ПРИЛОЖЕНИЕ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ В ОПЕРАТИВНО-ИЗДИРВАТЕЛНАТА ДЕЙНОСТ ЗА ПРОТИВОДЕЙСТВИЕ НА КИБЕРПРЕСТЪПНОСТТА

доц. д-р Андрей Джунин, Академия на МВР

Резюме:

Настоящата статия изследва приложението на изкуствения интелект (ИИ) в оперативно-издирвателната дейност за противодействие на киберпрестъпността в България. Изследването защитава тезата, че ИИ ще се утвърди като задължителен оперативен инструмент на системата за киберсигурност, а не като опционална технологична добавка. Анализът структурира аргументацията на три нива: количествено, качествено и времево. Разглеждат се теоретичните основи на ИИ, нормативната рамка и конкретните оперативно-технологични приложения, включително автоматизиран цифров анализ, предвиждащ анализ, детекция на аномалии и разузнаване от открити източници. В заключението се формулират условията, при които България може проактивно да реализира потенциала на тези технологии, и се набелязват ключовите институционални предпоставки.

Ключови думи: *изкуствен интелект, оперативно-издирвателна дейност, киберпрестъпност, киберсигурност.*

APPLICATION OF ARTIFICIAL INTELLIGENCE IN OPERATIONAL INVESTIGATIVE ACTIVITY FOR COUNTERING CYBERCRIME

Assoc. Prof. Andrey Dzhunin, PhD, Academy of the Ministry of Interior

Abstract:

This article examines the application of artificial intelligence (AI) in operational-investigative activities for countering cybercrime in Bulgaria. The study advances the thesis that AI will establish itself as a mandatory operational tool within the cybersecurity system, rather than as an optional technological addition. The analysis structures its argumentation across three dimensions: quantitative, qualitative, and temporal. It explores the theoretical foundations of AI, the regulatory framework, and specific operational-technological applications, including automated digital analysis, predictive analytics, anomaly detection, and open-source intelligence (OSINT). The conclusion outlines the conditions under which Bulgaria can proactively realize the potential of these technologies and identifies key institutional prerequisites.

Keywords: *artificial intelligence, operational-investigative activity, cybercrime, cybersecurity.*

Дигитализацията на обществения живот протича с ускорение, което не познава исторически аналог. В рамките на едно-две десетилетия глобалното киберпространство се превърна от академична концепция в критична инфраструктура, от която зависят финансовата система, здравеопазването, публичната администрация и националната сигурност. Тази трансформация не е само технологична. Тя е дълбоко социална и правна, тъй като преначертава границите между публичното и частното, между суверенното и извънтериториалното, между познатото и непознатото в престъпност-

та. Закономерно тази трансформация доведе до качествена промяна в характера на престъпността. Ако в края на XX век киберпрестъпленията са предимно индивидуални прояви с ограничен обхват, хакерство с демонстративна цел, примитивни вируси и дребни финансови измами, днес картината е принципно различна. Организирани престъпни структури и държавно спонсорирани субекти разполагат с ресурси и технологичен потенциал, съизмерими с тези на правоохранителните органи, а в редица случаи ги превъзхождат.

Атаките срещу критична инфраструктура – болници, електрически мрежи, водоснабдяване, финансови системи, вече не са хипотетични сценарии, а документирана реалност.

Именно тази асиметрия поставя ключовия въпрос на настоящото изследване – способни ли са традиционните методи на оперативно-издирвателната дейност да противостоят ефективно на противник, чиято технологична мощ нараства побързо, отколкото институционалният капацитет за адаптация? Анализът на тенденциите дава недвусмислен отговор.

Проучванията установяват, че електронно базираните киберпрестъпления съставляват преобладаващата част от всички регистрирани правонарушения, а за 2021 г. е отчетен ръст на кибератаките с 68% спрямо предходната година (Dunsin et al., 2024: 2). Тези данни не са изолирано явление. Те отразяват структурния проблем, че нападението в киберпространството е значително по-лесно от защитата. Този, който атакува, трябва да открие само една незащитена уязвимост, а този, който защитава, трябва да запуши всички. Нападателят определя времето и вектора на атаката, а защитникът трябва да реагира по всяко време и по всякакъв начин. Тази диспропорция се задълбочава с всяка нова технологична вълна и не може да бъде преодоляна с численост на персонала или административни мерки.

Тезата на настоящото изследване е, че изкуственият интелект не представлява опционален технологичен инструмент в арсенала на правоохранителните органи, а ще се утвърди като задължителен оперативен компонент на системата за киберсигурност. Тази необходимост произтича от три взаимосвързани фактора: безпрецедентния обем данни, подлежащи на анализ; скоростта на еволюция на заплахите; и обстоятелството, че самите киберпрестъпници вече разполагат и активно ползват ИИ инструменти. Всеки от тези фактори поотделно прави традиционните подходи недостатъчни, а в комбинация ги правят неприложими.

Изследването е структурирано в три основни раздела. Първият разглежда теоретичните основи на изкуствения интелект и неговата класификация, релевантна за приложенията в областта на сигурността. Вторият анализира киберпрестъпността от двоен ъгъл – като заплаха, която сама използва ИИ, и като правен феномен, регулиран от националната и международната нормативна уредба. Третият систематизира конкретните приложения на ИИ в оперативно-издирвателната дейност, съпроводени с анализ на правните и етичните предизвикателства при внедряването.

