FRI-2B.312-1-NS -01

ARTIFICIAL INTELLIGENCE IN SECURITY: BETWEEN STRATEGIC NECESSITY AND THE LEGAL BARRIERS OF THE CLASSIFIED INFORMATION PROTECTION ACT

Assoc. Prof. Milen Ivanov, DSc
Department of Security, Faculty of Law
„Angel Kanchev" University of Ruse
Phone: +359 82 888 736
E-mail: mivanov@uni-ruse.bg

Abstract: *The report examines the tension between the strategic necessity of using artificial intelligence (AI) in the security sector and the constraints imposed by the Classified Information Protection Act (CIPA). It highlights that AI is emerging as a key tool for accelerating analytical processes, while the requirements for isolation of automated information systems and strict access control create significant practical barriers to its use in processing classified information.*

Through normative and systemic analysis, the study identifies a structural paradox between operational efficiency and information security and proposes directions for a balanced approach, including the human-in-the-loop model, a specialized accreditation regime for AI components, and phased deployment in controlled environments.

Keywords: *artificial intelligence, CIPA, classified information, security sector, AI.*

ВЪВЕДЕНИЕ

Бурното развитие на изкуствения интелект (ИИ) през последното десетилетие доведе до качествено нов етап в обработката, анализа и интерпретацията на информацията. В сектор „Сигурност" ИИ все по-често се възприема като ключов ускорител на аналитичните процеси, позволяващ обработка на големи масиви данни, откриване на скрити зависимости и подпомагане на оперативното вземане на решения в почти реално време. В условията на нарастваща динамика на рисковете и хибридизация на заплахите подобни технологични възможности постепенно се превръщат не в опция, а в стратегическа необходимост за държавните институции.

Наред с това внедряването на ИИ поставя съществени въпроси пред действащата нормативна рамка за защита на информацията. Особено чувствителен е този проблем в контекста на Закона за защита на класифицираната информация (ЗЗКИ), който е изграден върху принципи като контрол на достъпа, проследимост на действията и ясна персонална отговорност. Част от съвременните AI-системи, характеризиращи се с висока степен на автоматизация и ограничена прозрачност на алгоритмичните решения, трудно се вписват в класическите модели на акредитация и контрол на автоматизираните информационни системи. Така се очертава нарастващо напрежение между технологичната логика на ИИ и нормативната логика на режима за защита на класифицираната информация.

Настоящият доклад има за цел да анализира взаимодействието между стратегическата необходимост от използване на изкуствен интелект в сектор „Сигурност" и ограниченията, произтичащи от действащия режим по ЗЗКИ. Основната изследователска теза е, че съществува структурно несъответствие между потенциала на съвременните AI-решения и нормативните механизми за защита на класифицираната информация, което изисква внимателно балансиране между оперативната ефективност и информационната сигурност.

ИЗЛОЖЕНИЕ

Методологически изследването се основава на нормативен анализ на разпоредбите на ЗЗКИ, системен подход към ролята на ИИ в процесите по сигурност и аналитично съпоставяне на технологичните възможности с регулаторните изисквания. Този подход позволява да бъдат откроени ключовите точки на напрежение и да се очертаят насоки за тяхното преодоляване.

1. Стратегическа необходимост от ИИ в сектор „Сигурност"

Динамиката на съвременната среда за сигурност се характеризира с нарастващ обем информация, висока скорост на събитията и усложняване на заплахите, включително хибридни, кибер и транснационални рискове. Традиционните аналитични подходи, основани преимуществено на човешка обработка на данни, все по-трудно отговарят на изискванията за навременност, дълбочина и мащаб на анализа. В този контекст изкуственият интелект постепенно се утвърждава като ключов технологичен инструмент за повишаване на ефективността на системите за сигурност.

Първият стратегически ефект от внедряването на ИИ е съществото ускоряване на аналитичния цикъл. Алгоритмите за машинно обучение и обработка на естествен език позволяват автоматизирано пресяване на големи информационни масиви, извличане на релевантни зависимости и предварително структуриране на данните за нуждите на анализатора. Това съкращава времето между събирането на информация и формирането на аналитичен продукт, което е критично важно в условия на бързо развиващи се оперативни ситуации. По този начин ИИ не заменя експертната преценка, а действа като мощен усилвател на човешкия аналитичен капацитет.

