https://doi.org/10.31891/2219-9365-2026-85-44
УДК 004

БОЙКО Микола
Приватний вищий навчальний заклад «Європейський університет»
https://orcid.org/0009-0006-6086-3969
САМОРАЙ Олег
Приватний вищий навчальний заклад «Європейський університет»
https://orcid.org/0009-0000-3384-155X
ДЕЙНИКОВСЬКИЙ Максим
Приватний вищий навчальний заклад «Європейський університет»
https://orcid.org/0009-0002-9519-0072

АВТОМАТИЗОВАНИЙ АУДИТ ЖУРНАЛІВ БАЗ ДАНИХ (AUDIT LOGS) З ВИКОРИСТАННЯМ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ: КОРЕЛЯЦІЯ ПОДІЙ, РЕКОНСТРУКЦІЯ ІНЦИДЕНТІВ ТА ПОЯСНЮВАНІ ВИСНОВКИ

У статті досліджується застосування великих мовних моделей (Large Language Models, LLM) та методів машинного навчання для автоматизованого аналізу журналів аудиту баз даних (audit logs). Основна увага приділяється інтелектуальній кореляції подій, реконструкції інцидентів безпеки та формуванню пояснюваних аналітичних висновків на основі журналів доступу до даних. Запропоновано концептуальну модель інтеграції LLM у процеси аудиту баз даних, що дозволяє автоматично виявляти аномалії, підозрілі сценарії доступу та потенційні порушення політик безпеки.

Дослідження охоплює аналіз структури журналів аудиту, методи їх попередньої обробки, використання NLP-моделей для інтерпретації SQL-операцій та алгоритми реконструкції інцидентів на основі часових послідовностей подій. Наведено приклади застосування LLM для аналізу SQL-запитів, ролей користувачів, контексту виконання операцій та їх потенційного впливу на безпеку інформаційної системи.

Отримані результати демонструють, що застосування LLM у процесах аудиту даних дозволяє підвищити ефективність систем моніторингу безпеки, скоротити час реагування на інциденти та забезпечити більш глибоке розуміння причин виникнення порушень політик доступу.

Ключові слова: аудит баз даних, audit logs, великі мовні моделі, кібербезпека, кореляція подій, реконструкція інцидентів, пояснюваний штучний інтелект.

BOIKO Mykola, SAMORAI Oleh, DEINYKOVSKYI Maksym
Private Higher Educational Institution "European University"

AUTOMATED DATABASE AUDIT LOG ANALYSIS USING LARGE LANGUAGE MODELS: EVENT CORRELATION, INCIDENT RECONSTRUCTION AND EXPLAINABLE INSIGHTS

The article investigates the application of large language models (LLMs) and machine learning techniques for automated analysis of database audit logs in modern information systems. As the volume and complexity of database transactions continuously increase, traditional manual analysis of audit records becomes inefficient and time-consuming. Therefore, the use of artificial intelligence technologies is considered a promising approach for improving the efficiency of database security monitoring and incident investigation. Particular attention in the study is given to event correlation, reconstruction of security incidents, and generation of explainable analytical insights based on database activity logs.

A conceptual framework for integrating LLMs into database auditing processes is proposed. The framework combines machine learning algorithms for anomaly detection with language-model capabilities for interpreting structured and semi-structured log data. This approach enables automatic detection of abnormal behavior, suspicious access patterns, privilege misuse, and potential violations of data access policies. The model can analyze sequences of database operations and identify complex relationships between user actions, system events, and security alerts.

The study includes a detailed analysis of typical audit log structures in relational database management systems, as well as preprocessing methods required for effective machine learning application. These methods include normalization of log records, feature extraction, temporal aggregation of events, and transformation of SQL queries into interpretable textual representations. Natural language processing techniques are used to interpret SQL commands, identify semantic similarities between queries, and detect unusual behavioral patterns in database access.

Examples of LLM-based analysis of user roles, SQL operations, privilege escalation attempts, and time-based sequences of events are presented. The experimental evaluation demonstrates that the integration of artificial intelligence into database auditing systems significantly improves the ability to detect complex attack scenarios and insider threats. In addition, the use of LLMs allows the generation of human-readable explanations of detected anomalies, which supports transparent and explainable decision-making in cybersecurity management.

The obtained results confirm that intelligent analysis of database audit logs can substantially enhance cybersecurity monitoring capabilities, accelerate incident investigation, and provide organizations with more reliable tools for protecting critical data assets in modern digital infrastructures.

Keywords: database audit, audit logs, large language models, cybersecurity, event correlation, incident reconstruction, explainable AI.

ISSN 2219-9365

Стаття надійшла до редакції / Received 10.01.2026
Прийнята до друку / Accepted 10.02.2026
Опубліковано / Published 05.03.2026

ПОСТАНОВКА ПРОБЛЕМИ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Сучасні інформаційні системи активно використовують бази даних для зберігання критично важливої інформації, зокрема персональних даних користувачів, фінансових операцій, конфіденційних документів та службових журналів діяльності системи. Для забезпечення контролю доступу до цих даних у більшості систем управління базами даних реалізовано механізми аудиту, які фіксують усі операції доступу до інформації.

