

Original Paper

Artificial Intelligence–Enabled Software Prototype to Inform Opioid Pharmacovigilance From Electronic Health Records: Development and Usability Study

Alfred Sorbello^{1*}, DO, MPH; Syed Arefinul Haque^{1*}, PhD; Rashedul Hasan^{1*}, PhD; Richard Jermyn^{2*}, DO; Ahmad Hussein^{2*}, MBS; Alex Vega^{2*}, BS; Krzysztof Zembruski^{2*}, BA; Anna Ripple^{3*}, MLS; Mitra Ahadpour^{1*}, MD

¹Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, United States

²Neuromuscular Institute, Rowan-Virtua School of Osteopathic Medicine, Stratford, NJ, United States

³Lister Hill National Center for Biomedical Communications, National Library of Medicine–National Institutes of Health, Rockville, MD, United States

*all authors contributed equally

Corresponding Author:

Rashedul Hasan, PhD

Center for Drug Evaluation and Research

US Food and Drug Administration

10903 New Hampshire Avenue

Silver Spring, MD, 20993

United States

Phone: 1 (888) 463 6332

Email: mdrashedul.hasan@fda.hhs.gov

Abstract

Background: The use of patient health and treatment information captured in structured and unstructured formats in computerized electronic health record (EHR) repositories could potentially augment the detection of safety signals for drug products regulated by the US Food and Drug Administration (FDA). Natural language processing and other artificial intelligence (AI) techniques provide novel methodologies that could be leveraged to extract clinically useful information from EHR resources.

Objective: Our aim is to develop a novel AI-enabled software prototype to identify adverse drug event (ADE) safety signals from free-text discharge summaries in EHRs to enhance opioid drug safety and research activities at the FDA.

Methods: We developed a prototype for web-based software that leverages keyword and trigger-phrase searching with rule-based algorithms and deep learning to extract candidate ADEs for specific opioid drugs from discharge summaries in the Medical Information Mart for Intensive Care III (MIMIC III) database. The prototype uses MedSpacy components to identify relevant sections of discharge summaries and a pretrained natural language processing (NLP) model, Spark NLP for Healthcare, for named entity recognition. Fifteen FDA staff members provided feedback on the prototype's features and functionalities.

Results: Using the prototype, we were able to identify known, labeled, opioid-related adverse drug reactions from text in EHRs. The AI-enabled model achieved accuracy, recall, precision, and F_1 -scores of 0.66, 0.69, 0.64, and 0.67, respectively. FDA participants assessed the prototype as highly desirable in user satisfaction, visualizations, and in the potential to support drug safety signal detection for opioid drugs from EHR data while saving time and manual effort. Actionable design recommendations included (1) enlarging the tabs and visualizations; (2) enabling more flexibility and customizations to fit end users' individual needs; (3) providing additional instructional resources; (4) adding multiple graph export functionality; and (5) adding project summaries.

Conclusions: The novel prototype uses innovative AI-based techniques to automate searching for, extracting, and analyzing clinically useful information captured in unstructured text in EHRs. It increases efficiency in harnessing real-world data for opioid drug safety and increases the usability of the data to support regulatory review while decreasing the manual research burden.

(JMIR AI 2023;2:e45000) doi: [10.2196/45000](https://doi.org/10.2196/45000)

KEYWORDS

electronic health records; pharmacovigilance; artificial intelligence; real world data; EHR; natural language; software application; drug; Food and Drug Administration; deep learning

Introduction

Postmarketing drug safety surveillance at the Center for Drug Evaluation and Research (CDER) of the US Food and Drug Administration (FDA) aims to detect, characterize, monitor, and prevent adverse drug reactions (ADRs) for FDA-approved drugs and therapeutic biologic products. Biomedical resources used to detect adverse drug event (ADE) safety signals include clinical trials, spontaneous adverse event (AE) reports submitted to the FDA Adverse Events Reporting System (FAERS), published scientific reports in the literature, and others. The FAERS database compiles AE and medication error reports submitted to the FDA to support postmarket drug safety monitoring. FAERS monitoring has yielded information on rare ADEs, but the information is limited by underreporting [1,2]. Multimodal approaches to pharmacovigilance using multiple biomedical resources may offer improved drug safety signal detection compared to reliance on single resources [3].

Electronic health records (EHRs) are a rich source of real-world information that may potentially serve as a new complementary drug safety resource. Although not specifically created to document ADEs, the EHR may provide information about product side effects, including those that occur a prolonged time following initial drug exposure [4], and may contribute to assessments of the safety of generic and pediatric drug products [5,6]. EHRs have been explored to complement ADE signal identification from spontaneous AE reports [7].

Published scientific reports describe various natural language processing (NLP) and artificial intelligence (AI)-based approaches to analyzing text from EHRs for ADE detection and pharmacovigilance. Named entity recognition (NER) to identify drug and AE mentions in text followed by extraction of the relationships between those entities is a critical technical challenge in building successful analytical algorithms. In general, keywords, rule-based algorithms, and machine learning methods have been used for case detection [8]. Some early studies used trigger phrases to screen the text of discharge summaries for AE concepts [9,10]. Established NLP algorithms applied to AE detection include MedLEE, which identifies clinical concepts and cross-maps them to Unified Medical Language System (UMLS) concepts [11]; MetaMap, which processes biomedical text and maps it to the UMLS [12]; and Clinical Text Analysis and Knowledge Extraction System (cTAKES), an NLP system that incorporates rules and machine learning [13]. More recent studies use multiple NLP models, including long short-term memory (LSTM), conditional random field (CRF), support vector machines (SVMs), and bidirectional encoder representations from transformers (BERTs) [14]. Shared task challenges designed to promote advances in NLP for drug safety and ADE detection from EHRs have been conducted in recent years, including the MADE 1.0 challenge [15] and the n2c2 Clinical Challenge [16]. Text analytic engines, such as Amazon Comprehend Medical, Microsoft Text Analytics for Health, and the Google Healthcare Natural Language application programming interface, are deep learning-based pretrained models. These models can perform a variety of general health care NLP tasks, such as NER, relation detection, entity disambiguation, and others [17]. We combine a similar deep

learning model with domain-specific, rule-based algorithms from domain expertise to detect opioid-related ADEs (ORADEs) from clinical notes.

Using novel AI methods, time-consuming manual chart review can be automated to provide active surveillance with enhanced detection of emerging product safety issues in near-real time. Opioids are one of the most frequently implicated drug classes for ADRs in hospitalized patients and are associated with confusion, constipation, respiratory depression, sedation, ileus, hypotension, and other ADRs [18]. One study reported an ORADE prevalence rate of 9.1% in previously opioid-free surgical patients [19]. In this manuscript, we report on the development of and user feedback for SPINEL (Supporting Pharmacovigilance by Leveraging Artificial Intelligence Methods to Analyze Electronic Health Records Data), a novel AI-enabled software prototype that analyzes unstructured text in discharge summaries to extract candidate ADEs for opioid drugs. FDA participants provide feedback on the serviceability of the prototype in meeting their needs to support drug safety, research, and regulatory decision-making.

Methods

Ethical Considerations

This study does not meet the requirements of research involving human subjects as defined by the US Department of Health and Human Services (45CFR46) for the following reasons: (1) there was no interaction or intervention with human subjects; (2) MIMIC is a free, publicly available database and the authors have completed the required Collaborative Institutional Training Initiative training and data use agreement; (3) all MIMIC III data were deidentified in accordance with Health Insurance Portability and Accountability Act requirements, including removal of 18 identifying data elements; (4) protected health information has been removed from free text fields; and (5) no personally identifiable information was available to the study investigators.

