https://doi.org/10.70265/DSHA6230

ИЗКУСТВЕНИЯТ ИНТЕЛЕКТ В КИБЕРСИГУРНОСТТА: ВЪЗМОЖНОСТИ И РИСКОВЕ

Цветомир Алексов

ARTIFICIAL INTELLIGENCE IN CYBERSECURITY: OPPORTUNITIES AND RISKS

Tsvetomir Aleksov

Резюме: *Статията е насочена към тенденциите при използването на изкуствен интелект в киберсигурността и неговото все по-голямо влияние. Обърнато е внимание на положителните и отрицателните страни при използването му, като са разгледани и регулаторните норми спрямо него.*

Ключови думи: *изкуствен интелект, киберсигурност, регулации, рискове, развитие, атаки*

Summary: *The article focuses on the trends in the use of artificial intelligence in cybersecurity and its growing impact. It highlights both the positive and negative aspects of its application and examines the regulatory frameworks related to it.*

Keywords: *artificial intelligence, cybersecurity, regulations, risks, development, attacks*

УВОД

Изкуственият интелект (ИИ) се е утвърдил като ключов фактор в съвременната киберсигурност, предлагайки нови и ефективни начини за подобряване на защитата срещу киберзаплахи, но също така създава възможности за злонамерени действия. ИИ може да се използва в различни роли и контекст. Във връзка с това от една страна ИИ помага на фирми и правоохранителни органи да откриват и да реагират на заплахи, а от друга предлага нови методи на нападателите, като те усъвършенстват и автоматизират своите атаки.

Целта на настоящата статия е да се анализират възможностите и рисковете, които съпътстват внедряването на ИИ в сферата на

киберсигурността с акцент върху необходимостта от балансиран подход между иновации и регулации.

Идеята на настоящата работа е да се види резултата от използването на изкуствен интелект в съвременната киберсигурност, а предмет на изследването са конкретните възможности на ИИ в киберсигурността, потенциалните ползи, свързаните рискове и предизвикателствата пред нормативната уредба.

Изкуственият интелект се превръща в неотменим компонент на съвременната киберсигурност, но неговата ефективна и етична интеграция изисква не само технологични решения, но и ясна регулаторна рамка, която да ограничава злоупотребите и да гарантира сигурността на цифровата среда.

Според изследване на една от водещите компании в сектора (Dhaliwal, 2022) много от фирмите за киберсигурност вече използват ИИ в своите продукти и бъдещи стратегии. Това подчертава растящото значение на технологиите в киберсферата. ИИ трансформира киберподходите, като подобрява откриването на заплахи, автоматизира реакциите и създава нови предизвикателства.

Настоящите приложения на ИИ в киберсигурността включват системи, които откриват, предотвратяват и отговарят на заплахи с ненадминатата прецизност и бързина. Тези тактически решения обхващат филтриране на спам, блокиране на зловреден софтуер, откриване на нарушения и поведенчески анализ. В отговорни сценарии ИИ позволява бързо ограничаване и отстраняване на проблемите, минимизирайки времето, щетите и демонстрирайки ключовата роля на ИИ за осигуряване на цялостност и устойчивост в информационните системи.

Въпреки тези тактически успехи съществува стратегическа празнина в цялостната оценка на рисковете в киберсигурността. Настоящите инструменти с ИИ често не предоставят холистичен поглед върху киберсигурността на организацията, което ограничава тяхната ефективност в подкрепа на планирането на рискове и препоръчване на оптимални контрамерки. Това от своя страна е от съществено за ефективното управление на рисковете. ИИ има потенциала да създаде детайлни рискови профили, стратегически насочвайки целенасочени действия в киберсигурността и осигурявайки съответствие между тактическата защита и специфичните заплахи и уязвимости на всяка компания (Schreiber, 2025).

