Forensic Science, 4'2026
ISSN: 3083-6085 (print), 3083-6093 (online)

УДК 343.98:004.8
DOI https://doi.org/10.32782/forensic.science.2026.4.11



Галина Кривоножко
кандидат технічних наук, старший науковий співробітник,
головний судовий експерт відділу інженерно-технічних досліджень,
Київське відділення Національного наукового центру
«Інститут судових експертиз ім. засл. проф. М. С. Бокаріуса»
Міністерства юстиції України
https://orcid.org/00000-0002-7635-541X



Олексій Собін
судовий експерт,
Науково-дослідний центр судової експертизи у сфері інформаційних
технологій та інтелектуальної власності Міністерства юстиції України
https://orcid.org/0000-0003-4653-5986

ЗАПОБІГАННЯ ХИБНОМУ КАТЕГОРИЧНОМУ ВИСНОВКУ ПРИ ЗАСТОСУВАННІ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ У СУДОВО-ЕКСПЕРТНІЙ ДІЯЛЬНОСТІ

Застосування великих мовних моделей у судово-експертній діяльності створює специфічний ризик: правдоподібна, але неперевірена відповідь може бути використана на наступних етапах дослідження і вплинути на формування категоричного висновку. Помилка може статися під час добору та тлумачення джерел, підготовки змісту, редагування, перевірки, технічного опрацювання або формування кінцевого цифрового документа. Тому проблема полягає не лише в окремій помилці мовної моделі, а в запобіганні тому, щоб неперевірена інформація стала підставою для остаточної категоричної відповіді. Метою статті є обґрунтування умов організації контролю, спрямованого на запобігання хибному категоричному висновку під час використання великих мовних моделей у судово-експертній діяльності. Методологічну основу становлять узагальнення практичного досвіду роботи з мовними моделями у складних аналітичних і документальних завданнях, аналіз типових помилок і причин, з яких наявний контроль їх не виявляв, а також зіставлення отриманих висновків із сучасними дослідженнями 2025–2026 років. Роботу з мовною моделлю запропоновано розглядати як послідовність професійно значущих етапів, перехід між якими допускається лише після підтвердження визначених вимог. Можливість відтворити перебіг роботи за журналом виконання відокремлено від професійної перевірки: журнал допомагає встановити послідовність дій і шукати причину помилки, а контрольна точка визначає, чи може результат бути використаний на наступній стадії експертного дослідження. Виокремлено шість типових груп помилок і відмов контролю: використання непідтвердженого об'єкта або джерела; неправильне застосування достовірної інформації; змішування джерел і контекстів; втрата перевіреного змісту під

час редагування; хибне підтвердження виконання перевірки; невідповідність між перевіреним змістом і фактично сформованим цифровим об'єктом. Обґрунтовано, що відповідь мовної моделі не може бути самостійною підставою для категоричного експертного висновку. Якщо інформація, істотно значуща для відповіді на поставлене питання, отримана або перетворена за допомогою мовної моделі й не може бути належно перевірена, вона не має включатися до логічної основи категоричного висновку. Запропонований підхід спрямований не на декларування безпомилковості мовної моделі, а на побудову запобіжного порядку роботи, який не допускає переходу неперевіреної інформації до категоричного експертного результату.

***Ключові слова**: велика мовна модель, категоричний висновок, перевірка, професійна помилка, судова експертиза, контроль якості.*

Постановка проблеми. Великі мовні моделі стали практичним інструментом для роботи з текстами, документами, аналітичними матеріалами, програмним кодом і даними. У судово-експертній сфері штучний інтелект і спеціалізовані цифрові інструменти розглядаються як допоміжний ресурс, використання якого не усуває необхідності професійної перевірки та не підміняє експертного дослідження [3, с. 184–188]. У журналі «Судова експертиза» також розглядалося використання чат-бота на основі мовної моделі в експертних сервісах із наголосом на нормативно сумісному функціоналі та вбудованих засобах перевірки [4, с. 4–9].

Водночас здатність мовної моделі правильно виконувати складне завдання не гарантує безпомилковості кожної окремої операції. Та сама система може коректно опрацювати значний масив вимог і водночас пропустити речення, неправильно перенести реквізит, змішати два джерела, застосувати достовірну інформацію поза межами її застосовності або помилково повідомити про виконання перевірки. Систематичний огляд застосування великих мовних моделей у цифровій криміналістиці також підкреслює суперечність між імовірнісною природою їх відповідей та вимогами до відтворюваності, перевіреності й доказової надійності цифрового дослідження [5].