Data Source

EHR Data

We limited our work to publicly accessible EHR databases and focused on the free text in discharge summaries from the Medical Information Mart for Intensive Care III (MIMIC III) [20]. This database contains EHRs from 2001 through 2012 from a single health care center; the records are encoded with codes in the International Classification of Diseases, Ninth Revision (ICD-9). We leveraged ICD-9 code E935.2, which indicates opioids and other narcotics causing AEs in therapeutic use, to prescreen discharge summaries that may contain information on ORADEs. We identified 227 summaries consisting of 227 unique hospital-event records for 226 unique patients. We planned to explore the more recently released MIMIC IV EHR database for additional cases, but the discharge summaries were not made publicly accessible until after this project was completed.

Reference Data Set for Testing and Training

Considering that ICD-9 codes have limited positive predictive value for drug safety surveillance [21], 2 medical students (AV, KZ) and a physician (AS) conducted independent manual reviews of the 227 discharge summaries identified by ICD-9 prescreening (as above) to manually assess for documentation of ORADEs in the text. We did not use a formal annotation guideline; positive assessments were based on specific textual mentions describing opioid drug exposure and adverse events either linked or potentially linked to the exposure irrespective of the severity or seriousness of the events. To create a reference data set of discharge summaries with true positive and negative cases, positive assessment for an ORADE required agreement among all 3 reviewers. Discrepancies were reconciled through joint discussion. The 3 reviewers had similar assessments for

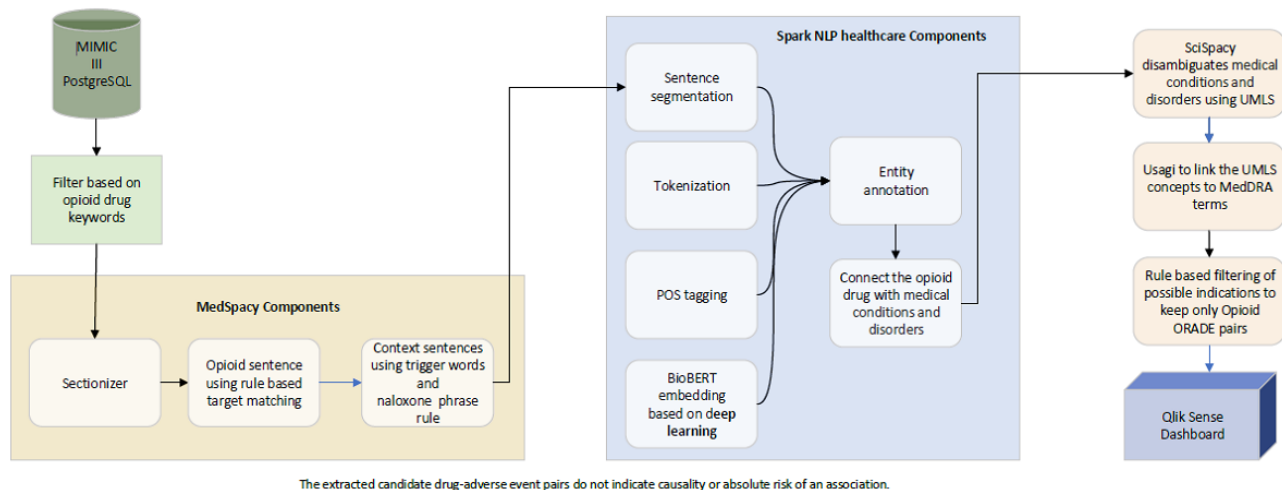
ORADE documentation for 174 (77%) of the 227 discharge summaries reviewed. We trained our AI-enabled model on 181 (80%) of the discharge summaries and used the remaining 46 (20%) for testing.

NLP Process

Detection of Sections in Discharge Summaries

Based on a manual review, we identified 3 sections with the highest frequency of ORADE mentions: “brief hospital course,” “hospital course,” and “history of present illness.” In our AI-enabled model (Figure 1), we used the Sectionizer module in the MedSpacy open-source Python library [22] to automate the identification of those component sections in the sample of discharge summaries.

Figure 1. The artificial intelligence-enabled model is depicted with natural language processing and rule-based algorithms, MedSpacy sectionizer components, Spark NLP for Healthcare entity recognition components, SciSpacy disambiguation of terms, Usagi interconnection of UMLS concepts with MedDRA terminology, and further filtering of ORADE pairs. A higher resolution version of this figure is available in [Multimedia Appendix 1](#). MIMIC: Medical Information Mart for Intensive Care; NLP: natural language processing; POS: part of speech; UMLS: Unified Medical Language System; MedDRA: Medical Dictionary for Regulatory Activities; ORADE: opioid-related adverse drug event.



Identifying ORADE Context Sentences Using Keywords, Trigger Phrases, and Rule-Based Algorithms

Using MedSpacy components, we divided the unstructured text in the 3 component sections of the discharge summaries into individual sentences. We identified the context sentences in 2 stages. In the first stage, we identified the sentences that contained one or more mentions of opioid-drug generic terms or opioid-drug brand names using keyword lists manually constructed by one of the team members (AS). The drug names were aligned with RxNorm terminology.

In the second stage, we used 2 rule-based approaches to identify context sentences with mentions of possible ORADEs. First, the trigger-phrase rule: We applied trigger phrases [23] to link mentions of an opioid drug with ADE terms using the MedSpacy context algorithm [24]. We curated 58 additional trigger phrases ([Multimedia Appendix 2](#)) from the training subset of the reference data set and included them in our analysis. To capture mentions of opioid drugs and ADEs that did not co-occur in the same sentence, we searched for those terms in the 3 sentences preceding and following the sentence of interest based on reported heuristics [23].

An example of a trigger-phrase rule is as follows: “It is noteworthy that the patient had received 0.5 mg Ativan x2 and morphine earlier in the afternoon and there is a concern that this may have contributed to his altered mental status.” In this context sentence, an opioid drug (“morphine”) is identified alongside a trigger phrase (“contributed to”). The Spark-NLP NER model identified the AE term as *altered mental status*. This term was resolved to the Medical Dictionary for Regulatory Activities (MedDRA) term *mental state abnormal* using Usagi (Observational Health Data Sciences and Informatics) and the corresponding UMLS concept, as in the section on disambiguation of the ORADEs below. The candidate ORADE pair generated from this information is *morphine-mental state abnormal*.

Second, the antidote-based ADE detection rule: We identified ORADE context sentences by identifying mentions of the drug naloxone, an FDA-approved medication that reverses an overdose caused by an opioid drug. To capture mentions of naloxone and opioid drugs that did not co-occur in the same sentence, we searched through the preceding and following 3 sentences. Antidote signals have been used in detecting ADRs in published literature reports [25,26].

An example of an antidote-based ADE detection rule is as follows: “He received dilaudid q 2 hr at 7:30 am, 9:30 am, 11:30 am. Code blue was called for respiratory arrest (unwitnessed). 0.4 mg of Narcan IV was administered followed by 1 mg of IV Narcan. This resulted in improvement of his respiratory status and regain of his consciousness.” In this example, the antidote-based detection rule captures mentions of naloxone in the context sentence, respiratory arrest in the preceding sentence, and the Dilaudid mention in the prior sentence to generate the candidate ORADE pair *Dilaudid-respiratory arrest*.