1. ВЪЗМОЖНОСТИ ЗА КИБЕРСИГУРНОСТ И АВТОМАТИЗАЦИЯ НА ПРОЦЕСИТЕ

ИИ промени изцяло киберпространството и наложи революционни методи за автоматизирано откриване и реагиране на

заплахи, което значително повишава ефективността. Според проучване (Dhaliwal, 2022) ИИ базираните системи могат да намалят значително времето за откриване на заплахи, което увеличава нивото на защитата. Ключовите предимства са своевременна реакция с анализ на мрежови модели, потребителско поведение, системни аномалии за милисекундно откриване и възможностите за мащабируемост. Пример за това е платформата Darktrace, която използва машинно обучение за идентифициране на аномалии в мрежовия трафик. Това позволява на организациите да откриват и да реагират на заплахи в реално време особено в сложни мрежови среди, където ръчният анализ би бил изключително трудоемък и бавен. Такава автоматизация е преди всичко полезна за големи корпорации, които генерират огромни обеми данни и се нуждаят от интелигентни решения за защита (Jackson, 2023).

Друг пример за конкретно приложение на ИИ в автоматизацията е използването на платформи като CylancePROTECT, които интегрират ИИ в своите антивирусни технологии, като анализират файлове и процеси в реално време. Те намаляват броя на фалшивите положителни и ускоряват откритията на нови заплахи (Dodt, 2022). Тези решения представляват иновативна платформа за предотвратяване на заплахи, базирана на изкуствен интелект. Платформата използва напреднали алгоритми за машинно обучение и модели за изкуствено зрение за превантивно откриване и блокиране на зловреден софтуер, които автоматично усъвършенстват детекционните си способности с течение на времето. Системата използва статичен анализ, а моделите са обучени върху милиарди файлове и разпознават шаблони свързани със злонамерено поведение. Според независими тестове (AV-TEST, 2020) CylancePROTECT има висока успеваемост при блокиране на zero-day атаки с фалшиво-положителни резултати под 0,1%. В допълнение няма зависимост от сигнатури и минимално въздействие върху производителността. В таблицата по-долу може да се види пример на успешни платформи за автоматизация.

Таблица 1. Платформи за автоматизация (оригинал изработен от автора)

Платформа	Технология	Приложение
CrowdStrike Falcon	Поведенчески анализ (EDR)	Откриване на злонамерени процеси в реално време
Palo Alto Cortex XDR	Изкуствен интелект за корпоративни мрежи	Консолидиран анализ на сигурността

IBM QRadar AI	NLP за анализ на логове	Откриване на сложни multi-vector атаки

Освен посочените плюсове от използване на ИИ няма как да не бъдат отбелязани и недостатъците:

– Зависимост от качеството на данните: Ако ИИ моделът е обучен върху недостатъчно или некоректни данни, може да пропусне заплахи.

– Риск от прекомерна автоматизация: Някои организации прекаляват с доверието си в ИИ, пренебрегвайки човешкия фактор прекалено много.

– Атаки срещу самия ИИ: Злонамерените лица могат да манипулират алгоритми.

2. ПРОГНОЗИРАНЕ НА БЪДЕЩИ ЗАПЛАХИ

Изкуственият интелект не просто реагира на заплахи – той предвижда бъдещи атаки чрез анализ на исторически данни, предлагайки революционен проактивен подход към сигурността. Това преобразява традиционната „реактивна“ парадигма в превантивна киберзащита. ИИ системите използват няколко ключови метода за прогнозиране:

– временни серии анализ;

– алгоритми като LSTM (Long Short-Term Memory) мрежи анализират исторически данни за атаки, за да откриват цикличност и тенденции в киберзаплахите;

– мрежи за анализ на взаимоотношения;

– прогнозиращо моделиране на риска – системите оценяват вероятността за конкретни типове атаки въз основа на корпоративния профил и индустрията.