Журнали аудиту (audit logs) містять інформацію про виконані SQL-операції, користувачів, час доступу до даних, IP-адреси, ролі доступу та результати виконання операцій. Ці журнали є ключовим джерелом інформації для аналізу інцидентів безпеки, проведення внутрішнього аудиту та забезпечення відповідності регуляторним вимогам.

Проте обсяг журналів у сучасних системах зростає надзвичайно швидко. У великих корпоративних системах щодня генеруються мільйони записів журналів аудиту. Аналіз такої кількості даних вручну або за допомогою традиційних правил є складним та неефективним.

Крім того, сучасні кібератаки часто складаються з багатьох взаємопов'язаних дій, які виконуються протягом тривалого часу. Традиційні системи моніторингу безпеки, що базуються на статичних правилах, не завжди здатні виявити такі складні сценарії атак.

Графік: кількість записів audit logs у великих системах залежно від часу зображено на Рис 1.

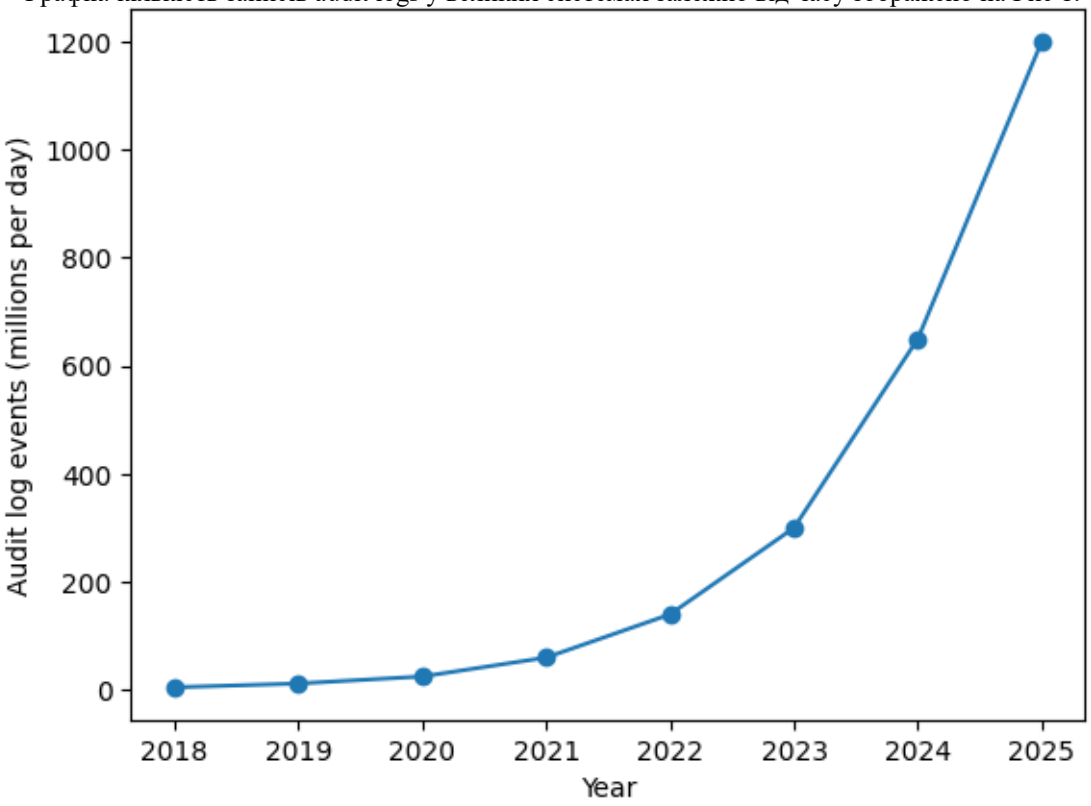


Рис. 1. Зростання обсягу журналів подій у сучасних інформаційних системах)

У цьому контексті виникає потреба у використанні інтелектуальних методів аналізу журналів аудиту, здатних автоматично встановлювати взаємозв'язки між подіями та реконструювати повну картину інциденту.

АНАЛІЗ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

У наукових дослідженнях останніх років активно розглядаються питання використання методів машинного навчання та штучного інтелекту для аналізу журналів подій у інформаційних системах.

Більшість досліджень у цій сфері зосереджена на таких напрямах:

- автоматичне виявлення аномалій у системних журналах;
- аналіз мережевого трафіку;
- виявлення вторгнень у комп'ютерні системи;

- автоматизація процесів реагування на інциденти.

Таблиця 1

Основні підходи до аналізу журналів подій

Підхід	Метод	Основна мета
Rule-based	сигнатури атак	виявлення відомих атак
Statistical	статистичний аналіз	пошук аномалій
Machine Learning	класифікація	аналіз поведінки
Deep Learning	нейронні мережі	виявлення складних шаблонів
LLM-based	NLP аналіз	контекстна інтерпретація подій

Останні дослідження демонструють перспективність використання великих мовних моделей для аналізу текстових журналів подій. LLM здатні інтерпретувати зміст записів журналів, аналізувати SQL-запити та формувати пояснення щодо потенційних інцидентів безпеки.