Вторият ключов аспект е способността на ИИ да работи с големи и хетерогенни масиви данни. Секторът „Сигурност“ генерира и използва информация от разнообразни източници — текстови масиви, комуникационни потоци, сензорни данни, изображения и видео. Класическите инструменти срещат сериозни ограничения при интегрирането и анализа на подобни обеми и формати. Съвременните AI-модели позволяват многомодална обработка, откриване на аномалии и идентифициране на слаби сигнали, които биха останали незабелязани при ръчен анализ. Това има пряко значение за ранното предупреждение, профилирането на рискове и разкриването на сложни зависимости между събития и субекти.

Третото стратегическо измерение е подпомагането на оперативното вземане на решения. Чрез предиктивна аналитика, сценарийно моделиране и автоматизирана оценка на вероятности ИИ може да предоставя на ръководните органи по-структурирана и навременна информационна основа за избор на действия. Особено ценни са тези възможности в ситуации на кризи, управление на инциденти и противодействие на динамични заплахи, където времевият фактор е критичен. При правилно внедряване ИИ функционира като система за интелигентна поддръжка на решенията (decision support), а не като автономен субект на управление.

Наред с предимствата следва да се отчете, че използването на ИИ във високочувствителни среди изисква внимателно управление на рисковете. Сред тях се открояват възможността за алгоритмична пристрастност, генериране на фалшиви положителни резултати, уязвимости към манипулиране на входните данни и зависимост от външни технологични доставчици. Тези рискове обаче не обезсилват стратегическата необходимост от внедряване на ИИ, а подчертават нуждата от адекватна регулаторна и организационна рамка.

В обобщение може да се приеме, че изкуственият интелект вече представлява функционална необходимост за модерните системи за сигурност. Неговото значение произтича не само от технологичния напредък, но и от обективното нарастване на информационната сложност на средата за сигурност. Именно поради това въпросът не е дали ИИ ще бъде използван в сектора, а при какви условия, гарантиращи едновременно оперативна ефективност и надеждна защита на чувствителната информация.

2. Ограничителният режим на ЗЗКИ спрямо ИИ

Внедряването на системи с елементи на изкуствен интелект в сектор „Сигурност“ неизбежно се пресича с режима за защита на класифицираната информация, уреден в Закона за защита на класифицираната информация (ЗЗКИ) и подзаконовата нормативна рамка. Макар законът да е технологично неутрален по своя замисъл, неговата логика е изградена върху предпоставки за контролируемост, проследимост и персонална отговорност, които в редица случаи трудно се съвместяват със спецификите на съвременните AI-системи. Това поражда структурни точки на напрежение, които следва да бъдат ясно идентифицирани.

2.1. Принципи на ЗЗКИ, релевантни за използването на ИИ

Режимът по ЗЗКИ се основава на няколко фундаментални принципа, които имат пряко значение за внедряването на изкуствен интелект.

На първо място стои принципът „необходимост да се знае“. Достъп до класифицирана информация се предоставя само на лица, за които е установена служебна необходимост и които

притежават съответното разрешение за достъп. При AI-системите възниква въпросът доколко автоматизираната обработка на данни от алгоритъм може да се разглежда като допустима форма на „достъп“, когато липсва човешки субект в класическия смисъл. Това е особено чувствително при модели, които обучават или дообучават своите параметри върху класифицирани масиви.

Вторият ключов принцип е строгият контрол на достъпа и разграничаването на правомощията. ЗЗКИ и подзаконовите актове изискват ясно дефинирани роли, нива на достъп и технически механизми за ограничаване на нерегламентирано използване. Съвременните AI-архитектури, особено тези с централизирано обучение или облачна инфраструктура, могат да създадат ситуации, при които обработката на информацията преминава през среди извън пряк организационен контрол, което поставя въпроси относно съответствието с режима за сигурност.

