Тања МИЛОШЕВСКА

УДК: 343.341:004.8(100)
Прегледен труд

ПОТЕНЦИЈАЛНА ЕКСПЛОАТАЦИЈА НА ГЕНЕРАТИВНА ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА ОД ТЕРОРИСТИ И НАСИЛНИ ЕКСТРЕМИСТИ

Кратка содржина:

Овој труд ќе анализира како и до кој степен терористите и насилните екстремисти досега комуницирале преку генеративна вештачка интелигенција и ќе ги идентификува потенцијалните начини на кои би можеле да ја злоупотребат генеративната вештачка интелигенција во иднина. Затоа е неопходно да се преиспитаат претпоставките дека терористичките и насилните екстремистички актери брзо ќе ја усвојат генеративната вештачка интелигенција само врз основа на проценката на способностите што таа може да ги понуди. Иако истражувањата покажаа дека технолошката способност и достапноста се клучни двигатели на терористичките иновации, терористичките и насилните екстремистички актери исто така ја проценуваат секоја нова технологија за нејзината компатибилност (и со нивниот начин на работа и со идеологија), релативната сложеност, трошоците и контекстот во кој тие работат. Разбирањето на овие фактори, и изолирано и во однос еден со друг, ќе биде клучен аспект во следењето на степенот до кој терористите и насилните екстремисти усвојуваат генеративни алатки за вештачка интелигенција. Притоа, трудот ќе се обиде да ја идентификува веројатната траекторија за злоупотреба на оваа технологија од страна на терористичките актери, како и да понуди рефлексија со некои првични препораки за креаторите на политиките.

Клучни зборови: *генеративна вештачка интелигенција, терористи, насилни екстремисти, пропаганда.*

Вовед

Со доаѓањето и брзото усвојување на софистицирани модели за длабоко учење како што е ChatGPT, постои зголемена загриженост дека терористите и насилните екстремисти би можеле да ги користат овие алатки за да ги подобрат своите операции на интернет и во реалниот свет. Големите јазични модели имаат потенцијал да им овозможат на терористите да учат, планираат и да ги пропагираат нивните активности со поголема ефикасност, точност и влијание од кога било досега. Како таква, постои значителна потреба да се истражат безбедносните импликации на овие модели за длабоко учење. Наодите од ова истражување ќе се насочат како составен дел на развојот на ефективни контрамерки за спречување и откривање на злоупотреба на овие платформи од страна на терористи и насилни екстремисти. Поточно, се анализираат потенцијалните импликации на командите што може да се внесат во овие системи кои ефективно го „нарушуваат" моделот, овозможувајќи му да отстрани многу од неговите стандарди и политики кои го спречуваат основниот модел да обезбедува екстремистичка, нелегална или неетичка содржина. Користејќи повеќе сметки, трудот ги истражуваше различните начини на кои екстремистите потенцијално би можеле да користат различни големи јазични модели за да ги поддржат нивните напори во обуката, спроведувањето на оперативно планирање и развивањето на пропаганда. Трудот ги разгледува потенцијалните импликации и предлага препораки за креаторите на политиките за решавање на овие прашања.

Терминологија

Генеративната вештачка интелигенција (GenAI) е тип на вештачка интелигенција (ВИ) која може да создаде широк спектар на податоци, како што се слики, видеа, аудио, текст и 3Д модели" (generativeai.net). Тоа го прави со учење на обрасци од постоечките податоци, а потоа го користи ова знаење за да генерира нови и уникатни резултати: „GenAI може да произведе високо реална и сложена содржина што ја имитира човечката креативност, што ја прави вредна алатка за многу индустрии како што се игри, забава и дизајн на производ" (Goaltide, 2023).

Индустријата GenAI се развива брзо, а моделите за основање (како што се моделите на опсежни јазични модели или LLM) се усвојуваат во речиси сите индустрии. Генерирањето текст вклучува користење на генеративни модели за учење со вештачка интелигенција за генерирање на нов текст врз основа на обрасци научени од постоечките текстуални податоци. Една од овие нови апликации е ChatGPT кој претставува четбот за генерирање текст, развиен од OpenAI и објавен во ноември 2022 година.

Оваа извонредна апликација може да се користи и за злонамерни цели, на пример, од терористи и насилни екстремисти.

Метаверзумот претставува слевање на физичкиот и на виртуелниот свет во дигиталната сфера, користејќи 3D технологии и уреди за онлајн комуникација како компјутери и паметни телефони, овозможувајќи им на луѓето да имаат интеракции и искуства во реално време на долги растојанија.

