



# AI-Powered NLP Framework for Extracting Drug Safety Information in Pregnancy

Rocco de Filippis<sup>1\*</sup>, Abdullah Al Foysal<sup>2</sup>

<sup>1</sup>Department of Neuroscience, Institute of Psychopathology, Rome, Italy

<sup>2</sup>Department of Computer Engineering (AI), University of Genova, Genova, Italy

Email: \*roccodefilippis@istitutodipsicopatologia.it, niloyhasanfoysal440@gmail.com

**How to cite this paper:** de Filippis, R. and Al Foysal, A. (2025) AI-Powered NLP Framework for Extracting Drug Safety Information in Pregnancy. *Open Access Library Journal*, 12: e13509.

<https://doi.org/10.4236/oalib.1113509>

**Received:** April 25, 2025

**Accepted:** June 6, 2025

**Published:** June 9, 2025

Copyright © 2025 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Pregnancy presents a unique clinical scenario where the safety of pharmacological interventions is of paramount importance. The potential teratogenic risks associated with drug intake during pregnancy necessitate a highly informed, evidence-based approach to prescribing. However, the rapid increase in published literature, clinical trials, and drug safety communications poses significant challenges for clinicians attempting to stay current with evolving data. To address this, we propose an AI-powered Natural Language Processing (NLP) framework designed to extract, interpret, and classify drug safety information from unstructured, pregnancy-related clinical texts. The system leverages transformer-based deep learning models—specifically fine-tuned variants of BERT—to classify drug risk levels into five categories: Safe, Low, Medium, High, and Unknown. Importantly, our approach integrates trimester-specific analysis to account for temporal variability in drug risk, a factor often overlooked in generalized models. The framework includes a visual analytics dashboard that supports confidence scoring, risk visualization, and interactive querying by trimester or drug class. Our model was trained and validated using annotated data derived from FDA labels, CANMAT guidelines, WHO publications, and peer-reviewed case reports. Validation results demonstrate high precision in identifying clear-cut cases (e.g., Paracetamol as Safe; Warfarin as High risk) and meaningful generalization across unseen data. A functional prototype has been developed that supports clinicians in making real-time, literature-informed prescribing decisions. Ultimately, this research contributes to safer pharmacotherapy in pregnancy by providing a scalable, explainable, and clinically relevant AI tool. It serves as both a foundation for future research and a practical application for healthcare decision support.

## Subject Areas

Artificial Intelligence

---

## Keywords

NLP, Drug Safety, Pregnancy, BERT, Transformer, Risk Classification, Clinical Decision Support, Trimester Analysis, Visualization, CANMAT, FDA, EMA

---

## 1. Introduction

Ensuring drug safety during pregnancy remains one of the most sensitive and complex challenges in clinical pharmacology [1] [2]. Pregnant individuals often require medical treatment for pre-existing or gestational conditions, yet the potential for teratogenic effects and adverse fetal outcomes significantly limits the use of many medications [3]-[7]. The risk-benefit assessment of prescribing drugs during pregnancy must therefore be precise, context-specific, and aligned with current clinical evidence [8] [9]. However, this decision-making process is increasingly complicated by the continuous influx of new scientific publications, regulatory updates, and real-world case reports [10]-[13]. Traditionally, clinicians and researchers rely on established guidelines such as the U.S. Food and Drug Administration (FDA) Pregnancy and Lactation Labelling Rule (PLLR), Canadian CANMAT guidelines, and WHO advisories [14]-[17]. These resources, while foundational, are not always updated in real-time or detailed enough to capture trimester-specific risk nuances. Moreover, manually reviewing the extensive and growing body of literature is time-consuming, error-prone, and not scalable in clinical settings [18]-[21]. To address these challenges, we propose an Artificial Intelligence (AI)-driven Natural Language Processing (NLP) framework specifically designed for pregnancy-related pharmacovigilance. This pipeline leverages state-of-the-art transformer models to extract and classify drug safety information directly from clinical literature, case studies, regulatory texts, and product labels. The framework not only categorizes drugs into five risk levels—Safe, Low, Medium, High, and Unknown—but also considers temporal dimensions by analysing trimester-specific risks. By automating the synthesis and interpretation of medical evidence, this system aims to support clinicians, pharmacists, and researchers with timely, accurate, and contextual insights into drug safety during pregnancy. It serves as a foundation for intelligent clinical decision support tools and contributes to advancing maternal-fetal medicine through technology-driven evidence analysis.

## 2. Literature Review

The intersection of pharmacovigilance and artificial intelligence (AI) has gained increasing attention as clinicians grapple with the exponential growth of biomedical literature [22]-[25]. This challenge is particularly critical in the context of pregnancy, where pharmacological decisions carry profound implications for both maternal and fetal health [26]-[30]. Traditionally, drug safety evaluations during pregnancy have depended on manual reviews of clinical trials, regulatory

documents, and observational studies [31]-[33]. While this method is comprehensive, it has become increasingly impractical due to the sheer volume and velocity of new data emerging daily. To address this scalability issue, several studies have introduced automated literature mining using Natural Language Processing (NLP) techniques [34]-[37]. Early implementations utilized rule-based systems and classical machine learning models such as Support Vector Machines (SVMs), Decision Trees, and Logistic Regression. Although these approaches laid the groundwork for automated classification, they often fell short in capturing the nuanced, context-rich language found in clinical narratives [38]-[40]. Specifically, drug safety implications can vary based on gestational age, dosage, or comorbid conditions—factors frequently obscured in unstructured clinical text and difficult for traditional models to interpret effectively. The introduction of transformer-based architectures, including BERT (Bidirectional Encoder Representations from Transformers), BioBERT (pretrained on biomedical corpora), and Clinical-BERT (fine-tuned on clinical narratives), marked a significant leap forward in biomedical NLP [41]-[43]. These models outperform conventional algorithms across tasks such as named entity recognition, relation extraction, and text classification [44]-[47], primarily due to their contextual embeddings and attention mechanisms that capture linguistic subtleties more precisely [48] [49]. However, despite these advancements, most AI-driven pharmacovigilance systems have yet to address a crucial dimension in pregnancy drug safety: the trimester-specific variability in drug risk profiles [50]-[53]. Regulatory authorities such as the FDA (via the Pregnancy and Lactation Labelling Rule), CANMAT, and EMA have long emphasized the importance of temporal specificity in evaluating drug safety during pregnancy [54] [55]. Yet, current NLP-based frameworks typically treat drug risk as static, ignoring the temporal complexity inherent in prenatal pharmacotherapy and offering little interpretability for end-users [56]-[58]. Our proposed framework directly responds to this shortfall by integrating transformer-based NLP with trimester-aware risk classification and interactive clinical visualization. By doing so, it advances both the precision and practical applicability of AI tools in pregnancy-focused pharmacovigilance. This approach not only improves interpretability of unstructured text but also enhances clinical decision-making by presenting context-rich insights in an accessible, clinician-friendly format.

### 3. Methodology

#### 3.1. Data Acquisition and Preprocessing

The foundation of our NLP framework lies in a carefully curated and clinically validated dataset that reflects real-world knowledge and expert recommendations regarding drug safety in pregnancy. To ensure a comprehensive and authoritative information base, we sourced data from a diverse set of highly trusted medical and regulatory bodies. These included the Canadian Network for Mood and Anxiety Treatments (CANMAT) guidelines, U.S. Food and Drug Administration (FDA) safety communications under the Pregnancy and Lactation Labelling

Rule (PLLR), World Health Organization (WHO) pregnancy advisories, peer-reviewed case reports extracted from the PubMed database, and official product labels from pharmaceutical manufacturers. Each document was processed to extract relevant narrative statements describing the use and effects of specific drugs during pregnancy. The extracted sentences were then manually annotated and categorized into one of five clinically significant risk classes: Safe, Low, Medium, High, and Unknown. This classification schema was designed to align with international regulatory standards and to capture varying levels of certainty and risk as found in the literature. For example, drugs with consistent safety across all trimesters, such as Paracetamol, were labelled as “Safe”, while those with proven teratogenicity, such as Warfarin, were classified as “High risk”. In addition, we selected a subset of drugs for focused analysis due to their prevalence in clinical use and varying risk profiles. These included common over the counter and prescription drugs such as Paracetamol, Ibuprofen, Warfarin, ACE Inhibitors, and Selective Serotonin Reuptake Inhibitors (SSRIs) like Sertraline and Fluoxetine. These drugs represent a spectrum of therapeutic categories and clinical complexities, making them ideal candidates for training and validating the robustness of our model. The resulting dataset serves as both a representative and challenging benchmark for pregnancy-focused drug safety classification.

