

Тези доповідей на пленарному засіданні конференції

Олексій КОСТЕНКО, завідувач наукової лабораторії імерсивних технологій і права наукового центру цифрової трансформації і права, Державна наукова установа «Інститут інформації, безпеки і права Національної академії правових наук України», доктор юридичних наук, старший дослідник

ШТУЧНИЙ ІНТЕЛЕКТ У ПРАВІ, ПСИХОЛОГІЇ ТА ПРАВООХОРОННИЙ ДІЯЛЬНОСТІ INTERPOL: ВІД МЕТАЗЛОЧИНІВ ДО ЦИФРОВОГО КОДЕКСУ

1. АКТУАЛЬНІСТЬ: ПРАВО У ПОСТЦИФРОВОМУ ПРОСТОРИ

Сучасна система права формувалася впродовж тисячоліть у фізичному, географічно локалізованому просторі. Людина як суб'єкт права, злочин як діяння у конкретному місці й часі, юрисдикція як суверенітет над визначеною територією — ці базові категорії відображали онтологічну реальність, в якій право функціонувало. Поява Web 3.0/4.0, генеративного штучного інтелекту та технологій розширеної реальності (XR) зруйнувала цю реальність, не запропонувавши натомість нових нормативних рамок.

Три паралельні тенденції формують новий правовий ландшафт. По-перше, ШІ-системи стають квазіавтономними учасниками правовідносин — не лише інструментами, а суб'єктами, чий рішення мають юридично значущі наслідки для конкретних осіб у сферах кримінального правосуддя, адміністративного управління та психологічних послуг. По-друге, простір правопорушень зміщується у Метавсесвіт: злочини вчиняються через аватари, без фізичного контакту, одночасно у декількох юрисдикціях, не залишаючи матеріальних слідів. По-третє, психологічна наука стикається з феноменом «алгоритмічного свідомості» — ШІ-систем, здатних маніпулювати поведінкою та психоемоційним станом осіб через персоналізований контент у XR-середовищах.

Реакція міжнародних інституцій підтверджує критичну серйозність цих тенденцій: ІНТЕРПОЛ у 2022 році запустив перший у світі правоохоронний Метавсесвіт, а у 2024 році оприлюднив White Paper з доктриною метазлочинів. Ці кроки перегукуються з авторськими концепціями цифрової юрисдикції та Цифрового кримінального кодексу, що розроблялися незалежно і паралельно у рамках вітчизняної правової науки [1; 2; 3].

2. ДОСВІД ІНТЕРПОЛ: ПРАВООХОРОННИЙ МЕТАВСЕСВІТ ТА ДОКТРИНА МЕТАЗЛОЧИНІВ

Поворотним моментом у визнанні Метавсесвіту як правоохоронного інструменту стала 90-та Генеральна Асамблея ІНТЕРПОЛ у Нью-Делі (жовтень 2022 року). Саме тоді відбулася сенсаційна презентація першого у світі Метавсесвіту, спеціально призначеного для правоохоронних органів [4]. Учасники Асамблеї через VR-гарнітури здійснили аватарний тур штаб-квартирою ІНТЕРПОЛ у Ліоні, пройшли навчання з верифікації документів у VR-класі та відпрацювали прикордонний контроль у симульованому аеропорту. Платформу розміщено на INTERPOL Secure Cloud — власній хмарній інфраструктурі організації, що гарантує нейтральність і захист від стороннього втручання.

Генеральний секретар ІНТЕРПОЛ Юрген Шток підкреслив: *«Метавсесвіт — це не абстрактне майбутнє. Він вже тут. Питання, які він ставить, — це питання, що завжди мотивували ІНТЕРПОЛ: підтримка держав-членів у боротьбі зі злочинністю та забезпечення безпеки світу — віртуального чи ні»*.

Водночас з метавсесвітом ІНТЕРПОЛ створив Metaverse Expert Group (MEG) — міжвідомчий орган, що об'єднав представників правоохоронних органів, урядів, приватного сектора, академічних установ та міжнародних організацій. Завдання MEG — розробка рекомендацій щодо протидії метазлочинам і забезпечення принципу «secure by design» у розвитку метавсесвіту [4].