Попри це, застосування LLM саме для аналізу журналів аудиту баз даних досі залишається недостатньо дослідженим напрямом.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Метою даного дослідження є розробка концептуальної моделі автоматизованого аудиту журналів баз даних із використанням можливостей великих мовних моделей, яка дозволить підвищити ефективність виявлення, аналізу та інтерпретації подій безпеки в інформаційних системах. У сучасних умовах стрімкого зростання обсягів даних та складності інформаційної інфраструктури традиційні підходи до аналізу журналів аудиту стають недостатньо ефективними, оскільки потребують значних людських ресурсів і часу. Застосування великих мовних моделей відкриває нові можливості для інтелектуального аналізу журналів, автоматизованої інтерпретації подій та формування аналітичних висновків, що можуть бути використані фахівцями з кібербезпеки та адміністрування баз даних.

Для досягнення поставленої мети передбачається вирішення комплексу взаємопов'язаних завдань. Насамперед необхідно дослідити структуру журналів аудиту баз даних, визначити їх основні елементи, типи подій, формат зберігання та особливості формування у різних системах управління базами даних. Це дозволить сформувати узагальнене уявлення про інформаційний зміст журналів та визначити ключові параметри, які можуть бути використані для подальшого аналізу.

Наступним важливим етапом є розробка методів попередньої обробки журналів аудиту. Такі методи повинні забезпечувати очищення даних, нормалізацію форматів, структурування інформації, а також виділення суттєвих ознак подій. Попередня обробка є необхідною умовою для ефективного застосування алгоритмів аналізу та використання великих мовних моделей, оскільки дозволяє зменшити обсяг шуму в даних та підвищити якість подальшої інтерпретації.

Важливою складовою дослідження є створення алгоритму кореляції подій, який дозволить встановлювати логічні та часові зв'язки між окремими записами журналів. Завдяки цьому стане можливим об'єднання розрізнених подій у більш складні ланцюжки дій користувачів або системних процесів. Такий підхід дозволяє краще зрозуміти контекст виконуваних операцій та підвищити точність виявлення потенційно небезпечних або аномальних активностей.

На основі отриманих результатів планується розробити модель реконструкції інцидентів, яка забезпечить відтворення послідовності подій, що призвели до виникнення певної ситуації або порушення безпеки. Реконструкція інцидентів дозволяє більш детально аналізувати причини виникнення проблем, визначати ключові етапи розвитку інциденту та оцінювати можливі наслідки для інформаційної системи.

Завершальним етапом дослідження є реалізація механізму формування пояснюваних аналітичних висновків на основі результатів автоматизованого аналізу. Використання великих мовних моделей дозволить не лише виявляти підозрілі події, але й генерувати зрозумілі текстові пояснення, що описують логіку виявлення інцидентів, їх можливі причини та рекомендації щодо подальших дій. Це сприятиме підвищенню прозорості системи аналізу та полегшить роботу фахівців, які здійснюють аудит та забезпечують безпеку баз даних.

Таким чином, реалізація зазначених завдань дозволить сформувати цілісну концептуальну модель автоматизованого аудиту журналів баз даних із використанням великих мовних моделей, що поєднуватиме методи обробки даних, алгоритми кореляції подій та інтелектуальні механізми інтерпретації результатів аналізу. Така модель може стати основою для створення сучасних систем моніторингу та аналізу подій безпеки в інформаційних системах.

ISSN 2219-9365

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

У межах проведеного дослідження було розроблено концептуальну архітектуру системи автоматизованого аудиту журналів баз даних, яка забезпечує інтелектуальний аналіз подій, що фіксуються у журналах аудиту. Запропонована архітектура орієнтована на автоматизацію процесів збору, обробки та інтерпретації інформації з журналів баз даних із використанням сучасних методів аналізу даних та можливостей великих мовних моделей.

Такий підхід дозволяє підвищити ефективність виявлення аномалій, інцидентів безпеки та підозрілої активності користувачів у системах управління базами даних.

Концептуальна модель системи включає декілька взаємопов'язаних компонентів, кожен з яких виконує окрему функцію в загальному процесі аналізу журналів. Взаємодія цих компонентів забезпечує послідовний перехід від отримання сирих журналів подій до формування аналітичних висновків та звітів для фахівців з інформаційної безпеки та адміністрування баз даних.

