pharmacotherapy: a scoping review. *Pharmacol Res & Perspect* (2020) 8(5):e00640. doi:10.1002/prp2.640
12. Rezaei Z, Ebrahimpour-Komleh H, Eslami B, Chavoshinejad R, Totonchi M. Adverse drug reaction detection in social media by deepm learning methods. *Cell J* (2020) 22(3):319–24. doi:10.22074/cellj.2020.6615
13. Li Y, Jimeno Yepes A, Xiao C. Combining social media and FDA adverse event reporting system to detect adverse drug reactions. *Drug Saf* (2020) 43(9):893–903. doi:10.1007/s40264-020-00943-2
14. Gattepaille LM, Hedfors Vidlin S, Bergvall T, Pierce CE, Ellenius J. Prospective evaluation of adverse event recognition systems in twitter: results from the web-RADR project. *Drug Saf* (2020) 43(8):797–808. doi:10.1007/s40264-020-00942-3
15. Dietrich J, Gattepaille LM, Grum BA, Jiri L, Lerch M, Sartori D, et al. Adverse events in twitter-development of a benchmark reference dataset: results from IMI WEB-RADR. *Drug Saf* (2020) 43(5):467–78. doi:10.1007/s40264-020-00912-9
16. van Stekelenborg J, Ellenius J, Maskell S, Bergvall T, Caster O, Dasgupta N, et al. Recommendations for the use of social media in pharmacovigilance: lessons from IMI WEB-RADR. *Drug Saf* (2019) 42(12):1393–407. doi:10.1007/s40264-019-00858-7
17. Golder S, Scantlebury A, Christmas H. Understanding public attitudes toward researchers using social media for detecting and monitoring adverse events data: multi methods study. *J Med Internet Res* (2019) 21(8):e7081. doi:10.2196/jmir.7081
18. Golder S, Chiuev S, Weissenbacher D, Klein A, O'Connor K, Bland M, et al. Pharmacoepidemiologic evaluation of birth defects from health-related postings in social media during pregnancy. *Drug Saf* (2019) 42(3):389–400. doi:10.1007/s40264-018-0731-6
19. Suragh TA, Lamprianou S, MacDonald NE, Loharikar AR, Balakrishnan MR, Benes O, et al. Cluster anxiety-related adverse events following immunization (AEFI): an assessment of reports detected in social media and those identified using an online search engine. *Vaccine* (2018) 36(40):5949–54. doi:10.1016/j.vaccine.2018.08.064
20. Smith K, Golder S, Sarker A, Loke Y, O'Connor K, Gonzalez-Hernandez G. Methods to compare adverse events in twitter to FAERS, drug information databases, and systematic reviews: proof of concept with adalimumab. *Drug Saf* (2018) 41(12):1397–410. doi:10.1007/s40264-018-0707-6
21. Patel R, Belousov M, Jani M, Dasgupta N, Winokur C, Nenadic G, et al. Frequent discussion of insomnia and weight gain with glucocorticoid therapy: an analysis of Twitter posts. *npj Digital Med* (2018) 1:20177. doi:10.1038/s41746-017-0007-z
22. Masino AJ, Forsyth D, Fiks AG. Detecting adverse drug reactions on twitter with convolutional neural networks and word embedding features. *J Health Inform Res* (2018) 2(1-2):25–43. doi:10.1007/s41666-018-0018-9
23. Lardon J, Bellet F, Aboukhamis R, Asfari H, Souvignet J, Jaulent MC, et al. Evaluating Twitter as a complementary data source for pharmacovigilance. *Expert Opin Drug Saf* (2018) 17(8):763–74. doi:10.1080/14740338.2018.1499724
24. Comfort S, Perera S, Hudson Z, Dorrell D, Meireis S, Nagarajan M, et al. Sorting through the safety data haystack: using machine learning to identify individual case safety reports in social-digital media. *Drug Saf* (2018) 41(6):579–90. doi:10.1007/s40264-018-0641-7
25. Caster O, Dietrich J, Kürzinger ML, Lerch M, Maskell S, Norén GN, et al. Assessment of the utility of social media for broad-ranging statistical signal detection in pharmacovigilance: results from the WEB-RADR project. *Drug Saf* (2018) 41(12):1355–69. doi:10.1007/s40264-018-0699-2
26. Pierce CE, Bouri K, Pamer C, Proestel S, Rodriguez HW, Van Le H, et al. Evaluation of Facebook and twitter monitoring to detect safety signals for medical products: an analysis of recent FDA safety alerts. *Drug Saf* (2017) 40(4):317–31. doi:10.1007/s40264-016-0491-0
27. Price J. What can big data offer the pharmacovigilance of orphan drugs? *Clin Ther* (2016) 38(12):2533–45. doi:10.1016/j.clinthera.2016.11.009