Для судового експерта найбільш небезпечними є не очевидно помилкові відповіді, а правдоподібні відхилення, які можуть бути прийняті за перевірену інформацію. Якщо таке відхилення переходить до наступної стадії як вихідна передумова, воно здатне впливати на проміжні результати, синтез дослідження та остаточну відповідь. Ризик зростає не тому, що будь-яка помилка мовної моделі автоматично створює хибний висновок, а тому, що неперевірена інформація може необґрунтовано набути професійного статусу.

У судово-експертній діяльності ця межа має принципове значення. Закон України «Про судову експертизу» та Інструкція про призначення та проведення судових експертиз і експертних досліджень визначають правові й організаційні рамки діяльності експерта [1; 2]. Запропонований у цій роботі порядок організації контролю не підміняє методику експертного дослідження, не змінює компетенції експерта й не встановлює допустимість доказів. Його призначення вужче: не допустити, щоб неперевірена відповідь мовної моделі стала фактичною або логічною підставою для категоричного висновку.

У межах статті під хибним категоричним висновком розуміється категорична відповідь судового експерта, на формування якої істотно вплинула помилкова або неперевірена інформація, отримана чи перетворена за допомогою великої мовної моделі. Отже, предметом дослідження є не окремий вид помилкової відповіді та не властивості конкретного програмного продукту, а запобіжний порядок роботи: де може виникнути помилка, як вона може перейти до професійно значущого етапу, які контрольні точки мають зупинити такий перехід і яким чином перевіряється кінцевий цифровий результат.

Мета статті. Метою статті є обґрунтування умов організації роботи, за яких використання великих мовних моделей у судово-експертній діяльності не призводить до включення неперевіреної інформації до логічної основи категоричного висновку, а також формування практичних принципів контролю такої інформації на професійно значущих етапах дослідження.

Аналіз останніх досліджень і публікацій. Попередня спільна робота Г. Кривоножко й О. Собіна була присвячена використанню

штучного інтелекту та спеціалізованих цифрових інструментів в експертному забезпеченні розслідування кіберзлочинів і цифрових правопорушень. У ній автоматизований пошук і автоматизоване опрацювання розмежовано з експертним висновком, який формується після професійної перевірки [3, с. 184–188]. Ця стаття продовжує зазначений напрям, але переносить акцент із загальної допоміжної ролі штучного інтелекту на конкретний ризик: недопущення хибного категоричного висновку внаслідок проходження неперевіреної інформації через багатоступеневий експертний процес.

У сучасній міжнародній літературі помітний перехід від оцінювання окремої відповіді до аналізу всієї послідовності її формування. В одному з досліджень багатоступеневого формування відповідей показано, що виявити непідтверджений зміст лише у фінальному результаті замало: важливо також з'ясувати, на якому проміжному етапі він міг з'явитися і як перевірений зміст пов'язаний із вихідним джерелом [6].

Водночас наявність журналу виконання ще не робить пошук причини помилки простим. За результатами дослідження журналів відмов багатомодельних систем установлено, що визначення точного вирішального кроку залишається складним навіть для спеціально побудованих методів [7, с. 76583–76599]. Для практики це означає, що контрольні точки можуть звузити ділянку можливого виникнення дефекту, але не гарантують встановлення точної миті помилки.

Особливе значення має якість самої перевірки. В іншому дослідженні методів автоматизованого контролю показано, що критерії оцінювання мають бути конкретними й не дублювати один одного, а перебіг виконання та фактичний кінцевий результат доцільно перевіряти окремо. Невдало сформульовані критерії самі можуть стати джерелом наступних помилок [8].

Ще одне обмеження полягає в тому, що багаторазова перевірка мовними моделями не є автоматично незалежною. За результатами дослідження понад 350 великих мовних моделей установлено, що різні моделі можуть систематично помилятися подібним чином [9, с. 30038–30066]. Отже, кількість повторних відповідей не може бути єдиним критерієм надійності.

Для документальної роботи особливо важливі результати дослідження тривалих циклів редагування документів. Автори зафіксували накопичення пошкоджень змісту, яке посилювалося зі збільшенням обсягу документа та тривалості взаємодії [10]. Це підтверджує практичний ризик: правильно сформований зміст може бути втрачений уже під час наступного редагування або «покращення» документа.

Сучасні практичні підходи дедалі частіше будують удосконалення системи за послідовністю помилка – причина – зміна – повторна перевірка. В одному з опублікованих практичних прикладів виправлення фахівця пов'язується із журналом виконання, перетворюється на контрольний випадок, після чого встановлена причина використовується для перевірюваного вдосконалення системи [11].