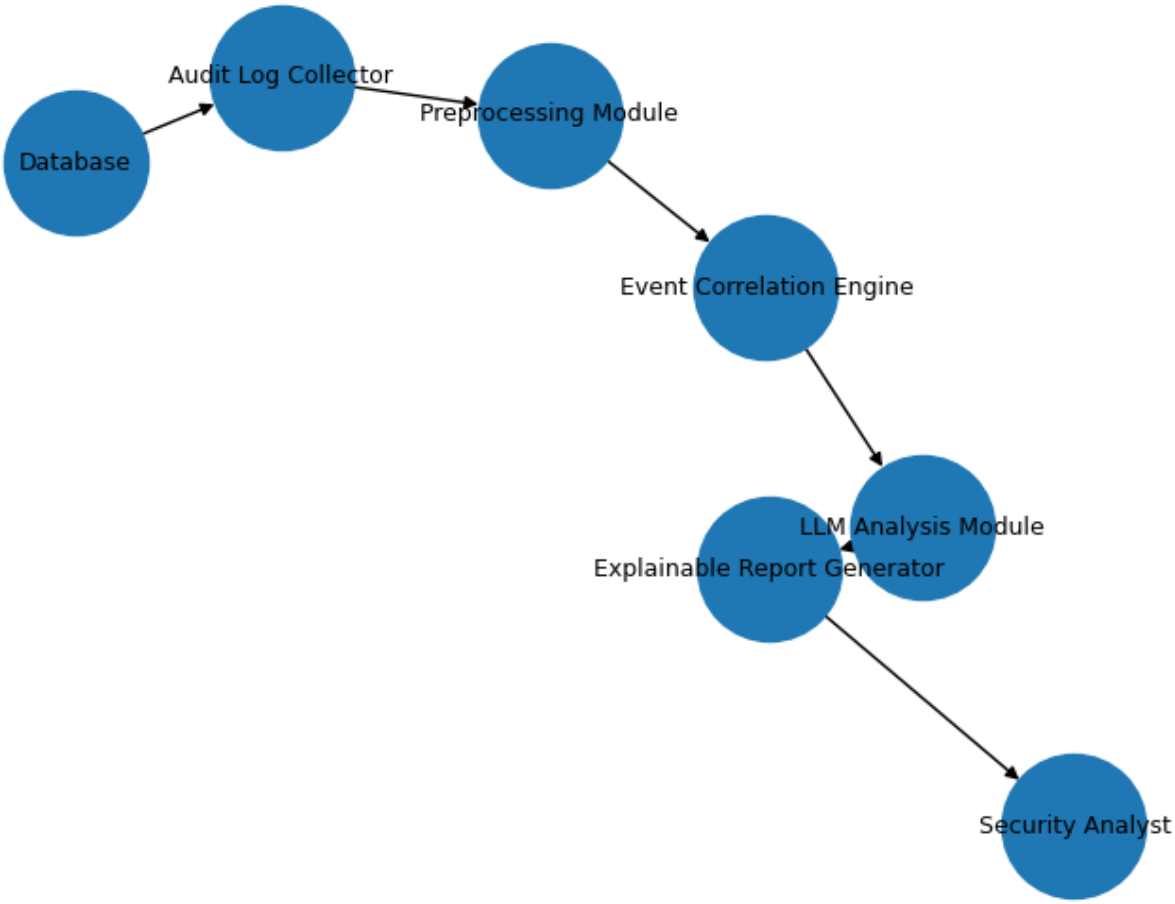


Рис. 2. Архітектура системи інтелектуального аудиту журналів БД

Першим елементом є модуль збору журналів, який відповідає за отримання та централізоване накопичення журналів аудиту з різних джерел. До таких джерел можуть належати системи управління базами даних, сервери додатків, системи авторизації та інші інформаційні компоненти. Основним завданням цього модуля є забезпечення безперервного збору подій, їх агрегування та передача до наступних компонентів системи для подальшої обробки.

Наступним компонентом є модуль попередньої обробки даних, який виконує підготовку отриманих журналів до подальшого аналізу. У межах цього модуля здійснюється очищення даних від зайвих або дубльованих записів, нормалізація форматів журналів, структурування інформації, а також виділення ключових параметрів подій, таких як час виконання операції, ідентифікатор користувача, тип запиту або виконаної дії. Попередня обробка дозволяє значно підвищити якість даних та створює основу для ефективного аналізу.

Третім компонентом системи є модуль кореляції подій, який призначений для встановлення логічних і часових зв'язків між окремими подіями журналів. Цей модуль аналізує послідовності дій користувачів або системних процесів та об'єднує окремі записи у взаємопов'язані ланцюжки подій. Завдяки цьому стає можливим виявлення складніших сценаріїв поведінки, які можуть свідчити про потенційні інциденти безпеки, порушення політик доступу або аномальні операції з даними.