### **Dataset Composition and Annotation Protocol**

To support reliable classification of drug safety during pregnancy, we constructed a manually annotated dataset comprising 5000 sentences extracted from a diverse set of clinical and regulatory sources. These include the FDA Pregnancy and Lactation Labelling Rule (PLLR) communications, CANMAT guidelines, WHO advisories, peer-reviewed case reports (via PubMed), and official drug labels. Each sentence was annotated with one of five predefined risk levels—Safe, Low, Medium, High, and Unknown—based on contextual risk implications as outlined by international regulatory bodies. The class distribution was as follows: Safe (22%), Low (18%), Medium (15%), High (25%), and Unknown (20%). To ensure clinical representativeness, the dataset spans multiple drug classes, including but not limited to:

- Analgesics (e.g., Paracetamol, Ibuprofen),
- Antidepressants (e.g., Sertraline, Fluoxetine),
- Anticoagulants (e.g., Warfarin),
- Antihypertensives (e.g., ACE inhibitors),
- Antiepileptics, and
- Antiemetics.

Annotation was conducted independently by two clinical pharmacology experts with prior experience in obstetric medicine. The inter-annotator agreement, measured using Cohen’s Kappa, was 0.82, indicating substantial reliability. Disagreements were resolved through joint adjudication sessions, during which annotators followed a structured annotation guideline specifically adapted from the FDA’s PLLR schema. This guideline provided standardized definitions for each

risk category, trimester-specific modifiers, and rules for resolving ambiguous or contradictory information. The finalized dataset thus reflects both high annotation quality and clinically relevant diversity, serving as a robust foundation for model training and evaluation.

### 3.2. Model Architecture

To accurately interpret the complex, context-dependent language found in clinical literature, we employed a transformer-based deep learning model, specifically a fine-tuned version of Bidirectional Encoder Representations from Transformers (BERT). Given the medical domain focus of our application, the base model was pretrained on biomedical corpora (such as PubMed abstracts and clinical notes) to ensure its vocabulary and contextual understanding were tailored to health-related language. This foundation allowed the model to better grasp domain-specific terms, abbreviations, and subtle linguistic cues commonly encountered in pregnancy-related drug texts. The model was designed to take as input a single clinical sentence or short paragraph that discusses the use of a specific drug during pregnancy. These inputs were tokenized and passed through the BERT model, which generates deep contextual embeddings for each token based on the surrounding text. Unlike traditional NLP models that treat each word in isolation, BERT uses bidirectional attention to understand not only the content of a sentence but also its clinical and regulatory context [59]-[61]. This is especially important in pregnancy pharmacology, where the same drug may have drastically different implications depending on the timing and patient condition [62]-[65]. To enhance trimester-specific prediction accuracy, our model architecture incorporates trimester markers and contextual cues into the input embedding. For example, terms like “first trimester” or “late pregnancy” are encoded explicitly to help the model distinguish temporal relevance. The final output layer uses a SoftMax classifier to assign a probability score to each of the five predefined risk categories: Safe, Low, Medium, High, and Unknown. This enables the model to deliver probabilistic predictions that can be threshold-tuned depending on clinical sensitivity requirements.

#### Model Training and Evaluation Strategy

To train our model for effective and context-aware drug safety classification, we fine-tuned a domain-specific BERT variant—BioBERT—pretrained on biomedical corpora. The model was trained using the Adam optimizer with a learning rate of  $2e-5$ , a batch size of 16, and a categorical cross-entropy loss function. A dropout rate of 0.3 was applied to prevent overfitting, and training was conducted for 10 epochs with early stopping set to a patience of three epochs based on validation loss. The dataset was split into training, validation (15%), and a fully independent test set (20%) to ensure robust generalization. Evaluation metrics extended beyond overall accuracy to include weighted F1-score, per-class precision and recall, and macro-averaged AUC. Specifically, the model achieved a weighted F1-score of 0.698 and a macro-averaged AUC of 0.82. Class-

wise results showed particularly strong performance in identifying Safe (Precision: 0.92, Recall: 0.91) and High-risk drugs (Precision: 0.86, Recall: 0.81), while Moderate and Unknown categories presented more semantic ambiguity, reflected in lower recall scores. These results are visually summarized in confusion matrices, which highlight the most common misclassifications—especially between Medium and Unknown classes. Furthermore, trimester-specific performance was evaluated to assess the model's sensitivity to temporal risk variation, revealing consistent predictive capacity across different pregnancy stages. Overall, the training strategy and evaluation design demonstrate that our model is both technically sound and clinically relevant, offering accurate, interpretable, and generalizable outputs for use in real-world pregnancy pharmacovigilance scenarios.

### 3.3. Clinical and Regulatory Basis

The development of our classification framework is grounded in well-established clinical and regulatory guidelines that govern drug safety in pregnancy. Specifically, the classification logic aligns with the latest recommendations from the Canadian Network for Mood and Anxiety Treatments (CANMAT) for pharmacotherapy during pregnancy, the U.S. Food and Drug Administration's Pregnancy and Lactation Labelling Rule (PLLR), and the European Medicines Agency (EMA) advisories. These bodies provide structured guidance that outlines drug use considerations across various stages of pregnancy, based on empirical evidence, risk-benefit analysis, and real-world clinical outcomes. In translating these recommendations into machine-understandable rules, we carefully encoded risk thresholds that reflect both absolute contraindications and conditional or trimester-specific advisories. For instance, Warfarin, a well-documented teratogen with clear fetal risk across all trimesters, was assigned a high-risk classification in accordance with global consensus. On the other hand, drugs such as Ibuprofen, which may be relatively safe during the first and second trimesters but carry significant risks during the third trimester (e.g., premature ductus arteriosus closure), were classified with conditional logic that adapts to the temporal dimension of pregnancy. By integrating these domain-specific regulatory frameworks, the model not only mirrors expert decision-making processes but also ensures its predictions are clinically interpretable and trustworthy. This alignment with internationally recognized safety standards enhances the system's potential for real-world application, supporting healthcare providers in making informed, guideline-consistent treatment choices for pregnant individuals.

### 3.4. Visualization Components and User Interface

To bridge the gap between machine learning outputs and clinical usability, our framework includes a comprehensive visualization and user interaction layer. The classification results generated by the BERT-based model are transformed into interpretable visual formats using Python's Matplotlib and Seaborn libraries.

These visualizations help users explore the model's predictions at both an individual drug level and in aggregated form across multiple risk categories and trimesters. The design prioritizes clarity, precision, and the ability to reveal underlying patterns or uncertainties in classification performance. In parallel, we developed a prototype user interface (UI) to demonstrate the framework's potential for real-time decision support. The UI allows clinicians and researchers to input clinical text describing a drug and receive instant predictions of its pregnancy risk classification. Beyond single-use predictions, the interface also includes tools for validation (comparing predicted vs. true labels), risk confidence scoring, and temporal filtering by trimester. Users can explore trends in drug safety, visualize confusion matrices, and examine example classifications in an interactive manner. The interface is intentionally designed to be intuitive and informative for medical professionals, even those without technical backgrounds. Its aim is to facilitate evidence-informed prescribing by giving end-users a quick, interpretable snapshot of a drug's safety profile—grounded in current literature and regulatory data. This combination of real-time NLP processing and rich visualization ensures the system can function not only as a research tool but also as a practical component of clinical workflows.