Фразата **насилен екстремизам** се користи во контексти кога екстремистичките погледи на светот се придружени со оправдување и употреба на екстремно насилство (како што се злосторствата) против оние кои не го делат истото верување или идеологија. Насилниот екстремизам може да се изрази од поединци или од групи преку говори или објави во медиумите, со извршување на изолирани акти на насилство во име на екстремистички идеологии или со физичко приклучување кон насилни групи (Aroua, 2018).

Дефинициите за **тероризам** варираат во различни национални јурисдикции и не постои универзално прифатена дефиниција за тероризам. За целите на овој труд, избрано е да се користи скратената верзија на дефиницијата од академски консензус на Шмид (2011), каде тероризмот е дефиниран како:

„1. Тероризмот се однесува, од една страна, на доктрина за претпоставената ефективност на посебна форма или тактика на генерирање страв, принудно политичко насилство и, од друга страна, на конспиративна практика на пресметано, демонстративно, директно насилно дејствување без законски или морални ограничувања, насочени главно кон цивили и неборци, извршени поради неговите пропагандистички и психолошки ефекти врз различна публика и конфликтни страни;

2. Тероризмот како тактика се користи во три главни контексти:

(i) нелегална државна репресија;

(ii) пропагандистичка агитација од недржавни актери во време на мир или надвор од зоните на конфликт; и

(iii) како недозволена тактика на нередовно војување употребена од државни и недржавни актери".

Потенцијална злоупотреба на генеративни платформи за вештачка интелигенција

Потенцијалната експлоатација на генеративната вештачка интелигенција од страна на терористите и насилните екстремисти привлече дел од предупредувања, вклучително и дека чет-ботите може да се користат за подведување или радикализирање на младите луѓе (Taher, 2023) или за зголемување на ризикот од биотероризам (Sandbrink, 2023).

Веќе во 2020 година, Kris McGuffie и Alex Newhouse (2020) го истакнаа потенцијалот за злоупотреба на генеративните јазични модели со оценување на GPT-3. Експериментирајќи со инструкции коишто претставуваат различни видови екстремистички содржини, тие откриja значителен

ризик за радикализација и регрутирање на интернет од големи размери. Во април 2023 година, Лабораторијата за иновации на ЕУРОПОЛ издаде извештај во кој се претставени некои од начините на кои LLM како што е ChatGPT може да се користат за извршување или за олеснување на криминалот, вклучувајќи имитирање, напади од социјален инженеринг и производство на злонамерен код што може да се користи во компјутерски криминал. LLM како ChatGPT имаат потенцијал значително да ги подобрат перформансите на чет-ботови, оддалечувајќи ги од однапред напишаните одговори засновани на правила и зголемувајќи ги нивните квалитети слични на луѓето. Ова доведе до дискусија во врска со потенцијалот за создавање на „терористички GPT", приспособен чет-бот кој може да ги охрабри поединците на патот кон радикализација.

Владините тела, исто така, изразија загриженост за потенцијалните злоупотреби на генеративните платформи за вештачка интелигенција, при што во извештајот на австралискиот комесар за безбедност на интернет кој беше објавен во август 2023 година, беа забележани многуте начини на кои терористите или другите насилни екстремисти би можеле да ја искористат оваа технологија (eSafety Commissioner, 2023). Во тој извештај, тие изразија загриженост дека терористите „потенцијално би можеле да ги користат овие модели за финансирање тероризам и за извршување измами и сајбер криминал;" дополнително, овие модели би можеле да им овозможат на „екстремистите да создаваат насочена пропаганда, да радикализираат и да таргетираат одредени поединци за регрутирање и да поттикнуваат насилство" (eSafety Commissioner, 2023).

На почетокот на ноември 2023 година, извештајот на Техниката против тероризмот, иницијатива поддржана од Извршниот директорат на Комитетот за борба против тероризмот на Обединетите нации, заклучи дека сè уште има релативно малку докази за генеративната вештачка интелигенција која е систематски искористена од терористи и насилни екстремисти, дефинирајќи дека нивниот ангажман со технологијата е „во нејзината експериментална фаза". И покрај прилично малиот сет на податоци, примерите во извештајот се илустративни за најочигледната употреба на генеративната вештачка интелигенција за терористите и насилните екстремисти - производството на пропаганда. Тие вклучуваат постери генерирани од вештачката интелигенција произведени од медиумски ентитет усогласен со Ал Каеда, слики и меминја генерирани од вештачка интелигенција на екстремно десничарски канал на Телеграм, и транскрипција на пропагандна порака на Исламската држава од арапски говор на арапски, индонезиски и англиски текст од поддржувач на ИСИС.

