

УДК 341.1/8

DOI <https://doi.org/10.24144/2307-3322.2025.90.5.18>

АВТОНОМНІ СИСТЕМИ З ШІ У ЗБРОЙНИХ КОНФЛІКТАХ Т А ПРАВОВЕ ЗАБЕЗПЕЧЕННЯ ДОТРИМАННЯ ПРИНЦИПУ РОЗРІЗНЕННЯ

Ігнатуша В.В.,
*аспірант відділу міжнародного права та права Європейського Союзу
Інституту держави і права імені В.М. Корецького НАН України.
e-mail: v.ihnatusha@gmail.com*

Ігнатуша В.В. Автономні системи з ШІ у збройних конфліктах та правове забезпечення дотримання принципу розрізнення.

Стаття присвячена аналізу проблем дотримання принципу розрізнення міжнародного гуманітарного права при використанні автономних систем зі штучним інтелектом під час збройних конфліктів. Досліджено технічні обмеження сучасних автономних систем із ШІ трьох рівнів автономності (Human-in-the-Loop, Human-on-the-Loop, Human-out-of-the-Loop), включаючи безпілотні літальні апарати, «рої» дронів, автономні системи наведення, роботизовані бойові платформи тощо. Виявлено критичну проблему «семантичного розриву» між алгоритмічним розпізнаванням об'єктів та розумінням їх значення в контексті збройного конфлікту, що унеможливлює надійну ідентифікацію комбатантів у складних бойових умовах. Проаналізовано досвід застосування таких систем у сучасних збройних конфліктах, зокрема випадки атак на цивільні об'єкти. Встановлено правові прогалини в регулюванні використання автономних систем з ШІ, зокрема невідомість механізмів встановлення відповідальності за порушення міжнародного гуманітарного права. Критично оцінено пропозиції наукової спільноти щодо повної заборони автономних систем із ШІ та обґрунтовано більш реалістичний підхід. Запропоновано трикомпонентну модель забезпечення дотримання принципу розрізнення при використанні автономних систем із ШІ: створення Міжнародної агенції з контролю автономних систем озброєння під егідою ООН/UNODA з детальною структурою та повноваженнями; розробку трирівневої процедури сертифікації систем з вимогою 99% точності розпізнавання; проект Конвенції про тимчасовий мораторій на повністю автономні системи з ШІ. Обґрунтовано необхідність збереження значущого контролю людиною над критичними функціями таких систем. Враховуючи ймовірну відсутність політичної волі провідних держав-розробників таких систем, запропоновано альтернативні механізми: Директиву ЄС, міжнародний стандарт ISO, керівництво МКЧХ, добровільний кодекс поведінки виробників. Розвинуто ідею створення спеціального правового режиму для миротворчих безпілотників з багаторівневою системою захисту через візуальні, електронні та програмні ідентифікатори.

Ключові слова: міжнародне гуманітарне право, міжнародне право, принцип розрізнення, автономні системи з ШІ, штучний інтелект, збройний конфлікт, людський контроль, міжнародні організації, ООН, міжнародний договір.

Ihnatusha V.V. Autonomous AI systems in armed conflicts and legal framework for ensuring compliance with the principle of distinction.

The article is devoted to the analysis of compliance issues with the principle of distinction of international humanitarian law when using autonomous systems with artificial intelligence during armed conflicts. The study examines technical limitations of contemporary autonomous AI systems across three levels of autonomy (Human-in-the-Loop, Human-on-the-Loop, Human-out-of-the-Loop), including unmanned aerial vehicles, drone swarms, autonomous targeting systems, robotic combat platforms, and others. The research identifies the critical problem of the «semantic gap» between algorithmic object recognition and understanding their significance within the context of armed conflict, which precludes reliable identification of combatants in complex combat conditions. The experience of deploying such systems in contemporary armed conflicts has been analyzed, including cases of attacks on civilian objects. Legal lacunae in the regulation of autonomous AI systems deployment have been

established, particularly the uncertainty of mechanisms for establishing accountability for violations of international humanitarian law. Proposals from the scientific community concerning a complete prohibition of autonomous AI systems have been critically assessed, and a more realistic approach has been substantiated. A three-component model for ensuring compliance with the principle of distinction when employing autonomous AI systems is proposed: establishing an International Agency for the Control of Autonomous Weapon Systems under the auspices of UN/UNODA with detailed structure and competencies; developing a three-tier certification procedure for systems with a requirement of 99% recognition accuracy; drafting a Convention on a temporary moratorium on fully autonomous AI systems. The necessity of maintaining meaningful human control over critical functions of such systems has been substantiated. Considering the probable absence of political will among leading states developing such systems, alternative mechanisms have been proposed: an EU Directive, international ISO standard, ICRC guidelines, and a voluntary code of conduct for manufacturers. The concept of establishing a special legal regime for peacekeeping unmanned aerial vehicles with a multi-level protection system through visual, electronic, and software identifiers has been developed.

Key words: international humanitarian law, international law, principle of distinction, autonomous AI systems, artificial intelligence, armed conflict, human control, international organizations, UN, international treaty.