1. Теоретични основи на изкуствения интелект.

1.1. Концепция, история и класификация.

Изкуственият интелект като научна дисциплина води началото си от средата на XX век, но теоретичните предпоставки за неговото зараждане са значително по-стари. Философските въпроси за природата на мисленето и възможността за неговото изкуствено възпроизвеждане са разисквани от Декарт и Лайбниц до Бул и Тюринг. Именно Алан Тюринг формулира през 1950 г. оперативния критерий за „машинна интелигентност“ в своята епохална статия „Компютърните машини и интелигентността“, в която предлага т.нар. тест на Тюринг – машина може да се смята за интелигентна, ако нейните отговори в текстова комуникация не могат да бъдат разграничени от тези на човек. Тюринг полага не само концептуалната, но и методоло-

гичната основа на дисциплината, задавайки въпроса по начин, допускащ емпирична проверка.

Официалното рождение на изкуствения интелект като самостоятелна изследователска програма обаче се свързва с конференцията в Дартмутския колеж от 1956 г., организирана от Джон Маккарти, Марвин Мински, Натаниел Рочестър и Клод Шанън. Джон Маккарти, приет за основоположник на термина, го определя като „науката и инженерството за създаване на интелигентни машини“ (Rigano, 2019: 1). Оттогава дисциплината преживява характерните за всяка революционна технология цикли на интензивно развитие и „зими на ИИ“ – периоди на застой, когато резултатите изостават от очакванията и финансирането намалява. Съвременният период е белязан от качествен пробив, дължащ се основно на три фактора: нарастващата изчислителна мощ; наличието на огромни масиви от данни за обучение; и усъвършенстването на алгоритмите особено в областта на дълбокото обучение. Тяхната комбинация обяснява защо след десетилетия на относителна стагнация ИИ постига впечатляващи резултати именно в периода след 2010 г.

От методологична гледна точка е полезно да се разграничат два основни типа ИИ системи – конвенционални и изчислителни. Конвенционалният тип работи на базата на явни знания: експертни системи[1], байесови мрежи[2] и дървета на решенията[3].

„Основните изчислителни парадигми на съвременния изкуствен интелект са: размити системи, включващи размита логика[4]; еволюционни пресмятания, включващи генетични алгоритми[5]; изкуствени невронни мрежи[6]; машинно обучение; вероятностни изводи[7]“ (Ivanov & Georgieva, 2018: 46–47).

За приложенията в правоохранителната дейност разграничението е съществено, тъй като конвенционалните системи генерират решения, обясними пред съд, докато изчислителните са по-мощни и адаптивни, но непрозрачни, обстоятелство, пораждащо конкретни процесуални предизвикателства при доказването.

1.2. Машинното обучение и дълбокото обучение като оперативен потенциал.

Машинното обучение представлява приложна подсистема на ИИ, при която алгоритъмът усвоява закономерности от данни, без тези закономерности да бъдат предварително програмирани. То може да работи в три основни режима. При обучение с „учител“ алгоритъмът се обучава върху маркирани примери. Например двоична класификация „злонамерен/незлонамерен“ трафик е типичен случай. При обучение „без учител“ системата сама открива структури в немаркирани данни, което е приложимо при разпознаване на нови видове атаки, за които липсват маркирани примери. При обучение с „подкрепление“ системата се обучава чрез взаимодействие със средата, оптимизирайки поведението си спрямо дефинирана функция на успех. Всеки от тези режими има специфично приложение в киберсигурността и изборът между тях е ключово методологично решение при проектирането на система за киберсигурност.

Дълбокото обучение разширява качествено възможностите на машинното обучение. То е негово подмножество и е базирано на многослойни невронни мрежи. Основното му предимство е способността да обработва сурови данни – мрежов трафик, логови файлове, изображения, без необходимост от предварителна ръчна инженерна обработка на характеристиките. Именно оттук произтича неговата особена приложимост при задачи, при които множеството релевантни характеристики е неизвестно предварително или е прекалено голямо за ръчно дефиниране.

Автоматизацията на ключови фази от прегледа на цифрови данни е една от областите, в които интеграцията на ИИ и машинното обучение демонстрира най-осезаем оперативен ефект. Анализът на данни, обработката на изображения и разпознаването на поведенчески образци, дейности, които традиционно изискват значителен човешки ресурс и са изложени

на риск от субективна грешка, могат да бъдат извършвани от автоматизирани системи с многократно по-висока скорост и последователност. При това системата не само ускорява вече познати аналитични процеси, но открива и корелации между разнородни масиви от данни, които при конвенционален ръчен преглед биха останали незабелязани поради обема и сложността на информацията. Това качествено разширява възможностите на оперативно-издирвателния орган да формира обоснована представа за хронологията и механизма на извършеното престъпление.

Практическото значение на тази констатация в оперативно-издирвателен контекст е голямо. ИИ системата може да идентифицира модели на поведение, характерни за киберпрестъпна дейност, дори когато конкретният способ е нов и непознат. Именно тази способност за разпознаване на новото по аналогия с поведенческите характеристики на познатото е ключовото предимство пред традиционния сигнатурен подход. Сигнатурата предполага предварително познание за заплахата, а невронната мрежа разпознава аномалията, независимо дали е виждала идентичен случай.