У вересні 2023 року ІНТЕРПОЛ розширив застосування Метавсесвіту, запустивши першу XR-програму навчання боротьби з порушеннями інтелектуальної власності — у рамках International IP Crime Investigators College (IPRIC). Слідчі відпрацьовують перевірку складів із контрафактом, вивчають ознаки фальсифікованих документів та здійснюють криміналістичний аналіз у повністю зануреному цифровому середовищі, без ризиків, пов'язаних із реальним обігом небезпечних товарів [7].

Ключовим теоретичним документом стала White Paper ІНТЕРПОЛ «Metaverse: A Law Enforcement Perspective», презентована у Давосі у рамках World Economic Forum (18 січня 2024 року). Документ запровадив концепцію «metacrimes» — метазлочинів: злочинних діянь, вчинюваних у метавсесвіті або за його посередництва, для яких чинні правові системи не мають адекватних норм реагування. Генеральний секретар Шток констатував: «Ми спостерігаємо, як Метавсесвіт і штучний інтелект створюють нові можливості для злочинної діяльності, до яких світ не готовий повною мірою» [5].

Ключовою ідеєю документа є також правоохоронний потенціал Метавсесвіту — не лише як середовища злочинності, але і як інструменту її розкриття та профілактики: можливість VR-реконструкції місць злочину для криміналістичного аналізу та судового представлення доказів; навчання першого ешелону реагування у симульованих середовищах; збереження та аналіз цифрових доказів; взаємодія з громадянами через цифрові поліцейські пункти у метавсесвіті. Слідчі ІНТЕРПОЛ прямо зазначають: можна уявити прокурора, який проводить суд присяжних через VR-реконструкцію місця злочину у Метавсесвіті — і це вже не наукова фантастика, а найближча технологічна перспектива.

У березні 2024 року ІНТЕРПОЛ, спільно з ЮНІСРІ, оприлюднив оновлений Toolkit відповідального застосування ШІ у правоохоронній практиці. Документ закріпив принципи пояснюваного штучного інтелекту (ХАІ), підзвітності алгоритмів та обов'язкового людського контролю над ШІ-рішеннями у поліцейській і слідчій практиці [6].

Хронологію та зміст ключових ініціатив ІНТЕРПОЛ у сфері Метавсесвіту та ШІ систематизовано у Табл. 1.

Таблиця 1

Хронологія ключових ініціатив ІНТЕРПОЛ у сфері метавсесвіту та ШІ (2022–2024)

Подія / документ	Дата	Ключовий зміст	Правове значення
INTERPOL Metaverse — перший поліцейський Метавсесвіт	20 жовтня 2022 р., 90-та Генеральна Асамблея ІНТЕРПОЛ, Нью-Делі	Запуск першого у світі Метавсесвіту для правоохоронних органів; 3D-тур штаб-квартирою в Ліоні через аватари; навчання верифікації документів та прикордонного контролю у VR-класі	Перший прецедент інституціалізації Метавсесвіту як правоохоронного інструменту; статус INTERPOL Secure Cloud забезпечує нейтральність платформи
Створення Metaverse Expert Group (MEG)	Жовтень 2022 р.	Міжнародна група експертів ІНТЕРПОЛ з Метавсесвіту: представники поліції, урядів, приватного сектора, академії та НУО; завдання — вироблення рекомендацій щодо протидії метазлочинам	Основа для міжнародного кримінально-правового регулювання Метавсесвіту; перший крок до формування lex metaverse

Подія / документ	Дата	Ключовий зміст	Правове значення
White Paper «Metaverse: A Law Enforcement Perspective»	18 січня 2024 р., WEF, Давос	Типологія метазлочинів (metacrimes); рекомендації щодо криміналістики, збору доказів з VR-гарнітур та серверів; інструментарій для реконструкції місць злочину у 3D	Доктринальна основа для криміналізації метазлочинів; підтверджує необхідність спеціалізованого кримінального законодавства (концепція Цифрового КК)
Імерсивне навчання для боротьби з ІР-злочинами	Вересень 2023 р.	Перша XR-програма навчання інспекторів інтелектуальної власності: VR-склади, місця злочину, криміналістичні лабораторії у Метавесвіті	Прецедент застосування Метавесвіту для спеціалізованої правоохоронної підготовки; формує стандарт immersive policing training
Оновлений INTERPOL AI Toolkit для правоохоронців	Березень 2024 р.	Ревізований інструментарій відповідального застосування ШІ в правоохоронній практиці; спільно з ЮНІСРІ; принципи ХАІ та підзвітності алгоритмів	Міжнародний стандарт ethi-compliant AI у поліцейській практиці; основа для імплементації рівня VCBR Цифрового Кодексу