Например IBM Watson for Cybersecurity използва когнитивни технологии за анализ на минали инциденти и прогнозиране на възможни бъдещи заплахи (IBM Security, 2021), също така IBM разглежда как ИИ вече е реален способ за преобразуване на бизнес операциите (Brodsky, 2025). Това предоставя стратегически предимства на организациите, като им позволява да се подготвят за бъдещи рискове и да разработват по-ефективни защитни механизми. Разширяването на ИИ решенията, които прогнозират атаки включва също така анализ на времето и мястото на нападенията. ИИ може да предсказва не само самите атаки, но и да анализира социалните медии и други публични източници за сигнали, които предшестват атаки. Такива иновации позволяват на екипите за киберсигурност да бъдат крачка напред и по-ефективни. В таблица 2 са посочени стратегическите предимства на някои от водещите организации:

Таблица 2. Предимства (оригинал изработен от автора)

Предимство	Описание	Бизнес ефект
Превантивна защита	Разполагане на патчове и правила преди атаката	Намаляване на downtime с 60-70%
Оптимизирани ресурси	Фокусиране върху най-вероятните рискове	До 40% икономия на SOC разходи
Конкурентно предимство	Защита на интелектуална собствена преди компрометиране	Запазване на пазарна позиция

Ето пример на някои водещи прогностични решения:

– Microsoft Cyber Signals – анализира огромен брой сигнали дневно;

– Google Chronicle – използва машинно обучение за прогнозиране на IoC (Indicators of Compromise);

– Splunk AIOps – използва изкуствен интелект и машинно обучение за автоматизирано откриване на аномалии, прогнозиране на инциденти и оптимизиране на реакциите в ИТ инфраструктури.

Тук също следва да се подчертае, че има и минуси, върху които трябва да се работи с цел подобряване. В това число влизат качеството на данните, етични дилеми и критична зависимост от експерти. Трябва да сме наясно, че ИИ все още не може да замени човешката интуиция напълно.

ИИ се използва и за подобряване на процесите за идентификация на потребителите включително чрез биометрични технологии. Биометричните системи, базирани на ИИ могат да анализират уникални характеристики на потребителите като лицето, гласа или дори поведението на потребителя, за да предотвратят неоторизиран достъп. За 2024 г. се забелязва увеличаване на компаниите използващи биометрия за физически достъп, което е увеличение спрямо 30% преди две години (Opiah, 2024). Глобалният пазар на биометрични системи се очаква да нарасне от $47.2 милиарда през 2024 г. до $84.5 милиарда през 2029 г., което показва значителен ръст в индустрията (MarketsandMarkets, 2024). Таблица 3 представя най-разпространените биометрични технологии.

Таблица 3. Биометрия (оригинал изработен от автора)

Биометрична технология	**Ориентировъчен пазарен дял в %**
Пръстови отпечатъци	59
Лицево разпознаване	32
Гласова биометрия	15
Ирисово разпознаване	7

Други	3

Важно е да се добави, че в трите процента (3%) на графа „Други“ от таблицата попадат: разпознаване на подпис, динамично разпознаване на ходене, както и биометрия на вените. Тези технологии предстои да бъдат развити (Mordor Intelligence, 2025).

Ключовите фактори за внедряване са:

- **Финансов сектор** – използването на биометрия в този сектор е широко разпространено, като се използват технологии като разпознаване на пръстови отпечатъци, лицево разпознаване и гласова биометрия за сигурна автентикация и предотвратяване на измами.

- Здравеопазване – внедряването на биометрични технологии в здравеопазването е нарастващо, като се използват за сигурен достъп до здравни досиета и предотвратяване на медицински идентификационни кражби.

- **Производство** – в този сектор биометрията се използва основно за контрол на достъпа до опасни зони и управление времето на служителите.

- **ИТ компании** – биометричните технологии се използват за защита на данни и предотвратяване на неоторизиран достъп.

Предизвикателства и съмнения:

– регулаторни ограничения (GDPR изисквания за биометрични данни);

– стойност на внедряване;

– съпротива от служители.

Бъдещи тенденции:

– поведенческа биометрия (анализ на клавиатурен почерк, движение на мишката) – очакван ръст;

– мултимодални системи (комбинация от 2+ биометрични метода) – очакван ръст;

– Blockchain за съхранение на биометрични шаблони – нова вълна от решения и възможности.