Третият съществен елемент е изискването за акредитация на автоматизираните информационни системи (АИС), обработващи класифицирана информация. Акредитационният процес предполага предвидимост на функциите, яснота на архитектурата и възможност за проверка на сигурността на системата. При AI-моделите, особено базирани на дълбоко обучение, често е налице ограничена обяснимост на вътрешните процеси („black box“ ефект), което затруднява традиционните процедури по оценка и удостоверяване на сигурността.

Четвъртият принцип е проследимостта и персоналната отговорност. Режимът по ЗЗКИ предполага възможност за установяване кой субект е осъществил конкретно действие с класифицирана информация. При високо автоматизирани AI-процеси, които генерират изводи или извършват предварителна аналитична обработка, възниква въпросът за атрибуцията на действията и за границите на човешката отговорност при оперирane със системата.

2.2. Критични точки на напрежение при използването на ИИ

На практика най-сериозните предизвикателства произтичат не от изрична забрана в закона, а от несъответствието между класическия модел на защитена информационна система и характеристиките на съвременните AI-решения.

Първата критична зона е свързана с т.нар. „черна кутия“ на алгоритмите. Много от модерните модели, особено невронните мрежи, предоставят резултати без пълна прозрачност на вътрешната логика. Това затруднява както проверката на надеждността, така и формалната акредитация на системата по действащите процедури. От гледна точка на ЗЗКИ ограничената обяснимост може да се разглежда като риск за управляемостта и контрола.

Втората зона на напрежение е автономната или полуавтономна обработка на класифицирана информация. Макар в повечето практически сценарии ИИ да функционира като инструмент за подпомагане на анализа, степента на автоматизация може да създаде сиви зони относно това дали се спазва принципът на човешки контрол. Особено чувствителни са случаите, при които системата извършва автоматизирано извличане, агрегиране или приоритизиране на класифицирани данни.

Ключов практически проблем, който все по-ясно се очертава в институционалната практика, е свързан с режима на изолираност на автоматизираните информационни системи. Съгласно изискванията за защита на класифицираната информация АИС, обработващи данни с ниво на класификация, по правило функционират в строго контролирани и изолирани среди и не могат да бъдат свободно свързвани с публични мрежи, включително Интернет. Съвременните високоефективни AI-решения обаче в преобладаващата си част разчитат на:

– мащабни външни изчислителни ресурси;
– периодично обновяване на модели;
– достъп до облачна инфраструктура;
– големи обучаващи корпуси.

В този контекст разработването и поддържането на напълно „вътрешен“ (суверенен) AI с висока степен на съвременност се оказва изключително ресурсно, технологично и организационно предизвикателство за повечето администрации. Така възниква **структурен парадокс**: от една страна, ИИ е обективно необходим за съвременните аналитични процеси, а от друга — действащият режим на изолираност и акредитация на АИС де факто и до голяма степен де юре затруднява или практически блокира използването на модерни AI-инструменти при обработването на класифицирана информация.

Третият проблемен кръг засяга използването на външни модели, облачни услуги и зависимости от доставчици. ЗЗКИ поставя високи изисквания към средата, в която се обработва класифицирана информация, включително по отношение на физическа, криптографска и организационна защита. Много съвременни AI-решения обаче са изградени върху разпределени архитектури и чуждестранни технологични платформи. Това поражда въпроси за юрисдикцията, контрола върху данните и възможността за сертифициране на средата.

Четвъртият съществен въпрос е свързан с отговорността при грешни или подвеждащи аналитични резултати. В класическия модел на ЗЗКИ отговорността е ясно персонализирана. При използване на AI-подпомогнат анализ обаче възниква необходимост от ясно разграничаване между ролята на системата и ролята на човешкия анализатор, особено когато алгоритъмът влияе съществено върху крайния аналитичен продукт.

В обобщение може да се приеме, че действащият режим по ЗЗКИ не съдържа изрична забрана за използване на изкуствен интелект, но неговите принципи — особено изискването за изолирани акредитирани среди и строг контрол върху АИС — създават висока практическа бариера пред внедряването на съвременни AI-решения при работа с класифицирана информация. Това налага внимателно адаптиране както на технологичните архитектури, така и на регулаторния подход, така че да се съхрани балансът между необходимата иновация и изискванията за надеждна защита на чувствителната информация.