NER to Identify ORADEs in Clinical Text

Having detected opioid drug terms, we used a pretrained NER model, Spark NLP for Healthcare, which uses deep learning-based NER to identify possible AE terms in sentences. The model is a biLSTM, convolutional neural network, character-based deep learning model trained using biomedical NER data sets such as AnatEM, BC5CDR, BC4CHEMD, BioNLP13CG, JNLPBA, Linnaeus, NCBI-Disease, and S800 [27]. After identifying the AE terms in the context sentences, we connected all opioid mentions in the context sentences with the AE terms to create candidate ORADE pairs.

Disambiguation of ORADEs

AE terms can appear with different spellings, spelling errors, or abbreviations; therefore, we used the UMLS to map the free text to standardized concepts. We used ScispaCy to map the raw phrase found in the discharge summary to the standard UMLS translation of the concept [28]. Furthermore, we used Usagi to obtain the MedDRA term for the UMLS concept. The identified MedDRA AE term is mapped to the opioid drug term to create a candidate ORADE pair that incorporates standardized MedDRA terminology, including preferred terms (PTs) or lower-level terms (LLTs).

Prototype User Testing and Feedback From Participants

We recruited 15 CDER staff members to assess the various features, functionalities, and graphic visualizations. They were

experienced in the use of web-based software tools but were not involved in the development of this prototype.

Testing Design and Conduct

A testing guide was provided that included login instructions, descriptions and screenshots of the application features and components, and instructions for exporting outputs. Test participants worked remotely, were not monitored, and were given 1 week to complete their testing. Participants were free to explore the application for their regulatory work.

For user testing, we extracted from the MIMIC III database a subset of discharge summaries filtered for an opioid drug keyword. The subset included 31,052 notes corresponding to 30,326 hospital admission events for 24,539 patients.

Metrics

Each participant completed an anonymous electronic survey covering technical operation, ease of navigating and interpreting various visualizations, and user satisfaction for drug safety and research (Multimedia Appendix 3).

Results

ORADE Detection

The prototype application successfully detected ORADEs that correspond to known opioid drug toxicities. The most commonly identified opioid drugs and the top 3 most frequent ORADEs per drug are summarized in Table 1.

To assess the contribution of keywords with trigger phrases and antidote (naloxone) signals for ORADE detection, we examined quantitative parameters for a filtered MIMIC III data subset, as shown in Table 2.

Table 2 shows that keywords with trigger phrases detect most unique AEs and candidate ORADEs in context discharge summaries. In comparison, the approach based on the antidote (ie, naloxone) makes a much smaller relative contribution to ORADE detection.

Table 1. Opioid-related adverse drug event detection from the text of the electronic health record discharge summaries.

Opioid drug class	Most frequently identified opioid drug	Top 3 most frequently identified opioid-related adverse drug events
Natural	Morphine	Hypotension; somnolence; nausea
Semisynthetic	Hydromorphone	Confusion; hypotension; agitation
Synthetic	Fentanyl	Hypotension; adverse reaction; hepatitis C

Table 2. Relative contribution of keywords with trigger phrase and antidote (ie, naloxone) signals for candidate opioid-related adverse drug event detection. International Classification of Diseases (Ninth Revision) code E935.2, which specifies opioids and other narcotics causing adverse effects in therapeutic use, was used to create a filtered subset of Medical Information Mart for Intensive Care III (MIMIC III) discharge summaries having at least one opioid-related adverse drug event pair.

	ORADE ^a detection based on keywords with trigger phrases	ORADE detection based only on antidote (naloxone) signals	ORADE detection based on both trigger phrases and antidote signals
Number of unique opioid drugs detected (n=12)	12 (100%)	6 (50%)	6 (50%)
Number of unique AEs ^b detected (n=117)	110 (94%)	8 (7%)	15 (13%)
Number of unique candidate ORADE pairs (n=219)	205 (94%)	13 (6%)	17 (8%)
Number of discharge summaries (n=101)	95 (94%)	8 (8%)	12 (12%)
Number of unique patients (n=101)	94 (94%)	8 (8%)	12 (12%)

^aORADE: opioid-related adverse drug event.

^bAE: adverse event.

Error Analysis

An error analysis was performed to characterize incorrect candidate ORADE pairs and is summarized with mitigation strategies in [Table 3](#).

Table 3. Error analysis of false-positive and false-negative candidate opioid-related adverse drug event pairs.

Category/type and relative frequency	Example	Mitigation strategy
False positive		
Drug indication pairs ^a	Text: “She was given fentanyl for the back pain with subsequent hypotension.” Incorrect candidate ORADE ^b pair: <i>fentanyl-back pain</i>	Condition terms that include “pain” are excluded.
Drug/medication change events ^c	Text: “She was changed from Percocet to Ultram due to nausea, which resolved.” Incorrect candidate ORADE pair: <i>Ultram-nausea</i>	The context sentence is scanned for the following phrases using regular expressions: “change to,” “switch to,” “change from drug X to drug Y,” or “switch from drug X to drug Y.” Candidate opioid drug-drug medication change event pairs so generated are excluded.
Negated ADE ^d mentions where the AE ^e is not due to a drug ^f	Text: “No further apneic events.” Incorrect candidate ADE: <i>apneic events</i>	The assertion module in Spark NLP ^g for Healthcare is used to detect negation so that any negated condition term is not included in a candidate ORADE pair.
False negative		
Concept fragmentation ^c	Text: “She had been treated with high dose fentanyl and benzodiazepines which were the most likely cause of delirium.... She was also found to be severely constipated. # Constipation: patient developed severe constipation related to pain medication. She was manually disimpacted and started on an aggressive [sic] bowel regimen.” Missed candidate ORADE pair: <i>opioid drug-constipation</i>	<i>Severe constipation</i> was detected, but the current model could not find which pain medication it was related to. To resolve, we will explore more data and consider other rules or models.
Entity not recognized as an AE	Text: “He does endorse decreased sleep latency, falling asleep in less than 5 minutes, and also questionable daytime hypersomnolence, but denies morning headaches. Of note, patient received prescription for Vicodin upon discharge from ED on [**2173-8-28**].” Missed candidate AE: <i>hypersomnolence</i>	To resolve, we will explore more data and consider other rules or models.
Entity not recognized as an opioid drug ^c	Text: “His hospital course was complicated by a respiratory code on the floor attributed to respiratory suppression from narcotics.” Missed candidate drug: <i>narcotics</i>	<i>Narcotics</i> could be added to the opioid keyword list. To resolve to a specific opioid drug, we will explore more data and consider other rules or models.

^aMost commonly encountered error.

^bORADE: opioid-related adverse drug event.

^cModerately encountered error.

^dADE: adverse drug event.

^eAE: adverse event.

^fRarely encountered error.

^gNLP: natural language processing.

Prototype Application Performance Metrics

We calculated the performance metrics accuracy, recall, precision, and F_1 -score using conventional mathematical formulas [14]. The AI-enabled model achieved accuracy, recall, precision, and F_1 -scores of 0.66, 0.69, 0.64, and 0.67, respectively, based on the test subset of 46 discharge summaries. Candidate ORADE pairs generated with this software prototype are hypothetical and do not indicate causality or absolute risk for an association. Further assessment is required by subject matter experts.