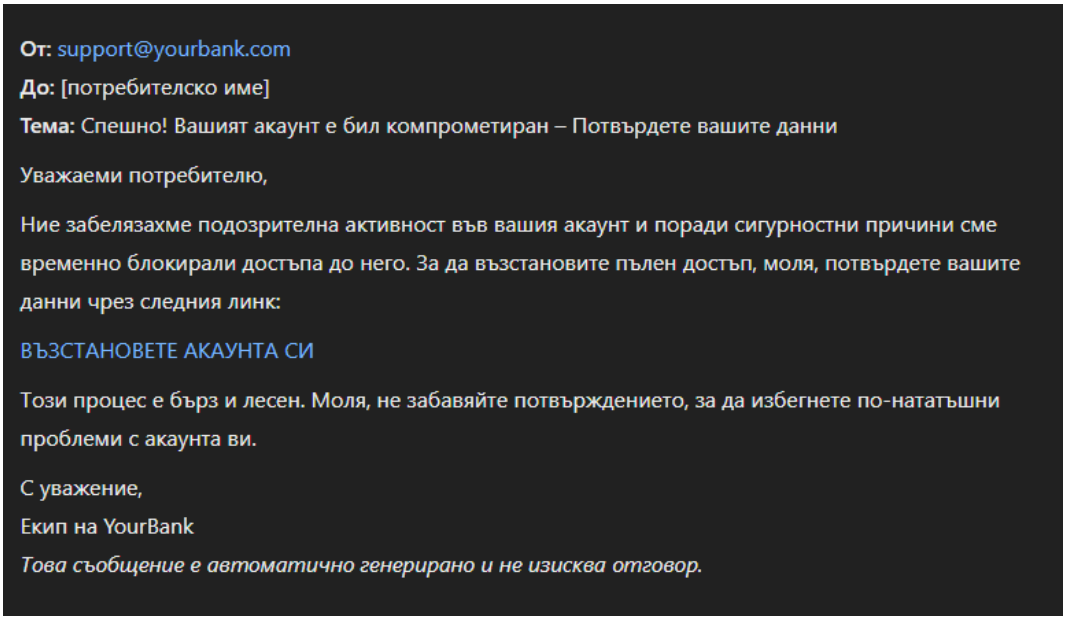
Един от успешните примери за биометрична идентификация е **Face++**, който използва ИИ за разпознаване на лицето и идентифициране на фалшиви идентичности в реално време. Такъв тип система се използва в банковата индустрия за предотвратяване на финансови измами и защита на личните данни (Face++, 2025).

3. РИСКОВЕ ОТ УПОТРЕБАТА НА ИИ

Злонамерените лица и групи също използват ИИ за автоматизация на кибератаките. От една страна модели като GPT-3/4, deepseek и др. могат да генерират фишинг имейли, които изглеждат много по-

убедителни и трудни за разпознаване от традиционните методи за защита.

Ето два примера за фишинг имейли, генерирани за няколко секунди без никакви допълнителни действия.



От: support@yourbank.com
До: [потребителско име]
Тема: Спешно! Вашият акаунт е бил компрометиран – Потвърдете вашите данни

Уважаеми потребителю,

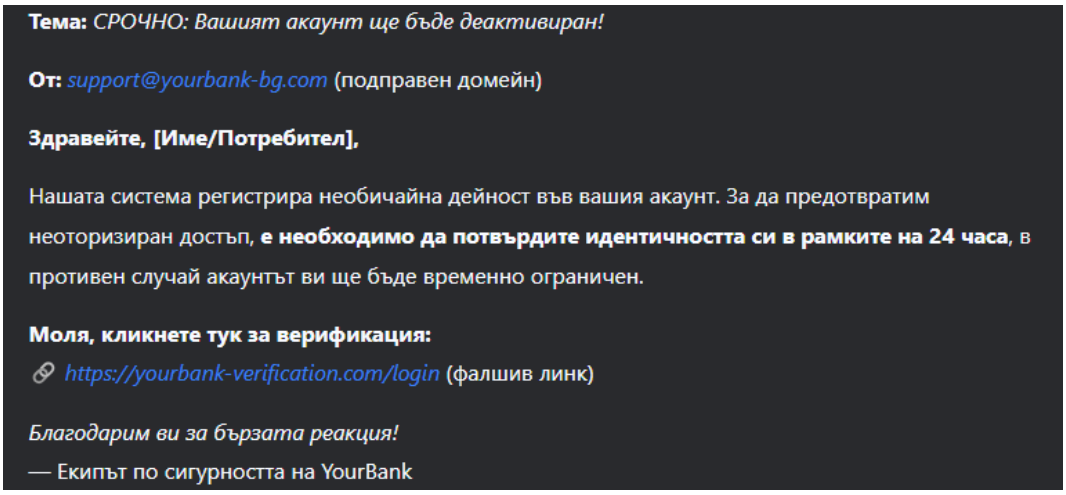
Ние забелязахме подозрителна активност във вашия акаунт и поради сигурностни причини сме временно блокирали достъпа до него. За да възстановите пълен достъп, моля, потвърдете вашите данни чрез следния линк:

ВЪЗСТАНОВЕТЕ АКАУНТА СИ

Този процес е бърз и лесен. Моля, не забавяйте потвърждението, за да избегнете по-нататъшни проблеми с акаунта ви.

С уважение,
Екип на YourBank
Това съобщение е автоматично генерирано и не изисква отговор.

Фигура 1. Примерен фишинг мейл. (Източник: ChatGPT, 2025)



Тема: *СРОЧНО: Вашият акаунт ще бъде деактивиран!*

От: *support@yourbank-bg.com* (подправен домейн)

Здравейте, [Име/Потребител],

Нашата система регистрира необичайна дейност във вашия акаунт. За да предотвратим неоторизиран достъп, **е необходимо да потвърдите идентичността си в рамките на 24 часа**, в противен случай акаунтът ви ще бъде временно ограничен.

Моля, кликнете тук за верификация:
https://yourbank-verification.com/login (фалшив линк)

Благодарим ви за бързата реакция!
— Екипът по сигурността на YourBank

Фигура 2. Примерен фишинг мейл. (Източник: DeepSeek, 2025)

От друга страна се наблюдава увеличаване на атаките с използване на **deepfake** технологии, които могат да имитират гласове и лица на публични личности, за да манипулират обществото или да извършват измами. Ето генерирана фигура от ИИ:

Фигура 3. Deepfake пример. (*Източник: ChatGPT, 2025*)

Пример за атака е инцидентът с **AI-generated phishing**, при който хакери създадоха фалшиви имейли от лица с високи постове в компанията, като използваха ИИ за да генерират много убедителни текстове, които подлъгват служители да разкрият своите данни за достъп.

ИИ също така позволява създаването на еволюиращ зловреден софтуер, който може да променя кода си в реално време, за да избегне открития от традиционни антивирусни системи. Например Generative Adversarial Networks (GANs) могат да генерират нови варианти на malware, които са оптимизирани да избегнат открития от традиционни системи за сигурност. Според доклад на Symantec атаките все повече използват ИИ и машинно обучение (Symantec, 2023).

4. ЕТИЧНИ И ПРАВНИ АСПЕКТИ НА ИЗПОЛЗВАНЕТО НА ИИ В КИБЕРСИГУРНОСТТА

Един от основните етични въпроси е кой носи отговорност, ако ИИ базирана система пропусне заплаха или генерира фалшиво положителен резултат. Въпросът за отговорността става особено важен, когато ИИ се използва за автономни атаки или при взаимодействие с критични инфраструктури. Според European Union Agency for Cybersecurity трябва да се разработят ясни регулаторни рамки за използването на ИИ в киберсигурността, които да осигуряват отговорност и прозрачност (ENISA, 2021).

Правителства и регулаторни органи предприемат различни стъпки за използването на ИИ в киберсигурността чрез законодателство и съдебни решения. Регламентът за ИИ, приет през 2024 г., е първия всеобхватен документ за ИИ в света. Регламентът:

– класифицира ИИ системите в 4 рискови категории;

– забранява определени практики;

– поставя изисквания за прозрачност, изисквайки прозрачност при приложенията в киберсигурността (European Parliament, 2024)

Таблица 4. ИИ в производството (оригинал изработен от автора)

Индустрия	Влияние
Финтех	Строги изисквания за ИИ в изкуствен интелект за откриване на измами
Здравеопазване	Задължителен аудит на диагностични ИИ инструменти
Умни градове	Забрана за използване на ИИ за масово наблюдение без съдебно решение