3. Анализ на конфликта: оперативна ефективност срещу информационна сигурност

Съпоставянето между стратегическата необходимост от внедряване на изкуствен интелект и ограниченията на режима за защита на класифицираната информация очертава класически управленски конфликт между ефективност и сигурност. От една страна, институциите в сектор „Сигурност“ са изправени пред обективен натиск за ускоряване на аналитичните процеси и повишаване на капацитета за обработка на данни. От друга страна, нормативната рамка, изградена върху принципа на максимална защита, въвежда строги ограничения, които забавят или възпрепятстват внедряването на нови технологични решения. Този конфликт не е формален, а структурен и произтича от различната логика на технологичното развитие и на режима за защита на информацията.

3.1. Рискове при прекомерни ограничения

Прекомерно рестриктивното тълкуване и прилагане на изискванията на ЗЗКИ води до няколко съществени оперативни дефицита.

На първо място се откроява рискът от технологично изоставане. В условията на глобална дигитална трансформация способността за използване на AI-инструменти постепенно се превръща в фактор на институционална конкурентоспособност. Организации, които не могат да интегрират автоматизирани аналитични решения, са принудени да разчитат на по-бавни и трудоемки процеси, което намалява тяхната адаптивност към динамични заплахи.

На второ място възниква проблемът със забавянето на аналитичния цикъл. При отсъствие на AI-подпомагане обработката на големи масиви информация остава силно зависима от човешкия ресурс. Това увеличава времето между събирането на данни и формирането на аналитичен продукт, което в кризисни ситуации може да доведе до пропуснати възможности за превенция или навременна реакция.

Третият негативен ефект е намалената дълбочина на анализа. Съвременните AI-модели са особено ефективни при откриване на слаби сигнали, латентни зависимости и нетривиални модели в данните. Ограничаването на достъпа до подобни инструменти повишава вероятността част от значимите индикатори да останат извън вниманието на анализаторите.

Следва да се отчете и рискът от неформално или заобиколно използване на AI-инструменти. Когато институционалната рамка не предоставя легитимен и регулиран механизъм за внедряване, съществува опасност отделни структури или служители да прибягват до нерегламентирани решения, което парадоксално може да увеличи риска за сигурността вместо да го намали.

3.2. Рискове при неконтролирано внедряване

Обратната крайност — прибързано или недостатъчно регулирано внедряване на ИИ в среди с класифицирана информация — също носи значителни заплахи.

На първо място стои рискът от компрометиране на класифицирана информация. Използването на външни AI-модели, облачни услуги или недобре защитени обучаващи среди може да доведе до неволно изтичане на чувствителни данни, включително чрез логове, телеметрия или вторично обучение на модели.

Вторият съществен риск е зависимостта от външни технологични доставчици. В контекста на националната сигурност подобна зависимост може да създаде уязвимости както от техническо, така и от геополитическо естество. Ограниченият контрол върху архитектурата на модела и върху веригата на доставки допълнително усложнява оценката на риска.

Третият проблем е свързан с надеждността на алгоритмичните резултати. AI-системите могат да генерират убедително изглеждащи, но фактически неточни или подвеждащи изводи. В среда на класифицирана информация подобни грешки могат да имат непропорционално сериозни последици, особено ако липсва достатъчно ефективен човешки контрол.

Не на последно място стои въпросът за правната и дисциплинарната отговорност. При силно автоматизирани аналитични процеси може да възникне размиване на отговорността между разработчик, администратор на системата и краен анализатор. Без ясно регламентиран модел на човешки контрол и валидиране това създава правна несигурност, несъвместима с философията на ЗЗКИ.

3.3. Необходимост от балансиран подход

Изложените аргументи показват, че нито крайно рестриктивният, нито прекалено либералният подход представлява устойчиво решение. Реалният управленски въпрос е намирането на балансиран модел, който едновременно:

– съхранява строгия режим на защита на класифицираната информация;

– позволява контролирано използване на AI като инструмент за подпомагане;

– гарантира проследимост и човешка отговорност;

– минимизира зависимостите от външни среди.