Prototype Application Analytics Dashboard

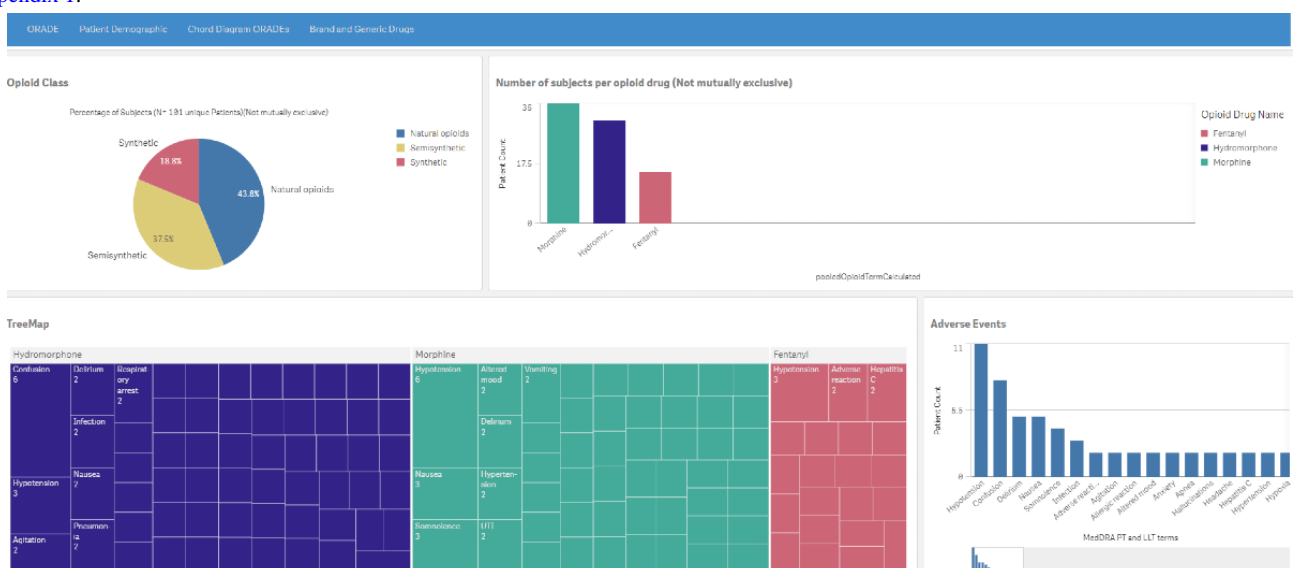
The Qlik Sense data analytics platform (QlikTech International AB) was used to implement the SPINEL dashboard with interactive graphics, visualizations, and line listings. The landing page (Figure 2) has 4 sheet tabs: ORADE, Patient Demographic, Chord Diagram ORADEs, and Brand and Generic Drugs. They are described below with morphine used as an arbitrarily selected opioid drug for the graphics and visualizations.

The ORADE tab (Figure 3) has four components: (1) a pie chart that shows subsets of the 3 classes of opioid drugs, (2) a histogram of all subjects per drug, (3) a tree map of the MedDRA PTs and LLTs for each drug, and (4) a second histogram of patient count by MedDRA PT and LLT for the selected drug(s) of interest.

Figure 2. The landing page for SPINEL (Supporting Pharmacovigilance by Leveraging Artificial Intelligence Methods to Analyze Electronic Health Records Data) depicting a pie chart (upper left) of the 3 opioid classes; a histogram (upper right) of the subject counts per opioid drug; a tree map (lower left) of the electronic health record–derived opioid-related adverse drug profiles, where the adverse events for each opioid drug are represented by nested rectangles and the size of the nested rectangle relates to the patient count per adverse event; and a histogram (lower right) of patient count by MedDRA (Medical Dictionary for Regulatory Activities) preferred term and lower-level term for the drugs. A higher resolution version of this figure is available in [Multimedia Appendix 1](#).



Figure 3. The opioid-related adverse drug page depicting a pie chart (upper left) and a histogram (upper right) of the 101 subjects who received at least one opioid class drug, a tree map (lower left) of the electronic health record–derived opioid-related adverse drug profile for the most frequently identified opioid class drugs, and a histogram (lower right) of patient count by MedDRA (Medical Dictionary for Regulatory Activities) preferred term and lower-level term for the top 3 most frequently identified opioid-related adverse drugs. A higher resolution version of this figure is available in [Multimedia Appendix 1](#).



The patient demographic tab (Figure 4) includes the following components: (1) a histogram of age, (2) a pie chart of gender, (3) another histogram of ethnicity, and (4) a line listing of the individual patients with AEs and associated demographics.

The chord diagram tab (Figure 5) displays a graphic to visually explore interconnections between opioid drugs and AE mentions.

The brand and generic drugs tab (Figure 6) includes multiple displays: (1) a pie chart with the percentage patient count by brand or generic drug type, (2) a stacked bar chart of patients by opioid class and drug type, and (3) a searchable, scrollable spreadsheet listing of the drug name, drug type, and adverse events associated with the subject IDs.

Figure 4. Patient demographics page depicting a histogram for morphine treated-patients by age (upper left), a pie chart for gender (upper right), a histogram for ethnicity (lower left), and a line listing (lower right) of the individual patients with adverse events and associated demographics. Morphine is an arbitrarily selected natural opioid drug. A higher resolution version of this figure is available in [Multimedia Appendix 1](#).

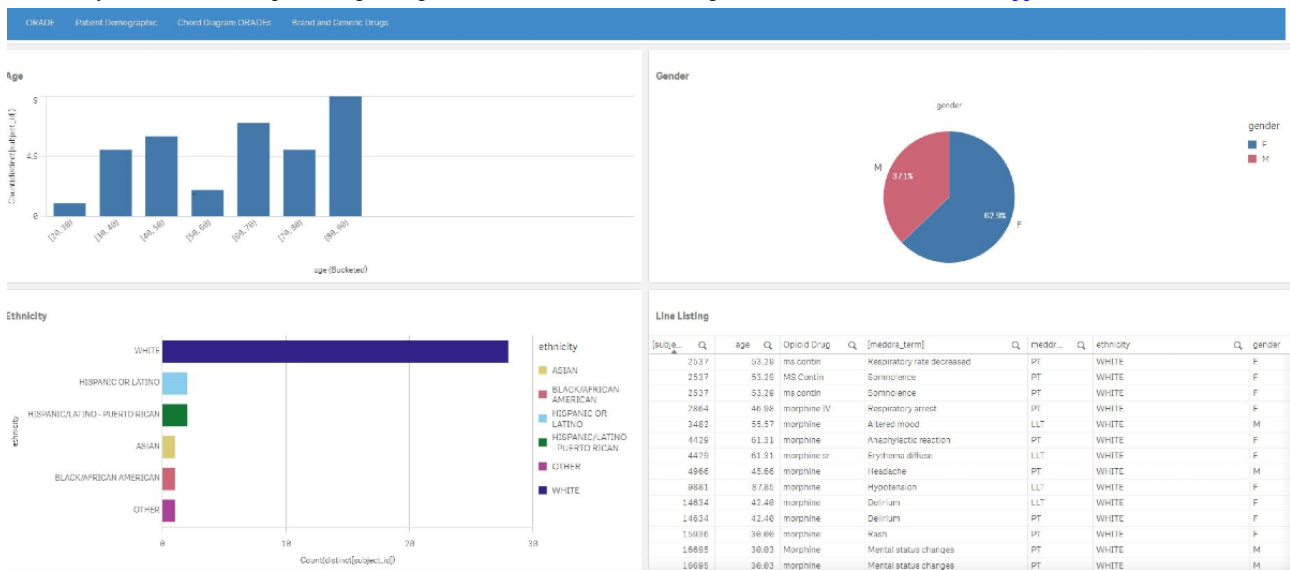


Figure 5. Cord diagram page visually depicting the interconnections between the opioid drug of interest (morphine in this example) and adverse event mentions as derived from the electronic health record discharge summaries. The larger the caliber of the connecting cord, the higher the adverse drug event frequency. Morphine is an arbitrarily selected natural opioid drug. A higher resolution version of this figure is available in [Multimedia Appendix 1](#).