Важен аспект е и въпросът за необходимите данни. Качеството на машинно обучените модели е функция от качеството и обема на обучителните данни. В контекста на киберсигурността системите, обучени на реални данни за кибератаки, са значително по-ефективни от обучените на синтетични набори. Това поставя въпроса за споделянето на данни между правоохранителните органи на национално и международно ниво, въпрос с конкретни правни и оперативно-издирвателни измерения, заслужаващ специално внимание в самостоятелно изследване.

2. Характеристики и правна рамка на противодействието на киберпрестъпността.

2.1. Изкуственият интелект в арсенала на киберпрестъпниците.

Изкуственият интелект е двупосочна технология и ефективното реагиране на оперативните предизвикателства изисква да се осмисли задълбочено този фундаментален факт. Той не е инструмент единствено на защита, а активно се използва и от престъпниците. Силата на подобна технология е именно в нейната универсалност. Алгоритъм, обучен да разпознава модели, може да бъде насочен срещу обекта или в негова защита в зависимост от намерението на субекта. Това превръща въпроса за ИИ в правоохранителната дейност от технологичен в стратегически.

Конкретните приложения на ИИ в престъпната дейност са разнообразни и вече са добре документирани. Например генеративните езикови модели се използват за масово производство на персонализирани фишинг съобщения, чиято убедителност е непостижима при ръчно изготвяне. Изпращаният текст е съобразен с конкретния получател, стила на неговата комуникация и актуалния контекст на обичайните му разговори.

Прогностичните модели пък се прилагат при рансъмуер атаки за определяне на оптималния размер на откупа, който ще се иска впоследствие. Алгоритъмът анализира публично достъпните данни за финансовото положение на пострадалия и предлага сума, която винаги е значителна, но в същото време е реалистична за плащане. Всичко това се подпомага от ботнет мрежи, управлявани с ИИ, които автоматично адаптират поведението си, за да се избегне засичане от системите за сигурност.

Съвременните тенденции включват използването на ИИ системи за извършване на вредоносни и незаконни действия, включително дълбоки фалшификати, което поражда сериозни предизвикателства пред правото и международните инструменти за противодействие на киберпрестъпността (Velasco, 2022: 109).

Дълбоките фалшификати (deepfakes) представляват особено тревожна категория заплаха. Технологията позволява генерирането на убедителни видео- и аудиозаписи, приписващи на реални лица думи и действия, които никога не са произнасяли или извършвали.

В контекста на финансовата престъпност тази технология вече е документирана при реални случаи на измамно нареждане на банкови преводи. Например изпълнителен директор на финансова институция „лично нарежда“ по видеовръзка превод на значителна сума, а служителят изпълнява нареждането, защото е убеден в неговата автентичност.

Категоричният изводът е, че ако правоохранителният орган не разполага с ИИ инструменти, се изправя в неравностойна битка срещу престъпниците. В случая едностранното технологично самоограничение се явява един вид капитулация, а не етична позиция. Поради тези обективни причини в развитите страни ИИ технологиите масово навлизат в дейността на органите, противодействащи на киберпрестъпността.

Р. Факир констатира широкото навлизане на технологиите на изкуствения интелект в правоохранителната дейност за противодействие на киберпрестъпността, особено в областта на лицевия и гласовия анализ за разпознаване на престъпни модели, като ролята на ИИ в наказателното правосъдие се разширява и в сферата на оценката на риска, решенията за задържане и съдебните определения (Faqir, 2023: 88). И тук въпросът е не дали ИИ да се прилага в правоохранителната дейност, а как да се прилага законосъобразно, ефективно и при гарантиран контрол.

2.2. Структурна асиметрия в киберпространството.

В киберпространството между атака и защита съществува голямата асиметрия, която има обективна технологична обусловеност. Тя не може да се преодолее с административни мерки или кадрово усилване, тъй като е вградена в самата архитектура на глобалната мрежа и в икономическата логика на разработването на софтуер. Сигурността традиционно се разглежда като разход, а не като инвестиция. Следствие от това е, че уязвимостите в кода се натрупват с темпа на разработката, докато тяхното отстраняване закономерно изостава.

Количествените параметри на този проблем илюстрират достатъчно отчетливо неговия мащаб. Когато ежедневно се създават стотици хиляди нови варианти на зловреден код, актуализацията на сигнатурни бази данни в системите за сигурност чрез традиционни методи е невъзможна. Времето между появата на нова уязвимост (например т.нар. zero-day vulnerability[8]) и нейната масирана експлоатация се е съкратило от месеци до часове.

Проектирането, тестването и разпространението на корекция с класически методи отнема дни до седмици. Този времеви интервал, т.нар. window of exposure[9], се явява принципното ограничение на конвенционалния подход към киберсигурността.

Качественото измерение на проблема също е значимо. Съвременните атаки от типа APT[10] са многоетапни, многовекторни и предназначени да останат незабелязани за продължителен период от време, понякога месеци или години. За да избегне засичане, атакуващият методично изгражда „плацдарм“ в мрежата, постепенно повишава привилегиите си и действа в рамките на нормалния трафик. Обикновено такива атаки не се разкриват от автоматизираните системи с традиционните методи, а от случайно наблюдение или след инцидент, който е причинил видима вреда. Обикновено до момента на разкриването атакуващият вече е реализирал целите си.