В този смисъл конфликтът между оперативна ефективност и информационна сигурност следва да се разглежда не като взаимно изключване, а като задача по институционален дизайн. Решението предполага адаптиране както на технологичните архитектури, така и на регулаторните механизми, така че изкуственият интелект да бъде интегриран като контролиран усилвател на аналитичния капацитет, без да се компрометират основните принципи на защитата на класифицираната информация.

ЗАКЛЮЧЕНИЕ И ПРЕПОРЪКИ

Извършеният анализ показва, че внедряването на изкуствен интелект в сектор „Сигурност“ вече не е въпрос на технологичен избор, а на стратегическа необходимост, обусловена от нарастващия обем данни, ускорената динамика на заплахите и повишените изисквания към аналитичния капацитет на институциите. Същевременно действащият режим по Закона за защита на класифицираната информация е изграден върху принципи, които при настоящата си интерпретация създават висока бариера пред интегрирането на съвременни AI-решения в среди, обработващи класифицирана информация.

Основният структурен проблем се проявява в несъответствието между изискването за изолирани и строго контролирани автоматизирани информационни системи и технологичната логика на модерните AI-модели, които често разчитат на мащабни външни ресурси, периодично обучение и динамична актуализация. В резултат на това се формира парадоксална ситуация, при която ИИ е обективно необходим за ефективното функциониране на съвременните процеси по сигурност, но едновременно с това де факто и до голяма степен де юре трудно може да бъде използван при обработването на класифицирана информация.

При тези условия устойчивото решение не се състои нито в запазване на изцяло рестриктивен подход, нито в прибързана либерализация, а в търсене на внимателно балансиран модел. Необходима е еволюция на регулаторната и организационната практика, която да съхрани високото ниво на защита на класифицираната информация, като същевременно позволи контролирано използване на AI като инструмент за подпомагане на човешкия анализ.

В тази връзка могат да бъдат формулирани следните принципни препоръки:

На първо място, следва ясно да се утвърди моделът **„human-in-the-loop“**, при който изкуственият интелект функционира изключително като система за подпомагане на анализа, а крайното решение и отговорност остават при оправомощено длъжностно лице с достъп до съответното ниво на класификация.

На второ място, необходимо е разработване на специализиран режим за акредитация на AI-компоненти в рамките на автоматизираните информационни системи, който да отчита спецификите на алгоритмичните модели, включително изисквания за обяснимост, проследимост на входните данни и контрол върху обучаващите корпуси.

На трето място, препоръчително е поетапно внедряване на AI-решения чрез пилотни среди с повишен контрол, което ще позволи натрупване на институционален опит и калибриране на механизмите за сигурност преди мащабно прилагане.

На четвърто място, следва да се насърчава изграждането на суверенни или строго контролирани изчислителни среди за чувствителни приложения, с цел ограничаване на зависимостта от външни доставчици и намаляване на риска от неволно изтичане на информация.

На пето място, необходимо е въвеждане на засилен одит и мониторинг на AI-подпомаганите аналитични процеси, включително механизми за верификация на резултатите и проследимост на човешката намеса.

В заключение може да се приеме, че интеграцията на изкуствения интелект в сектор „Сигурност“ е неизбежен етап от технологичното развитие на държавното управление. Предизвикателството пред правната и институционалната рамка не е да възпрепятства този процес, а да го канализира по начин, който едновременно повишава оперативната ефективност и гарантира надеждна защита на класифицираната информация. Само чрез такъв балансиран подход може да бъде преодолян очертаният структурен парадокс между стратегическата необходимост и нормативните ограничения.

REFERENCES

1. Република България. 2002. *Закон за защита на класифицираната информация*. Държавен вестник, бр. 45 от 30 април 2002 г., посл. изм. и доп.

2. **European Union. 2024.** *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng?utm_source=chatgpt.com

Докладът отразява резултати от работата по проект No 2025-ЮФ-01, финансиран от фонд „Научни изследвания“ на Русенския университет.“