Постановка проблеми. Стрімкий розвиток технологій штучного інтелекту (ШІ) та їх інтеграція у воєнну сферу створюють безпрецедентні виклики для міжнародного гуманітарного права. Автономні системи з ШІ є системами озброєння, що здатні самостійно виявляти, ідентифікувати та уражати цілі без безпосереднього людського втручання в момент прийняття рішення про застосування сили. До таких систем, зокрема, належать безпілотні літальні апарати з автономним режимом наведення та визначення цілі, такі як деякі версії дронів «Ланцет», «Shahed»; дрони, що працюють за системою «рою»; роботизовані бойові платформи; системи протиповітряної оборони; автономні підводні апарати та інші системи, що використовують алгоритми ШІ для прийняття рішень. Як зазначає Ж.-М. Ріклі, «новітні технології, такі як штучний інтелект та дешевші технологічні альтернативи традиційному озброєнню, такі як дрони, безсумнівно, починають змінювати поле бою та відіграватимуть дедалі більшу роль у майбутніх конфліктах» [1, с. 4]. Практичне застосування таких систем у сучасних збройних конфліктах демонструє серйозні проблеми з дотриманням принципу розрізнення. Показовими є випадки використання РФ у російсько-українському збройному конфлікті автономних систем з ШІ у вигляді дронів, що атакували цивільні автобуси замість військових цілей [2]. Воєнний аналітик Юліан Репке підкреслює, що ці дрони «практично некеровані» та «не здатні розрізнити військову ціль від цивільної» [3]. За даними Стокгольмського міжнародного інституту досліджень проблем миру, понад 80 країн світу активно розвивають воєнні технології з використанням ШІ [4, с. 89], що підкреслює нагальність міжнародно-правового регулювання таких систем. Однією з основних проблем є те, що автономні системи з ШІ на даному етапі розвитку обмежені технічними можливостями розпізнавання цілей, що робить їх наразі нездатними повною мірою дотримуватися принципу розрізнення, закріпленого в статті 48 Додаткового протоколу I до Женевських конвенцій. Як зазначають Дж. Квік та Т. Ван Енгерс, «відсутність передбачуваності та пояснюваності перешкоджає виконанню ключових вимог МГП до такої міри, що використання повністю непрозорих ШІ може порушувати міжнародне право» [5, с. 43].

Аналіз досліджень та нормативної бази. Проблема дотримання принципу розрізнення при використанні автономних систем із ШІ під час збройних конфліктів викликає інтерес у сучасних фахівців з міжнародного права. Окремі аспекти цієї проблематики досліджуються в роботах Ж.-М. Ріклі, Дж. Квіка, Т. Ван Енгерса, Д.А. Льюїса, М. Брістона, З. Піотровського, Дж.Г. Хьюса, Б. Дресп-Ленглі, Н. Сірегара, М. Саті, І.М. Проценко, З. Калленборна, Д. Адама, А.Г. Міхеля та інших. Нормативну основу дослідження складають Додатковий протокол I до Женевських конвенцій 1949 року [6], Консультативний висновок Міжнародного суду ООН щодо правомірності застосування ядерної зброї 1996 року [7], документи Міжнародного комітету Червоного Хреста [8, 9], а також документи Групи урядових експертів з летальних автономних систем зброї (GGE LAWS) [10, 11]. У 2019 році GGE LAWS сформулювала 11 керівних принципів щодо регулювання автономних систем з ШІ, де особливо підкреслюється, що «міжнародне гуманітарне право повною мірою застосовується до всіх систем озброєння, включаючи потенційну розробку та

використання летальних автономних систем» [10]. У її документах особлива увага приділяється концепції «значущого контролю людиною», яка розглядається як необхідна умова дотримання норм МГП. Як зазначається у звіті GGE LAWS 2021 року: «автономні системи зброї повинні функціонувати в рамках відповідального ланцюга командування та контролю, з чіткими лініями відповідальності людини за рішення про застосування сили» [11].

Водночас аналіз наукової літератури виявляє фрагментарність досліджень проблематики автономних систем із ШІ. Більшість авторів зосереджуються або на технічних аспектах функціонування таких систем (наприклад, Дж.Г. Хьюз, Б. Дресп-Ленглі), або на загальних правових питаннях (наприклад, Д.А. Льюїс, Н. Сірегар). Бракує практичних пропозицій щодо інституційного забезпечення контролю за автономними системами з ШІ, детальних процедур верифікації їх відповідності принципу розрізнення тощо. Існуючі дослідження також не завжди враховують досвід застосування таких систем у сучасних конфліктах, зокрема в російсько-українському збройному конфлікті.

Мета статті – на основі аналізу поточних технічних обмежень автономних систем із ШІ та прогалин у їх міжнародно-правовому регулюванні запропонувати модель забезпечення дотримання принципу розрізнення, що включатиме інституційні механізми, процедурні інструменти та альтернативні шляхи регулювання через регіональні та галузеві міжнародні організації.

