https://doi.org/10.70265/YVKC6923

ГЕНЕРАТИВЕН AI И СЪДЪРЖАНИЕ В СОЦИАЛНИТЕ МЕДИИ – КОМУНИКАЦИОННИ ПРЕДИЗВИКАТЕЛСТВА В СФЕРАТА НА СИГУРНОСТТА

Христо Чаушев

GENERATIVE AI AND SOCIAL MEDIA CONTENT – SECURITY COMMUNICATION CHALLENGES

Hristo Chaushev

Резюме: *Генеративният изкуствения интелект (AI) се налага като значителна трансформираща сила в социалните медии, подобряваща създаването на персонализирано съдържание и увеличаваща ангажираността на потребителите. Разработките и възприемането му обаче се фокусират върху предимствата и улесненията, за сметка на дискурса за възможно негативно влияние в сферата на сигурността чрез генериране на фалшиво, подвеждащо или провокиращо съдържание в социалните медии. Посочен е подходът на НАТО към изкуствения интелект.*

Ключови думи: *генеративен изкуствен интелект, социални медии, комуникации, НАТО, сигурност*

Summary: *Generative artificial intelligence (AI) is emerging as a significant transformative force in social media, improving personalized content creation and increasing user engagement. However, its development and uptake has focused on its benefits and facilitation, at the expense of discourse about its possible negative impact in the security sphere by generating false, misleading or provocative content on social media. NATO's approach to artificial intelligence is outlined.*

Key words: *generative artificial intelligence, social media, communications, NATO, security*

УВОД

Появата на изкуствения интелект (*англ*. Artificial Intelligence), последван от генеративен изкуствен интелект, предизвиква бурни и непрестанни промени във всички обществени сфери, включително и

тези, свързани с комуникациите и сигурността. Необходимо е да се отчете навлизането на AI в социалните медии, които се превърнаха в неизменна част от ежедневието ни, свързвайки ни с аудитории отвъд национални, икономически или политически граници. Възможностите за анализ на огромни количества данни и предоставяне на по-персонализирано и препоръчано съдържание от изкуствения интелект се очертава като комуникационно предизвикателство в сферата на сигурността, тъй като е възможно формиране на модерирано поведение на потребителите, разделянето им на целеви групи по предварително зададени критерии и насочване в определена посока на последващите им действия, които могат да бъдат заплаха за съществуващата архитектура за сигурност. Генеративният AI може почти безпроблемно да въздейства върху големи групи хора в социалните медии, приучавайки ги да приемат без особена съпротива изкуствено създадени предложения, представяйки ги като най-подходящи и съобразени с индивидуалните им потребности. Тези промени никак не са безобидни и представляват сериозно предизвикателство пред комуникационния процес. Проучването на ползите и проблемите от използването на генеративния изкуствен интелект в социалните медии, последващото формиране на нагласи и взаимодействия, които могат пряко да рефлектират върху средата за сигурност, са въпроси, които тепърва трябва да бъдат обект на задълбочен и целенасочен научноизследователски интерес.

1. СОЦИАЛНИ МЕДИИ – ОБЩА СТАТИСТИКА И ТЕНДЕНЦИИ

Влиянието на социалните медии върху хората в съвременното общество е повече от значимо. Те въздействат върху начина ни на живот, общуването, взаимодействието ни със семейството и с приятелите. Свързват ни с аудитории без оглед на етноконфесионална, религиозна, национална или друга принадлежност. В света има над 4 милиарда активни потребители на социални медии, а огромна част от населението на Земята използва поне една социална платформа, отчита Statista – глобална компания за данни от статистически проучвания. Организацията публикува информация за най-популярните социални мрежи в света, а данните към месец април 2024 г. са представени в Таблица 1 по брой месечно активни потребители:

Таблица 1. Най-популярните социални мрежи в света към м. април 2024 г. по брой на месечно активни потребители (Адаптирано по: Dixon, 2024)