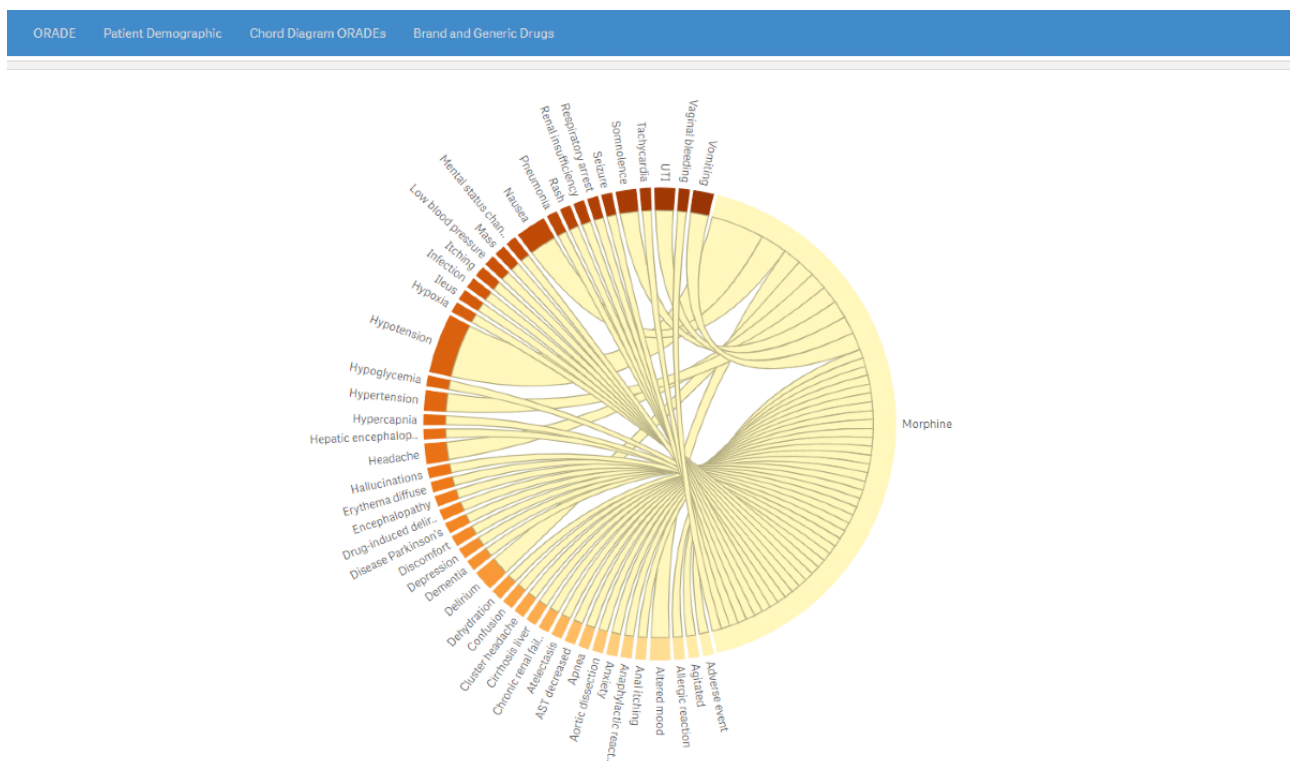
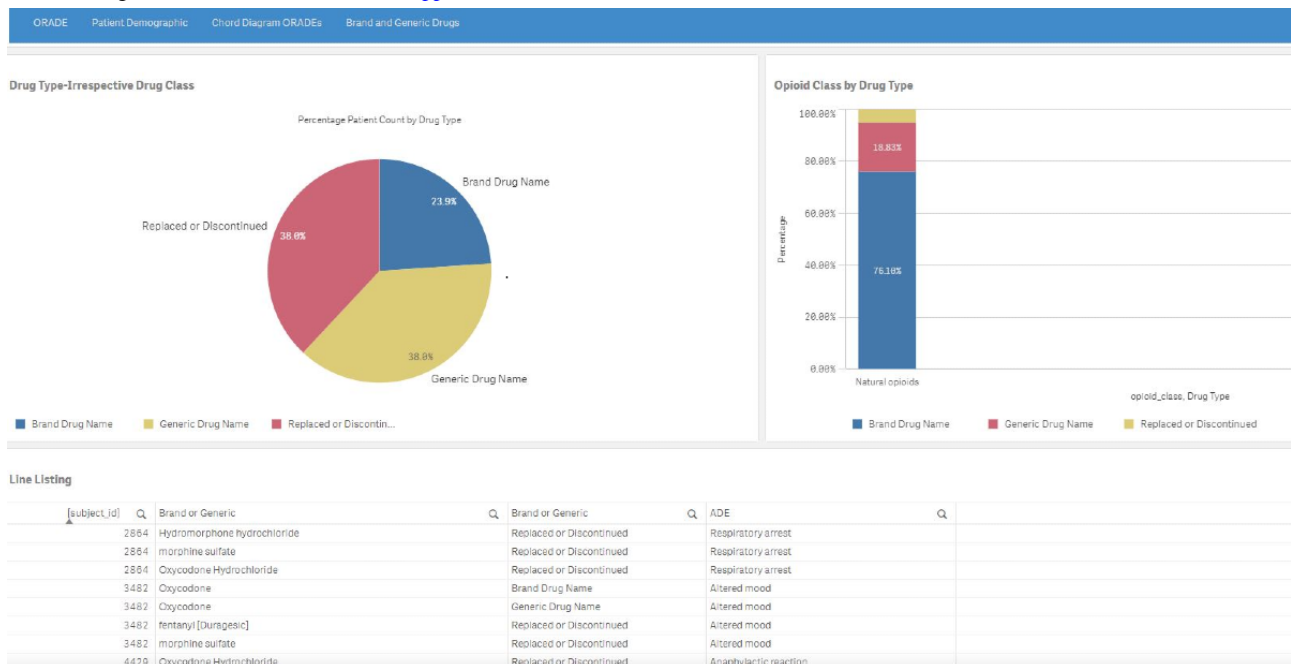


Figure 6. Brand and generic drugs page depicting a pie chart (upper left) of brand, generic, or replaced/discontinued drug type, a stacked bar chart (upper right) of patients by type, and a line listing (lower section) of the patients by drug name, drug type, and adverse events. A higher resolution version of this figure is available in [Multimedia Appendix 1](#).



Results of User Testing

SPINEL was assessed as a highly desirable prototype that satisfies end user needs for supporting opioid drug safety signal detection from EHR data. The application was easy to use, the visualizations enhanced detection of drug safety signals, and the prototype ranked high in saving time compared to manual chart review. Survey results were based on a Likert rating scale ([Multimedia Appendix 4](#)).