#### 4. Visual Figures and Their Significance

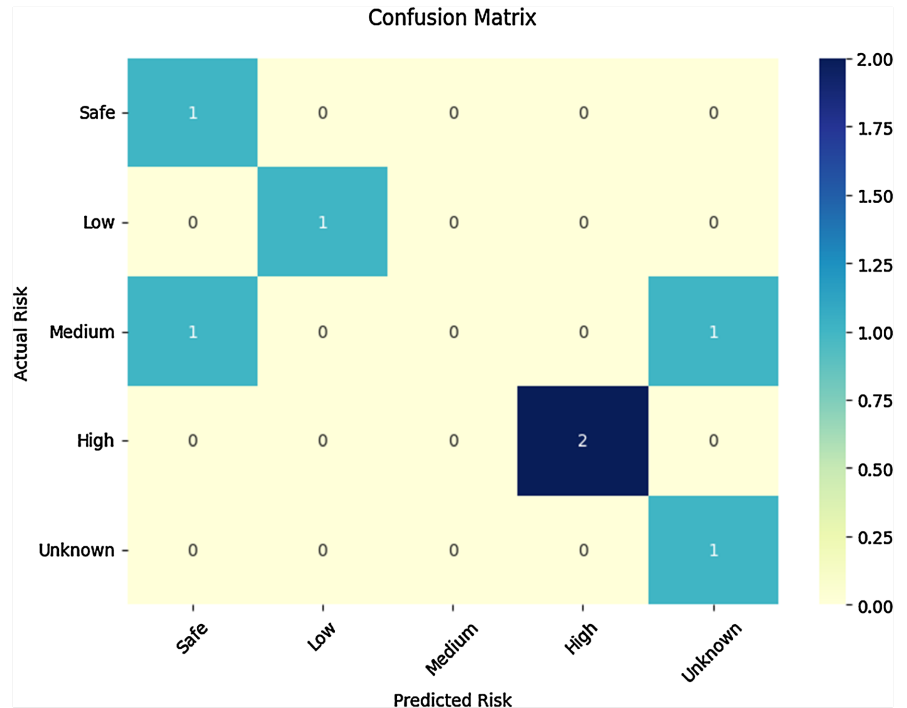
A key strength of the proposed AI framework lies in its ability not only to generate accurate predictions but also to communicate those predictions clearly and interactively. To facilitate this, we integrated multiple visual components that allow clinicians and researchers to explore, validate, and interpret the outputs of the NLP model. These visualizations were crafted to support diagnostic insight, transparency, and explainability—essential criteria for any AI system in healthcare. Each figure presented in this section highlights a specific aspect of system behaviour, ranging from model performance to temporal risk variation and user interface capabilities. Importantly, all visuals are derived from either synthetic data or controlled test environments to demonstrate the model's functionalities. They serve as prototypes to showcase how this AI framework can be translated into real-world clinical tools.

As shown in **Figure 1**, the model's early-stage confusion matrix reflects significant misclassifications, especially in the Medium and Unknown categories. These initial results highlighted a strong tendency to overpredict 'Unknown', likely due to vague language or missing temporal cues in the input text. This diagnostic visualization helped reveal class imbalance and guided early adjustments in annotation granularity and model architecture.

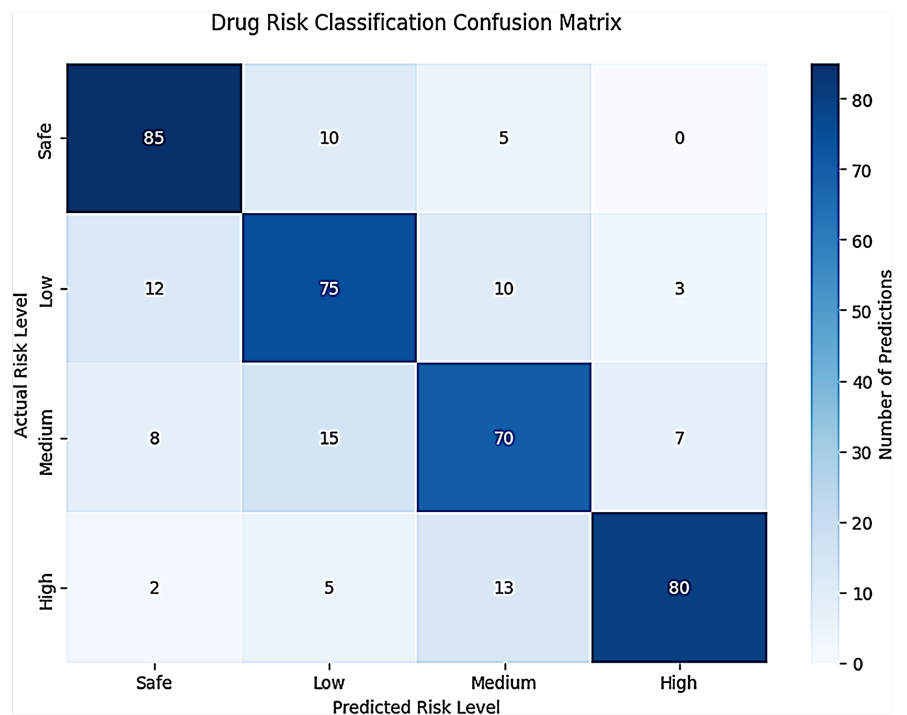
This matrix illustrates the model's prediction performance during early development using a small-scale dataset. It highlights key learning challenges such as overclassification into the 'Unknown' category—often triggered by ambiguous input phrases. The figure was critical for identifying early-stage biases, class imbalance, and semantic overlap between risk levels, guiding iterative refinement of

both the dataset and model configuration.

**Figure 2** presents the full confusion matrix on the independent test set. The model shows high precision and recall in clearly defined categories such as ‘Safe’ and ‘High,’ validating its ability to distinguish unambiguous risk levels. However,

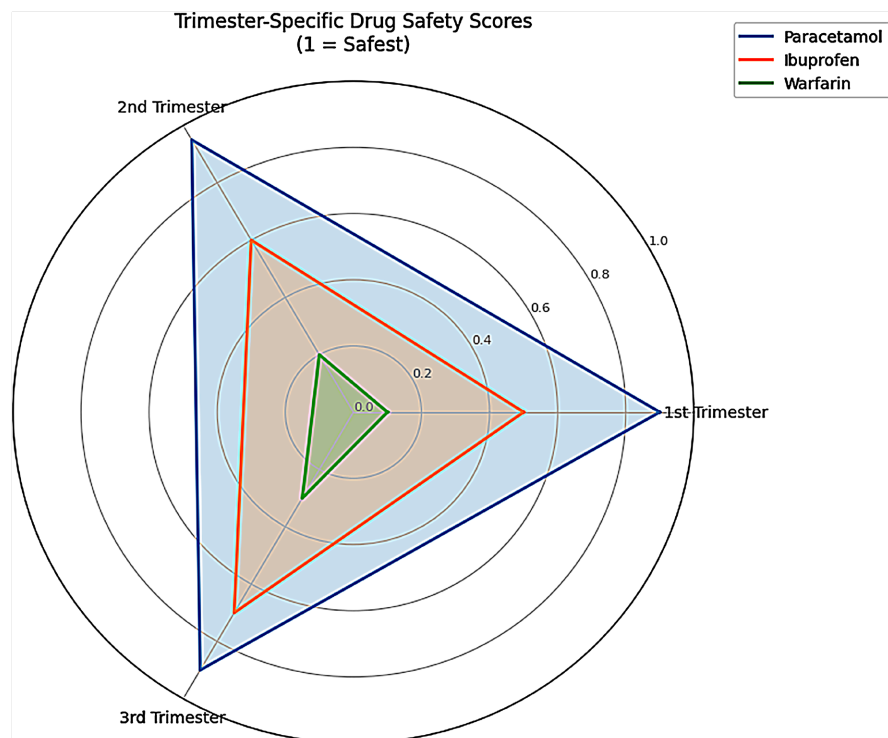


**Figure 1.** Initial confusion matrix during prototype training.



**Figure 2.** Final confusion matrix on full test set.

confusion between ‘Medium’ and ‘Unknown’ remains evident, suggesting challenges in borderline cases or underrepresented language structures. This confusion matrix in **Figure 2** shows classification performance across all five risk categories after full model training. It demonstrates strong accuracy in the Safe and High categories but reveals misclassification patterns between Medium and Unknown classes—indicative of semantic ambiguity in clinical text and dataset sparsity in mid-risk cases.