Регламентът GDPR (Общ регламент за защита на данните) и използването на ИИ са тясно свързани, особено при обработката на лични данни. Там са заложени законност и прозрачност. Системите използващи лични данни, трябва да имат ясно определена правна основа (съгласие, договор, законово изискване). Хората трябва да бъдат информирани (чрез Privacy Policy) как техните данни се използват за автоматизирано вземане на решения. При автоматизирани решения лицата имат право да получат обяснение как ИИ модела е взел решението (чл. 22 GDPR). Компаниите трябва да прилагат Privacy by Design в ИИ системите. Длъжностно лице по защита на данните трябва да надзирава проекти с висок риск.

През 2023 г. италианският орган за защита на данните глоби Clearview AI с 20 млн. евро за незаконна обработка на биометрични данни без съгласие, което подчертава рисковете от ИИ при обработката на лична информация (Mascellino, 2022). Европейският съд по правата на човека (ЕСПЧ) постанови, че автоматизираните системи за наблюдение трябва да спазват принципа на пропорционалност, особено при масово събиране на данни (ЕСПЧ, 2023).

ИИ носи и значителни рискове, които не са свързани с хакерски атаки. На първо място прекомерна зависимост от автоматизацията. Експертните екипи, които разчитат изцяло на ИИ могат да пропуснат аномалии извън обхвата на обучението на моделите. Автоматизираните системи могат да игнорират нови типове заплахи, които не са били предвидени при проектирането им. Също така, ако ИИ модел е обучен върху данни, които не отразяват различен тип заплахи може да пропуска атаки срещу определени групи. Използването на ИИ за автоматична защита може да доведе до неконтролируема ескалация между нападатели и защитници.

Ето и няколко категории обособени рискове при използването на ИИ:

• **Етични проблеми**: Значителни етични въпроси относно ролята на човешкото разсъждение в нападателните действия се поставят с разгръщането на автономни оръжия.

• **Надеждност и непредсказуемост**: В сложни, реални ситуации ИИ системите могат да действат неочаквано, което може да има непредвидени последици.

• **Уязвимост и подправяне**: Противниците могат да повредят ИИ системи, което може да обърне оръжията срещу техните оператори;

• **Зависимост от данни**: За да работят добре системите с ИИ изискват огромни обеми висококачествени данни, което не винаги е възможно в бойни условия.

• **Липса на човешко разсъждение**: ИИ може да има трудности при вземането на сложни решения, които изискват емпатия, културна осведоменост или внимателно разглеждане на етични въпроси.

• **Рискове от ескалация**: Ефективността и бързината на ИИ управляемото военно действие имат потенциал да доведат до бързо ескалиране на конфликти, увеличавайки възможността за голям конфликт.

• **Проблеми с разпространението**: Технологията на ИИ може да попадне в ръцете на лоши държави или недържавни актьори, тъй като става все по-достъпна.

• **Правна неяснота**: Съществуват сложни правни въпроси свързани с отговорността и спазването на международното право при разгръщането на автономни въоръжени системи (Mircheska, 2024, pp. 43-55).

ЗАКЛЮЧЕНИЕ

Изкуственият интелект представлява както огромна възможност, така и сериозна заплаха за киберсигурността. От една страна ИИ може да помогне за значително подобряване на защитата, като автоматизира откритията на заплахи и прогнозира бъдещи атаки, а от друга страна предоставя нови възможности за нападатели да създават по-ефективни и адаптивни кибератаки. Организациите трябва да инвестират в ИИ решения за киберсигурност, но също така трябва да бъдат наясно с рисковете, които тези технологии носят и да предприемат необходимите мерки за защита и етично използване на ИИ. Освен това регулирането на ИИ и сътрудничеството между различни държави ще бъде от решаващо значение за справяне с новите предизвикателства в киберсигурността.

Автоматизацията чрез ИИ вече е необходим стандарт в киберсигурността. Платформи като Darktrace, CrowdStrike и IBM QRadar демонстрират, че бъдещето на защитата е заложено в интелигентните, самообучаващи се системи. Въпреки това

организациите трябва да намерят баланс между автоматизация и човешки надзор, за да минимизират рисковете.