Виклад основного матеріалу. Автономні системи з ШІ функціонують на трьох рівнях автономності. По-перше, системи під прямим контролем людиною (Human-in-the-Loop) вимагають підтвердження оператора перед ураженням цілі. По-друге, системи під наглядом людини (Human-on-the-Loop), такі як системи C-RAM, діють автономно, але оператор може втрутитися та скасувати дії системи. По-третє, системи без безпосереднього людського контролю (Human-out-of-the-Loop) самостійно виявляють та уражають цілі відповідно до заданих параметрів [12, с. 378]. Технічні можливості цих систем базуються на алгоритмах ШІ, машинного навчання та комп'ютерного зору. Основу становлять системи, які використовують різноманітні сенсори (оптичні, інфрачервоні, радарні) для збору даних про оточуюче середовище. У дослідженні проєкту VGI-9 зазначається, що сучасні системи здатні обробляти великі обсяги даних у реальному часі та приймати рішення на основі цих даних. Зокрема, як підкреслюється в цих дослідженнях, «сукупні дані, миттєво оброблені ШІ, надають набагато більше можливостей для військових дій». Водночас, ними підкреслюються і істотні проблеми у використанні ШІ. Так, одним із основних викликів точності та надійності розпізнавання цілей автономними системами з ШІ є «використання недостовірних або неповних даних, отриманих через перешкоди чи погану видимість» та «зміни зовнішніх умов – погоди, рівня освітлення тощо» [13]. Також, МКЧХ підкреслює фундаментальну проблему семантичного розриву: алгоритми ШІ можуть ідентифікувати та класифікувати об'єкти на зображеннях, але не розуміють значення або концепції цих об'єктів, що може призводити до помилок, які людина ніколи б не зробила [9, с. 11]. Окрім того, такі системи «створюють власні правила на основі даних, яким вони піддаються, внаслідок чого вони є набагато менш передбачуваними, ніж попередньо запрограмовані системи» [9, с. 10].

Принцип розрізнення, закріплений у статті 48 Додаткового протоколу I, вимагає від сторін конфлікту «завжди розрізняти цивільне населення й комбатантів, а також цивільні й військові об'єкти та відповідно спрямовувати свої дії тільки проти військових об'єктів» [6]. Міжнародний суд ООН у Консультативному висновку 1996 року визнав принцип розрізнення одним із «непорушних принципів міжнародного звичаєвого гуманітарного права» [7]. Однак автономні системи з ШІ демонструють критичні обмеження в ідентифікації комбатантів, особливо в умовах асиметричних конфліктів, коли учасники бойових дій не носять військової форми та змішуються з цивільним населенням. Також, І.М. Проценко слушно вказує на випадок з дроном «KARGU-2» у Лівії, який вийшов з-під контролю і нібито діяв «повністю автономно від будь-якого стороннього управління» [14, с. 147]. Хоча інформація про вбивство людини цим дроном залишається офіційно непідтвердженою, на думку дослідниці, сам факт можливості такої атаки підкреслює небезпеку автономних систем.

Як слушно зазначає Д.А. Льюїс, «багато способів використання ШІ сторонами збройного конфлікту та їх потенційні наслідки ще не відомі» [15, с. 490]. Концепція «розумних сумнівів» щодо статусу цілі, закріплена в статті 50 Протоколу I, важко піддається алгоритмічній формалізації в автономних системах з ШІ. М. Саті аналізує питання відповідальності та доходить висновку, що термін «особа» у контексті МГП «використовується для позначення «людини» згідно із загальним

духом Женевських конвенцій», і тому «короткою відповіддю на питання про те, чи можуть військові технології ШІ мати право на статус комбатанта, буде ні» [16, с. 130]. Це створює «прогалину у відповідальності», коли неможливо встановити відповідального за порушення МГП, оскільки «ШІ функціонує як непрямий об'єкт, що ускладнює чітке встановлення відповідальності» [17, с. 192]. Н. Сірегар підсумовує, що «міжнародне гуманітарне право на даний час не має чітких норм для ШІ, що підкреслює нагальну потребу в новій регуляторній рамці, що встановлює обмеження на передову зброю, не вдаючись до повних заборон» [17, с. 202].

Запропоновані у науковій літературі підходи до міжнародно-правового регулювання автономних систем із ШІ, можна звести до двох основних напрямків:

По-перше, пропозиція заборони розробки та застосування автономних систем з ШІ. Так міжнародні неурядові організації, як Campaign to Stop Killer Robots, International Committee for Robot Arms Control та Human Rights Watch виступають за всеосяжну заборону всіх автономних систем озброєння [18]. Б. Дресп-Ленглі аргументує необхідність заборони розробки та/або розгортання автономних систем з ШІ через міжнародне право. Авторка підкреслює, що «етичні дебати самі по собі не можуть вирішити проблему багатьох загроз, які автономні системи озброєння представляють для людства» [19, с. 15]. І.М. Проценко підтримує цей підхід, пропонуючи «запровадити принцип повної заборони розробки, виготовлення, використання та передачі такої летальної зброї, яка може діяти автономно від оператора, зокрема, під час її запуску, визначенні цілей та, особливо, їх знищенні» [14, с. 149]. Однак, як визнає сама авторка, «такий підхід, безумовно, не буде підтриманий рядом держав, зокрема тими, що активно розробляють безпілотники» [14, с. 149]. Погоджуючись із думкою дослідниці, зазначимо, що підхід, який забороняє повну розробку автономних систем з ШІ, ймовірно, отримає значний супротив серед держав, які займаються розробкою та впровадженням зазначених технологій, у зв'язку з чим варто вести мову про тимчасову заборону повністю автономних систем з ШІ, при цьому допускаючи розробку систем з обов'язковим контролем людиною. Примітно, що на рівні ООН лунають політичні заяви про часткову заборону автономних систем із ШІ, які, однак, поки що не знайшли втілення в розробці відповідних міжнародно-правових механізмів. Так, Генеральний секретар ООН Антоніу Гутерреш заявляв, що «автономні машини з владою та дискреційними повноваженнями вибирати цілі та забирати життя без участі людини є політично неприйнятними, морально огидними і повинні бути заборонені міжнародним правом» [18], а повної заборони зброї, яка діє без людського нагляду планувалось досягти до 2026 року [20]. Схожої думки дотримується дослідник автономних систем озброєння З. Калленборн, зазначаючи, що «замість широкої заборони всіх автономних систем озброєння міжнародна спільнота повинна визначити та зосередити обмеження на зброї найвищого ризику: рої дронів та автономну хімічну, біологічну, радіологічну та ядерну зброю». Соціолог Л. Сачман підкреслює важливість того, щоб «людські агенти повинні візуально перевіряти цілі перед авторизацією ударів і повинні мати можливість відмінити удар, якщо умови на полі бою зміняться» [20].