На институционално ниво проблемът се изостря от значителния дефицит на квалифицирани специалисти в областта на киберсигурността. Глобалното търсене на кадри многократно надвишава предлагането. Прогнозите за незаетите позиции в сектора се измерват в милиони. При тези условия автоматизацията на рутинните аналитични задачи чрез ИИ не е само въпрос на ефективност, а на функционална устойчивост на системата за киберсигурност. Например оперативният работник, който е освободен от механичния преглед на логове, може да се фокусира върху аналитичното интерпретиране на алгоритмично маркираните аномалии.

Това е задача, изискваща именно онова, което ИИ не притежава – оперативен опит, контекстуална преценка и правна компетентност.

2.3. Нормативна рамка на оперативно-издирвателната дейност и мястото на изкуствения интелект в нея.

Правната рамка, в която се осъществява оперативно-издирвателната дейност в България, е изградена на няколко нормативни нива – национално, европейско и международно. Тяхното взаимодействие определя пространството на законосъобразно внедряване на ИИ технологии в правоохранителната практика.

Съгласно чл. 8, ал. 1 от Закона за Министерството на вътрешните работи оперативно-издирвателната дейност е съвкупност от явни и тайни действия на оперативно-издирвателните органи на МВР за противодействие на престъпността и заплахите за националната сигурност и за опазване на обществения ред. Тя се осъществява чрез специфични сили, средства и способи. Й. Пенев систематизира деветте оперативно-издирвателни способа по ЗМВР в три групи: подготвителни способи, способи за събиране на данни и комбинирани способи за събиране на данни и въздействие в обществено полезен интерес (Penev, 2020: 13).

Приложението на ИИ в оперативно-издирвателната дейност потенциално засяга почти всеки от тези способи – от автоматизирания анализ на данни и разпознаването на образи до подпомагането на разузнавателната беседа чрез анализ на реч и поведенчески модели.

На европейско равнище Директива (ЕС) 2013/40 хармонизира националните законодателства относно атаките срещу информационните системи. В националното право Глава девета „а“ от Наказателния кодекс (чл. 319а–319е) урежда наказателната отговорност за извършени компютърни престъпления, а Законът за киберсигурност (2018) създава специализиран координационен механизъм за организацията, управлението и контрола на киберсигурността, включително противодействието на киберпрестъпността, и определя правомощията и функциите на компетентните органи в тази област.

Основополагащ акт на международно ниво е Будапещенската конвенция за престъпления в кибернетичното пространство, ратифицирана от България, в сила от 2005 г. Конвенцията задължава страните да криминализират определен набор от деяния в киберпространството и да осигурят процесуален инструментариум за разследването им, включително разпоредби за запазване на данни, претърсване и изземване на компютърни системи и прихващане на данни.

В контекста на очертаната правна рамка В. Шестак и А. Волеводз обосновават, че новите технологии не изискват задължително създаването на специални или нови правни норми, тъй като „съществуващите правни понятия са достатъчно гъвкави и абстрактни, за да се адаптират към новите сценарии на технологично развитие“ (Shestak & Volevodz, 2019: 203). От тази гледна точка може да се приеме, че технологично неутралните разпоредби на правоохранителното законодателство, доколкото не изключват изрично използването на алгоритмични инструменти, могат да служат като правна основа за внедряване на ИИ системи в оперативно-издирвателната дейност.

Анализът на националното законодателство потвърждава тази констатация. ЗМВР е технологично неутрален и не ограничава изрично използването на автоматизирани системи за анализ.

Наказателно-процесуалният кодекс урежда изискванията за годност и допустимост на доказателствата, без да изключва такива, добити с помощта на технологични средства. Това, че законодателните промени не са задължителна предпоставка, е предимство, но едновременно и недостатък, тъй като липсват специфични гаранции, стандарти и процедури за ИИ базирани методи в оперативно-издирвателната и процесуалната дейност.

Все пак този нормативен вакуум е нужно да бъде запълнен.

3. Приложения на изкуствения интелект в оперативно-издирвателната дейност.

3.1. Автоматизиран цифров анализ на данни.

Цифровият анализ е областта, в която потенциалът на изкуствения интелект е най-непосредствено приложим и измерим. Всяко киберпрестъпление оставя цифрови следи под формата на логови файлове, мрежов трафик, метаданни, фрагменти от паметта, регистрационни записи на файловата система. Проблемът при разкриване на киберпрестъпление, за разлика например от престъпление от състава на конвенционалната престъпност, не е в липсата на следи, а в тяхното количество. То многократно надхвърля капацитета на ръчния анализ, а всяко забавяне в обработката им носи риск от тяхното унищожаване, презаписване или изтриване от обекта на наблюдение.

Средно голяма организация генерира ежедневно стотици гигабайти мрежов трафик. Пълният логов архив на корпоративна информационна система за един месец може да достигне петабайтов обем. При традиционния подход оперативният работник претърсва тези масиви ръчно или с помощта на прости скриптове. Резултатът е неизбежна непълнота на анализа и значително времезакъснение, в период, когато доказателствените следи могат да изчезнат необратимо.

Редица ИИ алгоритми и такива за машинно обучение могат да открият модели и аномалии, недостъпни непосредствено за човешкото възприятие, което е особено предимство при идентифицирането на укрито или прикрито доказателство, като по този начин се постигат по-точни и надеждни резултати (Dunsin et al., 2024: 3). Практическото значение на тази способност се изразява в конкретни оперативни ползи. Системите за автоматизиран цифров анализ могат да реконструират хронологията на събитията в компрометирана система, да идентифицират точката на проникване, да проследят движението на нападателя в мрежата и да установят данните, до които е осъществен неоторизиран достъп, и то в реално или почти реално време. При ръчен анализ тази задача би изисквала седмици специализирана работа.