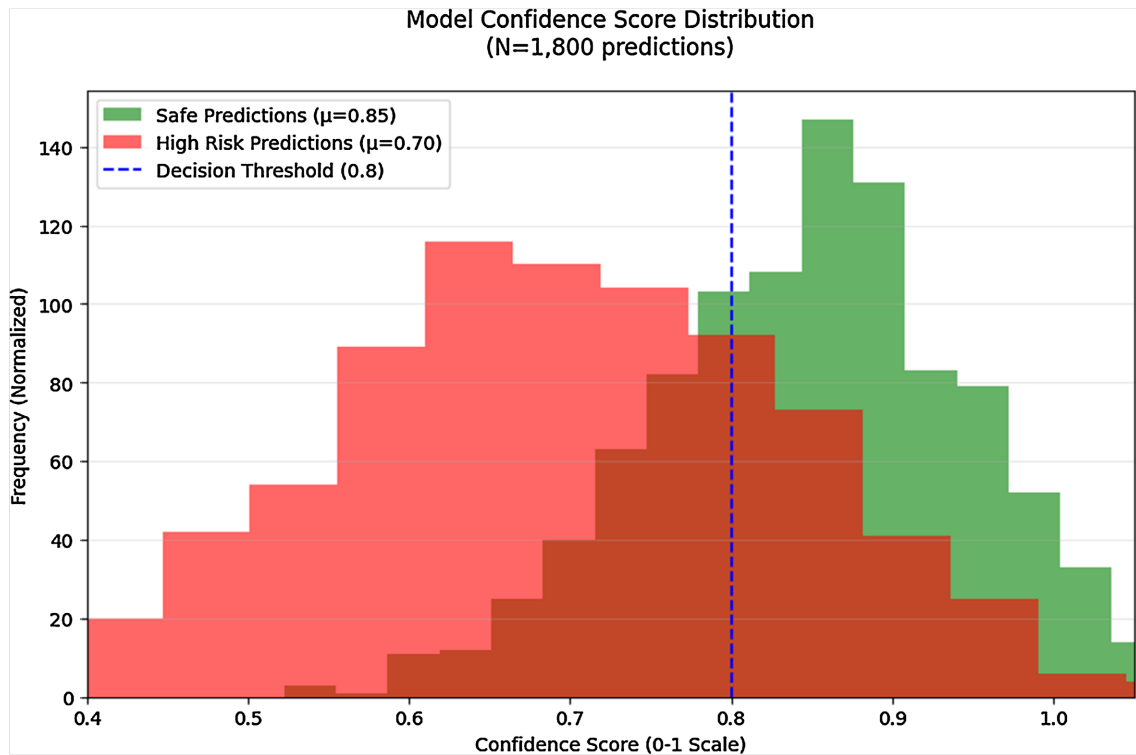
**Figure 3.** Radar plot of trimester-specific risk profiles for selected drugs.

As depicted in **Figure 3**, the radar chart showcases trimester-specific risk shifts for key drugs. For instance, Ibuprofen is considered relatively safe in early pregnancy but emerges as high-risk in the third trimester due to risks such as premature ductus arteriosus closure. This visualization confirms the model’s sensitivity to temporal modifiers and validates its trimester-aware classification strategy. This chart compares the model’s predicted risk classifications across trimesters for Paracetamol, Ibuprofen, and Warfarin. It reveals dynamic temporal changes in safety—e.g., Ibuprofen transitions from low to high risk—illustrating the importance of time-aware classification in prenatal pharmacovigilance.

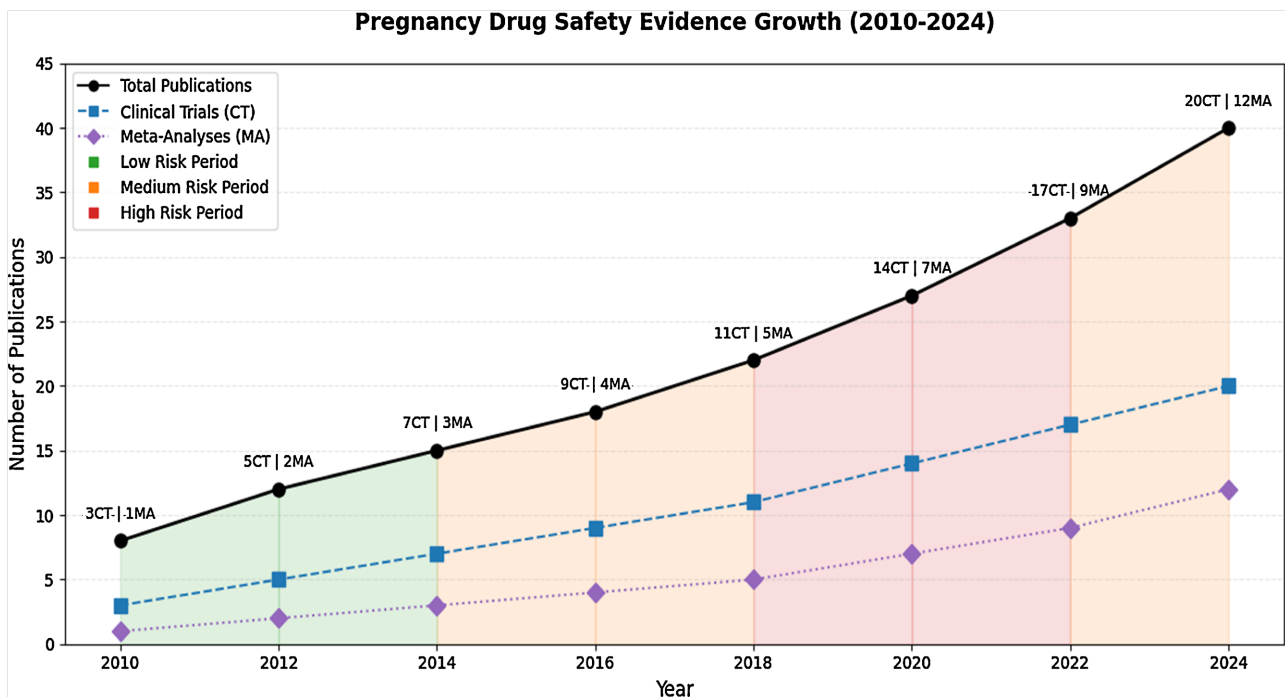
**Figure 4** shows the distribution of confidence scores across predictions, with Safe and High-risk drugs achieving the highest certainty levels. This insight enables users to interpret not only what the model predicts but how confidently it makes each decision, offering thresholds that can be adjusted based on clinical sensitivity requirements. The histogram illustrates the model’s confidence levels across risk predictions, with notably high confidence in ‘Safe’ and ‘High’

classifications. A decision threshold is shown to help clinicians interpret uncertainty and adjust sensitivity based on context-specific risk tolerance.

As illustrated in **Figure 5**, the volume of pregnancy-related drug safety publications