У проєкті рамкових рекомендацій Національного інституту стандартів і технологій США 2026 року також виходять із того, що перевірка систем штучного інтелекту має налаштовуватися відповідно до конкретного застосування, його цілей і ризиків [12]. Це підтримує підхід з урахуванням рівня ризику, але не дає підстав визначати одну універсальну кількість контрольних точок для всіх професійних завдань.

Виклад основного матеріалу. *Від окремого запиту до контрольованого судово-експертного процесу.* Коли мовна модель дає нестабільний результат, першим кроком зазвичай стає уточнення запиту. До початкової інструкції додаються правила, приклади, заборони, вимоги до самоперевірки, порядок роботи із джерелами та критерії оформлення результату. Добре сформульований запит справді здатний істотно підвищити якість виконання.

Однак зі зростанням складності завдання постійне додавання нових правил має межу. Одна інструкція повинна одночасно регулювати дедалі більшу кількість взаємопов'язаних умов. Виправлення одного класу помилок не захищає від виникнення іншого на пізнішій стадії. Частина критичних дефектів узагалі виникає вже після правильної первинної генерації – під час скорочення, стилістичного редагування, форматування, конвертації або технічного випуску.

Практика показує, що зі зростанням складності завдання об'єкт контролю зміщується

від тексту окремого запиту до послідовності професійно значущих етапів, контрольованих переходів, зворотного зв'язку, повторної перевірки після внесення змін і окремого контролю кінцевого цифрового результату. Ідеться не про відмову від удосконалення запиту, а про його підпорядкування ширшому запобіжному порядку роботи, мета якого – не дозволити неперевіреній відповіді мовної моделі вплинути на категоричну відповідь судового експерта.

Типові помилки мовної моделі, здатні вплинути на категоричний експертний висновок. Використання непідтвердженого об'єкта або джерела. Модель включає до подальшого аналізу документ, факт, реквізит або інший об'єкт, існування чи походження якого не підтверджено. Подальша логіка може бути бездоганною, але будуватися на хибній вихідній основі. Тому до оцінювання висновку необхідно підтвердити існування та походження критично важливих вхідних об'єктів.

Неправильне застосування достовірної інформації. Джерело може існувати і бути правильно відтворене, але застосовуватися поза межами своєї предметної, часової або ситуаційної застосовності. Перевірка має розрізняти існування джерела, його зміст, застосовність і зроблений на його основі висновок.

Змішування джерел і контекстів. У тривалій роботі модель може поєднати інформацію з різних документів або завдань. Якщо таке змішування не виявлено на у процесі, подальший аналіз фактично виконується над іншим інформаційним масивом, ніж той, який мав контролювати користувач.

Втрата правильного змісту під час подальшого редагування. Уже перевірений фрагмент може зникнути або змінитися під час скорочення, стилістичного поліпшення, реорганізації структури чи іншого перетворення. Тому нову версію необхідно оцінювати не лише за критерієм «чи стала вона кращою», а й за критерієм «чи збережено все, що вже було визнано правильним». Можливість такого накопичення пошкоджень підтверджується результатами дослідження тривалих циклів редагування документів [10].

Хибне підтвердження виконання перевірки. Мовна модель може сформулювати переконливе повідомлення про те, що якусь операцію виконано, хоча об'єктивного доказу цього немає. Особливо небезпечно це там, де перевірка потребує зовнішнього інструмента, доступу до конкретного файлу, бази даних або обчислення. Позитивний статус має ґрунтуватися на доказі перевірки, а не на мовному твердженні про її виконання.

Невідповідність між перевіреним змістом і фактично випущеним цифровим об'єктом. Перевірено одну версію, але передано іншу; після конвертації змінено елемент; у новому файлі залишилися небажані метадані або втрачено вкладення. Змістова правильність і правильність конкретного кінцевого файлу є різними властивостями.

Від локальної помилки мовної моделі до хибного категоричного висновку. Наведені класи помилок можуть виникати вздовж усього життєвого циклу результату, як-от: вихідні матеріали, аналіз, робочий результат, редагування, перевірка, кінцевий документ, зовнішнє використання. Критичним є не факт появи локального дефекту, а його подальший перехід у стан, який сприймається як перевірений і використовується для обґрунтування наступних висновків.

Тому запобігання хибному категоричному висновку має розривати ланцюг локальна помилка → неперевірена передумова → професійно значуще твердження → категорична відповідь. Завдання полягає не в тому, щоб гарантувати відсутність будь-яких помилок мовної моделі, а в тому, щоб помилкова або неперевірена інформація не могла без належного підтвердження стати частиною логічної основи висновку.