Особено ценна е способността на ИИ системите за корелация на данни от разнородни източници. Следите от киберпрестъплението обикновено са разпръснати в множество системи. Такива могат да бъдат крайни устройства, мрежови устройства, облачни услуги, системи за управление на идентичностите. Ръчното съпоставяне на тези данни е практически невъзможно при значителен обем. Автоматизираните SIEM[11] платформи с ИИ компонент изпълняват тази задача в реално време, генерирайки приоритизиран списък от инциденти. В този случай оперативно-издирвателният орган получава не суров масив от данни, а аналитично резюме с наредени по ранг, приоритет и степен на значимост индикатори. Това е трансформация, която качествено преструктурира самата оперативна работа.

3.2. Предвиждащ анализ и детекция на аномалии.

Освен ретроспективния цифров анализ ИИ предоставя и проспективни възможности за прогнозиране на заплахи, преди те да са се материализирали в конкретно престъпление. Тук машинното обучение работи не само като детектор на известни модели на атаки, но и като идентификатор на поведенчески отклонения, предшестващи атаката. Тази смяна на парадигмата от реактивна към превантивна е от принципно значение за оперативно-издирвателната дейност, чийто фундаментален принцип е именно настъпателността.

Обучен на историческа база данни за нормалното мрежово поведение на дадена среда, алгоритъмът за машинно обучение изгражда статистически профил на „нормата“ и сигнализира при отклонения от нея. Тази парадигма, детекция на аномалии, принципно преодолява ограничението на сигнатурния подход. Системата засича нови заплахи не защото ги познава, а защото те се отклоняват от установения модел. Атаките през уязвимости от нулев

ден, при които не съществува предварително известна сигнатура, стават принципно откриваеми.

Тук практическото предизвикателство е контролът на т.нар. „фалшиви позитивни“ резултати. Прекалено чувствителна система може да генерира огромен шум, който се състои от стотици сигнали дневно, поголямата част от които са естествени аномалии, като например нов служител с нетипично поведение, мигриране на данни, пикова натовареност и др.

От друга страна пък, прекалено консервативна система може да пропусне реални заплахи.

Оптималното калибриране на прага на системата е задача, изискваща комбинация от техническа компетентност и оперативен опит, и поради тази причина не може да бъде напълно автоматизирана.

Р. Факир констатира, че интеграцията на изкуствения интелект, машинното обучение и анализ на масиви от данни в правоохранителната дейност е осъществила революционна промяна в извличането на доказателства, идентифицирането на заподозрени и разкриването на цифрови престъпления (Faqir, 2023: 81).

За оперативно-издирвателната дейност превантивното измерение на тази технология има особена стойност. Системата може да идентифицира мрежова активност, характерна за разузнавателната фаза преди атаката, като сканиране на портове, опити за изброяване на ресурси, необичайни заявки за автентификация, преди атаката изобщо да е предприета. Времевото предимство, което това осигурява, може да бъде решаващо за нейното предотвратяване.

3.3. Разузнаване от открити източници (OSINT).

Информационното разузнаване от открити източници е традиционен оперативен метод с дълга история в практиката на правоохранителните органи. Неговите възможности се разширяват качествено с развитието на ИИ, като трансформацията не е просто количествена, а принципна. Автоматизацията променя самата логика на събиране на разузнавателна информация. Глобалната мрежа съдържа безпрецедентен обем публично достъпна информация – социални мрежи, форуми, тъмния уеб, новинарски сайтове, корпоративни регистри, съдебни бази данни. Тяхното систематично ръчно „претърсване“ е практически неосъществимо дори от голям специализиран екип. Новите информационни технологии качествено трансформират тази ситуация. Информация, изискваща преди дни или седмици на наблюдение, днес може да бъде събрана и обработена в рамките на минути чрез автоматизиран преглед на същите тези източници. Автоматизацията не само ускорява процеса, но и позволява едновременното обхващане на източници в мащаб, принципно недостъпен за анализатора от правоохранителните органи.

Е. Маданski установява, че информацията от явни източници може да бъде от решаващо значение за пълнотата и обективността на оперативното проучване, като данните от уеб базирани източници нерядко потвърждават или допълват съществуващи хипотези относно значими факти и личностни характеристики на проучвания (Madanski, 2021: 57).

ИИ системите за обработка на естествен език (NLP[12]) автоматизират OSINT процесите в качествено различен мащаб. Те могат да обхванат милиони публикации за минути, да извличат семантично свързани сигнали, да идентифицират псевдоними на едно лице в различни платформи и да проследяват промени в онлайн поведението на набелязан обект. Системите за разпознаване на изображения например добавят визуалното измерение, представляващо автоматична идентификация на лица, обекти и геолокация по фонова информация в снимки.

В контекста на противодействието на киберпрестъпления тези инструменти имат специфична приложимост.

Киберпрестъпниците, макар и да действат предимно анонимно, нерядко оставят следи в публичното пространство. Те могат да бъдат най-различни, като

например хвалба в специализирани форуми, демонстрация на технически умения, обсъждане на използвани инструменти и др.