№ по ред	**Име на мрежата**	**Брой потребители (в милиони)**
1	Facebook	3,065
2	YouTube	2,504
3	Instagram	2,000
4	WhatsApp	2,000
5	TikTok	1,582
6	WeChat	1,343
7	Facebook Messenger	1,010
8	Telegram	900
9	Snapchat	800
10	Douyin	755

Данните на Statista показват, че през месец февруари 2024 г. Facebook, когато отбелязва 20 години от своето създаване, има над три милиарда активни потребители месечно към април същата година и е най-използваната социална мрежа в света. Компанията Meta Platforms притежава четири от най-големите платформи за социални медии, всяка от които има повече от един милиард активни потребители месечно: Facebook (основна платформа), WhatsApp, Facebook Messenger и Instagram.

Водещите социални мрежи обикновено са достъпни на няколко езика и дават възможност на потребителите да се свързват с аудитории отвъд географски, политически или икономически граници. Очакванията на Statista са ползвателите непрекъснато да се увеличават, тъй като използването на мобилни устройства и на мобилни социални мрежи придобива все по-голяма популярност в региони, които досега не са били обслужвани.

2. ИЗКУСТВЕНИЯТ ИНТЕЛЕКТ В СОЦИАЛНИТЕ МЕДИИ

Изкуственият интелект в социалните медии се използва за анализиране на огромен обем от данни, за да се предостави персонализирано съдържание на потребителите и да се увеличи тяхната ангажираност. Генеративният AI разширява полето на въздействие и е способен да произвежда богата гама от творческо съдържание, включващо изображения, видео, реч, истории, музика и т.н., въз основа на събрани потребителски данни и по заявка на потребител. Без да навлизаме в полето на теоретичните дефиниции, свързани с

изкуствения интелект и съобразно целите на настоящата статия, възприемаме определението на една от най-големите световни технологични компании, а именно IBM, което е достъпно на интернет сайта на софтуерния гигант, че генеративния AI е изкуствен интелект, който може да създаде оригинално съдържание – като текст, изображения, видео, реч, аудио или софтуерен код, в отговор на заявка на потребителя (Stryker & Scapicchio, 2024). Създаденото ново съдържание може да се използва за публикация в социалните медии. От този факт произтичат съответните предимства и предизвикателства, които най-общо могат да бъдат представени по следният начин:

2.1 Предимства на AI в социалните медии

Използването на изкуствения интелект в социалните медии има множество потенциални ползи както за потребителите, така и за самите платформи (Bhavya, n.d.).

2.1.1 Препоръки за персонализирано съдържание – AI алгоритмите се учат от поведението на потребителите и предлагат публикации, страници и групи, които могат да ги заинтересуват.

2.1.2 Подобрено потребителско изживяване – предоставяне на интуитивен интерфейс.

2.1.3 Подобрено насочване на рекламите – AI предлага персонализирани реклами.

2.1.4 Поддръжка на чатботове – предоставят незабавна поддръжка на клиенти, подобрява се времето за реакция.

2.1.5 Улеснения за бизнеса – компаниите познават по-добре целевата си аудитория и разработват по-ефективни маркетингови стратегии.

2.2 Предизвикателства относно AI в социалните медии

Независимо, че приложението на AI в социалните медии е свързано с множество потенциални ползи, не са за пренебрегване опасенията относно въздействието му върху потребителите като цяло и в частност в сферата на сигурността.

2.2.1 Притеснения, свързани с поверителността – потенциалното нарушаване на поверителността на личните данни е предизвикателство, което изисква своевременни национални и общностни законодателни инициативи; въвеждане на етични практики; поемане от страна на компаниите на технологични гаранции за защита и сигурност на данните, които ангажименти могат да бъдат проверими от трети страни.