Наведена хронологія засвідчує системний, поетапний характер діяльності ІНТЕРПОЛ у цій сфері — від технологічної демонстрації (2022) до доктринальної кодифікації метазлочинів (2024). Ця траєкторія підтверджує достовірність і своєчасність авторських концепцій, що розвивалися паралельно у вітчизняній правовій науці.

3. ТИПОЛОГІЯ МЕТАЗЛОЧИНІВ: СПІВВІДНОШЕННЯ З ЦИФРОВИМ КОДЕКСОМ

White Paper ІНТЕРПОЛ (2024) встановила типологію метазлочинів, що включає: грумінг неповнолітніх, радикалізацію та вербування через XR-контент, кібер-фізичні атаки на критичну інфраструктуру, крадіжку 3D-власності та цифрових активів, самовільне проникнення у приватні VR-простори, пограбування аватарів, маніпуляції з NFT і шахрайство з цифровими активами [5]. Аналіз цієї типології у співвідношенні з авторським проектом Цифрового кримінального кодексу виявляє значний рівень концептуальної конвергенції — при тому, що обидва документи розроблялися незалежно (Табл. 2).

Таблиця 2

Типологія метазлочинів ІНТЕРПОЛ (2024) та відповідні склади Цифрового Кодексу

Тип метазлочину	Характеристика	Наявність у КК України	Цифровий КК (Костенко)
Grooming у метавесвіті	Цілеспрямований цифровий контакт педофілів із неповнолітніми через аватари та XR-середовища	Частково (ст. 156 КК — розбещення)	Окремий склад: «сексуальне грумінг-переслідування у цифровому середовищі»
Радикалізація та вербування в екстремістські організації через VR	Використання іммерсивного XR-контенту для психологічної обробки та рекрутування	Відсутній спеціальний склад	«Кібертероризм у метавесвіті»; посилена відповідальність за XR-вербування

Тип метазлочину	Характеристика	Наявність у КК України	Цифровий КК (Костенко)
Кібер-фізичні атаки на критичну інфраструктуру	Атаки з метавсесвіту на фізичні об'єкти через цифрові двійники промислових систем (SCADA)	Ст. 361 КК (часткове покриття)	«Кібер-фізичний напад»: окремий склад з підвищеною санкцією
Пограбування аватара / theft of virtual assets	Незаконне заволодіння цифровими активами, NFT, ігровими предметами через злам або шахрайство у XR	Відсутній; ст. 190 КК незастосовна через відсутність «майна»	«Цифрове пограбування»: розкрадання цифрових нематеріальних активів
Самовільне проникнення у приватний VR-простір	Несанкціонований доступ до приватних XR-просторів («trespassing in virtual spaces»)	Відсутній (ст. 162 КК — лише фізичне житло)	«Цифрове вторгнення»: порушення недоторканності цифрового простору особи
ШІ-deepfake із злочинною метою	Генерація реалістичного аудіо/відео/XR-контенту для шахрайства, дискредитації, маніпуляції правосуддям	Ст. 190, 351 ¹ КК — часткове покриття	«ШІ-шахрайство»: окремий склад із кваліфікуючою ознакою застосування GenAI

Порівняльний аналіз свідчить, що всі категорії метазлочинів, ідентифіковані ІНТЕРПОЛ, мають відповідники у авторській концепції Цифрового Кримінального Кодексу [1; 2]. Водночас авторська концепція є більш деталізованою у частині нейровтручання і ШІ-шахрайства з елементами GenAI, тоді як документ ІНТЕРПОЛ приділяє більшу увагу процедурним аспектам криміналістики (збір доказів із VR-гарнітур, серверів, блокчейн-аналітика) [5]. Це дозволяє стверджувати про взаємодоповнюваність двох підходів і доцільність їх синтезу при формуванні національного законодавства.