Комуникацията в по-слабо защитените платформи може да предостави оперативно значими данни, насочващи оперативните работници. Автоматизираното наблюдение на тези платформи от ИИ системи е значително по-ефективно от периодичното ръчно претърсване.

Използването на съвременни инструменти за придобиване на информация в киберпространството все повече навлиза в практиката на оперативно-издирвателните органи, а познаването на методите за работа с уеб базирани явни източници постепенно се превръща в задължителна квалификация на всеки оперативен работник (Madanski, 2024: 95).

3.4. Правни и етични предизвикателства при внедряването на изкуствения интелект.

Внедряването на ИИ в оперативно-издирвателната дейност поражда предизвикателства, които не бива нито да се подценяват, нито да служат като аргумент за бездействие. Признаването на предизвикателствата е условие за тяхното преодоляване. Три от тях са с приоритетно значение.

Първото е проблемът за прозрачността на алгоритмичното решение. Дълбоките невронни мрежи функционират като „черни кутии“, т.е. генерират резултат, без да предоставят разбираемо обяснение за логиката му. За оперативни цели понякога това не е проблем, но в контекста на наказателното производство, където доказателствата трябва да бъдат представени и обосновани пред съд, тази непрозрачност е съществен правен проблем.

В отговор на нея научният клон на „обяснимия ИИ“ разработва методи за генериране на разбираеми обяснения на алгоритмичните решения. Паралелно с технологичното решение е необходим и процесуален подход, който включва задължителен човешки надзор, при който ИИ системата предоставя хипотеза, а оперативният работник я верифицира и носи правна отговорност за нея.

Второто предизвикателство е рискът от т.нар. „алгоритмично пристрастие“. Система, обучена на исторически данни, неизбежно отразява историческия дисбаланс в правоприлагането. Ако историческите данни показват непропорционална насоченост на оперативните проверки към определени социални групи, алгоритъмът ще научи тази закономерност и ще я възпроизвежда и усилва с времето. Контролът изисква внимателен подбор и одит на обучителните данни, редовна оценка на резултатите и разнороден екип при разработката.

Третото предизвикателство е защитата на личните данни при масовото наблюдение. ИИ системите за мрежов анализ и OSINT по необходимост обработват данни за множество лица, по-голямата част от които не са обекти на оперативен интерес. Тази обработка трябва да отговаря на изискванията на Регламент (ЕС) 2016/679 (GDPR) и Директива (ЕС) 2016/680 за защита на данните при правоприлагането и националното законодателство. Тук ключовият критерий се явява пропорционалността. Намесата в правата на гражданите трябва да бъде съразмерна с преследваната цел.

К. Ригано посочва, че ИИ има потенциала да се превърне в постоянна част от системата на наказателното правосъдие, като осигурява помощ при разкриването и разследването и позволява на специалистите в областта да поддържат по-ефективно обществената сигурност (Rigano, 2019: 2).

Оптималният подход предполага ясно разграничение между оперативната и процесуалната функция на ИИ. В оперативен план системата служи за генериране на хипотези (оперативно-издирвателни версии) и насочване на дейността по разкриване, където изискванията за разбираемост са по-ниски, тъй като резултатите не влизат директно в доказателствения материал. В процесуален план всяко доказателство, получено с помощта на ИИ, трябва да бъде

верифицирано чрез традиционни методи и да отговаря на изискванията на НПК. Поддържането на тази граница е ключова институционална гаранция срещу злоупотреба.

Заключение

Настоящото изследване позволява формулирането на няколко извода с принципен характер, отнасящи се едновременно до теоретичната рамка и до практическата приложимост на изкуствения интелект в оперативно-издирвателната дейност за противодействие на киберпрестъпността.

Първо, интеграцията на ИИ в оперативно-издирвателната дейност не е технологична мода, а функционален императив, произтичащ от обективните параметри на съвременната киберпрестъпност. Количественото измерение, обхващащо обема на данните, подлежащи на анализ, надхвърля човешките възможности с порядъци. Качественото измерение, обхващащо способността да се открият скрити корелации в многомерни масиви, е структурно недостъпна за конвенционалните методи. Времевото измерение, обхващащо скоростта на адаптация към нови заплахи, изисква реакция, измерима в минути. Традиционните инструменти не отговарят на нито едно от тези изисквания при обема и темпото на развитие, характерни за съвременната киберпрестъпност.

Второ, нормативната рамка на България, включваща ЗМВР, НПК, Закона за киберсигурност и международните инструменти (Будапещенска конвенция, Директива 2013/40), е технологично неутрална и имплицитно допуска прилагането на ИИ инструменти в оперативно-издирвателната дейност. Това е необходима, но не достатъчна предпоставка. Необходима е разработката на специфична методология, включително стандарти за допустимост на алгоритмично генерирани данни като доказателство и процесуални гаранции, съразмерни с интензивността на намесата в основните права.

Трето, успешното внедряване на ИИ системи предполага решаване на конкретни методологични проблеми като осигуряване на разбираемост на алгоритмите, контрол на „пристрастията“ в обучителните данни, ясно разграничение между оперативна и процесуална функция на ИИ системите и квалификационна подготовка на служителите. Всичките тези задачи са решими в една или друга степен. Международният опит предоставя достатъчно референтни модели, а изследователското поле на „обяснимия ИИ“ предлага конкретни технически решения.