По-друге, розробка нових міжнародних договорів. Б. Дресп-Ленглі доходить висновку, що «міжнародна спільнота повинна встановити новий набір конвенцій для регулювання їх використання через міжнародне законодавство та договори» [19, с. 15]. Авторка зазначає, що «нові форми війни за допомогою зброї, контрольованої ШІ, з потенціалом масового знищення загрожують подальшій ефективності існуючих конвенцій, таких як Женевський протокол» [19, с. 15]. У свою чергу, Н. Сірегар зазначає, що існуючі міжнародні конвенції, що регулюють збройні конфлікти застаріли та потребують оновлення: «Гаазька конвенція 1907 року, яка регулює використання зброї у збройних конфліктах, є занадто широкою та застарілою, щоб враховувати складності, пов'язані зі зброєю, інтегрованою на основі штучного інтелекту» [17, с. 202]. Окрім того, Н. Сірегар наголошує на «нагальній потребі у всеосяжному регулюванні використання ШІ та автономних систем озброєння у війні» та вказує, що «специфічні норми мають бути встановлені для контролю та обмеження розгортання цих технологій» [17, с. 198]. Автор підкреслює, що «використання ШІ у війні має відповідати принципам міжнародного права, особливо принципам розрізнення, пропорційності та необхідності» [17, с. 198]. Однією з ключових його пропозицій є розробка норм, які б призначали відповідальну особу для кожної автономної системи з ШІ: «Існує нагальна потреба в нормативних актах, які призначатимуть відповідального наглядача за кожною зброєю на основі штучного інтелекту. Такі заходи є важливими для забезпечення відповідальності у випадках порушення міжнародного гуманітарного права» [17, с. 202].

Варто зазначити, що обидва напрямки недостатньо враховують динаміку розвитку ШІ та необхідність гнучкого реагування на нові виклики. Крім того, жоден з проаналізованих підходів не пропонує конкретної інституційної архітектури для забезпечення контролю над розробкою, впровадженням та використанням автономних систем із ШІ, обмежуючись загальними закликами до регулювання.

Погоджуючись з необхідністю забезпечення дотримання принципу розрізнення автономними системами з ШІ, та у розвиток ідей зазначених науковців, пропонуємо модель, що базується на трьох взаємопов'язаних компонентах. По-перше, створення спеціалізованої міжнародної інституції, що має забезпечити координацію зусиль держав-членів у питаннях нагляду і контролю за розробкою автономних систем із ШІ. По-друге, впровадження багаторівневої системи технічної сертифікації та моніторингу. По-третє, розробка міжнародного договору, що має стати правовим фундаментом для регулювання автономних систем із ШІ. У вигляді альтернативних механізмів розглядаємо можливість впровадження регулювання через регіональні та галузеві організації, що має створити паралельні шляхи формування міжнародних стандартів за умови відсутності універсального консенсусу. Кожен із цих компонентів є важливим у створенні ефективної системи контролю за дотриманням принципу розрізнення автономними системами з ШІ, при цьому доповнюючи один одного і компенсуючи потенційні слабкі місця окремих механізмів.

Інституційний механізм: створення міжнародної агенції. Центральним елементом запропонованої моделі є створення під егідою Організації Об'єднаних Націй, зокрема в рамках Управління ООН з питань роззброєння (UNODA), Міжнародної агенції з контролю автономних систем озброєння (МАКАСО, англ. – International Agency for the Control of Autonomous Weapons Systems, ICAWS), яка діяла б за моделлю МАГАТЕ. Структура агенції має включати Генеральну конференцію з участю всіх держав-членів агенції, що визначатиме загальну політику та пріоритети діяльності. Окрім того, пропонується створити виконавчий орган – Раду керівників у складі 35 членів, обраних на ротаційній основі від держав-членів. Такий склад обґрунтовується необхідністю забезпечити баланс між ефективністю прийняття рішень та представництвом різних регіонів і рівнів технологічного розвитку. Зокрема, модель із 35 членів дозволяє включити всі держави з розвиненими програмами автономних систем (згідно з даними міжнародної неурядової організації Human Rights Watch, щонайменше 8 країн заявляли публічно про розробку таких систем [21]) та забезпечити представництво всіх географічних регіонів ООН (по 7 представників від кожного з п'яти регіонів) та водночас зберегти керованість органу для оперативного прийняття рішень, на відміну від більш громіздких структур. Секретаріат на чолі з Генеральним директором має забезпечувати поточну діяльність агенції. Серед робочих органів МАКАСО пропонується створити Технічний комітет, що об'єднуватиме провідних експертів у галузі штучного інтелекту, міжнародного гуманітарного права, воєнних технологій та етики для розробки стандартів та рекомендацій, а також Інспекційний комітет, завданням якого буде проведення перевірок та моніторингу дотримання встановлених норм.