4. ШІ У ПРАВОСУДДІ ТА ЮРИДИЧНІЙ ПСИХОЛОГІЇ: СУЧАСНИЙ СТАН

Застосування ШІ у правосудді набуло глобального масштабу. Системи предиктивного правосуддя функціонують у десятках юрисдикцій: COMPAS (США) оцінює ризик рецидиву та застосовується при визначенні запобіжних заходів і умов дострокового звільнення; Prometea (Аргентина) автоматизує аналіз судових справ; ROSS Intelligence (США) здійснює семантичний пошук правових прецедентів; eJustice-AI (ЄС, пілотні програми) опрацьовує адміністративні справи. Суд штату Вісконсин у справі *State v. Loomis* (2016) став першим прецедентом оскарження вироку, постановленого з урахуванням ШІ-оцінки COMPAS — і визнав її конституційною, що викликало широку правову дискусію [13].

Ключовою проблемою є феномен «алгоритмічної упередженості» (algorithmic bias): дослідження ProPublica (2016) встановило, що COMPAS вдвічі частіше помилково класифікує темношкірих підсудних як «високоризикових» порівняно з білошкірими [10]. Це є прямим порушенням принципу рівності перед законом та права на справедливий суд, гарантованих ст. 6 ЄКПЛ. Дослідження ScienceDirect (2025) в галузі метавсесвітнього полісингу підтверджує, що аналогічні ризики будуть притаманні і ШІ-системам правоохоронної діяльності у XR-середовищах [11].

У галузі юридичної психології ШІ трансформує три ключові практики. По-перше, криміналістичний профайлінг: системи мультимодального аналізу (рух очей, голос, ЕЕГ-патерни, мікровирази обличчя) дозволяють оцінювати психоемоційний стан особи під час допиту або судового засідання без її відома. По-друге, ризик-оцінка рецидивізму: ШІ-системи

прогнозують поведінку засудженого на основі психологічних тестів, соціально-демографічних даних та поведінкової аналітики. По-третє, ІІІ-опосередкована психологічна реабілітація: особливо актуальна для ветеранів та осіб, що постраждали від наслідків збройного конфлікту, — але потребує ретельного правового регулювання в частині конфіденційності та відповідальності за шкоду.

Нейроправовий вимір цих трансформацій полягає у виникненні нового покоління прав людини — прав на когнітивну свободу, ментальну приватність, психологічну цілісність і психологічну тяглість (continuity). Ці права, обґрунтовані у рамках Neurorights Framework Колумбійського університету та частково закріплені у Конституції Чилі (2021), є прямою відповіддю на можливість технологічного втручання у ментальне середовище особи через XR-пристрої та ІІІ-платформи [12].

5. ЦИФРОВИЙ КОДЕКС ЯК АРХІТЕКТУРА РЕГУЛЮВАННЯ: ЗАСТОСУВАННЯ У ПРАВООХОРОННІЙ СФЕРІ

Концепція Цифрового Кодексу (Digital Code), розроблена автором у монографії «Metaverse та WEB 4.0» [1] та в дослідженні [2], є уніфікованою чотирирівневою нормативною платформою, що систематизує правовідносини у цифровому просторі. Кожен рівень платформи має специфічне значення для правоохоронної, кримінально-правової та психологічно-правової сфер (Табл. 3).

Таблиця 3

Архітектура Цифрового Кодексу (IVC/NMR/DLRE/VCBR) у правоохоронному, кримінально-правовому та психологічному контексті