2.2.2 Разпространение на фалшиви новини – социалните медии могат да бъдат употребени за злоумишлени цели, целящи бързо разпространение на неистинска и невярна информация. Невъзможността за потвърждаване на достоверността на информацията и разпространяваните послания е безпрецедентно съвременно комуникационно предизвикателство, особено в сферата на сигурността.

2.2.3 Дезинформация и deepfakes – генеративния изкуствен интелект може да създаде изключително убедителни дълбоки фалшиви изображения, видео, аудио или реч, които да въздействат много силно върху огромна група потребители или върху предварително селектирана аудитория. Хората могат да бъдат убедени, че виждат или чуват неща, които никога не са се случвали. Това са сред най-смразяващите примери за силата на генеративния AI, който може да бъде употребен със злонамерени намерения. В глобалната мрежа Интернет са достъпни редица публикации за deepfakes. Без претенции за изчерпателност, като примери може да бъдат посочени следните статии:

- *Най-големите фейкове на 2023 година*, в която са посочени само част от абсурдните фейкове, които се разпространяват в социалните мрежи през 2023 година като например, че Джо Байдън носел памперс, а украинският му колега Зеленски танцувал бели денс (Весоловски, 2023);

- *Българският президент стана жертва на фалшив видеоклип*, в която се съобщава, че фалшив видеоклип с образа и гласа на българския президент е свален от интернет пространството (БНР, 2023);

- *Elon Musk made a Kamala Harris deepfake ad go viral, sparking a debate about parody and free speech*, която е посветена на пародийната реклама на кампанията на Камала Харис за президент, използваща фалшив глас на Камала Харис, публикувана повторно от Илон Мъск, и която се превърна в точка на възпламеняване в ескалиращия дебат за това как да се третират манипулираните медии – често наричани дълбоки фалшиви съобщения (Tenbarge, 2024).

Тези и още много други примери извеждат на преден план извода, че обект на злонамерени атаки чрез генеративен AI може да бъде всеки човек, държава или международна организация. Подкопаването на утвърдените правила за сътрудничество в сферата на сигурността може да се осъществи чрез целенасочена комуникационна кампания в социалните медии, която да прерасне в хибридна дезинформационна война (Чаушев, Стоянова, и Йорданова, 2024, с. 38). Необходимо е да се подчертае, че генеративни технологии за манипулиране на видеосъдържание, известни като „дълбинни фалшификати“; използване на ботове за разпространение на разединяващо съдържание и провокиране на напрежение; кражба на данни; създаване на фалшиви профили и т.н., непрекъснато еволюират и се усъвършенстват. Осъзнавайки предизвикателствата, породени от генеративния AI, Организацията на Северноатлантическия договор (НАТО) обръща специално внимание на изкуствения интелект и съпътстващата промяна в глобалната среда за отбрана и сигурност. Нещо повече, НАТО разработва собствена стратегия за изкуствения интелект, която

трябва да отговори на съвременните изисквания и същевременно да укрепи сътрудничеството, основано на взаимно доверие и подкрепа (NATO, 2021).

2.2.4 Дискриминация – AI алгоритмите се базират на данни, върху които се обучават. Ако данните са манипулирани, самият алгоритъм ще бъде нарушен и ще създава предубеден модел на комуникация. Това може да провокира нарастване на дискриминационните нагласи, рефлектиращи върху сферата на сигурността.

2.2.5 Психологическа устойчивост – пристрастяването към социалните медии буди нарастващо безпокойство за психологическата устойчивост на потребителите. Генерираното съдържание може да ангажира до такава степен, че проявите на нервна превъзбуда да придобият непредвидими мащаби, при които потребителите реагират, без да проявяват критично мислене.

3. ПРЕДИЗВИКАТЕЛСТВА ПРЕД ИЗПОЛЗВАНЕТО НА AI В СИГУРНОСТТА

Навлизането на изкуствения интелект в сферата на сигурността е огромно предизвикателство. Ключов момент е постигането на гаранции, че AI ще работи за хората, тъй като всяка промяна в средата за сигурност може да предизвика непредвидими последствия.