Повноваження МАКАСО пропонується визначити наступним чином: технічна експертиза алгоритмів розпізнавання цілей з правом доступу до вихідного коду та навчальних даних систем; видача міжнародних сертифікатів відповідності автономним системам із ШІ після проходження комплексної перевірки; регулярні інспекції виробничих потужностей та дослідницьких центрів для забезпечення транспарентності розробок; моніторинг застосування систем у бойових умовах через систему обов'язкової звітності держав-членів та незалежних спостерігачів; координація міжнародної співпраці в розробці безпечних технологій.

Процедурний інструмент: система сертифікації. Ефективна діяльність міжнародної агенції неможлива без чітких процедурних механізмів оцінки відповідності автономних систем із ШІ принципу розрізнення. Саме тому другим компонентом моделі є трирівнева процедура сертифікації автономних систем із ШІ, покликана забезпечити перевірку на всіх етапах розробки та застосування таких систем. На першому рівні пропонується здійснювати попередню технічну оцінку на етапі проектування, що включатиме обов'язковий аудит алгоритмів ШІ незалежними експертами, перевірку наявності механізмів розпізнавання захищених об'єктів (лікарні, школи, культові споруди тощо), оцінку здатності системи ідентифікувати некомбатантів та осіб hors de combat, а також видачу попереднього протоколу з детальними рекомендаціями щодо вдосконалення системи. На другому рівні пропонується проводити польові випробування в контрольованих умовах з тестуванням у різних сценаріях, включаючи міську забудову, погані погодні умови, електро-

магнітні перешкоди, оцінкою точності розпізнавання комбатантів та цивільних осіб з вимогою досягнення 99% точності, перевіркою здатності системи припиняти атаку при появі цивільних та видачею сертифіката обмеженої готовності. Третій рівень передбачає постійний моніторинг через систему «чорних скриньок» із встановленням модулів фіксації всіх рішень системи, аналізом даних незалежними інспекторами, негайним розслідуванням будь-яких інцидентів та можливістю відкликання сертифіката при виявленні порушень.

Правова основа: проект Конвенції про тимчасову заборону повністю автономних систем із ШІ. Для легітимізації діяльності запропонованої міжнародної агенції та системи сертифікації, необхідна відповідна міжнародно-правова база. З цією метою пропонується розробити міжнародний договір – «Конвенцію про тимчасову заборону повністю автономних систем із ШІ», що має стати правовим фундаментом для регулювання автономних систем із ШІ.

Преамбула має відображати фундаментальні принципи та мотивацію прийняття Конвенції. У ній слід зазначити посилання на Женевські конвенції 1949 року та Додаткові протоколи до них, підкреслити непорушність принципів міжнародного гуманітарного права, особливо принципу розрізнення. Необхідно визнати, що сучасні технології ШІ не здатні гарантовано забезпечити розрізнення між комбатантами та цивільними особами, між військовими та цивільними об'єктами. Преамбула має містити визнання права держав на технологічний розвиток у сфері ШІ для мирних цілей та систем озброєння з контролем людиною, а також підтвердження необхідності збереження відповідальності людини за рішення про застосування сили.

Стаття I. Загальні зобов'язання повинна встановлювати категоричну заборону для кожної держави-учасниці: не розробляти, не виробляти, не купувати, не накопичувати, не розповсюджувати та не передавати повністю автономні системи з ШІ категорії Human-out-of-the-Loop, зокрема, але не виключно: «рої» дронів, автономні системи хімічної, біологічної, радіологічної та ядерної зброї з ШІ; не застосовувати такі системи в збройних конфліктах; не проводити польові випробування таких систем у бойових умовах; не допомагати іншим державам чи недержавним акторам у розробці таких систем. Також необхідно встановити норми, які призначатимуть відповідального наглядача за кожною автономною системою з ШІ для уникнення «прогалин у відповідальності». При цьому стаття має чітко дозволяти розробку та використання систем категорій Human-in-the-Loop та Human-on-the-Loop, які зберігають значущий контроль людиною.

Стаття II. Визначення та критерії має містити детальні дефініції ключових понять. «Повністю автономна система з ШІ» визначається як система, здатна без людського втручання після активації самостійно виявляти, ідентифікувати, відбирати та уражати цілі на основі власного аналізу даних. «Система Human-in-the-Loop» – система, що вимагає підтвердження людини-оператора перед кожним ураженням цілі. «Система Human-on-the-Loop» – система, яка діє автономно, але оператор зберігає можливість втрутитися та скасувати дії системи. «Значущий контроль людиною» означає збереження за людиною остаточного рішення щодо застосування летальної сили з можливістю оцінки контексту та наслідків.

Стаття III. Декларації. Має встановлювати обов'язок держав-учасниць протягом 90 днів після набуття чинності Конвенцією подати до Міжнародної агенції з контролю автономних систем озброєння: декларацію про всі існуючі програми розробки автономних систем із ШІ; перелік дослідницьких центрів та підприємств, залучених до таких розробок; інформацію про системи Human-in-the-Loop та Human-on-the-Loop, що перебувають на озброєнні або в розробці; плани щодо забезпечення відповідності існуючих систем вимогам Конвенції. Держави також мають щорічно оновлювати ці декларації.