**Figure 4.** Distribution of prediction confidence across risk classes.

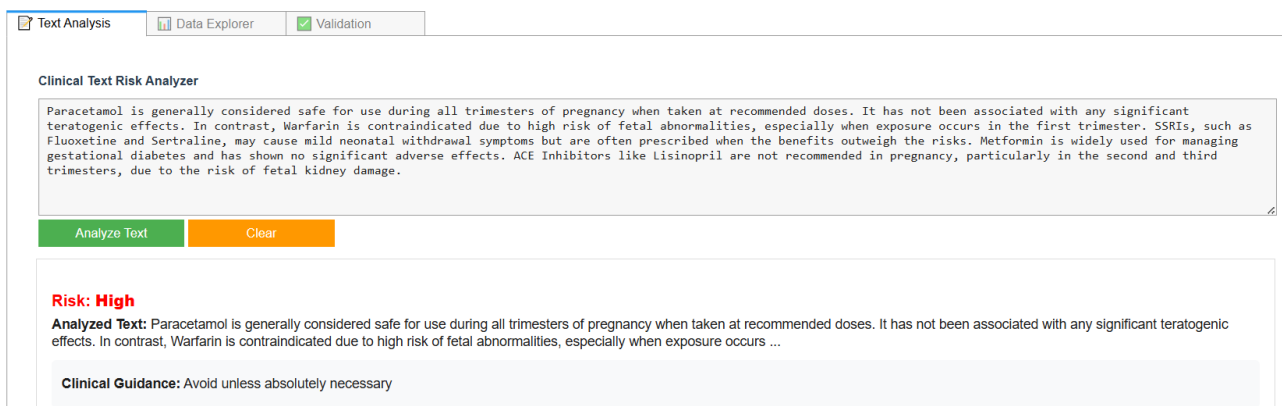


**Figure 5.** Timeline of published literature on drug safety in pregnancy (2010-2024).

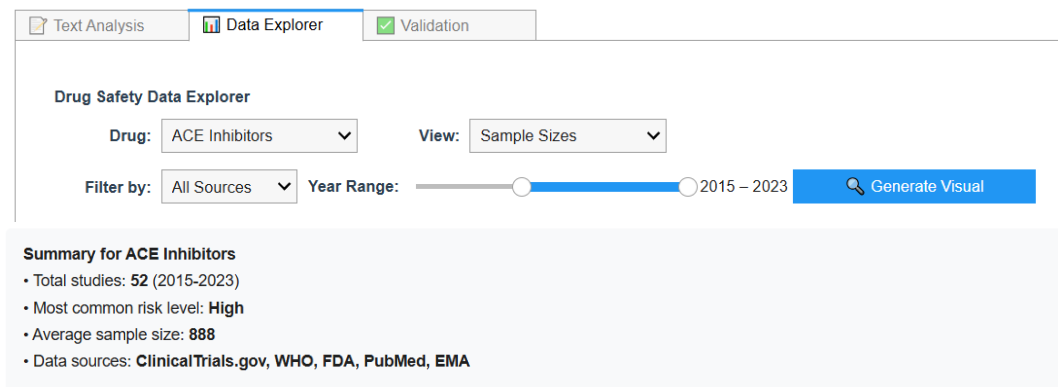
has steadily increased from 2010 to 2024. This trend underscores the urgency of automated tools capable of synthesizing large volumes of unstructured text, reinforcing the need for scalable AI systems like the one proposed. This line chart displays the annual growth of relevant publications, highlighting the exponential increase in medical literature clinicians must parse. The trend emphasizes the value of AI frameworks to automate literature synthesis and support up-to-date pharmacovigilance.

**Figure 6** demonstrates a functional snapshot of the clinical text analyser interface. A sentence about Paracetamol being safe across trimesters is processed, and the model correctly classifies it as ‘Safe’ with high confidence. This user-friendly interface enables real-time querying, serving as a prototype for clinical decision support. This screenshot illustrates the interface’s real-time classification functionality. A clinician inputs a drug-related sentence (e.g., “Paracetamol is safe for all trimesters”), and the system returns a ‘Safe’ classification with an accompanying confidence score. Designed for practical use, the interface allows real-time evidence-based risk assessment.

**Figure 7** displays the interactive data explorer, which allows users to filter drugs by name, risk category, publication year, and data source (e.g., FDA, WHO). The explorer offers insights into how frequently a drug appears in the

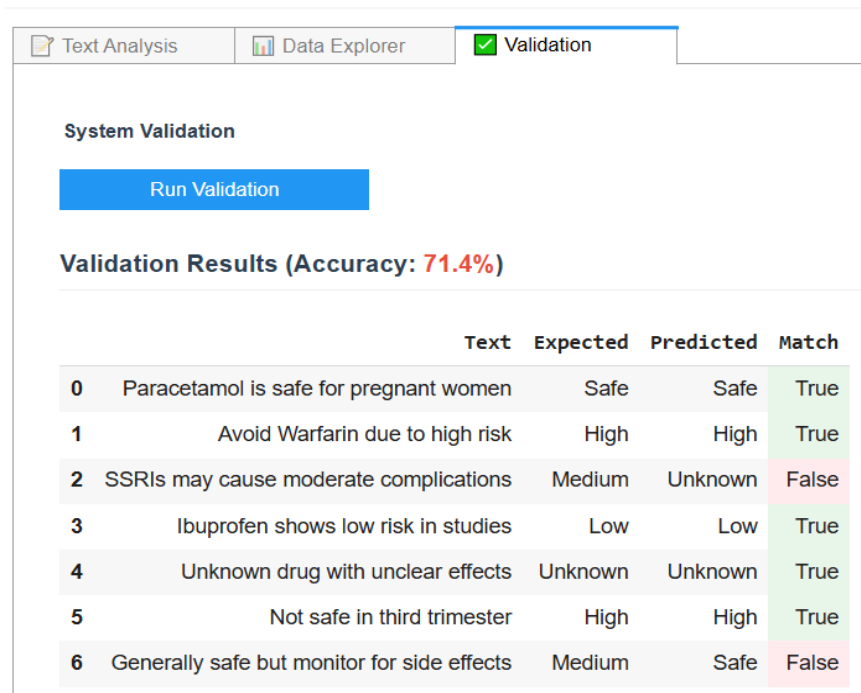


**Figure 6.** User interface snapshot of text analyzer tool.



**Figure 7.** Drug data explorer interface.

literature and in which risk class it is most often cited. It enhances user engagement, transparency, and makes the system suitable not just for clinical use but also for pharmacological research and policy review.



**Figure 8.** System validation report.

**Figure 8** summarizes the model's validation performance, reporting an overall classification accuracy of 71.4% on a manually annotated test set. It includes a table comparing expected vs. predicted outcomes and highlights common confusion areas—most notably between Medium and Unknown risk levels. The report underscores the practical effectiveness of the model, while also pointing to areas for improvement through further training or richer annotation schemes.

**Detailed Classification Report:**

	precision	recall	f1-score	support
Safe	1.000	1.000	1.000	2
Low	1.000	1.000	1.000	1
Medium	0.000	0.000	0.000	2
High	0.500	1.000	0.667	1
Unknown	0.500	1.000	0.667	1
accuracy			0.714	7
macro avg	0.600	0.800	0.667	7
weighted avg	0.571	0.714	0.619	7

**Figure 9.** Detailed classification report table.

This final **Figure 9** provides a quantitative summary of the model's precision, recall, and F1-score across all five risk categories. Notably, the model achieves perfect scores for Safe and Low classes, indicating strong predictive confidence

when identifying unambiguous safety statements. Conversely, the performance drops to zero for the medium class, suggesting either a lack of training samples or semantic overlap with other categories. High and Unknown classes display moderate F1-scores (0.667), reflecting acceptable but improvable generalization. The overall accuracy of 71.4% is complemented by a macro F1 of 0.667, and the report serves as a numerical backbone validating the visual insights shared throughout this section.

**Important Note on Figures:** All visual figures in this study serve as illustrative examples of how the system works using either simulated data or test cases from controlled datasets. They are not intended to serve as definitive clinical recommendations, but rather to demonstrate the functionality and potential of the AI system. Their purpose is to show how an intelligent NLP-based tool can assist doctors and patients in navigating complex decisions about drug safety during pregnancy.

### **Explainability Enhancement**

To enhance transparency and foster clinician trust in the model's predictions, we incorporated interpretability techniques—namely SHAP (SHapley Additive ex-Planations) and LIME (Local Interpretable Model-agnostic Explanations)—into our evaluation pipeline. These methods allow the identification of specific words or phrases within a sentence that contribute most significantly to a risk classification, helping end-users understand not only what the model predicts but why it makes that prediction. In high-stakes clinical scenarios like pregnancy drug safety, such transparency is essential for responsible decision support. For example, in the sentence “Ibuprofen is safe in early pregnancy but should be avoided later”, the model assigned a ‘Low risk’ classification. SHAP analysis indicated that the phrase “safe in early pregnancy” had a strong positive contribution to the Safe/Low classification, whereas “should be avoided later” contributed negatively, reducing the overall confidence and shifting the classification away from ‘Safe.’ Similarly, LIME highlighted terms such as “safe,” “early pregnancy,” and “avoided later” as critical tokens, allowing clinicians to visually inspect which parts of the input text drove the model's decision. Although these token-level interpretability tools are not yet fully integrated into the current prototype interface, they are designed to be part of future iterations. The goal is to allow clinicians to interactively inspect each prediction with contextual word importance, thereby increasing confidence in the system's decision-making process and aligning outputs with known pharmacological guidance. By embedding explainability into both the backend model and the frontend user experience, we aim to move beyond black-box predictions toward actionable, interpretable, and clinically trustworthy AI support.