Модуль	Предмет регулювання	Застосування у кримінальному праві та правосудді	Застосування у психології та правоохоронній практиці
IVC (ідентифікація та верифікація суб'єктності)	Правовий статус цифрової особи, аватара, ІІІ-агента; реєстр цифрових ідентичностей	Суб'єкт злочину у метавсесвіті; кримінальна відповідальність оператора ІІІ; «цифровий кіднепінг»	Верифікація особи у цифрових психологічних консультаціях; ідентифікація в ІІІ-профайлінгу
NMR (нормативно-метавсесвітні відносини)	Правовідносини у XR-просторі; lex protocolli; смарт-контрактне право; GLM	Юрисдикція над метазлочинами; транскордонне кримінальне переслідування; Цифровий КК	Правові рамки ІІІ-профайлінгу; нейроправовий захист у XR-сесіях
DLRE (цифрово-правове регулювання середовища)	Алгоритмічне управління; платформна відповідальність; аудит ІІІ-рішень	Відповідальність за предиктивне правосуддя; стандарти XAI у слідчій практиці; INTERPOL AI Toolkit	Регулювання ІІІ-психодіагностики; відповідальність за хибний психологічний діагноз ІІІ
VCBR (верифікація, відповідальність, права)	Цифровий due process; право на пояснення алгоритму; цифровий суд; омбудсмен	Апеляція на рішення COMPAS/предиктивних систем; захист від алгоритмічної дискримінації	Право на когнітивну свободу; захист нейроданих; INTERPOL immersive training standards

Особливого значення в контексті взаємодії з ініціативами ІНТЕРПОЛ набуває рівень NMR — нормативно-метавсесвітніх відносин, який включає механізм lex protocolli (право протоколу платформи як самостійне джерело цифрового права). Саме цей механізм може

забезпечити правову силу процедурним стандартам ІНТЕРПОЛ щодо збору і збереження цифрових доказів та правил поведінки в правоохоронному метавсесвіті на рівні, наближеному до обов'язкової норми для держав-членів організації.

Концепція Grand Lex Metaverse (GLM) — транснаціонального цифрового права як *lex specialis* для вирішення юрисдикційних конфліктів — прямо кореспондує з висновком White Paper ІНТЕРПОЛ (2024) щодо необхідності «багатостороннього підходу із залученням множинних юрисдикцій, вимірів та організацій» у протидії метазлочинам [5]. GLM може стати доктринальною основою такого підходу, подолавши роздробленість національних юрисдикцій без шкоди для принципу державного суверенітету [1].

6. ПРОБЛЕМИ ВПРОВАДЖЕННЯ В УКРАЇНІ ТА РЕКОМЕНДАЦІЇ

Україна перебуває у специфічній геополітичній і технологічній ситуації, що одночасно ускладнює і стимулює впровадження ІІІ у правоохоронну та юридично-психологічну сфери. Воєнний стан, масштабна цифрова агресія (понад 4 тисячі кібератак лише на критичну інфраструктуру за 2023–2024 роки за даними CERT-UA), необхідність документування воєнних злочинів у режимі реального часу та гостра потреба у психологічній реабілітації мільйонів ветеранів і цивільних — усі ці фактори роблять питання правового регулювання ІІІ не академічним, а невідкладно практичним.

Проблема 1. Юрисдикційний вакуум у кіберпросторі. Чинні КК і КПК України не містять норм, що регулювали б злочини, вчинені через Метавсесвіт або за допомогою цифрових аватарів. Будапештська конвенція про кіберзлочинність (2001), до якої приєдналась Україна, охоплює лише традиційні ІКТ-злочини і не поширюється на метавсесвіт як середовище злочинної діяльності.

Проблема 2. Відсутність правових стандартів ІІІ у правосудді. Проект Закону України «Про штучний інтелект» не містить спеціальних норм щодо застосування ІІІ у кримінальному судочинстві, предиктивній поліцейській діяльності або психологічній судовій експертизі. Це створює ризик безсистемного впровадження ІІІ-інструментів без необхідних гарантій права на захист (пор. підхід ЄС: EU AI Act [8] та стратегія Web 4.0 [9]).

Проблема 3. Відсутність захисту нейроданих. Чинний Закон України «Про захист персональних даних» не виокремлює нейрофізіологічні дані як окрему категорію, що потребує посиленого захисту. Водночас CERT-UA фіксує випадки витоку медичних і психологічних даних при атаках на заклади охорони здоров'я.

Проблема 4. Відсутність процесуального статусу VR-доказів. КПК України не містить положень щодо допустимості доказів, отриманих із VR-гарнітур, метаплатформ або генерованих ІІІ-системами. Це унеможливує ефективне кримінальне переслідування за метазлочини навіть за наявності технічних доказів.

З урахуванням окреслених проблем пропонуються такі рекомендації.