Стаття IV. Процедури верифікації та сертифікації мають детально описати запропоновану трирівневу систему сертифікації автономних систем з ШІ та визначати процедури відкликання сертифіката при виявленні порушень.

Стаття V. Міжнародна агенція з контролю автономних систем озброєння (МАКАСО) повинна передбачити заснування агенції як спеціалізованої установи під егідою ООН. Стаття має визначити структуру та повноваження агенції.

Стаття VI. Діяльність, що не забороняється Конвенцією має гарантувати право держав на: розробку та використання ШІ для мирних цілей; створення автономних систем для невійськових застосувань; розробку систем озброєння зі збереженням контролю людиною; проведення досліджень у сфері ШІ для покращення дотримання МГП. Стаття також має передбачати заходи щодо запобігання перешкоджанню технологічному розвитку та міжнародному співробітництву в мирних цілях.

Стаття VII. Національні заходи щодо імплементації повинна зобов'язати держави-учасниці: прийняти національне законодавство, що, відповідно до Конвенції, забороняє розробку повністю автономних систем з ШІ; створити національні органи контролю за дотриманням Конвенції; встановити кримінальну відповідальність за порушення заборон; забезпечити контроль за експортом технологій подвійного призначення; співпрацювати з МАКАСО та надавати необхідну інформацію.

Стаття VIII. Консультації та врегулювання спорів. Має передбачати систему механізмів для вирішення питань, пов'язаних з дотриманням Конвенції: консультацій між державами, що стосуються можливих порушень Конвенції; звернення до МАКАСО; розгляд спірних питань тощо. Особливу увагу слід приділити процедурам реагування на інциденти з виготовлення, випробування, використання чи поширення автономних систем з ШІ, що підпадають під тимчасову заборону.

Стаття IX. Технічна допомога та співробітництво повинна створити механізми: обміну кращими практиками забезпечення контролю людиною автономних систем із ШІ; спільної розробки стандартів безпеки для автономних систем із ШІ; допомоги державам, що розвиваються, у створенні національних систем контролю; створення міжнародного фонду для підтримки переходу на системи з контролем людиною.

Стаття X. Конференції з перегляду має встановити, що через 5 років після набуття чинності скликається перша оглядова конференція для оцінки: технологічного прогресу у сфері ШІ; ефективності заходів контролю; необхідності коригування технічних стандартів; можливості продовження або скасування мораторію. Наступні конференції проводяться кожні 5 років або за рішенням 2/3 держав-учасниць.

Стаття XI. Тимчасовий характер заборони повинна чітко визначити: початковий термін дії мораторію; умови можливого дострокового скасування (досягнення технологією рівня 99% точності розпізнавання); процедури продовження мораторію за рішенням 2/3 держав-учасниць; перехідний період для адаптації після закінчення мораторію.

Стаття XII. Заходи реагування на порушення має передбачати градуйовану систему заходів: попередження та вимога усунути порушення; обмеження доступу до технічної співпраці; призупинення прав та привілеїв у рамках Конвенції; економічні санкції через Раду Безпеки ООН.

Стаття XIII. Поправки повинна встановлювати процедури внесення змін: пропозиції можуть вносити держави-учасниці або МАКАСО; розгляд поправок на конференції держав-учасниць; прийняття поправок 2/3 голосів; набуття чинності після ратифікації половиною держав-учасниць.

Стаття XIV. Підписання, ратифікація, набуття чинності має визначити: відкриття для підписання протягом 12 місяців; процедуру ратифікації; можливість приєднання після закриття для підписання; неприпустимість застережень до ключових статей.

Стаття XV. Вихід з Конвенції повинна передбачати: право виходу з повідомленням за 6 місяців; збереження зобов'язань щодо вже сертифікованих систем; процедури передачі документації до МАКАСО; неможливість виходу під час збройного конфлікту.

Запропонована структура Конвенції має забезпечити комплексне правове регулювання, чіткі механізми контролю та гнучкість для адаптації до технологічного прогресу, зберігаючи при цьому головну мету – запобігання застосуванню автономних систем з ШІ без належного контролю людиною до досягнення технологією необхідного рівня відповідності принципу розрізнення.

Альтернативні механізми. Запропонована трикомпонентна модель є оптимальним, але довгостроковим сценарієм, реалізація якого може зайняти роки. Тому критично важливо паралельно розвивати альтернативні механізми регулювання, які не замінюють, а доповнюють основну модель, створюючи проміжні етапи на шляху до прийняття обов'язкових норм. Ці механізми виконують три ключові функції: по-перше, формують регіональну та галузеву практику, яка згодом може бути масштабована на глобальний рівень; по-друге, створюють прецеденти технічних стандартів та процедур верифікації, які можуть бути інтегровані в майбутню універсальну систему; по-третє, демонструють практичну можливість регулювання автономних систем із ШІ, поступово долаючи опір скептично налаштованих держав.

На регіональному рівні пропонується розпочати розробку відповідних міжнародно-правових актів, зокрема, на рівні Європейського Союзу – відповідної Директиви, в якій має бути передбачена обов'язкова сертифікація для всіх автономних систем із ШІ, що використовуються державами-членами ЄС. Паралельно слід розробити міжнародний стандарт ISO для тестування автономних систем із ШІ на відповідність МГП, Керівництво МКЧХ з кращих практик для розробників

автономних систем з ШІ та Добровільний кодекс поведінки виробників автономних систем озброєння з етичними принципами розробки.