28. Powell GE, Seifert HA, Reblin T, Burstein PJ, Blowers J, Menius JA, et al. Social media listening for routine post-marketing safety surveillance. *Drug Saf* (2016) 39(5):443–54. doi:10.1007/s40264-015-0385-6
29. Coloma PM, Becker B, Sturkenboom MC, van Mulligen EM, Kors JA. Evaluating social media networks in medicines safety surveillance: two case studies. *Drug Saf* (2015) 38(10):921–30. doi:10.1007/s40264-015-0333-5
30. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug Saf* (2014) 37(5):343–50. doi:10.1007/s40264-014-0155-x
31. Bian J, Topaloglu U, Yu F. Towards large-scale twitter mining for drug-related adverse events. *SHB12 (2012)* (2012) 2012:25–32. doi:10.1145/2389707.2389713
32. Khademi Habibabadi S, Palmer C, Dimaguila GL, Javed M, Clothier HJ, Buttery J. Australasian institute of digital health summit 2022-automated social media surveillance for detection of vaccine safety signals: a validation study. *Appl Clin Inform* (2023) 14(1):1–10. doi:10.1055/a-1975-4061
33. Golder S, O'Connor K, Wang Y, Gonzalez Hernandez G. The role of social media for identifying adverse drug events data in pharmacovigilance: protocol for a scoping review. *JMIR Res Protoc* (2023) 12:e47068. doi:10.2196/47068
34. Fisher A, Young MM, Payer D, Pacheco K, Dubeau C, Mago V. Automating detection of drug-related harms on social media: machine learning framework. *J Med Internet Res* (2023) 25:e43630. doi:10.2196/43630
35. Zhang T, Lin H, Xu B, Yang L, Wang J, Duan X. Adversarial neural network with sentiment-aware attention for detecting adverse drug reactions. *J Biomed Inform* (2021) 123:103896. doi:10.1016/j.jbi.2021.103896
36. Black JC, Margolin ZR, Olson RA, Dart RC. Online conversation monitoring to understand the opioid epidemic: epidemiological surveillance study. *JMIR Public Health Surveill* (2020) 6(2):e17073. doi:10.2196/17073
37. Calac AJ, Haupt MR, Li Z, Mackey T. Spread of COVID-19 vaccine misinformation in the ninth inning: retrospective observational infodemic study. *JMIR Infodemiology* (2022) 2(1):e33587. doi:10.2196/33587
38. Curtis JR, Chen L, Higginbotham P, Nowell WB, Gal-Levy R, Willig J, et al. Social media for arthritis-related comparative effectiveness and safety research and the impact of direct-to-consumer advertising. *Arthritis Res Ther* (2017) 19(1):48. doi:10.1186/s13075-017-1251-y
39. Schellack N, Strydom M, Pepper MS, Herd CL, Hendricks CL, Bronkhorst E, et al. Social media and COVID-19-perceptions and public deceptions of ivermectin, colchicine and hydroxychloroquine: lessons for future pandemics. *Antibiotics (Basel)* (2022) 11(4):445. doi:10.3390/antibiotics11040445
40. Golder S, Weissenbacher D, O'Connor K, Hennessy S, Gross R, Hernandez GG. Patient-reported reasons for switching or discontinuing statin therapy: a mixed methods study using social media. *Drug Saf* (2022) 45(9):971–81. doi:10.1007/s40264-022-01212-0
41. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Tsagkaris C, Parvanov ED, et al. Medical and health-related misinformation on social media: bibliometric study of the scientific literature. *J Med Internet Res* (2022) 24(1):e28152. doi:10.2196/28152
42. Yamaguchi T, Shimizu J, Oya Y, Horio Y, Hida T. Drug-Induced liver injury in a patient with nonsmall cell lung cancer after the self-administration of fenbendazole based on social media information. *Case Rep Oncol* (2021) 14(2):886–91. doi:10.1159/000516276
43. Härmark L, Weits G, Meijer R, Santoro F, Norén GN, van Hunsel F. Communicating adverse drug reaction insights through patient organizations: experiences from a pilot study in The Netherlands. *Drug Saf* (2020) 43(8):745–9. doi:10.1007/s40264-020-00932-5
44. Sinha MS, Freifeld CC, Brownstein JS, Donneyong MM, Rausch P, Lappin BM, et al. Social media impact of the Food and drug administration's drug safety communication messaging about zolpidem: mixed-methods analysis. *JMIR Public Health Surveill* (2018) 4(1):e1. doi:10.2196/publichealth.7823