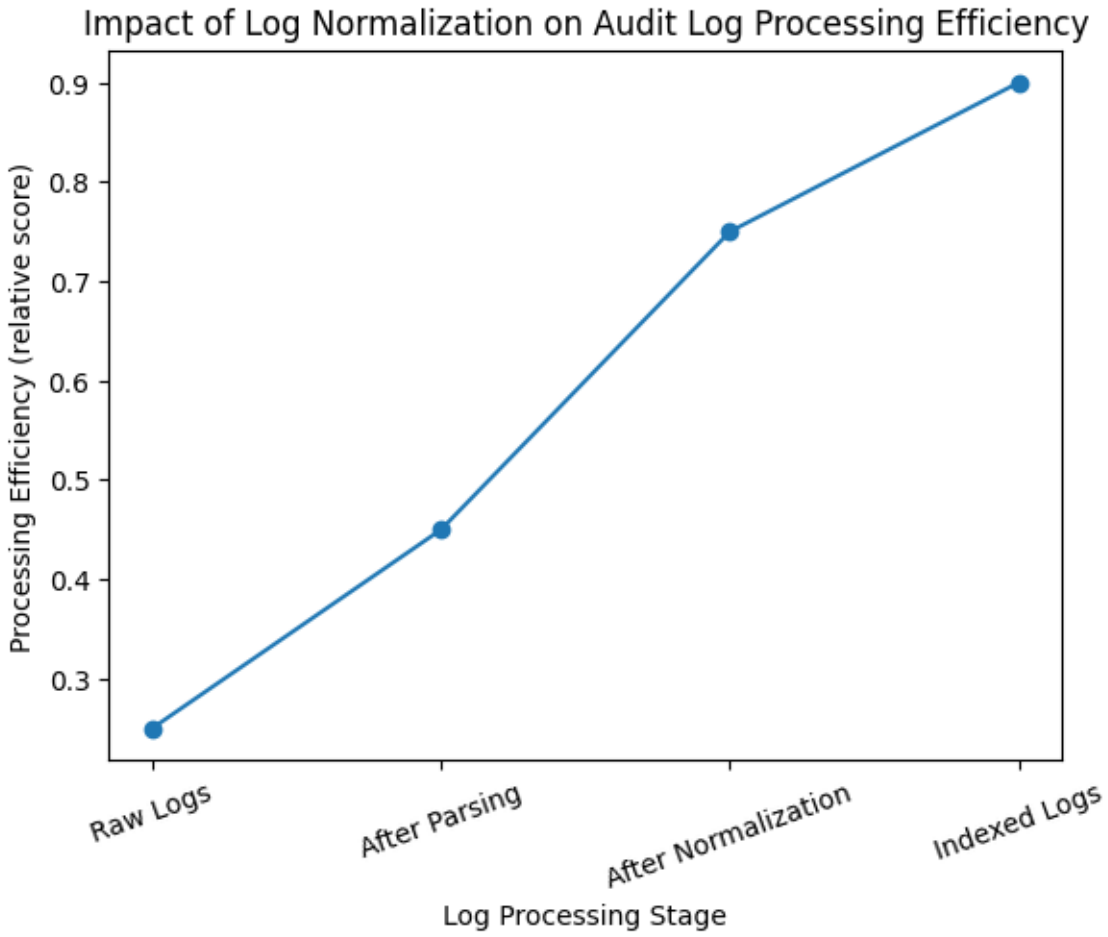
Ключовим елементом запропонованої архітектури є аналітичний модуль LLM, який використовує можливості великих мовних моделей для інтелектуального аналізу отриманих даних. У цьому модулі здійснюється інтерпретація корельованих подій, виявлення потенційно небезпечних сценаріїв, а також формування узагальнених аналітичних висновків. Великі мовні моделі дозволяють аналізувати контекст подій, визначати їх логічні взаємозв'язки та формувати пояснення щодо виявлених аномалій або інцидентів.

Завершальним компонентом системи є система формування звітів, яка забезпечує представлення результатів аналізу у зрозумілій для користувача формі. Цей модуль формує аналітичні звіти, що можуть містити опис виявлених інцидентів, часові лінії подій, оцінку ризиків та рекомендації щодо реагування. Звіти можуть використовуватися адміністраторами баз даних, аудиторами або фахівцями з кібербезпеки для прийняття управлінських рішень та вдосконалення політик безпеки.

Важливим етапом функціонування системи інтелектуального аудиту є попередня обробка журналів баз даних. На цьому етапі виконується підготовка сирих журналів аудиту до подальшого аналітичного опрацювання. Журнали, що надходять із систем управління базами даних, зазвичай мають різні формати, можуть містити дубльовані записи, технічні повідомлення або неповні дані. Тому перед проведенням аналітичних операцій необхідно виконати їх структурування та стандартизацію.

У межах цього етапу здійснюється очищення журналів, яке передбачає видалення дубльованих записів, технічних або службових повідомлень, а також некоректних або пошкоджених даних. Це дозволяє зменшити обсяг інформаційного шуму та підвищити якість подальшого аналізу.

Наступною процедурою є нормалізація структури записів, що полягає у приведенні журналів різних форматів до єдиної стандартизованої структури. Така нормалізація дозволяє уніфікувати представлення подій, що значно спрощує подальшу обробку даних, їх індексацію та аналіз алгоритмами кореляції.



Графік 1. Вплив нормалізації журналів на ефективність обробки audit logs

Також на цьому етапі виконується класифікація подій, під час якої записи журналу розподіляються за типами операцій. До таких типів можуть належати події авторизації, виконання SQL-запитів, зміни структури бази даних, операції читання або модифікації даних. Класифікація дозволяє швидше виявляти потенційно небезпечні або аномальні дії користувачів.

У результаті попередньої обробки формується структурований запис журналу аудиту, що містить набір ключових параметрів події.

Таблиця 2

Структура запису журналу аудиту

Поле	Опис
Timestamp	Час Події
User	Користувач
Role	Роль Доступу
Query	Sql-Запит
Object	Таблиця
Result	Результат

Така структура дозволяє стандартизувати інформацію про події та створює основу для подальшого аналізу й кореляції записів журналів.

Після завершення етапу попередньої обробки виконується кореляція подій журналу аудиту. Основною метою цього етапу є встановлення логічних та часових взаємозв'язків між окремими записами журналів. Окремі записи часто відображають лише частину дій користувача або системного процесу, тому їх ізольований аналіз може бути недостатнім для виявлення складних сценаріїв поведінки.

Модуль кореляції аналізує послідовності подій, що відбуваються у певному часовому інтервалі, та об'єднує їх у логічні ланцюги. При цьому враховуються такі параметри, як ідентифікатор користувача, роль доступу, тип виконаної операції, а також об'єкти бази даних, до яких здійснювався доступ.