Важливим елементом здійснення перевірок та моніторингу дотримання встановлених запропонованою міжнародною агенцією норм може стати створення спеціального правового режиму для окремих категорій безпілотників, які виконують гуманітарні та миротворчі функції. У цьому контексті особливої уваги заслуговує пропозиція І.М. Проценко розробити «окреме міжнародно-правове регулювання правового режиму використання безпілотних літальних апаратів миротворчими силами» [14, с. 150] з системою спеціальних розпізнавальних знаків за аналогією з символікою Міжнародного Руху Червоного Хреста і Червоного Півмісяця. Ця ідея заслуговує на подальшу конкретизацію. Розвиваючи цю ідею, пропонуємо створити комплексну систему ідентифікації таких безпілотників, що включатиме кілька рівнів. По-перше, візуальні розпізнавальні знаки – спеціальне забарвлення корпусу (наприклад, яскраво-блакитний колір ООН з білими смугами) та нанесення символів миротворчих місій, видимих з усіх ракурсів. По-друге, електронні ідентифікатори – транспондери, що постійно передають зашифрований сигнал із кодом місії на всіх стандартних частотах військового зв'язку, систему IFF (Identification Friend or Foe) з унікальними кодами для миротворчих апаратів та GPS-маяки з відкритою трансляцією координат у режимі реального часу. По-третє, програмні протоколи розпізнавання – обов'язкове включення в алгоритми всіх автономних систем із ШІ модулів розпізнавання миротворчих ідентифікаторів, автоматична заборона на ураження об'єктів з такими ідентифікаторами та негайне припинення атаки при виявленні миротворчих сигналів. По-четверте, правовий захист – криміналізація навмисних атак на миротворчі безпілотники, створення спеціального реєстру миротворчих апаратів під контролем Департаменту миротворчих операцій ООН/МАКАСО та обов'язкове розслідування всіх інцидентів з миротворчими дронами незалежною комісією. Така багаторівнева система має забезпечити захист миротворчих безпілотників та сприятиме ефективному виконанню ними моніторингових і верифікаційних функцій у зонах конфліктів.

Висновки. Дослідження виявило фундаментальне протиріччя між вимогами принципу розрізнення МГП та технічними можливостями сучасних автономних систем з ШІ. Встановлено три ключові проблеми: (1) технічна неспроможність на поточному етапі розвитку автономних систем з ШІ надійно розрізнити цілі через «семантичний розрив» між алгоритмічним розпізнаванням об'єктів та розумінням їх значення в контексті збройного конфлікту; (2) «прогалина у відповідальності» через відсутність правових механізмів встановлення відповідальності за порушення МГП автономними системами з ШІ; (3) відсутність політичної волі провідних держав-розробників до прийняття обов'язкових обмежень.

Запропонована трикомпонентна модель міжнародно-правового регулювання автономних систем з ШІ передбачає: створення Міжнародної агенції під егідою ООН/UNODA зі структурою, аналогічною МАГАТЕ (Генеральна конференція, Рада керівників з 35 членів, Секретаріат, Технічний та Інспекційний комітети); впровадження трирівневої системи сертифікації (попередня оцінка на етапі проектування, польові випробування з вимогою 99% точності розпізнавання, постійний моніторинг через «чорні скриньки»); розробку міжнародного договору щодо тимчасового мораторію на повністю автономні системи з ШІ категорії Human-out-of-the-Loop зі збереженням можливості розробки систем з контролем людиною.

Через відсутність політичної волі провідних держав-розробників автономних систем із ШІ до прийняття обов'язкових обмежень у розробці та використанні таких систем, найбільш реалістичним є паралельний поетапний підхід до запровадження таких обмежень через альтернативні механізми: регіональне регулювання, міжнародний стандарт ISO, керівництво МКЧХ та добровільний кодекс поведінки виробників таких систем. Розвинуто концепцію системи захисту миротворчих безпілотників через багаторівневу ідентифікацію, що може стати пілотним проектом для відпрацювання механізмів інтеграції правових вимог у технічну архітектуру автономних систем із ШІ.