45. Mackey TK, Liang BA, Strathdee SA. Digital social media, youth, and nonmedical use of prescription drugs: the need for reform. *J Med Internet Res* (2013) 15(7):e143. doi:10.2196/jmir.2464
46. Mackey TK, Liang BA. Global reach of direct-to-consumer advertising using social media for illicit online drug sales. *J Med Internet Res* (2013) 15(5):e105. doi:10.2196/jmir.2610
47. Liang BA, Mackey TK. Prevalence and Global Health implications of social media in direct-to-consumer drug advertising. *J Med Internet Res* (2011) 13(3):e64. doi:10.2196/jmir.1775
48. Li J, Xu Q, Shah N, Mackey TK. A machine learning approach for the detection and characterization of illicit drug dealers on Instagram: model evaluation study. *J Med Internet Res* (2019) 21(6):e13803. doi:10.2196/13803
49. Smith BP, Hoots B, DePadilla L, Roehler DR, Holland KM, Bowen DA, et al. Using transformer-based topic modeling to examine discussions of delta-8 tetrahydrocannabinol: content analysis. *J Med Internet Res* (2023) 25:e49469. doi:10.2196/49469
50. Guo M, Ma Y, Eworuke E, Khashei M, Song J, Zhao Y, et al. Identifying COVID-19 cases and extracting patient reported symptoms from Reddit using natural language processing. *Sci Rep* (2023) 13(1):13721. doi:10.1038/s41598-023-39986-7



51. Godinez H, Xu Q, McMan TJ, Li J, Mackey TK. Analysis of online user discussions on Reddit associated with the transition of use between HIV PrEP therapy. *Front Public Health* (2023) **11**:1073813. doi:10.3389/fpubh.2023.1073813
52. Pepper JK, Zitney LV, Preiss A, Bukowski T, Szczypka G, Kim AE. Can social media monitoring help identify the next EVALI? An examination of Reddit posts about vitamin E acetate and Dank Vapes. *Drug and Alcohol Dependence* (2022) **230**:109193. doi:10.1016/j.drugalcdep.2021.109193
53. Sharp KJ, Vitagliano JA, Weitzman ER, Fitzgerald S, Dahlberg SE, Austin SB. Peer-to-Peer social media communication about dietary supplements used for weight loss and sports performance among military personnel: pilot content analysis of 11 Years of posts on Reddit. *JMIR Form Res* (2021) **5**(10):e28957. doi:10.2196/28957
54. Lavertu A, Hamamsy T, Altman RB. Quantifying the severity of adverse drug reactions using social media: network analysis. *J Med Internet Res* (2021) **23**(10):e27714. doi:10.2196/27714
55. Szczypka G, Pepper JK, Kim A. Weed light: an exploratory study of delta-8 THC conversations on Reddit. In: *RTI press research brief*. Research Triangle Park (NC): RTI Press (2014).
56. Hansen C, Interrante JD, Ailes EC, Frey MT, Broussard CS, Godoshian VJ, et al. Assessment of YouTube videos as a source of information on medication use in pregnancy. *Pharmacoepidemiol Drug Saf* (2016) **25**(1):35–44. doi:10.1002/pds.3911
57. Mical C, Golder S, O'Connor K, Weissenbacher D, Gross R, Hennessy S, et al. Patient-reported reasons for antihypertensive medication change: a quantitative study using social media. *Drug Saf* (2024) **47**(1):81–91. doi:10.1007/s40264-023-01366-5
58. Park S, Choi SH, Song YK, Kwon JW. Comparison of online patient reviews and national pharmacovigilance data for tramadol-related adverse events: comparative observational study. *JMIR Public Health Surveill* (2022) **8**(1):e33311. doi:10.2196/33311
59. Roche V, Robert JP, Salam H. A holistic AI-based approach for pharmacovigilance optimization from patients behavior on social media. *Artif Intelligence Med* (2023) **144**:102638. doi:10.1016/j.artmed.2023.102638
60. Powell G, Kara V, Painter JL, Schifano L, Merico E, Bate A. Engaging patients via online healthcare fora: three pharmacovigilance use cases. *Front Pharmacol* (2022) **13**:901355. doi:10.3389/fphar.2022.901355
61. Xu Q, McMan TJ, Li J, Wenzel C, Mackey TK. Characterization of COVID-19 vaccine clinical trial discussions on the social question-and-answer site Quora. *Trials* (2023) **24**(1):790. doi:10.1186/s13063-023-07837-5