Рекомендація 1. Розробити та прийняти Концепцію Цифрового Кодексу України як першого кроку до системного регулювання цифрових правовідносин, включно із спеціальним розділом щодо застосування ІІІ у правосудді та правоохоронній діяльності.

Рекомендація 2. Ратифікувати Другий Додатковий протокол до Будапештської конвенції про кіберзлочинність (2022) та ініціювати розширення предметної сфери конвенції для охоплення метазлочинів.

Рекомендація 3. Розпочати розробку процесуального регламенту щодо допустимості та оцінки доказів, отриманих із метавсесвіту та XR-пристроїв, у рамках оновлення КПК України.

Рекомендація 4. Закріпити нейродані як окрему категорію персональних даних підвищеного захисту у Законі «Про захист персональних даних» та Законі «Про штучний інтелект».

Рекомендація 5. На базі Наукової лабораторії імерсивних технологій і права ДНУ ШБП НАПрН України та Академії МВС створити Центр компетенцій з правоохоронного Метавсесвіту, орієнтований на підготовку слідчих, прокурорів і суддів до роботи з доказами метазлочинів і ІІІ-рішень.

7. ВИСНОВКИ

Проведене дослідження дозволяє сформулювати такі узагальнені висновки.

1. Стрімкий розвиток ІІІ та XR-технологій формує принципово новий правовий ландшафт, в якому традиційні категорії кримінального права, юрисдикції та доказування потребують кардинального переосмислення.
2. Ініціативи ІНТЕРПОЛ — правоохоронний метавсесвіт (2022), Metaverse Expert Group (2022), XR-навчання (2023) і White Paper з метазлочинів (2024) — є незаперечним міжнародним підтвердженням актуальності авторської концепції цифрової юрисдикції та Цифрового Кодексу.
3. Концепція Цифрового Кодексу (IVC/NMR/DLRE/VCBR) є найбільш комплексним вітчизняним доктринальним інструментом для правового регулювання ІІІ-опосередкованих відносин у праві, психології та правоохоронній діяльності.
4. Типологія метазлочинів ІНТЕРПОЛ демонструє значний рівень конвергенції з авторським проектом Цифрового КК, що свідчить про методологічну коректність останнього та відкриває можливості для їх синтезу при розробці вітчизняного законодавства.
5. Україна має унікальний шанс — на підставі власного кризового досвіду цифрових воєнних злочинів, кібератак та масштабної необхідності в ІІІ-підтримці правоохоронної та психологічної практики — сформулювати передове законодавство у цій сфері, що може стати орієнтиром для регіону Центральної та Східної Європи.

Список використаних джерел

1. Костенко О. В. Metaverse та WEB 4.0 : монографія. Київ : Фенікс, 2026. 380 с.
2. Костенко О. В. Цифрова юрисдикція: модель. *Інформація і право* . 2025. № 4(55). DOI: [https://doi.org/10.37750/2616-6798.2025.4\(55\).346297](https://doi.org/10.37750/2616-6798.2025.4(55).346297) .
3. Костенко О. В. Електронна юрисдикція, метавсесвіт, штучний інтелект, цифрова особистість, цифровий аватар, нейронні мережі: теорія, практика, перспективи. *Наукові інновації та передові технології* . 2022. № 2(4). URL: <https://www.researchgate.net/publication/358486613> (дата звернення: 08.06.2026).
4. INTERPOL launches first global police Metaverse : офіц. повідомл. INTERPOL. 20 жовт. 2022 р. URL: <https://www.interpol.int/en/News-and-Events/News/2022/INTERPOL-launches-first-global-police-Metaverse> (дата звернення: 08.06.2026).
5. Metaverse: A Law Enforcement Perspective. Use Cases, Crime, Forensics, Investigation, and Governance: White Paper / INTERPOL Metaverse Expert Group. Davos: INTERPOL, Jan. 2024. 27 р. URL: <https://www.interpol.int/en/content/download/20828/file/Metaverse%20-%20a%20law%20enforcement%20perspective.pdf> (дата звернення: 08.06.2026).
6. Revised toolkit empowers law enforcement with responsible AI practices : офіц. повідомл. INTERPOL. 6 берез. 2024 р. URL: <https://www.interpol.int/en/News-and-Events/News/2024/Revised-toolkit-empowers-law-enforcement-with-responsible-AI-practices> (дата звернення: 08.06.2026).
7. Capitalizing on immersive learning to fight IP crime : офіц. повідомл. INTERPOL. Верес. 2023 р. URL: <https://www.interpol.int/en/News-and-Events/News/2023/Capitalizing-on-immersive-learning-to-fight-IP-crime> (дата звернення: 08.06.2026).
8. Регламент (ЄС) 2024/1689 Європейського Парламенту та Ради від 13 червня 2024 року, що встановлює гармонізовані правила щодо штучного інтелекту (Закон про штучний інтелект). *Офіційний журнал Європейського Союзу* . 2024. L 1689. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_2024_1689 (дата звернення: 08.06.2026).
9. Веб 4.0 та віртуальні світи: лідерство у наступному технологічному переході: Повідомлення Комісії до Європейського Парламенту, Ради, Європейського економічного та соціального комітету та Комітету регіонів. COM(2023) 442 final. Брюссель: Європейська Комісія, 2023. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2023:442:FIN> (дата звернення: 08.06.2026).