Терористите и насилните екстремисти се покажаа како неверојатно приспособливи во користењето на онлајн платформите за да ги унапредат своите цели (Weimann, 2005, 2015). Од појавата на екстремистичките веб-страници во доцните 1990-ти, до новите платформи за социјални медиуми како што се Фејсбук, Јутјуб, Твитер, Инстаграм и ТикТок, овие групи

брзо ги усвоија и ги искористија новите случувања во сајбер просторот. Во поново време, тие исто така почнаа да прифаќаат шифрирани апликации за пораки, како што се Telegram, TikTok и TamTam. Тие користат анонимни платформи за складирање облак, па дури и Dark Net, нагласувајќи ги нивните континуирани обиди да ги искористат најновите достигнувања и еволуции во дигиталниот свет. „Од своја страна, многу терористи го променија начинот на дејствување, усвојувајќи ги овие нови технологии и спроведувајќи различни оперативни безбедносни мерки дизајнирани да ги избегнат или поразат софистицираните операции за собирање на разузнавачки информации“ (Wagner, 2007). За терористите, овие технологии нудат можност за комуникација и координирање на операциите низ целиот свет со разумни очекувања за приватност и безбедност. Вештачката интелигенција можеше да ги искористи новите технологии за поединци и групи, правејќи ја заканата од сајбер напади и шпионажа присутна од кога било досега (Esmailzadeh, 2023). Таа има потенцијал да биде и алатка и закана во контекст на терористички и екстремистички групи.

Поимот за вештачка интелигенција и тероризам најмногу се фокусираше на потенцијалните употреби на вештачката интелигенција за контратероризам или спротивставување на насилен екстремизам (McKendrick, 2019). Во 2021 година, Канцеларијата за борба против тероризмот на Обединетите нации (2021) објави специјален извештај во кој се разгледуваат можностите понудени од вештачката интелигенција за борба против тероризмот на интернет. Навистина, неколку студии се фокусираа на употребата на вештачка интелигенција во контратероризмот (Verhelst, at all, 2020). Сепак, малку внимание е посветено на истражување на другата страна: како терористите и насилните екстремисти можат да ги користат технологиите базирани на вештачка интелигенција за да шират омраза, пропаганда и да влијаат на ранливите поединци кон нивните идеологии. Неодамна, Глобалниот интернет-форум за борба против тероризмот објави извештај за заканите од екстремистичката/терористичката употреба на GenAI (GIFCT Red Team Working Group, 2023). Потенцијалните употреби на вештачката интелигенција од страна на екстремистичките групи вклучуваат:

- ✓ Пропаганда: Вештачката интелегенција може да се користи за генерирање и дистрибуција на пропагандна содржина побрзо и поефикасно од кога било досега. Ова може да се користи за регрутирање или за ширење говор на омраза и радикални идеологии. Ботови со вештачка интелигенција, исто така, можат да ја засилат оваа содржина, што го отежнува откривањето и реагирањето.
- ✓ Интерактивно регрутирање: чет-ботови напојувани со вештачка интелигенција можат да комуницираат со потенцијалните регрути обезбедувајќи им приспособени информации засновани на нивните

интереси и верувања, со што пораките на екстремистичката група ќе изгледаат порелевантни за нив.

- ✓ Автоматски напади: Терористите можат да користат вештачка интелигенција за поефикасно и поефективно извршување на напади - на пример, со користење дронови или други автономни возила.
- ✓ Експлоатација на социјалните медиуми: Вештачката интелегенција може да се користи и за манипулирање со социјалните медиуми и со други дигитални платформи за ширење пропаганда и за регрутирање следбеници.
- ✓ Сајбер напади: Вештачката интелигенција може да ја користат екстремистичките групи за да ја подобрат нивната способност да лансираат сајбер напади врз цели, потенцијално предизвикувајќи значителна штета.

Индоктринација и регрутирање

Регрутирањето и ангажирањето преку интернет се заштитен знак на современиот екстремизам. Метаверзумот ризикува да ја прошири оваа способност така што ќе им олесни на поединците да се социјализираат и да се поврзуваат (Elson at all, 2022).