Повний автоматизований процес може містити сотні технічних операцій. Перетворення кожної з них на окремий об'єкт людського погодження означало б фактичне дублювання автоматизованої роботи. Водночас контроль лише фінального документа є запізнілим, якщо неперевірене положення вже було використане на попередніх стадіях. Практичним рішенням є контроль професійно значущих станів – тобто етапів, на яких результат набуває нового значення для експертного дослідження.

До таких станів залежно від конкретного завдання можуть належати: підтверджений корпус вихідних матеріалів; робоча версія результату; перевірені професійно значущі твердження; зафіксована змістовна контрольна версія; сформований кінцевий

цифровий документ; перевірений саме той об'єкт, який фактично передається адресату. Перелік не є універсальним: він визначається характером дослідження та ризиком помилки.

Контрольна точка в такій системі не є механічним повторним перечитуванням. Її функція – встановити, чи має професійно значуща інформація достатню доказову основу для використання на наступній стадії. Якщо обов'язкова передумова не підтверджена, результат може залишатися робочою версією або гіпотезою, але не повинен входити до обґрунтування категоричного висновку.

Немає достатнього підтвердження професійно значущої інформації – немає її допуску до категоричного висновку. Контрольні точки, журнал виконання та професійна перевірка. Журнал виконання, якщо він технічно доступний, і професійна перевірка вирішують різні завдання. Журнал дає змогу відтворити, що відбувалося між двома контрольними етапами. Професійна перевірка відповідає на інше питання: чи належно підтверджено отриманий результат, щоб використовувати його далі.

Ці функції не дублюють одна одну. Наявність детального журналу не означає, що експерт має переглядати кожний технічний запис. Водночас контрольна точка не замінює журнал: якщо виявлено відхилення, можливість простежити перебіг виконання створює інформаційну основу для подальшого встановлення причини.

Важливо не перебільшувати можливості такого підходу. Контрольні точки не гарантують точного визначення миті виникнення дефекту. Результати дослідження багатомодельних систем показують складність такої локалізації навіть у спеціально побудованих системах [7, с. 76583–76599]. Коректніше стверджувати, що послідовно перевірені етапи звужують ділянку можливого виникнення відхилення, після чого журнал використовується для подальшого аналізу причини.

Журнал виконання допомагає відтворити перебіг роботи; професійна перевірка визначає, чи може отриманий результат впливати на висновок. Категоричний висновок як критична межа процесу. Не всі помилки мають однакове значення для експертного висновку. Неправильне твердження у внутрішній чернетці може бути виявлено та вилучено без наслідків. Те саме твердження, використане як обґрунтована передумова категоричної відповіді, створює принципово інший ризик. Отже, критичність визначається не лише змістом помилки, а й тим, на якій стадії вона була допущена до подальшого використання.

Для визначення жорсткості контролю доцільно враховувати щонайменше чотири чинники: значущість можливих наслідків; зворотність наслідків після виявлення помилки; потенціал впливу на подальші проміжні результати; ступінь остаточності – від робочої гіпотези до категоричного висновку.

Ці чинники не є математичним індексом. Вони задають принцип: чим ближче неперевірена інформація до логічної основи категоричного висновку і чим суттєвішими можуть бути наслідки її помилковості, тим сильнішим має бути підтвердження. Такий підхід узгоджується із проєктом рамкових рекомендацій Національного інституту стандартів і технологій США, який станом на вересень 2026 року має статус первинного публічного проєкту [12].

Для судового експерта із цього випливає принципова межа: категорична відповідь не може ґрунтуватися на переконливості тексту, згенерованого мовною моделлю. Професійно значущі положення мають бути підтверджені дослідженими об'єктами й матеріалами, спеціальними знаннями, застосовними методами та перевіреним інструментарієм у межах компетенції експерта [1; 2]. Якщо невизначеність стосується істотно значущого елемента й не може бути усунена перевіркою, вона не повинна маскуватися категоричною формою відповіді.

Чому повторна перевірка мовною моделлю не є незалежним підтвердженням. Інтуїтивно простим способом підвищити довіру є повторно доручити мовній моделі перевірити власний або чужий результат. Такі повторні перевірки можуть бути корисними: вони виявляють суперечності, пропуски й альтернативні тлумачення. Однак їх не можна автоматично прирівнювати до незалежної перевірки.

За результатами дослідження понад 350 великих мовних моделей установлено, що їхні помилки можуть істотно збігатися [9, с. 30038–30066]. Отже, кілька повторних відповідей моделей можуть відтворювати

спільну помилкову передумову й створювати лише видимість незалежного підтвердження.

Надійнішим є контроль, який спирається на різні за природою підстави: первинний документ, зовнішнє фактичне джерело, формальне обчислення, інструментальний тест, порівняння із зафіксованою контрольною версією, перевірку властивостей конкретного файлу або професійну оцінку. Дослідження методів автоматизованої перевірки також показує, що якість контролю значною мірою залежить від правильності критеріїв і від окремого оцінювання перебігу виконання та кінцевого результату [8].