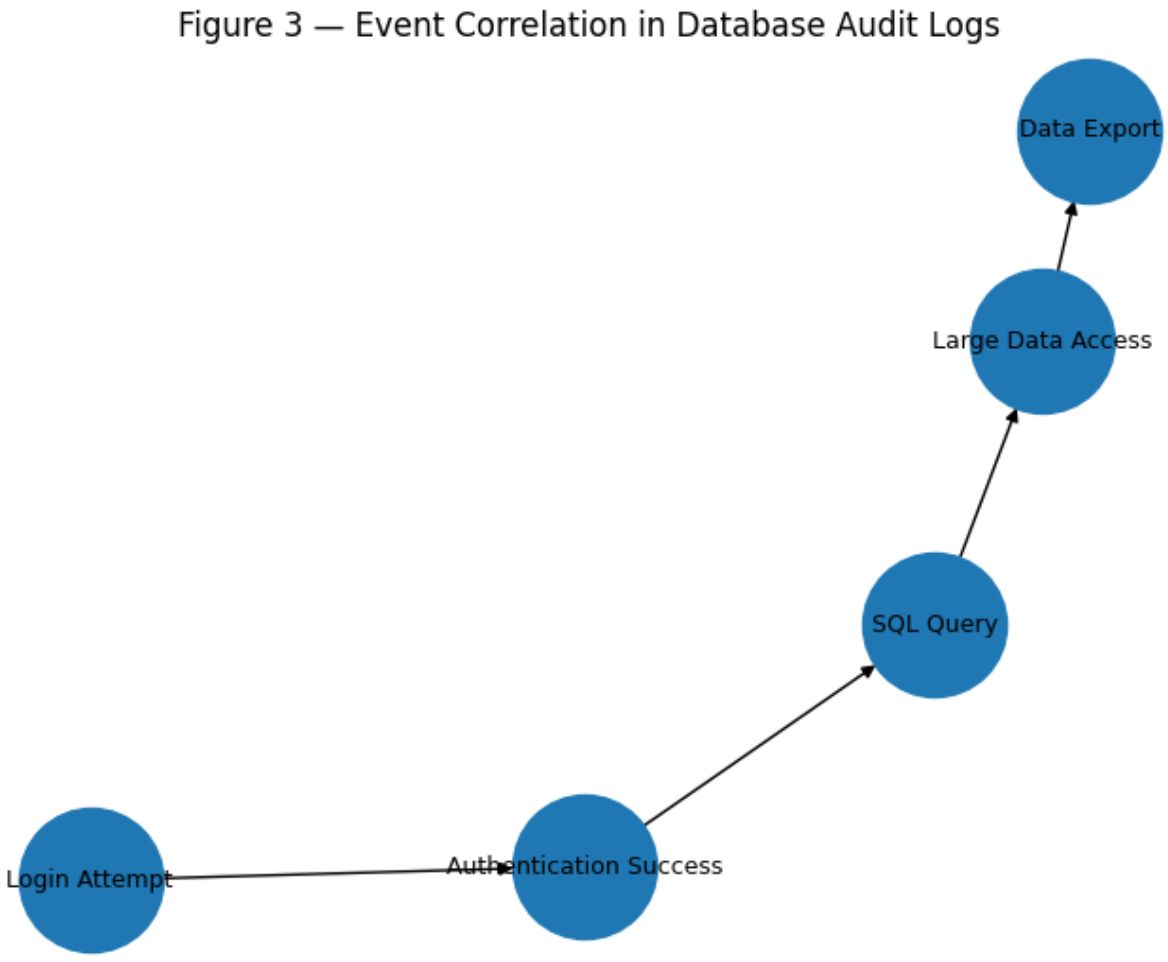


Рис. 3. Приклад кореляції подій у журналі аудиту

Наприклад, система може встановити зв'язок між такими подіями:

login → SELECT → EXPORT DATA

У цьому випадку послідовність подій може свідчити про потенційний сценарій витоку даних, коли після авторизації користувач виконує запит на отримання інформації з таблиці, а потім експортує отримані дані. Кореляція подій дозволяє виявляти подібні сценарії та передавати їх до наступного етапу аналізу.

ISSN 2219-9365

Наступним етапом системи є реконструкція інцидентів, яка здійснюється на основі часових послідовностей корельованих подій. Основною метою цього етапу є відтворення повної картини подій, що призвели до виникнення потенційного інциденту безпеки або аномальної активності в базі даних.

Під час реконструкції інциденту система аналізує зібрані ланцюги подій, визначає їх послідовність, взаємозв'язки та контекст виконання операцій. На основі цього формується модель інциденту, яка дозволяє зрозуміти, які саме дії були виконані користувачем або системним процесом, у якій послідовності вони відбувалися та які об'єкти бази даних були задіяні.