Голем дел од досегашната анализа се фокусираше на тоа како генеративната вештачка интелигенција може да помогне во создавањето и ширењето на терористичка и насилна екстремистичка пропаганда. Најзначајно, генеративната вештачка интелигенција овозможува создавање нови слики или адаптација на постоечките на размер и со брзина што претходно не беше возможна. Слично на тоа, актерите сега можат да користат такви алатки за да генерираат синтетичко видео и аудио, вклучувајќи монтирани фалсификати на познати или на значајни поединци. Иако доверливоста и квалитетот на видеопродукцијата вообичаено се неконзистентни, лансирањето на OpenAI's Sora во февруари 2024 година, кое може да генерира видеа врз основа на текстуални инструкции, укажува на големата брзина со која се развива оваа технологија (Montgomery, 2024).

Конечно, различни LLM кои постојано се подобруваат можат да креираат текст користејќи различни стилови, формати и, што е најважно, јазици. Претходно, терористичките групи мораа да се потпираат на мануелни и често релативно слаби преводи на пропаганден материјал, а овој процес во голема мера се потпираше на вештините на неколку поединци. Генеративната вештачка интелигенција теоретски може да се користи за создавање и транскрипција на видео и на аудиопропаганда, или генерирање на пропаганда базирана на текст, речиси моментално и на повеќе јазици.

Во комбинација, овие случувања создаваат потенцијал за зголемување на обемот и квалитетот на терористичкиот или насилен екстремистички пропаганден материјал.

Во моментов, технолошките компании можат да го споделат „дигиталниот отпечаток од прст“ или „хаш“ на терористичката содржина меѓу себе, овозможувајќи нејзино навремено отстранување и/или спречување да се постави на изворот (GIFCT's Hash-Sharing Database, 2023). Употребата на генеративна вештачка интелигенција за манипулирање со слики може да го промени овој дигитален хаш без суштинска промена на датотеката, ефикасно „уништувајќи го споделувањето хаш како решение“ (Gilbert, 2023). Иако главните платформи можат да идентификуваат и отстранат терористичка содржина на други начини - вклучително и употреба на обработка на природен јазик за да се идентификуваат нови содржини што се слични, но не и идентични со постоечката терористичка содржина (United Nations Office of Counter-Terrorism and United Nations Interregional Crime and Justice Research Institute, 2021) - споделувањето хаш е суштинско во напорите на меѓуплатформските напори за спротивставување на терористичката содржина од 2016 година (Meta, 2016). Нејзиното потенцијално деградирање како решение е опишано како „голем ризик“ (Gilbert, 2023).

Важно е да се нагласи дека создавањето терористичка содржина е само првиот дел од процесот. Терористичките актери, исто така, треба да најдат начин со сигурност да складираат и да споделуваат содржина на интернет. Благодарение на комбинација од регулатива, иновативни акции и јавно-приватно партнерство, ова во моментов е тешко да се направи на повеќето големи платформи, со терористички актери наместо да се потпираат на помали, помалку регулирани опции (Wells, 2022). Сепак, клучно е што генеративната вештачка интелигенција им нуди на терористичките актери потенцијална способност да го оптимизираат нивното избегнување на главните контрамерки на платформата, особено употребата на т.н. hash-sharing.

Како може метаверзумот да биде искористен од насилни екстремисти и насилни екстремистички организации?

Метаверзумот може да стане нова територија за терористичка активност, перспективна платформа за подобрување и унапредување на нивните онлајн активности, вклучувајќи радикализација, регрутирање, обука, собирање средства и координација на нападите (Debuire, 2022).

Комбинирањето на вештачката интелигенција со зголемената реалност во метаверзумот ќе им овозможи на екстремистичките лидери да се состанат и да се сретнат со своите поддржувачи, да развиваат и одржуваат виртуелни идеалистички општества и да ги зголемат нивните сфери на влијание. Поради екстремната емоционална средина

овозможена од метаверзумот, можеби е предизвик за некои поединци да направат разлика помеѓу реалниот живот и виртуелната реалност (Council of the European Union, 2022). Некои корисници можеби сметаат дека она што се случува во метаверзумот не е фактичко дури и ако има реални последици по нивните животи. Со спојување на вештачката интелигенција и проширената реалност во метаверзумот, онлајн регрутерите за терористички или насилни екстремистички групи ќе можат да се сретнат во виртуелна просторија со потенцијалните следбеници и да ги примамат со визии за иднината.

Прво, метаверзумот ќе им обезбеди на екстремистите уникатна средина за регрутирање нови членови бидејќи тие можат да создадат свои приватни сервери, да допрат до поширока публика и да симулираат мрежи за регрутирање лично. Овие сервери ќе бидат тешки до умерени, а сегашните регулаторни упатства и технолошки механизми не го спречуваат доволно производството на малициозна содржина и однесувања. Како резултат на тоа, метаверзумот ќе послужи како идеален инкубатор за регрутирање на екстремисти.