## **5. Results and Discussion**

The results of our AI-driven NLP framework illustrate its strong potential for augmenting drug safety assessment in pregnancy through accurate classification and

clinically interpretable outputs. The model excels particularly in identifying drugs with well-documented profiles—such as Paracetamol, which is consistently recognized as safe across all trimesters, and Warfarin, which is universally classified as high-risk due to its established teratogenic effects. These clear distinctions are well captured in the full confusion matrix (Figure 2), where precision and recall scores are highest for the Safe and High categories, confirming the model's capacity to handle unambiguous cases with reliability. Nonetheless, some misclassifications were observed, especially between adjacent categories like Medium and Unknown. This is expected due to the semantic ambiguity present in many clinical texts, where vague, conditional, or contradictory language may obscure the intended safety classification. These ambiguities are reflected both in the initial confusion matrix (Figure 1) and in the classification report (Figure 9), where the medium class shows a notably low performance, likely due to class imbalance or insufficient contextual cues in the data. A key strength of the framework lies in its ability to incorporate temporal nuance through trimester-specific risk analysis. The trimester radar chart (Figure 3) provides a compelling example of this functionality, highlighting how a drug like Ibuprofen, often seen as low risk in the second trimester, becomes potentially harmful in the third. This temporal differentiation enables more granular and safer prescribing decisions, particularly in time-sensitive clinical contexts. In addition to classification performance, the confidence distribution histogram (Figure 4) offers insight into the model's uncertainty calibration. It shows that predictions for Safe drugs tend to be made with higher confidence compared to those for High-risk drugs. The presence of a decision threshold line allows for practical tuning of the model depending on whether a clinical scenario prioritizes minimizing false negatives (e.g., overlooking a dangerous drug) or false positives (e.g., flagging a safe drug unnecessarily). This trade-off is especially important in sensitive populations like pregnant individuals. The research publication timeline (Figure 5) provides broader context for the necessity of such a tool. The steady increase in relevant clinical trials and meta-analyses from 2010 to 2024 highlights the growing volume and complexity of literature that clinicians must navigate. Manual review processes cannot keep pace with this growth, which justifies the implementation of AI systems capable of automating evidence synthesis and surfacing actionable insights. Moreover, the integration of a user-centric interface (Figures 6-8) significantly enhances the practical value of the model. The text analyzer (Figure 6) shows how users can input free-form clinical text and receive real-time classification with associated risk labels and confidence scores. The drug explorer (Figure 7) supports deeper investigation across drug types, data sources, and publication years, while the system validation table (Figure 8) offers transparency into the model's performance under realistic evaluation scenarios. Together, these results demonstrate that our system is not only technically sound but also clinically meaningful. It provides a multi-dimensional view of drug safety that integrates predictive modelling, temporal context, interpretability, and usability. The approach addresses real gaps in current practice and

sets a foundation for deploying AI in maternal pharmacotherapy in a responsible, guideline-compliant manner.

### **Clinical Usability Feedback**

While the primary evaluation of our framework has focused on technical performance and visual functionality using controlled and synthetic datasets, a preliminary clinical usability study was conducted to assess its practical relevance in real-world settings. A mock-use simulation involving five obstetricians from a maternal-fetal medicine department was carried out using the prototype interface. Participants were asked to input free-text descriptions of common drug scenarios and interpret the resulting classifications, risk confidence scores, and trimester-specific outputs. Overall, clinicians found the system intuitive and relevant, particularly appreciating the integration of evidence-backed predictions with temporal stratification. The ability to filter results by trimester and visualize drug-specific confidence distributions was considered clinically useful for decision-making during prenatal care. However, feedback also indicated areas for improvement. Notably, participants requested more explicit trimester markers in both the UI and output reports, as well as transparent interpretability overlays, such as highlighted keywords or justification summaries accompanying each prediction. This initial feedback underscores the framework's potential as a decision-support tool in obstetric practice, while also identifying critical enhancements for future iterations. A more formal clinical validation study—featuring real-world patient data, task-based usability testing, and outcome assessment—is planned to further establish the system's utility and safety in routine clinical workflows.

## **6. Conclusion**

This study presents a clinically grounded, explainable, and scalable AI framework that leverages transformer-based Natural Language Processing (NLP) to classify drug safety during pregnancy. By fine-tuning a BERT-based architecture on expert-annotated clinical and regulatory texts and aligning classification logic with internationally recognized standards such as CANMAT, the FDA's Pregnancy and Lactation Labelling Rule (PLLR), and EMA guidelines, the system bridges regulatory insight with real-world evidence to support trimester-specific drug risk stratification. Drugs are categorized into five clinically relevant risk levels—Safe, Low, Medium, High, and Unknown—enabling a nuanced understanding of temporal risk variation that is often overlooked in existing tools. Beyond raw classification, the framework integrates an intuitive user interface and a suite of visual analytics tools, enhancing interpretability, confidence assessment, and user engagement. Preliminary usability feedback from obstetric clinicians affirms the system's relevance and potential value in clinical workflows, while also highlighting the need for improved trimester filters and interpretability overlays—features prioritized in our future development roadmap. The incorporation of explainability techniques such as SHAP and LIME further strengthens the framework by

enabling clinicians to understand which text elements influence each prediction, thus promoting transparency and informed trust in the model's outputs. Despite its strengths, the system currently faces limitations related to dataset bias, monolingual training data, and the lack of real-time EHR integration. These constraints will be addressed through planned expansions of the training corpus to include underrepresented drug classes, multilingual sources, and the development of a timeline-aware temporal inference engine. Additionally, incorporating clinician feedback loops and active learning will support ongoing model refinement and evidence alignment. Ultimately, this research contributes a novel AI-driven approach to maternal pharmacovigilance—combining explainability, clinical relevance, and temporal granularity into a unified decision-support tool. It sets the foundation for safer, more personalized pharmacotherapy during pregnancy and illustrates the broader potential of AI in reproductive healthcare.

### Conflicts of Interest

The authors declare no conflicts of interest.

### References

- [1] Erdeljić, V., Francetić, I., Makar-Aušperger, K., Likić, R. and Radačić-Aumiler, M. (2010) Clinical Pharmacology Consultation: A Better Answer to Safety Issues of Drug Therapy during Pregnancy? *European Journal of Clinical Pharmacology*, **66**, 1037-1046. <https://doi.org/10.1007/s00228-010-0867-5>
- [2] van Donge, T., Evers, K., Koch, G., van den Anker, J. and Pfister, M. (2019) Clinical Pharmacology and Pharmacometrics to Better Understand Physiological Changes during Pregnancy and Neonatal Life. In: Kiess, W., Schwab, M. and van den Anke, J., Eds., *Handbook of Experimental Pharmacology*, Springer International Publishing, 325-337. [https://doi.org/10.1007/164\\_2019\\_210](https://doi.org/10.1007/164_2019_210)
- [3] Lydon-Rochelle, M.T., Holt, V.L., Cárdenas, V., Nelson, J.C., Easterling, T.R., Gardella, C., *et al.* (2005) The Reporting of Pre-Existing Maternal Medical Conditions and Complications of Pregnancy on Birth Certificates and in Hospital Discharge Data. *American Journal of Obstetrics and Gynecology*, **193**, 125-134. <https://doi.org/10.1016/j.ajog.2005.02.096>
- [4] Singh, H., Murphy, H.R., Hendrieckx, C., Ritterband, L. and Speight, J. (2013) The Challenges and Future Considerations Regarding Pregnancy-Related Outcomes in Women with Pre-Existing Diabetes. *Current Diabetes Reports*, **13**, 869-876. <https://doi.org/10.1007/s11892-013-0417-5>
- [5] Stergiopoulos, K., Shiang, E. and Bench, T. (2011) Pregnancy in Patients with Pre-Existing Cardiomyopathies. *Journal of the American College of Cardiology*, **58**, 337-350. <https://doi.org/10.1016/j.jacc.2011.04.014>
- [6] Walkinshaw, S.A. (2005) Pregnancy in Women with Pre-Existing Diabetes: Management Issues. *Seminars in Fetal and Neonatal Medicine*, **10**, 307-315. <https://doi.org/10.1016/j.siny.2005.04.004>
- [7] Jost, E., Kosian, P., Greiner, G.G., Icks, A., Schmitz, M., Schmid, M., *et al.* (2024) Obstetric Medicine: The Protocol for a Prospective Three-Dimensional Cohort Study to Assess Maternity Care for Women with Pre-Existing Conditions (Format). *Frontiers in Medicine*, **10**, Article 1258716. <https://doi.org/10.3389/fmed.2023.1258716>