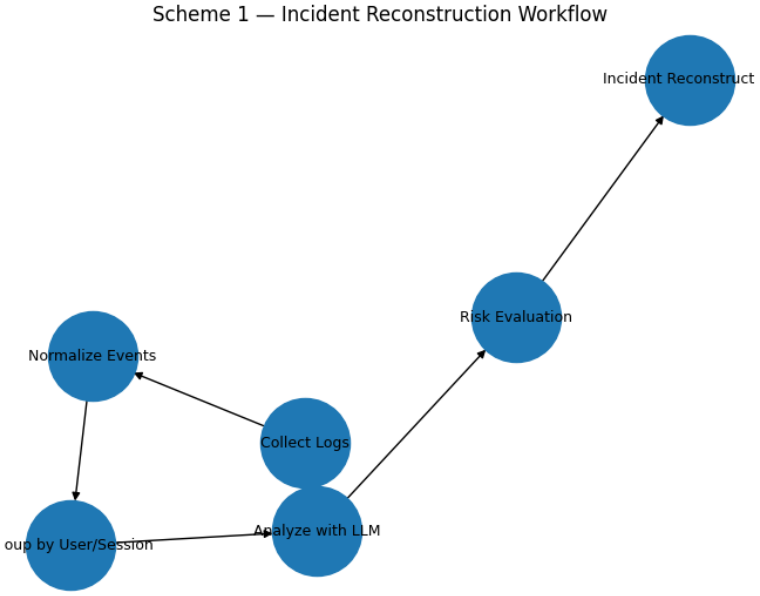


Схема 1. Модель реконструкції інциденту

Процес реконструкції інциденту включає декілька основних етапів:

групування подій — об'єднання взаємопов'язаних записів журналу у єдиний ланцюг подій на основі часових інтервалів та ідентифікаторів користувачів;

аналіз ролей користувача — визначення прав доступу користувача та перевірка відповідності виконаних дій наданим привілеям;

оцінка ризику — визначення рівня потенційної загрози на основі характеру виконаних операцій та їх послідовності;

формування сценарію інциденту — побудова узагальненого опису події або інциденту, який може бути використаний для подальшого аналізу або формування звітів.

Реконструкція інцидентів дозволяє не лише виявляти підозрілі події, але й відтворювати повний контекст їх виникнення, що є важливим для проведення аудиту, аналізу інцидентів безпеки та прийняття рішень щодо подальших заходів захисту інформаційних систем.

LLM аналізує SQL-операції та визначає їх потенційний ризик.

```
for event in audit_log:
    context = build_context(event)
    risk_score = LLM.evaluate(context)
    if risk_score > threshold:
        flag_incident(event)
```

Для оцінки ефективності системи використовуються такі показники:

Формула 1 — точність виявлення інцидентів

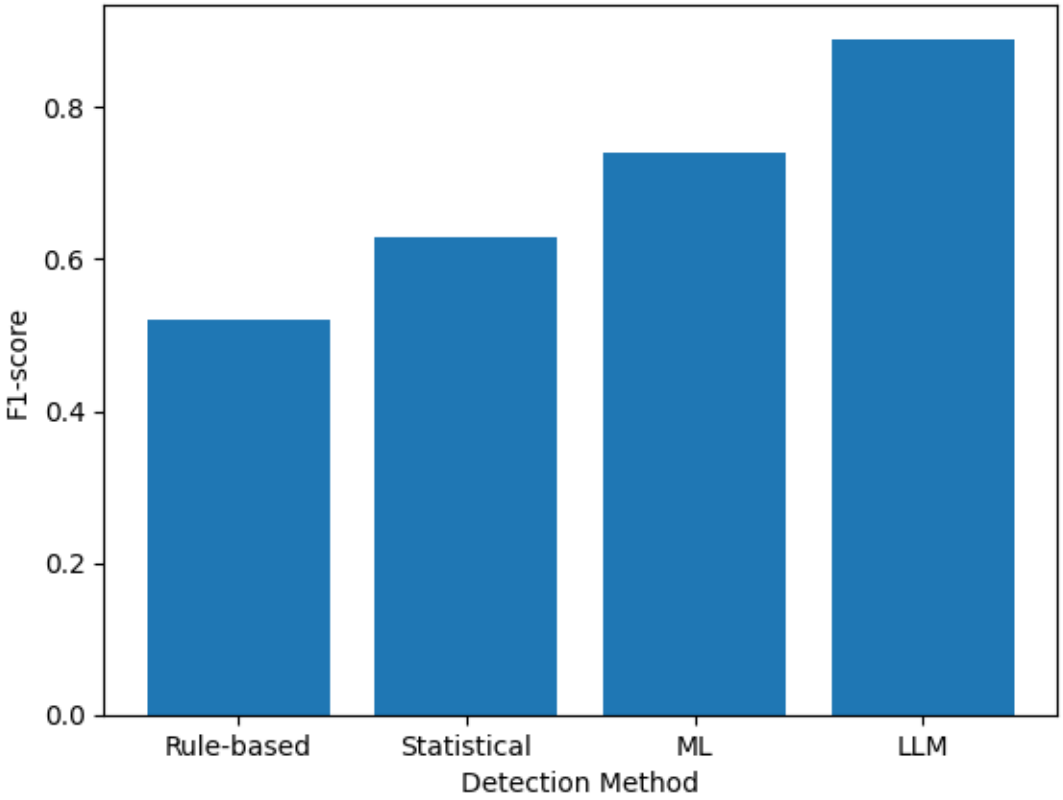
$$Precision = TP / (TP + FP)$$

(Формула 2 — повнота виявлення)

$$Recall = TP / (TP + FN)$$

(Формула 3 — F1-score)

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall)$$



Графік 2. Порівняння ефективності традиційних методів і LLM-аналізу

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

У статті досліджено можливості застосування великих мовних моделей для автоматизованого аналізу журналів аудиту баз даних. Актуальність такого підходу зумовлена постійним зростанням обсягів журналів подій, які формуються сучасними системами управління базами даних, а також необхідністю швидкого виявлення потенційних загроз безпеці інформаційних систем. Традиційні методи аналізу журналів часто вимагають значних часових і людських ресурсів, тоді як використання великих мовних моделей дозволяє автоматизувати значну частину аналітичних процесів, підвищуючи швидкість і якість обробки інформації.