Fifteen FDA staff completed the survey questionnaire with 11 providing observational feedback. Participant feedback uncovered a few minor bugs and indicated the following areas for potential improvement: (1) enlarge the tabs and visualizations, (2) enable more flexibility and customizations to fit each end user's needs, (3) provide additional instructional resources to enhance learning about the various features and functionalities, (4) add multiple graph export functionality, and (5) add project summaries. Possible mitigation strategies include adding a slider bar with zoom function for the more complex visualizations, providing an instructional video on the application's features and functionalities, providing tool-tip pop-ups and a supplemental "user tips" guide to highlight key features or functionality, modifying the export function to accommodate multiple graphics, and developing a customizable user portal to include project summaries.

Discussion

Principal Results

The AI-enabled SPINEL prototype successfully detects known opioid drug toxicities from free text in EHRs and provides a framework to uncover emerging safety data that could potentially augment regulatory review and decision-making. Automated processing and analysis of EHR data reduces the

research burden compared to manual chart review, saving considerable time and effort. The prototype expedites the quick perusal of data for trends and patterns reflecting drug toxicities while facilitating drilling down into the data to patient-level line listing information. FDA participants conveyed high satisfaction ratings for this prototype and acknowledged its potential to add value in harnessing unstructured text in EHRs for pharmacovigilance.

In applying our AI-based model, we limited our analysis to discharge summaries because published studies confirm that discharge summaries are the best subsection of the EHR for gathering information about ADEs reported by physicians [29-31]. In reviewing the discharge summaries, we observed considerable heterogeneity in the quality of reporting and the depth of detail conveyed about possible ORADEs, which could affect the accuracy and other performance metrics for the software application. We applied 2 rule-based algorithms to enhance ORADE detection from discharge summaries. Our results demonstrate that the majority of candidate ORADE pairs and context discharge summaries are detected using keywords with trigger phrases. As described in published literature [32], this approach to searching for drug safety signals is best for uncovering ADEs potentially related to specific drug products as delineated in the keyword list (opioids in our use case). As new drug products are approved by the FDA, the keyword list would need manual updating to keep it current. However, for broader and more generalized searching, this could become cumbersome, as new keyword lists would need to be manually compiled for each drug grouping or class of interest.

The accuracy, recall, and precision of this prototype will need to be improved to better align with established NLP processors. Two steps to be considered in future work to improve performance are (1) leveraging information from established drug databases, such as the DailyMed database of the most

recent FDA-approved drug product labels to filter out false positive ORADEs due to drug-indication pairs and (2) using large language models (LLMs) such as GPT-4 [33], BioGPT [34], or GatorTron [35] to improve capture of mentions of opioid drugs and ADE terms that may be separated by multiple paragraphs.

Limitations

This project encountered three main challenges and limitations. First, patient cohort identification: Use of ICD-9 codes to prescreen discharge summaries for potential cases of ORADEs could be impacted by selection and misclassification biases resulting in a subset that may not reflect the total number of ORADE cases in the MIMIC III data. These biases could result in a skewed patient sample wherein there may be missed patients with ORADEs or patients incorrectly classified as having an ORADE due to erroneous coding. In addition, in focusing only on the free-text discharge summaries, we may have missed patients whose ORADEs were captured only in other text reports that we did not explore, such as physician notes, nursing notes, and consultation reports. Together, these issues may prevent us from capturing the full extent and scope of patients experiencing ORADEs from the MIMIC III EHRs. In future work, a more robust approach to identifying patients with ORADEs will be considered, including use of a standardized annotation guideline and reporting of interannotator agreement scores related to development of a reference data set; possible inclusion of objective components for case ascertainment, such as laboratory or medical imaging abnormalities; and expanding the scope of reports assessed to include physician notes, nursing notes, and consultation reports, where available, in addition to discharge summaries. Second was the use of MIMIC III. The single-center MIMIC III EHR database may not reflect the broad diversity of the US population, which could limit generalizability for

drug safety surveillance to larger and more diversified domains and lend to potentially biased assessments. Third, the lack of a publicly available reference standard data set hindered efforts to evaluate the NLP component of our AI-enabled model in detecting ADE safety signals from text in EHRs. The small size of our reference data set risked overfitting and biased assessments.

There were limitations inherent in the user testing procedures. User testing was unmonitored and conducted without prespecified tasks. This approach accommodated participants working in remote locations to explore the software in their regulatory work. However, direct observation by a facilitator may have enabled us to gather more details about end-user experience. Additionally, the sample size of intended users was small. Feedback from a larger group of CDER regulatory staff may be more informative about the potential impact on their regulatory work and decision-making.

Conclusions

SPINEL, our novel AI-enabled software, extracts ORADEs from free-text discharge summaries in EHRs, streamlines workflow, and augments access to real world data for pharmacovigilance. Detecting opioid safety signals from EHRs enhances the capacity to harness an important yet underutilized resource of clinically relevant information for regulatory review and decision-making.

Future work will explore detecting newly emerging opioid drug safety issues using a larger and more diversified EHR database, investigating various methods to improve NLP performance, resolving application features per FDA participant feedback, and integrating knowledge graphs to interconnect information from EHRs with reports published in the literature.

Acknowledgments

This project was supported in part by appointment to the research participation program at the Center for Drug Evaluation and Research (CDER), which is administered by the Oak Ridge Institute for Science and Education for the US Food and Drug Administration (FDA) and by the Lister Hill National Center for Biomedical Communications of the National Library of Medicine of the National Institutes of Health. Funding support was received from the FDA, CDER, and the Opioid Research Program. We also acknowledge Henry Francis, MD, and Mahmudul Hasan, PhD, for their contributions to the project.

Disclaimers

This publication reflects the views of the authors and should not be construed to represent US Food and Drug Administration views or policies. The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the US Department of Health and Human Services.

Authors' Contributions

AS and RH contributed to project concept and design; SAH and RH contributed to data processing; AS, AV, and KZ contributed to the reference standard; SAH and RH contributed to the natural language processing and artificial intelligence-enabled models; RH contributed to dashboard configuration; AS, RH, SAH, RJ, AH, AV, KZ, and AR contributed to interpretation of the results; AS and SAH contributed to drafting of the manuscript; AS, RH, AR, RJ, and MA contributed to manuscript revision; and AS, RH, AR, RJ, and MA contributed to final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

High resolution versions of Figures 1-6.

[\[PDF File \(Adobe PDF File\), 1379 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

List of 58 newly curated trigger phrases.

[\[PDF File \(Adobe PDF File\), 44 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Anonymous user survey questions.

[\[PDF File \(Adobe PDF File\), 81 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

User Survey: Likert scale ratings from user testing.