Четвърто, забавянето в адаптирането на тези технологии не е неутрална позиция. Докато правоохранителните органи действат с конвенционални инструменти, техните „противници“ ускоряват технологичното „въоръжаване“. Ако не бъде преодолян своевременно, в средносрочна перспектива разривът може да стане необратим. Асиметрията в киберпространството не се преодолява с численост на служителите или административни мерки, тя се преодолява с технологично адекватен и законово уреден инструментариум.

Пето, въпросът не е дали ИИ ще се утвърди като задължителен компонент в противодействието на киберпрестъпността. Въпросът е дали институциите ще го признаят навреме, за да инвестират проактивно, вместо да реагират принудително след понесени „поражения“. Историята на технологичните трансформации в правоприлагането показва, че тези, които изчакват окончателна правна уредба, преди да предприемат действия, плащат значително по-висока цена от онези, които действат с разумна предпазливост в нормативно неопределена среда.

Бележки

[1] Компютърни програми, които имитират знанията и разсъжденията на човешки експерт в дадена специализирана област. Знанията на един или повече експерти, лекар, юрист, инженер и др., се кодифицират под формата на правила от типа „АКО... ТОГАВА...“. Например „АКО пациентът има температура над 38° и кашли-

ца и болки в гърлото, ТОГАВА вероятната диагноза е грип“. Системата след това прилага тези правила към конкретен случай и стига до заключение.

[2] Математически модели за вземане на решения при непълна или несигурна информация, базирани на теоремата на Байес. Системата работи с вероятности, а не с категорични твърдения. Тя изчислява колко вероятно е дадено събитие да е настъпило, като отчита наличните доказателства и актуализира вероятността при всяко ново доказателство.

[3] Модел за вземане на решения, при който проблемът се разгражда на поредица от последователни въпроси с разклонения, точно като разклонено дърво. От „корена“ на дървото системата задава въпрос, отговорът на който води до следващ въпрос по съответния „клон“, и така нататък, докато се стигне до крайно „листо“, т.е. до решение или класификация.

[4] За разлика от класическата логика, при която нещо е или вярно (1), или невярно (0), размитата логика допуска междинни стойности, например 0.7 или 0.3, т.е. тя работи с понятия като „частично вярно“, „до голяма степен“, „приблизително“.

[5] Метод за решаване на сложни оптимизационни задачи, вдъхновен от биологичната еволюция. Вместо да се търси решение по права линия, системата генерира множество случайни решения, т.нар. „популация“. След това, подобно на естествения подбор в природата, „оцеляват“ и се „размножават“ най-добрите от тях. Чрез „кръстосване“ и случайни „мутации“ на тези решения се получава нова, по-добра популация. Процесът се повтаря многократно, докато се стигне до оптимално или достатъчно добро решение.

[6] Изчислителни модели, вдъхновени от структурата на човешкия мозък. Състоят се от изкуствени „неврони“ (математически функции), наредени на слоеве и свързани помежду си. Когато мрежата се „обучава“ върху данни, теглата на връзките между невроните се коригират, докато системата не започне да разпознава правилно.

[7] Методи, основани на теорията на вероятностите, при които системата работи с несигурност и непълна информация. Вместо категорични отговори генерира вероятностни оценки, които се актуализират при постъпване на нови данни.

[8] Уязвимост от нулев ден, за която все още не съществува официална корекция (updates) от разработчика и която може да бъде използвана за атаки, преди проблемът да бъде публично разкрит или решен.

[9] Прозорец на уязвимост. Периодът от време, през който дадена система остава уязвима за атака, след като дадена слабост е възникнала или е станала известна, но преди да бъде приложена ефективна защитна мярка като инсталиране на корекция или промяна в конфигурацията. През този прозорец нападателите могат да експлоатират уязвимостта, без системният администратор или разработчик да имат възможност да предотвратят атаката.

[10] Advanced Persistent Threat – усъвършенствана устойчива заплаха.

[11] Security Information and Event Management – управление на информацията и събитията по сигурността. Това е клас софтуерни системи, които в реално време събират, агрегират и анализират регистрационни записи (логове) и данни за сигурността от множество източници в дадена IT инфраструктура.

[12] Natural Language Processing – това е подобласт на изкуствения интелект, която се занимава с взаимодействието между компютрите и човешкия език. NLP системите са способни да четат, разбират, интерпретират и генерират текст на естествен човешки език, за разлика от формалните програмни езици.

Библиография

Budapeshtenska konventsiya za prestapleniyata v kibernetichnoto prostranstvo [Будапещенска конвенция за престъпленията в кибернетичното пространство]. (2001, ноември 23). Съвет на Европа. Ратифицирана от България. (2005). https://www.coe.int/en/web/cybercrime/the-budapestconvention?utm_source=chatgpt.com

Direktiva (ES) 2013/40 na Evropeyskiya parlament i na Saveta ot 12 avgust 2013 g. otnosno atakite sreshtu informatsionnite sistemi [Директива (ЕС) 2013/40 на Европейския парламент и на Съвета от 12 август 2013 г. относно атаките срещу информационните системи]. (2013). *Ofitsialen vestnik na ES*, L 218, 14.08.2013 г. https://eur-lex.europa.eu/eli/dir/2013/40/oj