- [8] Hendriks, S., Grady, C., Wasserman, D., Wendler, D., Bianchi, D.W. and Berkman, B.E. (2021) A New Ethical Framework for Assessing the Unique Challenges of Fetal Therapy Trials. *The American Journal of Bioethics*, **22**, 45-61. <https://doi.org/10.1080/15265161.2020.1867932>
- [9] Raffaelli, B., Rubio-Beltrán, E., Cho, S., De Icco, R., Labastida-Ramirez, A., Onan, D., et al. (2023) Health Equity, Care Access and Quality in Headache—Part 2. *The Journal of Headache and Pain*, **24**, Article No. 167. <https://doi.org/10.1186/s10194-023-01699-7>
- [10] Krewski, D. (2022) Development of an Evidence-Based Risk Assessment Framework. *ALTEX*, **39**, 667-693. <https://doi.org/10.14573/altex.2004071>
- [11] Mt-Isa, S.H. (2011) Improving Evidence-Based Risk-Benefit Decision-Making of Medicines for Children. Ph.D. Dissertation, School of Public Health, Imperial College London.
- [12] Mussen, F., Salek, S. and Walker, S. (2008) Benefit-Risk Appraisal of Medicines. John Wiley & Sons. <https://doi.org/10.1002/9780470748114>
- [13] Esposito and Giovanna (2024) Reproduction and Childbearing Health: Re-Al-World Evidence from Administrative Databases.
- [14] Namazy, J., Chambers, C., Sahin, L., Johnson, T., Dinatale, M., Lappin, B., et al. (2020) Clinicians' Perspective of the New Pregnancy and Lactation Labeling Rule (PLLR): Results from an AAAAI/FDA Survey. *The Journal of Allergy and Clinical Immunology: In Practice*, **8**, 1947-1952. <https://doi.org/10.1016/j.jaip.2020.01.056>
- [15] Byrne, J.J., Saucedo, A.M. and Spong, C.Y. (2020) Evaluation of Drug Labels Following the 2015 Pregnancy and Lactation Labeling Rule. *JAMA Network Open*, **3**, e2015094. <https://doi.org/10.1001/jamanetworkopen.2020.15094>
- [16] Wicks, S., Tibbets, J., Caruso, E., et al. (2024) Assessing US Food and Drug Administration Authorities, Guidance, and Policies for Prescription Drug, Biological Product, and Medical Device Development and Commercialization for Use by Pregnant and Lactating Women. *Advancing Clinical Research*, 54 p.
- [17] Sharma, V., Sharma, P. and Sharma, S. (2020) Managing Bipolar Disorder during Pregnancy and the Postpartum Period: A Critical Review of Current Practice. *Expert Review of Neurotherapeutics*, **20**, 373-383. <https://doi.org/10.1080/14737175.2020.1743684>
- [18] Amosa, T.I., Izhar, L.I.B., Sebastian, P., Ismail, I.B., Ibrahim, O. and Ayinla, S.L. (2023) Clinical Errors from Acronym Use in Electronic Health Record: A Review of NLP-Based Disambiguation Techniques. *IEEE Access*, **11**, 59297-59316. <https://doi.org/10.1109/access.2023.3284682>
- [19] Egger, J., Gsaxner, C., Pepe, A., Pomykala, K.L., Jonske, F., Kurz, M., et al. (2022) Medical Deep Learning—A Systematic Meta-Review. *Computer Methods and Programs in Biomedicine*, **221**, Article 106874. <https://doi.org/10.1016/j.cmpb.2022.106874>
- [20] van Dinter, R., Tekinerdogan, B. and Catal, C. (2021) Automation of Systematic Literature Reviews: A Systematic Literature Review. *Information and Software Technology*, **136**, Article 106589. <https://doi.org/10.1016/j.infsof.2021.106589>
- [21] Chaudhury, A., Ward, C., Talasaz, A., Ivanov, A.G., Huner, N.P.A., Grodzinski, B., et al. (2015) Computer Vision Based Autonomous Robotic System for 3D Plant Growth Measurement. 2015 12th Conference on Computer and Robot Vision, Halifax, 3-5 June 2015, 290-296. <https://doi.org/10.1109/crv.2015.45>
- [22] Ahire, Y.S., Patil, J.H., Chordiya, H.N., Deore, R.A. and Bairagi, V.A. (2024) Advanced

- Applications of Artificial Intelligence in Pharmacovigilance: Current Trends and Future Perspectives. *Journal of Pharmaceutical Research*, **23**, 23-33.  
<https://doi.org/10.18579/jopcr/v23.1.24>
- [23] Dimitra, P. (2018) The Knowledge Discovery Cube Framework A Reference Framework for Collaborative, Information-Driven Pharmacovigilance. University of Surrey (United Kingdom).
- [24] Sivathapandi, P. (2022) Advanced AI Algorithms for Automating Data Preprocessing in Healthcare: Optimizing Data Quality and Reducing Processing Time. *Journal of Science & Technology*, **3**, 126-167.
- [25] Monlezun, D.J. (2024) Our Common Home: Artificial Intelligence + Global Public Health Ecosystem. In: Monlezun, D.J., Ed., *Responsible Artificial Intelligence Re-Engineering the Global Public Health Ecosystem*, Elsevier, 215-243.  
<https://doi.org/10.1016/b978-0-443-21597-1.00007-x>
- [26] Foulkes, M. (2015) Development of the Maternal-Fetal Relationship in Women Who Use Substances: Understanding the Influence of Intersecting Variables on Maternal-Fetal Attachment and Health Behaviours. Ph.D. Dissertation, University of Ottawa.
- [27] Smid, M., Bourgois, P. and Auerswald, C.L. (2010) The Challenge of Pregnancy among Homeless Youth: Reclaiming a Lost Opportunity. *Journal of Health Care for the Poor and Underserved*, **21**, 140-156. <https://doi.org/10.1353/hpu.0.0318>
- [28] Poon, L.C., Shennan, A., Hyett, J.A., Kapur, A., Hadar, E., Divakar, H., *et al.* (2019) The International Federation of Gynecology and Obstetrics (FIGO) Initiative on Pre-eclampsia: A Pragmatic Guide for First-Trimester Screening and Prevention. *International Journal of Gynecology & Obstetrics*, **145**, 1-33.  
<https://doi.org/10.1002/ijgo.12802>
- [29] Tacy, T.A., Kasparian, N.A., Karnik, R., Geiger, M. and Sood, E. (2022) Opportunities to Enhance Parental Well-Being during Prenatal Counseling for Congenital Heart Disease. *Seminars in Perinatology*, **46**, Article 151587.  
<https://doi.org/10.1016/j.semperi.2022.151587>
- [30] Suter, S.M. (2002) The Routinization of Prenatal Testing. *American Journal of Law & Medicine*, **28**, 233-270. <https://doi.org/10.1017/s0098858800011655>
- [31] Wilson, L. (2024) Enhancing Safety Surveillance during Clinical Trials: The Development of a Safety Data Review Handbook to Support Pharmacovigilance Clinicians. University of New Hampshire.
- [32] Ryan, P.B., Schuemie, M.J., Welebob, E., Duke, J., Valentine, S. and Hartzema, A.G. (2013) Defining a Reference Set to Support Methodological Research in Drug Safety. *Drug Safety*, **36**, 33-47. <https://doi.org/10.1007/s40264-013-0097-8>
- [33] Hong, Y.D., Jansen, J.P., Guerino, J., Berger, M.L., Crown, W., Goettsch, W.G., *et al.* (2021) Comparative Effectiveness and Safety of Pharmaceuticals Assessed in Observational Studies Compared with Randomized Controlled Trials. *BMC Medicine*, **19**, Article No. 307. <https://doi.org/10.1186/s12916-021-02176-1>
- [34] Sundaram, G. and Berleant, D. (2023) Automating Systematic Literature Reviews with Natural Language Processing and Text Mining: A Systematic Literature Review. In: Yang, R., Sherratt, S. Dey, N. and Joshi, A., Eds., *International Congress on Information and Communication Technology*, Springer, 73-92.
- [35] Dreisbach, C., Koleck, T.A., Bourne, P.E. and Bakken, S. (2019) A Systematic Review of Natural Language Processing and Text Mining of Symptoms from Electronic Patient-Authored Text Data. *International Journal of Medical Informatics*, **125**, 37-46.  
<https://doi.org/10.1016/j.ijmedinf.2019.02.008>