[\[PDF File \(Adobe PDF File\), 35 KB-Multimedia Appendix 4\]](#)

References

1. Alatawi YM, Hansen RA. Empirical estimation of under-reporting in the U.S. Food and Drug Administration Adverse Event Reporting System (FAERS). *Expert Opin Drug Saf* 2017 Jul;16(7):761-767 [doi: [10.1080/14740338.2017.1323867](https://doi.org/10.1080/14740338.2017.1323867)] [Medline: [28447485](https://pubmed.ncbi.nlm.nih.gov/28447485/)]
2. La Grenade L, Graham DJ, Nourjah P. Underreporting of hemorrhagic stroke associated with phenylpropanolamine. *JAMA* 2001 Dec 26;286(24):3081 [doi: [10.1001/jama.286.24.3081](https://doi.org/10.1001/jama.286.24.3081)] [Medline: [11754672](https://pubmed.ncbi.nlm.nih.gov/11754672/)]
3. Harpaz R, DuMouchel W, Schuemie M, Bodenreider O, Friedman C, Horvitz E, et al. Toward multimodal signal detection of adverse drug reactions. *J Biomed Inform* 2017 Dec;76:41-49 [FREE Full text] [doi: [10.1016/j.jbi.2017.10.013](https://doi.org/10.1016/j.jbi.2017.10.013)] [Medline: [29081385](https://pubmed.ncbi.nlm.nih.gov/29081385/)]
4. Coloma PM, Trifirò G, Patadia V, Sturkenboom M. Postmarketing safety surveillance: where does signal detection using electronic healthcare records fit into the big picture? *Drug Saf* 2013 Mar;36(3):183-197 [doi: [10.1007/s40264-013-0018-x](https://doi.org/10.1007/s40264-013-0018-x)] [Medline: [23377696](https://pubmed.ncbi.nlm.nih.gov/23377696/)]
5. Spence MM, Nguyen LM, Hui RL, Chan J. Evaluation of clinical and safety outcomes associated with conversion from brand-name to generic tacrolimus in transplant recipients enrolled in an integrated health care system. *Pharmacotherapy* 2012 Nov;32(11):981-987 [doi: [10.1002/phar.1130](https://doi.org/10.1002/phar.1130)] [Medline: [23074134](https://pubmed.ncbi.nlm.nih.gov/23074134/)]
6. Nie X, Yu Y, Jia L, Zhao H, Chen Z, Zhang L, et al. Signal detection of pediatric drug-induced coagulopathy using routine electronic health records. *Front Pharmacol* 2022;13:935627 [FREE Full text] [doi: [10.3389/fphar.2022.935627](https://doi.org/10.3389/fphar.2022.935627)] [Medline: [35935826](https://pubmed.ncbi.nlm.nih.gov/35935826/)]
7. Harpaz R, Vilar S, Dumouchel W, Salmasian H, Haerian K, Shah NH, et al. Combining signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *J Am Med Inform Assoc* 2013 May 01;20(3):413-419 [FREE Full text] [doi: [10.1136/amiainl-2012-000930](https://doi.org/10.1136/amiainl-2012-000930)] [Medline: [23118093](https://pubmed.ncbi.nlm.nih.gov/23118093/)]
8. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc* 2016 Sep;23(5):1007-1015 [FREE Full text] [doi: [10.1093/jamia/ocv180](https://doi.org/10.1093/jamia/ocv180)] [Medline: [26911811](https://pubmed.ncbi.nlm.nih.gov/26911811/)]
9. Murff HJ, Forster AJ, Peterson JF, Fiskio JM, Heiman HL, Bates DW. Electronically screening discharge summaries for adverse medical events. *J Am Med Inform Assoc* 2003;10(4):339-350 [FREE Full text] [doi: [10.1197/jamia.M1201](https://doi.org/10.1197/jamia.M1201)] [Medline: [12668691](https://pubmed.ncbi.nlm.nih.gov/12668691/)]
10. Cantor MN, Feldman HJ, Triola MM. Using trigger phrases to detect adverse drug reactions in ambulatory care notes. *Qual Saf Health Care* 2007 Apr;16(2):132-134 [FREE Full text] [doi: [10.1136/qshc.2006.020073](https://doi.org/10.1136/qshc.2006.020073)] [Medline: [17403760](https://pubmed.ncbi.nlm.nih.gov/17403760/)]
11. Friedman C, Shagina L, Lussier Y, Hripscak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc* 2004;11(5):392-402 [FREE Full text] [doi: [10.1197/jamia.M1552](https://doi.org/10.1197/jamia.M1552)] [Medline: [15187068](https://pubmed.ncbi.nlm.nih.gov/15187068/)]
12. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc* 2010;17(3):229-236 [FREE Full text] [doi: [10.1136/jamia.2009.002733](https://doi.org/10.1136/jamia.2009.002733)] [Medline: [20442139](https://pubmed.ncbi.nlm.nih.gov/20442139/)]
13. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010;17(5):507-513 [FREE Full text] [doi: [10.1136/jamia.2009.001560](https://doi.org/10.1136/jamia.2009.001560)] [Medline: [20819853](https://pubmed.ncbi.nlm.nih.gov/20819853/)]
14. Murphy RM, Klopotoska JE, de Keizer NF, Jager KJ, Leopold JH, Dongelmans DA, et al. Adverse drug event detection using natural language processing: A scoping review of supervised learning methods. *PLoS One* 2023;18(1):e0279842 [FREE Full text] [doi: [10.1371/journal.pone.0279842](https://doi.org/10.1371/journal.pone.0279842)] [Medline: [36595517](https://pubmed.ncbi.nlm.nih.gov/36595517/)]