Второ, метаверзумот ќе го поедностави собирањето средства и сајбер криминалот за насилните екстремисти. Преку употреба на криптовалути, компјутерски криминал и перење пари, метаверзумот ќе обезбеди дополнителни слоеви на анонимност и ќе спречи потенцијално откривање од органите за спроведување на законот и од финансиските институции.

Трето, метаверзумот им овозможува на екстремистите способност да го направат поголемиот дел од нивното планирање пред нападот и виртуелното собирање на разузнавачки информации. Со ограничување на бројот на пати кога мора да посетат потенцијална цел и минимизирање на времето поминато на интернет за спроведување на разузнавачки информации со отворен код, насилните екстремисти ќе можат да дејствуваат дискретно, а сепак да одржуваат високо ниво на ефикасност. Дополнително, доколку добијат пристап до серверите што ги користат органите за спроведување на законот за да спроведуваат сопствени виртуелни обуки, насилните екстремисти ќе имаат можност да развијат поефикасни планови за напади и вонредни ситуации.

Конечно, како што технологијата на метаверзумот станува пореална, екстремистите можат да вежбаат создавање и ракување со експлозиви, оружје и спроведување симулации на напади со различни потенцијални нарушувања или промени и да воспостават методи на најдобра практика за да ги постигнат своите клучни цели (Levin, 2024).

Виртуелен тренинг - планирање пред напад

Еден аспект за кој насилниот екстремист може да го искористи метаверзумот е проширување на нивните способности за планирање на напад. Според традиционалното разбирање на активностите пред нападот,

поединците кои планираат да извршат акти на насилен екстремизам развиваат разновидни различни однесувања за да ги планираат своите напади. Во извештајот објавен од Канцеларијата на директорот на заедничкиот тим за антитеристичка проценка на националното разузнавање, нивните истражувачи идентификуваа циклус на планирање напади кои се повторуваат. Според извештајот, насилните екстремисти генерално планираат напади во фази што може да се набљудуваат, иако конкретните детали, редоследот и времето може многу да се разликуваат и да се менуваат со текот на времето. Покрај тоа, тие забележаа дека одредени активности полесно се забележуваат од другите. Надзорот пред нападот, обуката и пробите често се набљудуваат и можат да понудат можности за идентификување на заговори и спречување напади (ODNI, 2020). Сепак, една голема грижа за метаверзумот е тоа што може да ја отстрани потребата насилните екстремисти лично да го спроведат најголемиот дел од нивното собирање на разузнавачки информации пред нападот. На пример, доколку насилните екстремисти станат способни да ја комбинираат метаверзната технологија со други модерни технологии како видеонадзор, лични камери, објави на социјалните мрежи и други можности за пренос во живо, тие можеби никогаш нема да мораат да излезат во јавноста за да соберат разузнавачки информации што им се потребни за извршување на нападите на одредени цели или локации.

Друга област на загриженост во врска со планирањето на напад од насилните екстремистички групи е нивната способност да користат метаверзни игри и алатки за обука кои се веќе достапни за спроведување на законот и вооружените сили за да се спротивстават на потенцијалните проблеми за време на нивните напади.

Пример за алатките кои се веќе достапни и би можеле да се менуваат за незаконски цели е AUGGMED (Автоматски генератор на сценарија за сериозни игри за обука за мешана реалност). Целта на AUGGMED беше да развие сериозна платформа за игри за да овозможи обука на крајните корисници базирана на еден тим со различни нивоа на експертиза од различни организации кои реагираат на закани од тероризам и организиран криминал. Платформата автоматски генерира нелинеарни сценарија приспособени да одговараат на потребите на индивидуалните учесници со резултати од учењето кои го подобруваат стекнувањето на емоционално управување, аналитичко размислување, решавање проблеми и вештини за донесување одлуки. Сценаријата на игра вклучуваат напредни симулации на оперативни средини, агенти, телекомуникациски закани и може да се испорачаат преку виртуелна реалност и средини со мешана реалност преку мултимодални интерфејси. Покрај тоа, платформата AUGGMED ќе вклучува алатки за обучувачите кои ќе им овозможат да постават цели за учење, да дефинираат сценарија, да ги следат сесиите за обука, да ги менуваат сценаријата и да даваат повратни информации во реално време, како и да ја оценуваат работата на практикантите и да поставуваат наставни

програми за обука за поединечен персонал во фаза по сесијата за обука (European Commission, 2022). Дополнително, штом овие виртуелни сценарија беа креирани и функционираат без никакви технички проблеми, трошоците за реплицирање и споделување на софтверот со други агенции се минимални. Ова овозможува повторна употреба и повторување по ниска цена (Herath & Jarnecki, 2022).