Основните предизвикателства, свързани с изкуствения интелект в сигурността, са следните (Mircheska, 2024):

3.1 Етични въпроси – използването на AI в сигурността е свързано с редица етични въпроси, най-важният сред които е възможността за взимане на решения между живота и смъртта, без човешка оторизация. В сферата на сигурността крайното решение трябва да бъде на хората!

3.2 Надеждност – данните, върху които се обучава изкуственият интелект трябва да бъдат надеждни и проверими във всеки един момент. При необходимост да съществува протокол за тяхната промяна, гарантиращ надеждност на базовата информация, върху която се обучава AI.

3.3 Киберсигурност – защитата от хакерски атаки е основно задължение. Необходимо е гарантиране на мрежова и информационна сигурност, противодействие на киберпрестъпността и киберотбрана, които са ключови компоненти на киберсигурността.

3.4 Обществено възприемане – хората не крият безпокойствието си от използването на изкуствен интелект в сигурността, смятайки, че рискът за населението ще се увеличи.

4. РЕГУЛАТОРНИ ПРЕДИЗВИКАТЕЛСТВА

Възможността за регулиране на генеративния изкуствен интелект в социалните медии е сложна и отговорна задача. Постигането на

баланс между процесите на въвеждане на иновации и прилагане на система за регулации, е ключов момент, който трябва да предостави както гаранция за сигурността на информацията, така и да стимулира по-нататъшните разработки, свързани с генеративния AI. Създаването на експертен капацитет за навременно разпознаване на фалшиви новини, дезинформация, провокиращо съдържание и т.н., както и утвърждаване на протокол за последващото им противодействие, е интердисциплинарна задача, която трябва да бъде решавана и на ниво международни организации.

В контекста на сигурността, предизвикателствата, които произтичат от използването на генеративен AI в социалните медии, ще се увеличават с разширяване на приложното поле на изкуствения интелект. Необходимо е технологичните модели занапред да бъдат проектирани, развивани и прилагани и при необходимост коригирани, с ясното съзнание за сериозното негативно въздействие в сферата на сигурност. Към настоящия момент разработките и възприемането на изкуствения интелект се фокусират върху предимствата и улесненията, които се предоставят, а се пренебрегват случаите, в които може да се използва спрямо уязвими индивиди и общества (Велков, 2023).

Всяка държава може да предприеме законодателни инициативи, които да гарантират усъвършенстване процеса на мониторинг върху съдържанието в социалните медии, които да гарантират неразпространение на публикации, създаващи предпоставки за нарушаване на сигурността, но при строго спазване на законодателството. Съхраняване на данни за трафик и местонахождение, IP адреси, данни за самоличността на създателите на съдържание в социалните мрежи, са сред възможните действия, които обаче трябва да се извършват след разрешение на компетентните органи (Стоичков, 2023). Преди да се предприемат действия е необходимо да се извърши специфична оценка на риска за сигурността, който по своята същност представлява многоаспектен анализ на конкретна заплаха (Милушев, 2022). Изкуствения интелект може да подпомогне дейностите по организиране и класифициране на информацията за целите на анализа, но ролята му трябва да бъде второстепенна. Взимането на решение в сигурността е процес, финалът на който трябва да бъде изцяло в прерогативите на експертният ръководител.