62. Song J, Cui Y, Song J, Lee C, Wu M, Chen H. Evaluation of the needs and experiences of patients with hypertriglyceridemia: social media listening infosurveillance study. *J Med Internet Res* (2023) **25**:e44610. doi:10.2196/44610
63. Duh MS, Cremieux P, Audenrode MV, Vekeman F, Karner P, Zhang H, et al. Can social media data lead to earlier detection of drug-related adverse events? *Pharmacoepidemiol Drug Saf* (2016) **25**(12):1425–33. doi:10.1002/pds.4090
64. Adams DZ, Gruss R, Abrahams AS. Automated discovery of safety and efficacy concerns for joint & muscle pain relief treatments from online reviews. *Int J Med Inform* (2017) **100**:108–20. doi:10.1016/j.ijmedinf.2017.01.005
65. Gartland A, Bate A, Painter JL, Casperson TA, Powell GE. Developing crowdsourced training data sets for pharmacovigilance intelligent automation. *Drug Saf* (2021) **44**(3):373–82. doi:10.1007/s40264-020-01028-w
66. de Vries E, Denig P, de Vries ST, Monster TBM, Hugtenburg JG, Mol PGM. Drug safety issues covered by lay media: a cohort study of direct healthcare provider communications sent between 2001 and 2015 in The Netherlands. *Drug Saf* (2020) **43**(7):677–90. doi:10.1007/s40264-020-00922-7
67. Larson HJ, Wilson R, Hanley S, Parys A, Paterson P. Tracking the global spread of vaccine sentiments: the global response to Japan's suspension of its HPV vaccine recommendation. *Hum Vaccin & Immunother* (2014) **10**(9):2543–50. doi:10.4161/21645515.2014.969618
68. Matsuda S, Aoki K, Tomizawa S, Sone M, Tanaka R, Kuriki H, et al. Analysis of patient narratives in disease blogs on the internet: an exploratory study of social pharmacovigilance. *JMIR Public Health Surveill* (2017) **3**(1):e10. doi:10.2196/publichealth.6872
69. Zhu XM, Yan W, Sun J, Liu L, Zhao YM, Zheng YB, et al. Patterns and influencing factors of COVID-19 vaccination willingness among college students in China. *Vaccine* (2022) **40**(22):3046–54. doi:10.1016/j.vaccine.2022.04.013
70. Grundmann O. Patterns of Kratom use and health impact in the US-Results from an online survey. *Drug and Alcohol Dependence* (2017) **176**:63–70. doi:10.1016/j.drugalcdep.2017.03.007
71. Brosch S, de Ferran AM, Newbould V, Farkas D, Lengsavath M, Tregunno P. Establishing a framework for the use of social media in pharmacovigilance in europe. *Drug Saf* (2019) **42**(8):921–30. doi:10.1007/s40264-019-00811-8
72. Bulcock A, Hassan L, Giles S, Sanders C, Nenadic G, Campbell S, et al. Public perspectives of using social media data to improve adverse drug reaction reporting: a mixed-methods study. *Drug Saf* (2021) **44**(5):553–64. doi:10.1007/s40264-021-01042-6
73. Ghosh R, Lewis D. Aims and approaches of Web-RADR: a consortium ensuring reliable ADR reporting via mobile devices and new insights from social media. *Expert Opin Drug Saf* (2015) **14**(12):1845–53. doi:10.1517/14740338.2015.1096342
74. Lengsavath M, Dal Pra A, de Ferran AM, Brosch S, Härmark L, Newbould V, et al. Social media monitoring and adverse drug reaction reporting in pharmacovigilance: an overview of the regulatory landscape. *Ther Innov Regul Sci* (2017) **51**(1):125–31. doi:10.1177/2168479016663264
75. Wysota CN, Le D, Clausen ME, Ciceron AC, Fuss C, Bennett B, et al. Young adults' knowledge, perceptions and use of cannabidiol products: a mixed-methods study. *Health Education Res* (2022) **37**(6):379–92. doi:10.1093/her/cyac030
76. Nguyen C, Moeller KE, McGuire M, Melton BL. Consumer perception, knowledge, and uses of cannabidiol. *Ment Health Clinician* (2023) **13**(5):217–24. doi:10.9740/mhc.2023.10.217
77. Abdellaoui R, Foulquié P, Texier N, Faviez C, Burgun A, Schück S. Detection of cases of noncompliance to drug treatment in patient forum posts: topic model approach. *J Med Internet Res* (2018) **20**(3):e85. doi:10.2196/jmir.9222