Екстремистите или се приклучуваат на службата со цел да добијат борбена и логистичка обука, или се регрутираат откако нивната служба ќе заврши од радикални групи кои се засноваат на комбинацијата од траума, губење на целта и заедницата што често ги погодува ветераните (Ware, 2023). Накратко, пристапот до поранешните органи за спроведување на законот, персоналот на вооружените сили и нивните информации за протоколот за одговор ќе го подигнат успехот на насилното екстремистичко планирање пред нападот во метаверзумот и нивните оперативни способности.

Последната област на загриженост кога станува збор за насилните екстремисти кои го користат метаверзумот за планирање на напад веќе може да се демонстрира со онлајн игрите како што се Minecraft, Roblox и Sandbox. Врз основа на јавно достапни планови и распореди, онлајн информации, слики и други технологии за мапирање, корисниците можеа да создадат плејада на виртуелни рекреации на реалната инфраструктура во рамките на овие платформи. Како резултат на тоа, насилните екстремисти веројатно ќе имаат способност да соберат разузнавачки информации едноставно со создавање, преземање или интеракција со овие виртуелни рекреации за да снимаат белешки и да ги идентификуваат сите точки на ранливост кои би можеле да ги искористат за време на напад. Ова ќе биде особено загрижувачко за владините згради, критичната енергетска инфраструктура, јавните места, универзитетските кампуси, спортските стадиони и сите други локации кои можат да станат потенцијални цели за насилен екстремизам.

Симулација на напад

Последната активност која е алармантна и бара итно внимание е зголемената способност на насилните екстремисти да симулираат напади. Како што споменавме погоре, војската со децении ги користи моќите на симулацијата на виртуелна и проширена реалност. Иако технолошките можности на висококвалитетната метаверзна графика сè уште се во развој, софтверските компании се обидуваат да создадат програми кои се сметаат за толку реални што ќе го прекинат неверувањето дека тоа е симулација и ќе ги измамат нивните умови да мислат дека физички се на друго место. Ова ќе се преведе во софтверски програми кои ќе им овозможат на корисниците да манипулираат со сложеноста на животот, како што се различни фреквенции на цивили, различни модели на инфраструктура и сообраќај, промена на теренот и различни временски услови. Пример

за тоа како ова веќе се манифестира во американските вооружени сили е синтетичката средина за обука, која беше развиена од софтверската компанија Bohemia Interactive Simulations, која создава симулации со висока верност за војниците да тренираат насекаде во светот. Оваа верзија на метаверзумот се нарекува One World Terrain и софтверот комбинира тродимензионални податоци собрани од сателити, сензори или скенери и ги комбинира со дополнителни информации за да направи симулации на терен со висока верност. Понатаму, со дигиталниот свет може да се манипулира со вештачка интелигенција и машинско учење за да се постигнат специфични вежби и резултати за обука (Easley, 2022).

Доколку насилните екстремисти можат да добијат пристап до овие софтверски програми, постои бесконечна количина на веродостојни сценарија во кои тие ќе манипулираат со алгоритмите до точните спецификации на претстојниот напад и ќе ги симулираат своите активности под големи слоеви на анонимност.

Втората клучна компонента на симулацијата на напад во рамките на метаверзумот што претставува зголемена закана за јавната безбедност е нивната способност да вежбаат со виртуелно оружје и експлозиви. Сега се појавува алтернативна опција каде што корисниците би имале пристап до слични нивоа на анонимност, само што сега ќе биде многу поимпресивна од Tor мрежата. Во рамките на метаверзните сервери, организираниот криминал, како што е трговијата со оружје, може да се одвива релативно лесно под тековниот недостаток на технолошко разбирање, регулација или надзор (INTERPOL, 2022). Што се однесува до конкретно тестирање и купување огнено оружје, постојат компании кои веќе имаат изградено контролери за емулирање на огнено оружје кои се спаруваат со слушалки за виртуелна реалност кои се шокантно реални. Со оваа технологија, корисниците на метаверзумот и насилните екстремисти ќе можат да симулираат употреба на кој било број на различни огнени оружја, да ја тестираат нивната ефикасност во голем број симулирани ситуации и да донесат одлука каде преферираат да ја користат во реалниот живот.