15. Jagannatha A, Liu F, Liu W, Yu H. Overview of the first natural language processing challenge for extracting medication, indication, and adverse drug events from electronic health record notes (MADE 1.0). *Drug Saf* 2019 Jan;42(1):99-111 [FREE Full text] [doi: [10.1007/s40264-018-0762-z](https://doi.org/10.1007/s40264-018-0762-z)] [Medline: [30649735](https://pubmed.ncbi.nlm.nih.gov/30649735/)]
16. Henry S, Buchan K, Filannino M, Stubbs A, Uzuner O. 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. *J Am Med Inform Assoc* 2020 Jan 01;27(1):3-12 [FREE Full text] [doi: [10.1093/jamia/ocz166](https://doi.org/10.1093/jamia/ocz166)] [Medline: [31584655](https://pubmed.ncbi.nlm.nih.gov/31584655/)]
17. McKnight W, Dolezal J. Healthcare natural language processing. GigaOm. URL: <https://research.gigaom.com/report/healthcare-natural-language-processing/> [accessed 2023-03-21]
18. Davies EC, Green CF, Taylor S, Williamson PR, Mottram DR, Pirmohamed M. Adverse drug reactions in hospital in-patients: a prospective analysis of 3695 patient-episodes. *PLoS One* 2009;4(2):e4439 [FREE Full text] [doi: [10.1371/journal.pone.0004439](https://doi.org/10.1371/journal.pone.0004439)] [Medline: [19209224](https://pubmed.ncbi.nlm.nih.gov/19209224/)]
19. Urman RD, Seger DL, Fiskio JM, Neville BA, Harry EM, Weiner SG, et al. The burden of opioid-related adverse drug events on hospitalized previously opioid-free surgical patients. *J Patient Saf* 2021 Mar 01;17(2):e76-e83 [doi: [10.1097/PTS.0000000000000566](https://doi.org/10.1097/PTS.0000000000000566)] [Medline: [30672762](https://pubmed.ncbi.nlm.nih.gov/30672762/)]
20. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016 May 24;3:160035 [FREE Full text] [doi: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35)] [Medline: [27219127](https://pubmed.ncbi.nlm.nih.gov/27219127/)]
21. Hougland P, Xu W, Pickard S, Masheter C, Williams SD. Performance of International Classification Of Diseases, 9th Revision, Clinical Modification codes as an adverse drug event surveillance system. *Med Care* 2006 Jul;44(7):629-636 [doi: [10.1097/01.mlr.0000215859.06051.77](https://doi.org/10.1097/01.mlr.0000215859.06051.77)] [Medline: [16799357](https://pubmed.ncbi.nlm.nih.gov/16799357/)]
22. Eyre H, Chapman AB, Peterson KS, Shi J, Alba PR, Jones MM, et al. Launching into clinical space with medspaCy: a new clinical text processing toolkit in Python. *AMIA Annu Symp Proc* 2021;2021:438-447 [FREE Full text] [Medline: [35308962](https://pubmed.ncbi.nlm.nih.gov/35308962/)]
23. Tang Y, Yang J, Ang PS, Dorajoo SR, Foo B, Soh S, et al. Detecting adverse drug reactions in discharge summaries of electronic medical records using Readpeer. *Int J Med Inform* 2019 Aug;128:62-70 [doi: [10.1016/j.ijmedinf.2019.04.017](https://doi.org/10.1016/j.ijmedinf.2019.04.017)] [Medline: [31160013](https://pubmed.ncbi.nlm.nih.gov/31160013/)]
24. Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: an algorithm for determining negation, experienter, and temporal status from clinical reports. *J Biomed Inform* 2009 Oct;42(5):839-851 [FREE Full text] [doi: [10.1016/j.jbi.2009.05.002](https://doi.org/10.1016/j.jbi.2009.05.002)] [Medline: [19435614](https://pubmed.ncbi.nlm.nih.gov/19435614/)]
25. Handler SM, Hanlon JT, Perera S, Roumani YF, Nace DA, Fridsma DB, et al. Consensus list of signals to detect potential adverse drug reactions in nursing homes. *J Am Geriatr Soc* 2008 May;56(5):808-815 [FREE Full text] [doi: [10.1111/j.1532-5415.2008.01665.x](https://doi.org/10.1111/j.1532-5415.2008.01665.x)] [Medline: [18363678](https://pubmed.ncbi.nlm.nih.gov/18363678/)]
26. Kane-Gill SL, Bellamy CJ, Verrico MM, Handler SM, Weber RJ. Evaluating the positive predictive values of antidote signals to detect potential adverse drug reactions (ADRs) in the medical intensive care unit (ICU). *Pharmacoepidemiol Drug Saf* 2009 Dec;18(12):1185-1191 [doi: [10.1002/pds.1837](https://doi.org/10.1002/pds.1837)] [Medline: [19728294](https://pubmed.ncbi.nlm.nih.gov/19728294/)]
27. Kocaman V, Talby D. Accurate Clinical and Biomedical Named Entity Recognition at Scale. *Software Impacts* 2022 Aug;13:100373 [FREE Full text] [doi: [10.1016/j.simpa.2022.100373](https://doi.org/10.1016/j.simpa.2022.100373)]
28. Neumann M, King D, Beltagy I. ScispaCy: fast and robust models for biomedical natural language processing. In: *Proceedings of the 18th BioNLP Workshop and Shared Task*. 2019 Presented at: 18th BioNLP Workshop and Shared Task; Aug 1, 2019; Florence, Italy p. 319-327 [doi: [10.18653/v1/W19-5034](https://doi.org/10.18653/v1/W19-5034)]
29. Tinoco A, Evans RS, Staes CJ, Lloyd JF, Rothschild JM, Haug PJ. Comparison of computerized surveillance and manual chart review for adverse events. *J Am Med Inform Assoc* 2011;18(4):491-497 [FREE Full text] [doi: [10.1136/amiajnl-2011-000187](https://doi.org/10.1136/amiajnl-2011-000187)] [Medline: [21672911](https://pubmed.ncbi.nlm.nih.gov/21672911/)]
30. Anthes AM, Harinstein LM, Smithburger PL, Seybert AL, Kane-Gill SL. Improving adverse drug event detection in critically ill patients through screening intensive care unit transfer summaries. *Pharmacoepidemiol Drug Saf* 2013 May;22(5):510-516 [doi: [10.1002/pds.3422](https://doi.org/10.1002/pds.3422)] [Medline: [23440931](https://pubmed.ncbi.nlm.nih.gov/23440931/)]
31. Aguilera C, Agustí A, Pérez E, Gracia RM, Diogène E, Danés I. Spontaneously reported adverse drug reactions and their description in hospital discharge reports: A retrospective study. *J Clin Med* 2021 Jul 26;10(15):3293 [FREE Full text] [doi: [10.3390/jcm10153293](https://doi.org/10.3390/jcm10153293)] [Medline: [34362076](https://pubmed.ncbi.nlm.nih.gov/34362076/)]
32. Luo Y, Thompson WK, Herr TM, Zeng Z, Berendsen MA, Jonnalagadda SR, et al. Natural language processing for EHR-based pharmacovigilance: A structured review. *Drug Saf* 2017 Nov;40(11):1075-1089 [doi: [10.1007/s40264-017-0558-6](https://doi.org/10.1007/s40264-017-0558-6)] [Medline: [28643174](https://pubmed.ncbi.nlm.nih.gov/28643174/)]
33. GPT-4 technical report. OpenAI. URL: <https://cdn.openai.com/papers/gpt-4.pdf> [accessed 2023-06-19]
34. Luo R, Sun L, Xia Y, Qin T, Zhang S, Poon H, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Brief Bioinform* 2022 Nov 19;23(6):bbac409 [doi: [10.1093/bib/bbac409](https://doi.org/10.1093/bib/bbac409)] [Medline: [36156661](https://pubmed.ncbi.nlm.nih.gov/36156661/)]
35. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. *NPJ Digit Med* 2022 Dec 26;5(1):194 [FREE Full text] [doi: [10.1038/s41746-022-00742-2](https://doi.org/10.1038/s41746-022-00742-2)] [Medline: [36572766](https://pubmed.ncbi.nlm.nih.gov/36572766/)]

Abbreviations

ADE: adverse drug event

ADR: adverse drug reaction
AE: adverse event
AI: artificial intelligence
BERT: bidirectional encoder representations from transformers
CDER: Center for Drug Evaluation and Research
CNN: convolutional neural network
CRF: conditional random field
cTAKES: Clinical Text Analysis and Knowledge Extraction System
EHR: electronic health record
FAERS: FDA Adverse Events Reporting System
FDA: US Food and Drug Administration
GPT: generative pretrained transformer
ICD-9: International Classification of Diseases, Ninth Revision
LLM: large language model
LLT: lower-level term
LSTM: long short-term memory
MedDRA: Medical Dictionary for Regulatory Activities
MIMIC: Medical Information Mart for Intensive Care
NER: named entity recognition
NLP: natural language processing
ORADE: opioid-related adverse drug event
POS: part of speech
PT: preferred term
SPINEL: Supporting Pharmacovigilance by Leveraging Artificial Intelligence Methods to Analyze Electronic Health Records Data
SVM: support vector machine
UMLS: Unified Medical Language System

Edited by K El Emam, B Malin; submitted 12.12.22; peer-reviewed by Q Ye, J Shi, D Chrimes; comments to author 06.02.23; revised version received 29.03.23; accepted 02.06.23; published 18.07.23

Please cite as:

*Sorbello A, Haque SA, Hasan R, Jermyn R, Hussein A, Vega A, Zembrzski K, Ripple A, Ahadpour M
Artificial Intelligence–Enabled Software Prototype to Inform Opioid Pharmacovigilance From Electronic Health Records: Development and Usability Study
JMIR AI 2023;2:e45000
URL: <https://ai.jmir.org/2023/1/e45000>
doi: [10.2196/45000](https://doi.org/10.2196/45000)
PMID: [37771410](https://pubmed.ncbi.nlm.nih.gov/37771410/)*

©Alfred Sorbello, Syed Arefinul Haque, Rashedul Hasan, Richard Jermyn, Ahmad Hussein, Alex Vega, Krzysztof Zembrzski, Anna Ripple, Mitra Ahadpour. Originally published in JMIR AI (<https://ai.jmir.org>), 18.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.ai.jmir.org/>, as well as this copyright and license information must be included.