78. Karapetiantz P, Audeh B, Lillo-Le Louët A, Bousquet C. Discrepancy between personal experience and negative opinion with human papillomavirus vaccine in web forums. *Stud Health Technol Inform* (2020) **272**:417–20. doi:10.3233/SHTI200584
79. Karapetiantz P, Bellet F, Audeh B, Lardon J, Leprovost D, Aboukhamis R, et al. Descriptions of adverse drug reactions are less informative in forums than in the French pharmacovigilance database but provide more unexpected reactions. *Front Pharmacol* (2018) **9**:439. doi:10.3389/fphar.2018.00439
80. Vaughan Sarrazin MS, Cram P, Mazur A, Ward M, Reisinger HS. Patient perspectives of dabigatran: analysis of online discussion forums. *Patient* (2014) **7**(1):47–54. doi:10.1007/s40271-013-0027-y
81. Botsis T, Kreimeyer K. Improving drug safety with adverse event detection using natural language processing. *Expert Opin Drug Saf* (2023) **22**(8):659–68. doi:10.1080/14740338.2023.2228197
82. Dong F, Guo W, Liu J, Patterson TA, Hong H. BERT-based language model for accurate drug adverse event extraction from social media: implementation, evaluation, and contributions to pharmacovigilance practices. *Front Public Health* (2024) **12**:1392180. doi:10.3389/fpubh.2024.1392180
83. Zhang J, Wang X, Zhou Y. Comparative analysis of semaglutide induced adverse reactions: insights from FAERS database and social media reviews with a focus on oral vs subcutaneous administration. *Front Pharmacol* (2024) **15**:1471615. doi:10.3389/fphar.2024.1471615
84. Audeh B, Bellet F, Beyens MN, Lillo-Le Louët A, Bousquet C. Use of social media for pharmacovigilance activities: key findings and recommendations from the Vigi4Med project. *Drug Saf* (2020) **43**(9):835–51. doi:10.1007/s40264-020-00951-2
85. Karapetiantz P, Audeh B, Redjäl A, Tiffet T, Bousquet C, Jaulent MC. Monitoring adverse drug events in web forums: evaluation of a pipeline and use case study. *J Med Internet Res* (2024) **26**:e46176. doi:10.2196/46176
86. Li Y, Tao W, Li Z, Sun Z, Li F, Fenton S, et al. Artificial intelligence-powered pharmacovigilance: a review of machine and deep learning in clinical text-based adverse drug event detection for benchmark datasets. *J Biomed Inform* (2024) **152**:104621. doi:10.1016/j.jbi.2024.104621
87. Destere A, Marchello G, Merino D, Othman NB, Gérard AO, Lavrut T, et al. An artificial intelligence algorithm for co-clustering to help in pharmacovigilance before and during the COVID-19 pandemic. *Br J Clin Pharmacol* (2024) **90**(5):1258–67. doi:10.1111/bcp.16012
88. Carpenter KA, Altman RB. Using GPT-3 to build a lexicon of drugs of abuse synonyms for social media pharmacovigilance. *Biomolecules* (2023) **13**(2):387. doi:10.3390/biom13020387
89. Nasralah T, El-Gayar O, Wang Y. Social media text mining framework for drug abuse: development and validation study with an opioid crisis case analysis. *J Med Internet Res* (2020) **22**(8):e18350. doi:10.2196/18350
90. Chen X, Faviez C, Schuck S, Lillo-Le Louët A, Texier N, Dahamna B, et al. Mining patients' narratives in social media for pharmacovigilance: adverse effects and misuse of methylphenidate. *Front Pharmacol* (2018) **9**:541. doi:10.3389/fphar.2018.00541



91. Sarker A, O'Connor K, Ginn R, Scotch M, Smith K, Malone D, et al. Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from twitter. *Drug Saf* (2016) **39**(3):231–40. doi:10.1007/s40264-015-0379-4

92. Allen MR, Wightman GP, Zhu Z, Poliak A, Smith DM, Dredze M, et al. Pharmacovigilance in the age of legalized cannabis: using social media to monitor drug-drug interactions between immunosuppressants and cannabis-derived products. *Drug Saf* (2024) **48**:99–105. doi:10.1007/s40264-024-01481-x

93. Vilar S, Friedman C, Hripcsak G. Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media. *Brief Bioinform* (2018) **19**(5):863–77. doi:10.1093/bib/bbx010

94. Correia RB, Li L, Rocha LM. Monitoring potential drug interactions and reactions via network analysis of Instagram user timelines. *Pac Symp Biocomput* (2016) **21**:492–503. doi:10.48550/arXiv.1510.01006