Проведене дослідження показало, що інтеграція великих мовних моделей у системи аналізу журналів аудиту відкриває нові можливості для інтелектуального опрацювання подій, що відбуваються у базах даних. Зокрема, застосування LLM дозволяє автоматизувати процес кореляції подій, встановлюючи логічні та часові взаємозв'язки між окремими записами журналів. Це дає змогу об'єднувати розрізнені події у послідовні ланцюги дій користувачів або системних процесів, що значно спрощує аналіз поведінки у системі.

Крім того, використання великих мовних моделей сприяє більш ефективній реконструкції інцидентів безпеки. На основі аналізу послідовностей подій система може відтворювати сценарії дій користувачів, визначати потенційно небезпечні операції та встановлювати можливі причини виникнення інцидентів. Такий підхід дозволяє не лише виявляти аномалії, але й формувати повнішу картину розвитку подій у системі. Ще однією важливою перевагою використання LLM є можливість формування пояснюваних аналітичних висновків. Великі мовні моделі здатні генерувати текстові пояснення результатів аналізу, що дозволяє фахівцям з інформаційної безпеки швидше інтерпретувати отримані результати та приймати обґрунтовані рішення щодо реагування на потенційні загрози.

Отримані результати свідчать про перспективність використання великих мовних моделей у системах аудиту та моніторингу подій баз даних. Водночас подальші дослідження у цьому напрямі можуть бути спрямовані на розширення функціональних можливостей запропонованого підходу. Зокрема, перспективними напрямами є інтеграція розробленої системи з SIEM-платформами для централізованого управління подіями безпеки, розробка спеціалізованих мовних моделей, оптимізованих для аналізу SQL-операцій та журналів баз даних, а також застосування методів explainable AI, які дозволять підвищити прозорість і зрозумілість процесів автоматизованого аналізу.

Використання великих мовних моделей у системах аналізу журналів аудиту баз даних створює передумови для підвищення ефективності моніторингу подій, автоматизації процесів аудиту та покращення загального рівня безпеки інформаційних систем.

References

1. Du M., Li F., Zheng G., Srikumar V. DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning. Proceedings of the ACM Conference on Computer and Communications Security, 2017. https://doi.org/10.1145/3133956.3134015
2. He P., Zhu J., Zheng Z., Lyu M. Drain: An Online Log Parsing Approach with Fixed Depth Tree. Proceedings of the IEEE International Conference on Web Services, 2017. https://doi.org/10.1109/ICWS.2017.13
3. Meng W., Tischhauser E., Wang Q., Wang Y., Han J. When Intrusion Detection Meets Blockchain Technology: A Review. IEEE Access, 2022. https://doi.org/10.1109/ACCESS.2018.2799854
Zhang Q., Fang C., Xie Y., Zhang Y., Yang Y., Chen Z. A Survey on Large Language Models for Software Engineering. arXiv preprint arXiv:2312.15223, 2023. https://doi.org/10.48550/arXiv.2312.15223
4. Ahmed I. Artificial Intelligence for Software Engineering: The Journey and LLM Integration. ACM Computing Surveys, 2025.
5. Esposito M., Li X., Moreschini S., Ahmad N., Cerny T., Lenarduzzi V., Taibi D. Generative AI for Software Architecture: Applications, Trends, Challenges and Future Directions. arXiv preprint arXiv:2503.13310, 2025. https://doi.org/10.48550/arXiv.2503.13310
6. Nyaga F. AI-Driven Software Engineering: A Systematic Review of Machine Learning's Impact and Future Directions. Preprints.org, 2025. https://doi.org/10.20944/preprints202504.0174.v1
7. Behl J., Behl A. Explainable Artificial Intelligence in Cybersecurity: Techniques and Applications. Journal of Cybersecurity and Privacy, 2023.
8. Brown T. et al. Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems (NeurIPS), 2020.
9. OpenAI. GPT-4 Technical Report. arXiv preprint arXiv:2303.08774, 2023.
10. Kim G., Lee S., Kim H. Log-Based Anomaly Detection Using Deep Learning Methods. IEEE Transactions on Network and Service Management, 2022.
11. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey. ACM Computing Surveys, 2009.
12. Xu W., Huang L., Fox A., Patterson D., Jordan M. Detecting Large-Scale System Problems by Mining Console Logs. Proceedings of the ACM Symposium on Operating Systems Principles, 2009. https://doi.org/10.1145/1629575.162958
13. Sommer R., Paxson V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. IEEE Symposium on Security and Privacy, 2010.