Надійність перевірки визначається не кількістю повторних модельних відповідей, а достатністю та різнорідністю доказових підстав, на яких експерт ухвалює рішення про використання інформації. Удосконалення запобіжного процесу після виявлення помилки. Виявлена помилка створює два різні завдання: виправити конкретний результат і встановити, чи свідчить випадок про недолік самого процесу. Механічне додавання нового правила після кожної помилки призводить до накопичення історичних нашарувань, винятків і суперечностей. Якоїсь миті надмірно складна інструкція сама може стати джерелом нестабільності.

Тому доцільна така послідовність: виявлена помилка → встановлення першопричини → оцінювання наявного контролю → визначення причини його не спрацювання → встановлення прогалини контролю → мінімально необхідна зміна → перевірка проблемного випадку → повторна перевірка раніше стабільних властивостей → прийняття або відхилення зміни.

Ключовим є принцип мінімально необхідної зміни. Нове правило не має додаватися лише тому, що воно теоретично виглядає корисним. Підставою має бути встановлена прогалина. Після зміни потрібно повторно перевірити не лише проблемний випадок, а й раніше стабільні властивості процесу.

Такий підхід уже застосовується у практичних системах. В одному з опублікованих прикладів професійне виправлення пов'язують із журналом виконання, перетворюють на контрольний випадок і використовують для подальшого перевірюваного вдосконалення системи [11]. У цій роботі таку логіку перенесено з рівня програмного продукту на рівень повсякденної організації роботи судового експерта з мовною моделлю.

Змістова перевірка та контроль кінцевого цифрового об'єкта. Особливої уваги потребує остання стадія документального процесу. Перевірений текст і перевірений кінцевий цифровий документ – не тотожні об'єкти. Після змістовного погодження можуть виконуватися форматування, зміна структури, створення таблиць, конвертація, очищення метаданих, комплектування додатків, перейменування або збереження в інший файл. Кожна така операція може створити новий дефект.

Тому змістовно погоджену версію доцільно фіксувати як контрольну, а технічне формування кінцевого файлу розглядати окремо. Якщо файл після перевірки змінено, попередній позитивний результат перевірки не повинен автоматично переноситися на нову версію. Ризик втрати вже правильного змісту під час послідовних перетворень підтверджено дослідженням тривалих циклів редагування документів [10].

Цей принцип особливо важливий у судово-експертній роботі, де зовнішньому адресату передається не абстрактний «текст», а конкретний документ, електронний файл, додаток або інший цифровий об'єкт. Саме він має бути предметом останньої перевірки.

Перевіряти необхідно саме той цифровий об'єкт, який фактично буде передано, оприлюднено або використано. Практичні орієнтири для судового експерта. У прикладному вигляді запропонований підхід зводиться до низки обов'язкових запитань: які джерела дозволено використовувати; які результати мовної моделі можуть вплинути на відповідь експерта; які критерії підтверджують їх правильність і застосовність; які невідповідності повинні блокувати подальше використання; який вид зовнішнього доказу є належним для конкретної властивості; як фіксується перевірена версія; яким чином перевіряється кінцевий цифровий об'єкт; що відбувається після виявлення нового класу помилки?

Ключовим є не тотальний контроль кожної внутрішньої операції мовної моделі, а недопущення проходження неперевіреної інформації через ті межі, після яких вона впливає на експертну відповідь. Такий порядок створює додатковий запобіжний механізм

навколо професійного рішення експерта, але не є альтернативою експертній методиці й не формує висновок автоматично.

Обговорення. Проведене зіставлення показує, що аналіз послідовності дій, пошук місця виникнення помилки, повторна перевірка після внесення змін, використання різних способів контролю та перевірка з урахуванням рівня ризику вже розробляються в сучасній літературі [6–12]. Тому внесок цієї роботи не полягає в заяві про новизну кожного окремого механізму.

Науково-практичний результат полягає в узгодженні цих відомих підходів навколо конкретної судово-експертної мети – запобігання хибному категоричному висновку під час використання великих мовних моделей. Запропонована модель розглядає локальну помилку через її можливий вплив на професійно значущі етапи; відокремлює можливість відтворити перебіг роботи за журналом виконання від професійної перевірки; пов'язує жорсткість контролю з можливими наслідками помилки; потребує різних за природою доказових підстав; передбачає окрему перевірку кінцевого цифрового об'єкта та контрольоване вдосконалення порядку роботи після виявленої помилки.