Подальші дослідження у даній сфері мають зосередитися на розробці методології оцінки алгоритмів ШІ на відповідність МГП та механізмів забезпечення «пояснюваності» рішень автономних систем з ШІ для збереження відповідальності за дотримання норм міжнародного гуманітарного права.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Rickli, J.-M., & Mantellassi, F. (2024). The War in Ukraine: Reality Check for Emerging Technologies and the Future of Warfare. Geneva Paper 34/24. С. 4. URL: <https://www.gcsp.ch/publications/war-ukraine-reality-check-emerging-technologies-and-future-warfare> (дата звернення 15.07.2025).
2. Росіяни атакували рейсовий автобус на Сумщині: 9 загинули. KyivPost. URL: <https://www.kyivpost.com/uk/post/52826> (дата звернення 17.07.2025).
3. РФ збільшила дальність повністю автономного дрона-камікадзе до 100 км. Defense Express. URL: https://defence-ua.com/weapon_and_tech/rf_zbilshila_dalnist_povnistju_avtonomnogo_drona_kamikadze_do_100_km (дата звернення 17.07.2025).
4. Стокгольмський міжнародний інститут досліджень проблем миру. Щорічний звіт про розвиток військових технологій. 2024. С. 89–112. URL: <https://www.sipri.org/yearbook/2024> (дата звернення 18.07.2025).
5. Kwik, J., & Van Engers, T. (2021). Algorithmic fog of war: When lack of transparency violates the law of armed conflict. *Journal of Future Robot Life*, 2. С. 43–66. URL: <https://doi.org/10.3233/FRL-200019> (дата звернення 19.07.2025).
6. Додатковий протокол до Женевських конвенцій від 12 серпня 1949 року, що стосується захисту жертв міжнародних збройних конфліктів (Протокол I). (1977). URL: <https://ips.ligazakon.net/document/MU77003?ap=454582> (дата звернення 19.07.2025).
7. Міжнародний суд ООН. (1996). Консультативний висновок щодо правомірності загрози або застосування ядерної зброї. URL: <http://www.worldlii.org/int/cases/ICJ/1996/3.html> (дата звернення 19.07.2025).
8. Міжнародний Комітет Червоного Хреста. (2014). Автономні системи озброєння: технічні, військові, правові та гуманітарні аспекти. Звіт експертної групи МКЧХ. С. 15–20. URL: https://www.icrc.org/sites/default/files/document/file_list/4221-002-autonomous-weapons-systems-full-report.pdf (дата звернення 20.07.2025).
9. ICRC. (2019). Artificial intelligence and machine learning in armed conflict: A human-centred approach. С. 12. URL: https://www.icrc.org/sites/default/files/document_new/file_list/ai_and_machine_learning_in_armed_conflict-icrc.pdf (дата звернення 20.07.2025).
10. Звіт Групи урядових експертів з летальних автономних систем зброї (GGE LAWS). (2019). CCW/GGE.1/2019/3. URL: https://documents.unoda.org/wp-content/uploads/2020/09/CCW_GGE.1_2019_3_E.pdf (дата звернення 20.07.2025).
11. Звіт Групи урядових експертів з летальних автономних систем зброї (GGE LAWS). (2021). CCW/GGE.1/2021/CRP.1. URL: <https://documents.un.org/doc/undoc/gen/g22/264/50/pdf/g2226450.pdf> (дата звернення 20.07.2025).
12. Prezelj, I. (2024) «The Use of Artificial Intelligence-Enabled Systems by Modern Armed Forces and Some Related Concerns», in Zombory, K. and Szilágyi, J.E. (eds.) *Shielding Europe with the Common Security and Defence Policy: The EU Legal Framework for the Development of an Innovative European Defence Industry in Times of a Changing Global Security Environment*. Miskolc--Budapest: Studies of the Central European Professors' Network, pp. 357–394. URL: https://doi.org/10.54237/profnet.2024.zkjeszcodef_8 (дата звернення 22.07.2025).
13. Розумні дрони України: роль штучного інтелекту та машинного зору на фронті (2025). VGI-9. URL: <https://vgi.com.ua/rozumni-drony-ukrayiny-rol-mashynno-go-zoru-i-shi-na-fronti/> (дата звернення 22.07.2025).
14. Проценко І.М. (2022). Новітні проблеми міжнародного гуманітарного права та можливі шляхи їх вирішення. Сучасні аспекти модернізації науки: стан, проблеми, тенденції розвитку. Матеріали XVII-ої Міжнародної науково-практичної конференції (07 січня 2022 р., м. Ларнака, Кіпр). С. 147–152. URL: <http://perspectives.pp.ua/public/site/conferency/conf-17.pdf> (дата звернення 22.07.2025).
15. Lewis, D.A. (2022). On 'Responsible AI' in War: Exploring Preconditions for Respecting International Law in Armed Conflict. *The Cambridge Handbook of Responsible Artificial Intelligence*. Cambridge University Press. С. 488–506. URL: <https://doi.org/10.1017/9781009207898.037> (дата звернення 23.07.2025).
16. Sati, M.C. (2024). The Attributability of Combatant Status to Military AI Technologies under International Humanitarian Law. *Global Society*, 38(1). С. 130. URL: <http://dx.doi.org/10.1080/13600826.2023.2251509> (дата звернення 23.07.2025).

17. Siregar, N., Aini, D.C., Rehulina, Subandi, A.Y., & Mirza, I.M. M. (2024). The Use of Artificial Intelligence in Armed Conflict under International Law. *Hasanuddin Law Review*, 10(2). С. 189–205. URL: <http://dx.doi.org/10.20956/halrev.v10i2.5267> (дата звернення 25.07.2025).
18. Zachary Kallenborn. A Partial Ban on Autonomous Weapons Would Make Everyone Safer. *Foreign Policy*, 2020. URL: <https://foreignpolicy.com/2020/10/14/ai-drones-swarms-killer-robots-partial-ban-on-autonomous-weapons-would-make-everyone-safer> (дата звернення 25.07.2025).
19. Dresch-Langley, B. (2023). The weaponization of artificial intelligence: What the public needs to be aware of. *Frontiers in artificial intelligence*. URL: <https://doi.org/10.3389/frai.2023.1154184> (дата звернення 29.07.2025).
20. David Adam. Lethal AI weapons are here: how can we control them? *Nature*, 2024. URL: <https://doi.org/10.1038/d41586-024-01029-0> (дата звернення 29.07.2025).
21. Stopping Killer Robots. Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control, Human Rights Watch, 2020, URL: <https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and> (дата звернення 14.08.2025).