Direktiva (ES) 2016/680 na Evropeyskiya parlament i na Saveta ot 27 april 2016 g. otnosno zashtitata na fizicheskite litsa pri obrabotvaneto na lichni danni ot kompetentni organi za tselite na predotvratyavaneto, razsledvaneto, razkrivaneto ili nakazatelnoto presledvane na prestapleniya [Директива (ЕС) 2016/680 на Европейския парламент и на Съвета от 27 април 2016 г. относно защитата на физическите лица при обработването на лични данни от компетентни органи за целите на предотвратяването, разследването, разкриването или наказателното преследване на престъпления]. (2016). *Ofitsialen vestnik na ES*, L 119, 04.05.2016 г. https://eur-lex.europa.eu/eli/dir/2016/680/oj

Dunsin, D., Ghanem, M. C., Ouazzane, K., & Vassilev, V. (2024). A comprehensive analysis of the role of artificial intelligence and machine learning in modern digital forensics and incident response. *Forensic Science International: Digital Investigation*, *48*, Article 301675. https://doi.org/10.1016/j.fsidi.2023.301675

Faqir, R. S. A. (2023). Digital criminal investigations in the era of artificial intelligence: A comprehensive overview. *International Journal of Cyber Criminology*, *17*(2), 77–94. https://cybercrimejournal.com/menuscript/index.php/cybercrimejournal/article/view/189

Ivanov, A., & Georgieva, P. V. (2018). Classification with convolutional neural networks [Иванов, А., & Георгиева, П. В. (2018). Класификация с конволюционни невронни мрежи]. *Kompyutarni nauki i komunikatsii [Компютърни науки и комуникации]*, *7*(1), 46–52.

Madanski, E. (2021). Izpolzvane na yavni iztochnitsi v operativnoto prouchvane na litsa [Мадански, Е. (2021). Използване на явни източници в оперативното проучване на лица]. *Byuletin na Fakultet „Politsia“ AMVR [Бюлетин на факултет „Полиция“ AMBP]*, *5*, 47–59.

Madanski, E. (2024). Efektivni instrumenti i pohvati za pridobivane na operativna informatsiya ot ueb bazirani yavni iztochnitsi [Мадански, Е. (2024). Ефективни инструменти и похвати за придобиване на оперативна информация от уеб базирани явни източници]. *Byuletin na Fakultet „Politsia“ AMVR [Бюлетин на факултет „Полиция“ AMBP]*, *15*, 89–95.

Nakazatelen kodeks na Republika Bulgaria, Glava deveta „a“ – Kompyutarni prestapleniya (chl. 319a–319e) [Наказателен кодекс на Република България, Глава девета „а“ – Компютърни престъпления (чл. 319а–319е)]. (2002). Обн. ДВ, бр. 26, 2 април 2002 г. https://lex.bg/laws/ldoc/-1589654529

Nakazatelno-protsesualen kodeks na Republika Bulgaria [Наказателно-процесуален кодекс на Република България]. (2006). Обн. ДВ, бр. 86, 28 октомври 2005 г. https://lex.bg/laws/ldoc/2135512224

Penev, Y. (2020). Politseyski operativno-izdirvatelni sposobi i organizatsionni formi [Пенев, Й. (2020). Полицейски оперативно-издирвателни способи и организационни форми]. *Byuletin na Fakultet „Politsia“ AMVR [Бюлетин на факултет „Полиция“ AMBP]*, *2*(40), 9–20.

Reglament (ES) 2016/679 na Evropeyskiya parlament i na Saveta ot 27 april 2016 g. otnosno zashtitata na fizicheskite litsa pri obrabotvaneto na lichni danni i otnosno svobodnoto dvizhenie na takiva danni (Obsht reglament otnosno zashtitata na dannite) [Регламент (ЕС) 2016/679 на Европейския парламент и на Съвета от 27 април 2016 г. относно защитата на физическите лица при обработването на лични данни и относно свободното движение на такива данни (Общ регламент относно защитата на данните – GDPR)]. (2016). *Ofitsialen vestnik na ES*, L 119, 04.05.2016 г. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Rigano, C. (2019). Using artificial intelligence to address criminal justice needs. *NIJ Journal*, *280*, 1–10. https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs

Shestak, V. A., & Volevodz, A. G. (2019). Modern requirements of the legal support of artificial intelligence: A view from Russia [Шестак, В. А., & Волеводз, А. Г. (2019). Современные потребности правового обеспечения искусственного интеллекта: взгляд из России]. *Russian Journal of Criminology*, *13*(2), 197–206. https://doi.org/10.17150/2500-4255.2019.13(2).197-206

Velasco, C. (2022). Cybercrime and artificial intelligence: An overview of the work of international organizations on criminal justice and the international applicable instruments. *ERA Forum*, *23*, 109–126. https://doi.org/10.1007/s12027-022-00702-z

Zakon za kibersigurnost na Republika Bulgaria [Закон за киберсигурност на Република България]. (2018). Обн. ДВ, бр. 21, 13 март 2018 г. https://lex.bg/bg/laws/ldoc/-2137188253Zakon za Ministerstvoto na vatreshnite raboti [Закон за Министерството на вътрешните работи]. (2014). Обн. ДВ, бр. 53, 27 юни 2014 г. https://lex.bg/laws/ldoc/-2136243824

Данни за автора: *доц. д-р Андрей Атанасов Джунин, катедра „Оперативно-издирвателна дейност“, факултет „Полиция“, Академия на МВР, гр. София, бул. „Александър Малинов“ № 1, телефон: +359885031321, e-mail: junin@abv.bg*