Последната активност и критично важна област на загриженост е како метаверзумот ќе им овозможи на насилните екстремисти да ги вежбаат своите напади. Откако ќе го завршат планирањето пред нападот, ќе ги соберат сите потребни материјали, следниот чекор во циклусот на планирање напад е пробата. Насилните екстремисти често го вежбаат сценариото за напад за да ги потврдат претпоставките за планирање, да ги подобрат тактиките и да ги практикуваат патиштата за бегство. Тие, исто така, може да предизвикаат инцидент на целната локација за да ја тестираат реакцијата на безбедносниот персонал и првите лица што реагираат. Додека ја проценуваат ефикасноста на нивните планови пред нападот, насилните екстремисти, исто така, се обидуваат да ги утврдат условите кои фаворизираат највисока стапка на успех и најмал ризик. Факторите што тие вообичаено ги разгледуваат вклучуваат елемент

на изненадување, избор на време и место, употреба на тактики за пренасочување и начини да се попречат мерките за одговор. Освен ако насилниот екстремист не планира самоубиствен напад, патеките за бегство и плановите за вонредни состојби се исто така внимателно планирани (ODNI, 2020). До неодамна овие активности најчесто се извршуваа лично. Сепак, со додавање на метаверзум технологија, насилните екстремисти би можеле да имаат неспоредливи можности за вежбање (Elson et al., 2022). На пример, насилните екстремисти можат да симулираат различни потенцијални нарушувања или симулирани ситуации и да воспостават методи на најдобра практика за сепак да ги постигнат своите клучни цели. Ова ќе им овозможи на насилните екстремисти да ги реплицираат силите за спроведување на законот и цивилните одговори, да научат остварливи и ефикасни патеки, да ги координираат алтернативните рути доколку некои од нив се блокираат и да воспостават повеќе планови за вонредни ситуации доколку дојде до прекини.

Заклучок

Терористичките и насилните екстремистички групи и поединци покажаа дека можат многу добро да се прилагодат и значително да еволуираат во текот на деценииите. Тие покажаа потенцијал за иновација, на пример, во нивната организациска структура, станувајќи децентрализирани, франшизирани и глобални.

Импликациите од овој труд покажуваат дека метаверзумот наликува на Пандорината кутија; технологијата што се појавува е многу посложена отколку што можат да се справат нашите сегашни системи, а последиците би можеле да бидат поразителни. Треба да се следи еволуцијата на усвојувањето на вештачката интелигенција од страна на терористички групи и поединци.

Зголемената соработка помеѓу приватниот и јавниот сектор, меѓу академската заедница, високата технологија и безбедносната заедница, би ја зголемила свеста за потенцијалната злоупотреба на платформите засновани на вештачка интелигенција од страна на насилните екстремисти, поттикнувајќи го развојот на пософистицирана заштита и контрамерки. Треба да се истражи употребата на вештачка интелигенција и поврзаните нови технологии за спротивставување на терористичките закани овозможени со вештачка интелигенција, особено за да се спротивставиме на терористичката радикализација и да се шират позитивни наративи.

БИБЛИОГРАФИЈА:

- All Things Generative AI," generativeai.net, n.d.
- Aroua, A. (2018). *Addressing Extremism and Violence. The Importance of Terminology*. Geneva: The Cordoba Foundation of Geneva. Available at: https://www.cordoue.ch/images/pdf/Papers/CFG_ConflictTransformationPerspective.pdf [Accessed 16 April 2022].
- "Considerations of the Impacts of Generative AI on Online Terrorism and Extremism. (2023). GIFCT Red Team Working Group.
- "GIFCT's Hash-Sharing Database," (2023). Global Internet Forum to Counter Terrorism. Available at: https://gifct.org/hsdb/.[Accessed 10 May 2024].
- ChatGPT. (2023). The impact of Large Language Models on Law Enforcement. The Hague: EUROPOL Innovation Lab.
- Council of the European Union. (2022). "The Metaverse in the Context of the Fight Against Terrorism", Special Report. Available at:https://data.consilium.europa .eu/ doc/document/ST-9292-2022-INIT/en/pdf. [Accessed 3 March 2024].
- Countering Terrorism Online With Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia," United Nations Office of Counter-Terrorism and United Nations Interregional Crime and Justice Research Institute. (2021). Available at: https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf. [Accessed 12 April 2024].
- Debuire, D. (2022) "Terror-ist use of the Metaverse: new opportunities and new challenges", *The Security Distillery*. Available at: https://thesecuritydistillery.org/all-articles/terrorism-and-the-metaverse-new-opportunities-and-new-challenges. [Accessed 5 April 2024].
- Early terrorist experimentation with generative artificial intelligence services. (2023). Tech Against Terrorism. Available at: https://techagainstterrorism.org/hubfs /Tech % 20Against%20Terrorism%20Briefing%20-%20 Early%20terrorist% 20 experi mentation%20with %20generative% 20artificial% 20intelligence%20services.pdf. [Accessed 5 April 2024].
- Easley, M. (2022). "How Interconnected, Simulated Worlds Could Transform Military Training." *NDIA Business & Technology Magazine*. Available at: www.nationaldefensemagazine.org/articles/2022/11/23/how-interconnected-simulated-worlds-could-transform-military-training. [Accessed 5 May 2024].