Такий підхід розвиває попередню спільну авторську позицію про допоміжний характер ШІ та необхідність професійної перевірки [3], але ставить інше питання: за яких умов інформація, отримана або перетворена великою мовною моделлю, не має бути допущена до формування категоричної відповіді судового експерта?

Обмеження дослідження. Запропонована модель не гарантує нульової імовірності помилки. Неправильно визначена контрольна точка сама може пропустити дефект. Помилка може статися між двома контрольними етапами й не бути виявленою одразу. Повторні перевірки різними мовними моделями також не гарантують незалежності, оскільки їхні помилки можуть збігатися [9, с. 30038–30066]. Надмірна кількість контрольних точок здатна збільшити тривалість роботи та створити формальний контроль замість змістовного.

Робота має характер якісного практичного дослідження організації робочого процесу й не містить контрольованого статистичного порівняння з підходом, що спирається лише на складний текст запиту. Тому не стверджується кількісно доведене зменшення частоти помилок, математична достатність визначеної кількості контрольних точок або універсальна перевага конкретної реалізації.

Подальшого емпіричного дослідження потребують критерії вибору контрольних етапів, співвідношення витрат на додаткову перевірку і зниження ризику, ефективність різних видів зовнішнього підтвердження, а також порівняльне тестування контрольованого процесу й роботи, що спирається переважно на один складний запит, на однакових професійних завданнях.

Висновки:

1. Результат великої мовної моделі не є самостійною фактичною або методичною підставою для категоричного висновку судового експерта. Будь-яке положення, отримане або перетворене за допомогою мовної моделі й істотно значуще для відповіді на поставлене питання, має бути підтверджене незалежно від самої моделі – за об'єктами, матеріалами, джерелами, методами або інструментальними даними, придатними для конкретного дослідження.

2. Якщо істотно значуща для відповіді інформація, отримана або перетворена за допомогою мовної моделі, не може бути належно перевірена, вона не може входити до логічної основи категоричного висновку. Невирішену невизначеність щодо такого елемента не можна приховувати переконливістю або повторенням тексту, сформованого моделлю.

3. Запобігання хибному категоричному висновку потребує не збільшення кількості однакових повторних перевірок мовними моделями, а контролю на професійно значущих етапах. Правило має бути жорстким: немає достатнього підтвердження – немає допуску неперевіреної інформації до наступного етапу, до категоричного висновку й поготів. Журнал виконання допомагає відтворити перебіг роботи та встановлювати причини помилок, але не замінює професійної перевірки експертом.

4. Перевірці підлягає не лише зміст, а й конкретний кінцевий цифровий об'єкт, який фактично передається або використовується. Будь-яка зміна після змістовного погодження потребує повторної перевірки пов'язаних із нею властивостей. Якщо виявлено

системну помилку, необхідно встановити її першопричину, внести мінімально необхідну зміну та перевірити, чи не пошкодила ця зміна раніше правильні елементи.

Отже, за використання великих мовних моделей у судово-експертній діяльності категоричний висновок є методично обґрунтованим тоді, коли всі професійно значущі положення, що впливають на відповідь, мають перевірену доказову основу.

Декларація про використання штучного інтелекту.

Під час підготовки рукопису автори використовували OpenAI як допоміжний засіб для структурного опрацювання чернетки. Англомовну анотацію підготовлено особисто. Концепцію дослідження, твердження, добір джерел і остаточний зміст статті визначено авторами, які несуть повну відповідальність за її точність, оригінальність і академічну доброчесність.

Список використаної літератури:

1. Про судову експертизу : Закон України від 25.02.1994 р. № 4038-XII. *База даних «Законодавство України» / Верховна Рада України*. URL: https://zakon.rada.gov.ua/go/4038-12 (дата звернення: 09.09.2026).
2. Про затвердження Інструкції про призначення та проведення судових експертиз та експертних досліджень та Науково-методичних рекомендацій з питань підготовки та призначення судових експертиз та експертних досліджень : наказ Міністерства юстиції України від 08.10.1998 р. № 53/5. *База даних «Законодавство України»*. URL: https://zakon.rada.gov.ua/go/z0705-98 (дата звернення: 09.09.2026).
3. Кривоножко Г., Собін О. Штучний інтелект і спеціалізовані цифрові інструменти в експертному забезпеченні розслідування кіберзлочинів та цифрових правопорушень. *Криміналістика майбутнього: цифрові інновації, штучний інтелект, глобальні виклики та загрози :* збірник матеріалів Міжнародної науково-практичної конференції, м. Одеса, 5 червня 2026 р. / заг. ред. С. Ківалова, М. Аракеляна. Одеса : НУ «ОЮА», 2026. С. 184–188. DOI: https://doi.org/10.32837/11300.33533
4. Свінцицький А. Функціонал замовлення судової експертизи за допомогою інтегрованого чат-бота ChatGPT. *Судова експертиза*. 2025. № 2. С. 4-9. DOI: https://doi.org/10.32782/forensic.science.2025.2.1
5. Chernyshev M., Baig Z., Syed N., Doss R., Shore M. Large language models in digital forensics: capabilities, challenges and future directions. *Forensic Science International: Digital Investigation*. 2026. Vol. 56. Article 302043. DOI: https://doi.org/10.1016/j.fsidi.2025.302043
6. Metropolitansky D., Larson J. VeriTrail: Closed-Domain Hallucination Detection with Traceability. International Conference on Learning Representations (ICLR 2026). 2026. URL: https://proceedings.iclr.cc/paper_files/paper/2026/ hash/42ec51d1031d25eca238ccc67559fe84-Abstract-Conference.html (дата звернення: 09.09.2026).
7. Zhang S. et al. Which Agent Causes Task Failures and When? On Automated Failure Attribution of LLM Multi-Agent Systems. *Proceedings of the 42nd International Conference on Machine Learning. PMLR*. 2025. Vol. 267. P. 76583–76599. URL: https://proceedings.mlr.press/v267/zhang25cq.html (дата звернення: 09.09.2026).
8. Rosset C., Sharma P., Zhao A., González-Fernández M., Awadallah A. The Art of Building Verifiers for Computer Use Agents. *Conference on Language Modeling (COLM 2026)*. 2026. URL: https://www.microsoft.com/en-us/research/publication/the-art-of-building-verifiers-for-computer-use-agents/ (дата звернення: 09.09.2026).
9. Kim E. M., Garg A., Peng K., Garg N. Correlated Errors in Large Language Models. *Proceedings of the 42nd International Conference on Machine Learning. PMLR*. 2025. Vol. 267. P. 30038–30066. URL: https://proceedings.mlr.press/v267/kim25e.html (дата звернення: 09.09.2026).
10. Laban P., Schnabel T., Neville J. LLMs Corrupt Your Documents When You Delegate. arXiv:2604.15597. 2026. DOI: https://doi.org/10.48550/arXiv.2604 .15597
11. Srinivasan A., Shamdasani S., Araujo A. F., de Wasseige J. Building self-improving tax agents with Codex. *OpenAI*. 27.05.2026. URL: https://openai.com/index/building-self-improving-tax-agents-with-codex/ (дата звернення: 09.09.2026).
12. National Institute of Standards and Technology. The TEVV-Athlon Framework for Evaluating AI Systems. Initial Public Draft NIST AI 200-2. 07.08.2026. DOI: https://doi.org/10.6028/NIST.AI.200-2.ipd

References:

1. Pro sudovu ekspertyzu: Zakon Ukrainy vid 25.02.1994 № 4038–XII [On Forensic Examination: Law of Ukraine of February 25, 1994 № 4038–XII]. Verkhovna Rada of Ukraine. https://zakon.rada.gov.ua/go/4038-12 [in Ukrainian].
2. Pro zatverdzhennia Instruktsii pro pryznachennia ta provedennia sudovykh ekspertyz ta ekspertnykh doslidzhen ta Naukovo-metodychnykh rekomendatsii z pytan pidhotovky ta pryznachennia sudovykh ekspertyz ta ekspertnykh doslidzhen: nakaz Ministerstva yustytsii Ukrainy vid 08.10.1998 № 53/5 [On approval of the Instruction on appointment and conduct of forensic examinations and expert studies and methodological recommendations: Order of the Ministry of Justice of Ukraine № 53/5 of October 8, 1998]. https://zakon.rada.gov.ua/go/z0705-98 [in Ukrainian].
3. Kryvonozhko, H., & Sobin, O. (2026). Shtuchnyi intelekt i spetsializovani tsyfrovi instrumenty v ekspertnomu zabezpechenni rozsliduvannia kiberzlochyniv ta tsyfrovykh pravoporushen [Artificial intelligence and specialized digital tools in expert support for cybercrime and digital offence investigations]. In Kryminalistyka maibutnoho: tsyfrovi innovatsii, shtuchnyi intelekt, hlobalni vyklyky ta zahrozy [Forensics of the future: Digital innovations, artificial intelligence, global challenges and threats] (pp. 184–188). Odesa: National University “Odesa Law Academy”. https://doi.org/10.32837/11300.33533 [in Ukrainian].
4. Svintsytskyi, A. (2025). Funktsional zamovlennia sudovoi ekspertyzy za dopomohoiu intehrovanoho chat-bota ChatGPT [Functionality for ordering forensic examination using an integrated ChatGPT chatbot]. Sudova ekspertyza [Forensic Science], (2), 4–9. https://doi.org/10.32782/forensic.science.2025.2.1 [in Ukrainian].
5. Chernyshev, M., Baig, Z., Syed, N., Doss, R., & Shore, M. (2026). Large language models in digital forensics: Capabilities, challenges and future directions. Forensic Science International: Digital Investigation, 56, 302043. https://doi.org/10.1016/j.fsidi.2025.302043
6. Metropolitansky, D., & Larson, J. (2026). VeriTrail: Closed-domain hallucination detection with traceability. International Conference on Learning Representations (ICLR 2026). https://proceedings.iclr.cc/paper_files/paper/2026/hash/42ec51d1031d25eca238ccc67559fe84-Abstract-Conference.html
7. Zhang, S., Yin, M., Zhang, J., Liu, J., Han, Z., Zhang, J., Li, B., Wang, C., Wang, H., Chen, Y., & Wu, Q. (2025). Which agent causes task failures and when? On automated failure attribution of LLM multi-agent systems. Proceedings of the 42nd International Conference on Machine Learning, 267, 76583–76599. https://proceedings.mlr.press/v267/zhang25cq.html
8. Rosset, C., Sharma, P., Zhao, A., González-Fernández, M., & Awadallah, A. (2026). The art of building verifiers for computer use agents. Conference on Language Modeling (COLM 2026). https://www.microsoft.com/en-us/research/publication/the-art-of-building-verifiers-for-computer-use-agents/
9. Kim, E. M., Garg, A., Peng, K., & Garg, N. (2025). Correlated errors in large language models. Proceedings of the 42nd International Conference on Machine Learning, 267, 30038–30066. https://proceedings.mlr.press/v267/kim25e.html
10. Laban, P., Schnabel, T., & Neville, J. (2026). LLMs corrupt your documents when you delegate. arXiv. https://doi.org/10.48550/arXiv.2604.15597
11. Srinivasan, A., Shamdasani, S., Araujo, A. F., & de Wasseige, J. (2026, May 27). Building self-improving tax agents with Codex. OpenAI. https://openai.com/index/building-self-improving-tax-agents-with-codex/
12. National Institute of Standards and Technology. (2026). The TEVV-Athlon framework for evaluating AI systems: Initial public draft NIST AI 200-2. https://doi.org/10.6028/NIST.AI.200-2.ipd

Halyna Kryvonozhko, Oleksii Sobin. Preventing false categorical conclusions when using large language models in forensic expert practice

The use of large language models in forensic expert practice creates a specific reliability risk: a plausible but unverified model output may be inherited by subsequent stages of an examination and influence a categorical expert conclusion. An error may arise during source selection and interpretation, content generation, revision, verification, technical transformation, or release of the final digital document. The problem is therefore not limited to isolated hallucinations or incorrect answers; it concerns the conditions under which unverified model-derived information may acquire professional significance. The purpose of this article is to substantiate a process-based approach to preventing false categorical conclusions when large language models are used in forensic expert activity. The study combines qualitative synthesis

of practical experience with complex analytical and document-oriented tasks, retrospective analysis of recurrent operational errors and control failures, process modelling, and comparison with research published in 2025–2026. Professional work with an LLM is treated as a sequence of professionally significant states. Transition between states is permitted only when predefined criteria are confirmed. Machine traceability and professional verification are distinguished: execution logs provide a basis for reconstructing the trajectory and diagnosing failures, whereas a control point determines whether a result may influence the next stage of the expert examination. Six practically significant risk classes are identified: reliance on an unsupported object or source; incorrect application of valid information; contamination between sources or contexts; loss of previously verified content during revision; a false claim that verification has been completed; and inconsistency between verified content and the digital object actually released. The article argues that an LLM output cannot serve as an independent basis for a categorical expert conclusion. If information obtained or transformed using an LLM is materially significant to the answer and cannot be properly verified independently of the model, it should not enter the logical basis of a categorical conclusion. The proposed process is not intended to guarantee an error-free language model. Its purpose is to prevent unverified model-derived information from being promoted to a categorical forensic conclusion without an adequate evidentiary and methodological basis.

***Keywords**: large language model, categorical conclusion, verification, professional error, forensic examination, quality control.*

Дата першого надходження статті до видання: 10.09.2026
Дата прийняття статті до друку після рецензування: 09.10.2026
Дата публікації (оприлюднення) статті: 18.11.2026

Стаття поширюється на умовах ліцензії
відкритого доступу (CC BY 4.0)