ЗАКЛЮЧЕНИЕ

Ролята на изкуствения интелект в социалните медии непрекъснато се увеличава и има потенциала да революционизира начина, по който взаимодействаме с тези платформи. Използването на генеративен AI поставя нови изисквания за критичност по отношение автентичността на създаденото съдържание. Тази критичност изисква разработване и

прилагане на адекватни механизми за организация и управление на съдържанието, което важи с особена сила за сферата на сигурността. Свидетели сме на високотехнологичен бум, при който досега действащите разбирания за ролята и мястото на изкуствения интелект в социалните медии изглеждат като анахронизъм. Трябва да се активизира процесът на формиране на необходимия публичен дискурс, да се мобилизира научния капацитет, да се повиши осведомеността за рисковете, свързани със сферата на сигурността. Налице е необходимост от решаване на редица задачи, които да гарантират, че генерираното от изкуствения интелект съдържание в социалните медии няма да се превърне в заплаха в сигурността. Налага се извода, че въпросите за генеративния AI, неговата роля в социалните мрежи и възможните комуникационни предизвикателства в сферата на сигурността трябва да заемат полагащото им се място в научния дебат и да се превърнат в обект на самостоятелни и целенасочени научни изследвания, които да отговорят на съвременните предизвикателства, поставени от тази технология.

ЛИТЕРАТУРА:

БНР. (2023, декември 09). *Българският президент стана жертва на фалшив видеоклип*. Извлечено октомври 12, 2024 г., от https://bnr.bg/radiobulgaria/post/101919866/balgarskiat-prezident-stana-jertva-na-falshiv-videoklip

Велков, С. (Декември 2023). Методи за контрол на изкуствения интелект. *Сигурност и отбрана,* (2), 199-211. https://institute.nvu.bg/sites/default/files/inline-files/2023-2-15-velkov.pdf // Velkov, Sv. (Dekemvri 2023). Artificial Intelligence Control Methods. *Sigurnost i otbrana,* (2), 199-211. https://institute.nvu.bg/sites/default/files/inline-files/2023-2-15-velkov.pdf

Весоловски, К. (2023, декември 31). Най-големите фейкове на 2023 година. Deutsche Welle (DW). Извлечено октомври 12, 2024 г., от https://www.dw.com/bg/najgolemite-fejkove-na-2023-godina/a-67831665

Милушев, Л. (Декември 2022). Специфики при изграждане на система за сигурност в обекти с масово пребиваване на хора. *Сигурност и отбрана,* (2), 166-184. https://institute.nvu.bg/sites/default/files/inline-files/2022-2-11-milushev.pdf

Стоичков, О. (Декември 2023). Контрол на електронната комуникация за защита на националната сигурност. *Сигурност и отбрана,* (2), 247-260. https://institute.nvu.bg/sites/default/files/inline-files/2023-2-18-stoichkov.pdf

Чаушев, Х., Стоянова, Д., Йорданова, С. (2024). Стратегическите комуникации като ефективен фактор за противодействие на хибридните войни и съхраняване на националната идентичност. Академично издателство „За буквите – О писменехъ“

Bhavya M. (n.d.). *AI in Social Media: The Benefits and Risks*. InspiritAI. Retrieved October 10, 2024, from https://www.inspiritai.com/blogs/ai-student-blog/ai-in-social-media

Dixon, S. J. (2024, July 10). *Most popular social networks worldwide as of April 2024, by number of monthly active users*. Statista. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Mircheska, S. (June 2024). The Impact of Artificial Intelligence on the Defense Sector. *Security and Defense,* (1), 62-74. https://institute.nvu.bg/sites/default/files/inline-files/2024-1-04-mircheska.pdf

NATO. (2021, October 22). *Summary of the NATO Artificial Intelligence Strategy*. Retrieved October 12, 2024, from https://www.nato.int/cps/en/natohq/official_texts_187617.htm#top

Stryker, C., & Scapicchio, M. (2024, March 22). *What is generative AI?* IBM. https://www.ibm.com/topics/generative-ai

Tenbarge, K. (2024, August 1). *Elon Musk made a Kamala Harris deepfake ad go viral, sparking a debate about parody and free speech*. NBCNews. Retrieved October 12, 2024, from https://www.nbcnews.com/tech/misinformation/kamala-harris-deepfake-shared-musk-sparks-free-speech-debate-rcna164119