ИИ базираните системи често изискват големи количества данни за обучение, което може да наруши личната неприкосновеност. Проблемите със съхранението на чувствителни данни и спазването на регулации като GDPR в Европейския съюз поставят бариери пред широкото използване на ИИ в някои региони, които изискват стриктно спазване на законите за защита на личните данни.

Технологичният прогрес в областта на изкуствения интелект и квантовите изчисления се очаква коренно да преобрази киберсигурността. Квантовият ИИ може да доведе до експоненциално по-бърза обработка на данни и прецизно откриване на заплахи, докато автономните системи ще могат да реагират на атаки в реално време без човешка намеса.

Въпреки възможностите, регулаторните изисквания ще стават по-строги, особено що се отнася до прозрачността на алгоритмите при вземане на решения, за да се гарантира отговорност при критични инциденти. Иновации като Blockchain + ИИ и квантово машинно обучение ще подобрят сигурността, но същевременно ще поставят нови предизвикателства пред бизнеса и законодателите.

ЛИТЕРАТУРА:

AV-TEST (2020). *Independent Evaluation of AI-Based Endpoint Protection*. Retrieved July August 2020, from https://www.av-test.org/en/antivirus/business-windows-client/windows-10/august-2020/cylance-protect-2.1-203207/

Brodsky, S. (2025). *AI gets down to business at IBM Think 2025*. IBM https://www.ibm.com/think/news/ai-business-think-2025

ChatGPT. (2025). chatgpt.com.

Dhaliwal, J. (2022). *The what, why, and how of AI and threat detection*. McAfee. https://www.mcafee.com/blogs/internet-security/the-what-why-and-how-of-ai-and-threat-detection/

DeepSeek. (2025). deepseek.com

Dodt, C. (2022). *CylancePROTECT – Product overview*. Cylance. https://www.infosecinstitute.com/resources/general-security/cylanceprotect-product-overview/

European Union Agency for Cybersecurity [ENISA]. (2021). *Artificial Intelligence in Cybersecurity.* ENISA. https://www.enisa.europa.eu

European Parliament. (2024). *EU AI Act: first regulation on artificial intelligence*. June 03 2025, from https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Face++. (2025). *Face Detection*. https://www.faceplusplus.com/face-detection/

IBM Security (2021). *Cognitive Cybersecurity with Watson*. White Paper. November 16 2016, from https://research.ibm.com/publications/cognitive-cyber-security-assistants-computationally-deriving-cyber-intelligence-and-course-of-actions?utm_source=chatgpt.com

Jackson, A. (2023). *Darktrace AI offers real-time cyber resilience for cloud*. October 30, 2023, from https://cybermagazine.com/cloud-security/darktrace-ai-offers-real-time-cyber-resilience-for-cloud

MarketsandMarkets. (2024, August 07). Biometric System Market worth $84.5 billion by 2029 - Exclusive Report by MarketsandMarkets. https://www.prnewswire.com/news-releases/biometric-system-market-worth-84-5-billion-by-2029---exclusive-report-by-marketsandmarkets-302216426.html?utm_source=chatgpt.comdeepseek.com

Mircheska, S. (2024). Perspectives for the Development of Artificial Intelligence: Security Dimension of Artificial Intelligence in the Military Sector. *Security and Defense, (2*), 43-55. https://doi.org/10.70265/ETNV6785

Opiah, A. (2024). *Businesses increasingly turn to biometrics for physical access and visitor management*. August 26 2024, from https://www.biometricupdate.com/202408/businesses-increasingly-turn-to-biometrics-for-physical-access-and-visitor-management?utm_source=chatgpt.com

Schreiber, A., & Schreiber, I. (2025). AI for cyber-security risk: harnessing AI for automatic generation of company-specific cybersecurity risk profiles. *Information and Computer Security, Vol. ahead-of-print No. ahead-of-print*. https://doi.org/10.1108/ICS-08-2024-0177

Symantec (2023). *Symantec Security Center*. June 27 2023, from https://docs.broadcom.com/doc/ai-automation-and-cybersecurity