10. Angwin J., Larson J., Mattu S., Kirchner L. Machine Bias. *ProPublica* . 2016. 23 May. URL: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (дата звернення: 08.06.2026).
11. Ієнка М., Андорно Р. На шляху до нових прав людини в епоху нейротехнологій: когнітивна свобода, ментальна конфіденційність, ментальна цілісність та психологічна безперервність. *Науки про життя, суспільство та політика* . 2017. Том 13, № 5. DOI: <https://doi.org/10.1186/s40504-017-0050-1> .
12. Haber E. The Criminal Metaverse. *Indiana Law Journal*. 2024. Vol. 99, iss. 3. P. 843–891. URL: indiana.edu (дата звернення: 08.06.2026).
13. State v. Loomis. 371 Wis. 2d 235, 881 N.W.2d 749 (Wis. 2016). URL: <https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690> (дата звернення: 08.06.2026).

Tim D. WILLIAMS
BSc, MSc, MBA, PGCHE
FBCS CITP, MCIIS, MIET, GMBPsS
Independent researcher
ORCID: 0000-0002-1788-5736

WHY AI LEGAL ASSISTANTS HALLUCINATE IN NEW FIELDS: A MATHEMATICAL LIMIT, WITH POST-QUANTUM CRYPTOGRAPHY (PQC) AS EXAMPLE

1. Introduction and motivation

Some AI errors in legal practice cannot be removed by better engineering. They follow from a mathematical limit. Practitioners need to know which legal questions sit inside that limit.

AI assistants are now part of everyday legal work. Lawyers use them for research, drafting and first-pass review. When these assistants make confident mistakes, as in the US case *Mata v. Avianca* [1], the usual response is to call for better training data, better prompts and better engineering. This paper offers a different perspective. One specific kind of AI error cannot be removed by engineering. It follows from a mathematical proof. The proof applies to any AI assistant that gives confidence-marked answers, both now and in the future. The professional task is therefore not to wait for AI to become reliable. The task is to learn which legal questions fall inside the mathematical limit, and to avoid reliance on AI outside its known limits of reliability. Post-quantum cryptography (PQC) is used here as a worked example. PQC is a new legal field. The public record is thin. The stakes are high.

2. The Kalai–Vempala lower bound

A 2024 mathematical result shows that any properly calibrated AI assistant must make a minimum rate of errors. The minimum rate depends on how often facts appear in the training data.

In 2024, Kalai and Vempala proved a result about AI language models [2]. The result is a lower bound, which means a minimum that cannot be reduced. This bound applies to any AI model that is well-calibrated, meaning a model whose stated confidence matches how often it is correct. This bound says: the model must make errors at a rate at least equal to the share of facts that appear exactly once in its training data. Xu, Jain and Kankanhalli then strengthened the result [3]. They showed that some errors are unavoidable as a matter of computability, not training effort.

The practical meaning is simple. An AI assistant is not equally reliable across all domains of knowledge. The model is reliable where the same fact appears many times, in many sources. The model is unreliable where the authoritative source appears only once. This is not a problem of current models. It is a proven limit on every future calibrated model. The professional task is to identify which legal domains fall on the unreliable side of the limit.