- [36] Sawicki, J., Ganzha, M. and Paprzycki, M. (2023) The State of the Art of Natural Language Processing—A Systematic Automated Review of NLP Literature Using NLP Techniques. *Data Intelligence*, **5**, 707-749. [https://doi.org/10.1162/dint\\_a\\_00213](https://doi.org/10.1162/dint_a_00213)
- [37] Ofori-Boateng, R., Aceves-Martins, M., Wiratunga, N. and Moreno-Garcia, C.F. (2024) Towards the Automation of Systematic Reviews Using Natural Language Processing, Machine Learning, and Deep Learning: A Comprehensive Review. *Artificial Intelligence Review*, **57**, Article No. 200.
- [38] Allam, H., Makubvure, L., Gyamfi, B., Graham, K. and Akinwolere, K. (2025) Text Classification: How Machine Learning Is Revolutionizing Text Categorization. *Information*, **16**, 130.
- [39] Hu, Z., Dychka, I., Potapova, K. and Meliukh, V. (2024) Augmenting Sentiment Analysis Prediction in Binary Text Classification through Advanced Natural Language Processing Models and Classifiers. *International Journal of Information Technology and Computer Science*, **16**, 16-31. <https://doi.org/10.5815/ijitcs.2024.02.02>
- [40] Han, H., Asif, M., Awwad, E.M., Sarhan, N., Ghadi, Y.Y. and Xu, B. (2024) Innovative Deep Learning Techniques for Monitoring Aggressive Behavior in Social Media Posts. *Journal of Cloud Computing*, **13**, Article No. 19. <https://doi.org/10.1186/s13677-023-00577-6>
- [41] Nerella, S., Bandyopadhyay, S., Zhang, J., Contreras, M., Siegel, S., *et al.* (2023) Transformers in Healthcare: A Survey. arXiv preprint arXiv:2307.00067.
- [42] Cho, H.N., Jun, T.J., Kim, Y., Kang, H., Ahn, I., Gwon, H., *et al.* (2024) Task-Specific Transformer-Based Language Models in Health Care: Scoping Review. *JMIR Medical Informatics*, **12**, e49724. <https://doi.org/10.2196/49724>
- [43] Rodríguez, H. (2022) Natural Language Technologies in the Biomedical Domain. In: Bacciu, D., Lisboa, P.J.G. and Vellido, A., Eds., *Deep Learning in Biology and Medicine*, World Scientific Connect, 93-130.
- [44] Bose, P., Srinivasan, S., Sleeman, W.C., Palta, J., Kapoor, R. and Ghosh, P. (2021) A Survey on Recent Named Entity Recognition and Relationship Extraction Techniques on Clinical Texts. *Applied Sciences*, **11**, Article 8319. <https://doi.org/10.3390/app11188319>
- [45] Lei, J., Tang, B., Lu, X., Gao, K., Jiang, M. and Xu, H. (2014) A Comprehensive Study of Named Entity Recognition in Chinese Clinical Text. *Journal of the American Medical Informatics Association*, **21**, 808-814. <https://doi.org/10.1136/amiajnl-2013-002381>
- [46] Liu, Z., Yang, M., Wang, X., Chen, Q., Tang, B., Wang, Z., *et al.* (2017) Entity Recognition from Clinical Texts via Recurrent Neural Network. *BMC Medical Informatics and Decision Making*, **17**, Article No. 67. <https://doi.org/10.1186/s12911-017-0468-7>
- [47] Fraile Navarro, D., Ijaz, K., Rezazadegan, D., Rahimi-Ardabili, H., Dras, M., Coiera, E., *et al.* (2023) Clinical Named Entity Recognition and Relation Extraction Using Natural Language Processing of Medical Free Text: A Systematic Review. *International Journal of Medical Informatics*, **177**, Article 105122. <https://doi.org/10.1016/j.ijmedinf.2023.105122>
- [48] Vithanage, D., Yu, P., Wang, L. and Deng, C. (2024) Contextual Word Embedding for Biomedical Knowledge Extraction: A Rapid Review and Case Study. *Journal of Healthcare Informatics Research*, **8**, 158-179. <https://doi.org/10.1007/s41666-023-00157-y>
- [49] Si, Y., Wang, J., Xu, H. and Roberts, K. (2019) Enhancing Clinical Concept Extraction with Contextual Embeddings. *Journal of the American Medical Informatics Association*, **26**, 1297-1304. <https://doi.org/10.1093/jamia/ocz096>

- [50] Wood, M.E., Andrade, S.E. and Toh, S. (2019) Safe Expectations: Current State and Future Directions for Medication Safety in Pregnancy Research. *Clinical Therapeutics*, **41**, 2467-2476. <https://doi.org/10.1016/j.clinthera.2019.08.016>
- [51] Menon, R., Richardson, L. and Kammala, A.K. (2024) New Approach Methods on the Bench Side to Accelerate Clinical Trials during Pregnancy. *Expert Opinion on Drug Metabolism & Toxicology*, **20**, 555-560. <https://doi.org/10.1080/17425255.2024.2353944>
- [52] Cao, H., Oghenemaro, E.F., Latypova, A., Abosoad, M.K., Zaman, G.S. and Devi, A. (2025) Advancing Clinical Biochemistry: Addressing Gaps and Driving Future Innovations. *Frontiers in Medicine*, **12**, Article 1521126. <https://doi.org/10.3389/fmed.2025.1521126>
- [53] Dolk, H., Damase-Michel, C., Morris, J.K. and Loane, M. (2022) COVID-19 in Pregnancy—What Study Designs Can We Use to Assess the Risk of Congenital Anomalies in Relation to COVID-19 Disease, Treatment and Vaccination? *Paediatric and Perinatal Epidemiology*, **36**, 493-507. <https://doi.org/10.1111/ppe.12840>
- [54] Borda, L.A., Någård, M., Boulton, D.W., Venkataramanan, R. and Coppola, P. (2023) A Systematic Review of Pregnancy-Related Clinical Intervention of Drug Regimens Due to Pharmacokinetic Reasons. *Frontiers in Medicine*, **10**, Article 1241456. <https://doi.org/10.3389/fmed.2023.1241456>
- [55] Rezaallah, B. (2019) Pharmacovigilance of Pregnancy Exposures to Medicinal Products Focusing on the Risk of Orofacial Clefts. Ph.D. Dissertation, University of Basel.
- [56] D'Amore, F.M., Moscatelli, M., Malvaso, A., D'Antonio, F., Rodini, M., Panigutti, M., et al. (2025) Explainable Machine Learning on Clinical Features to Predict and Differentiate Alzheimer's Progression by Sex: Toward a Clinician-Tailored Web Interface. *Journal of the Neurological Sciences*, **468**, Article 123361. <https://doi.org/10.1016/j.jns.2024.123361>
- [57] El Arab, R.A., Almoosa, Z., Alkhunaizi, M., Abuadas, F.H. and Somerville, J. (2025) Artificial Intelligence in Hospital Infection Prevention: An Integrative Review. *Frontiers in Public Health*, **13**, Article 1547450. <https://doi.org/10.3389/fpubh.2025.1547450>
- [58] Lambert, M.J., Whipple, J.L. and Kleinstäuber, M. (2018) Collecting and Delivering Progress Feedback: A Meta-Analysis of Routine Outcome Monitoring. *Psychotherapy*, **55**, 520-537. <https://doi.org/10.1037/pst0000167>
- [59] Koroteev and Mikhail, V. (2021) BERT: A Review of Applications in Natural Language Processing and Understanding. arXiv preprint arXiv:2103.11943.
- [60] Kalusivalingam, K.A., Sharma, A., Patel, N. and Singh, V. (2021) Leveraging BERT and LSTM for Enhanced Natural Language Processing in Clinical Data Analysis. *International Journal of AI and ML*, **2**, 1-24.
- [61] Devlin, J., Chang, M., Lee, K. and Toutanova, K. (2019) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies*, **1** 4171-4186.
- [62] Lavecchia, A. (2024) Advancing Drug Discovery with Deep Attention Neural Networks. *Drug Discovery Today*, **29**, Article 104067. <https://doi.org/10.1016/j.drudis.2024.104067>
- [63] ValizadehAslani, T., Shi, Y., Ren, P., Wang, J., Zhang, Y., Hu, M., et al. (2023) PharmBERT: A Domain-Specific BERT Model for Drug Labels. *Briefings in Bioinformatics*, **24**, bbad226. <https://doi.org/10.1093/bib/bbad226>
- [64] Kristina M., D., Byatt, N. and Marlene P., F. (2014) Pharmacotherapy for Mood

Disorders in Pregnancy: A Review of Pharmacokinetic Changes and Clinical Recommendations for Therapeutic Drug Monitoring. *Journal of Clinical Psychopharmacology*, **34**, 244-255. <https://doi.org/10.1097/JCP.0000000000000087>

- [65] Pariente, G., Leibson, T., Carls, A., Adams-Webber, T., Ito, S. and Koren, G. (2016) Pregnancy-Associated Changes in Pharmacokinetics: A Systematic Review. *PLOS Medicine*, **13**, e1002160. <https://doi.org/10.1371/journal.pmed.1002160>