- Elson, J at all. (2022). "The metaverse offers a future full of potential – for terrorists and extremists", *The Conversation*. Available at: https://theconversation.com/the-metaverse-offers-a-future-full-of-potential-for-terrorists-and-extremists-too-173622. [Accessed 5 April 2024].
- Esmailzadeh, Y. (2023). "Potential Risks of ChatGPT: Implications for Counterterrorism and International Security," *International Journal of Multicultural and Multireligious Understanding* 10:4.
- European Commission. (2022). "Automated Serious Game Scenario Generator for Mixed Reality Training." *CORDIS*. Available at: https://cordis.europa.eu /project/ id /653590. [Accessed 5 April 2024].
- Gilbert, D (2023). "Here's How Violent Extremists Are Exploiting Generative AI Tools," *Wired*. Available at: https://www.wired.com/story/generative-ai-terrorism-content/[Accessed 5 April 2024].
- Herath, C., Jarnecki, J. (2022). "Securing Future Realities: What Can We Expect from the Metaverse?" *RUSI,* Available at:https://rusi.org/explore-our-research/ publications /commentary/securing-future-realities-what-can-we-expect-metaverse. [Accessed 5 April 2024].
- INTERPOL. (2022). "INTERPOL launches first global police Metaverse.". Available at: https://www.interpol.int/en/News-and-Events/News/2022 / INTERPOL-launches-first-global-police-Metaverse. [Accessed 10 May 2024].
- Levin, J. (2024). Violent Extremists in the Metaverse*, Global Affairs Review*, The Center for Global Affairs, New York University. Available at: https:// wp.nyu.edu /schoolofprofessionalstudies-ga_review/violent-extremists-in-the-metaverse/. [Accessed 15 April 2024].
- McGuffie, K., Newhouse, A. (2020). "The Radicalization Risks of GPT-3 and Advanced Neural Language Models," available via Arxiv.
- McKendrick, K. (2019). "Artificial Intelligence Prediction and Counterterrorism," International Security Department, *Royal Institute of International Affairs*.
- Montgomery, B. (2024). "Sora: OpenAI launches tool that instantly creates video from text," *The Guardian*. Available at:https://www.theguardian.com /technology /2024 /feb/15/openai-sora-ai-model-video. [Accessed 19 April 2024].
- Partnering to Help Curb Spread of Terrorist Content," (2016). Meta. Available at: https://about.fb.com/news/2016/12/partnering-to-help-curb-spread-of-online-terrorist-content/.[Accessed 12 April 2024].
- Sandbrink. J. (2023). "ChatGPT could make bioterrorism horrifyingly easy," *Vox.* Available at: https://www.vox.com/future-perfect/23820331/

chatgpt-bioterrorism-bioweapons-artificial-inteligence-openai-terrorism. [Accessed 2 April 2024].

- Schmid, A (Ed.) (2011). *The Routledge Handbook of Terrorism Research*. London and New York: Routledge.
- Taher, A. (2023). "AI chatbots could be 'easily be programmed' to groom young men into launching terror attacks, warns top lawyer," *Daily Mail*. Available at: https://www.dailymail.co.uk/sciencetech/article-11952997/amp/AI-chatbots-easily-programmed-groom-young-men-terror-attacks-warns-lawyer.html?s=03 [Accessed 5 February 2024].
- Tech Trends Position Statement – Generative AI. (2023). eSafety Commissioner.
- US Office of the Director of National Intelligence (ODNI). (2020) "Counter-terrorism Guide for Public Safety Personnel." *JCAT*. Available at: https://www.dni.gov/nctc/jcat/index.html. [Accessed 5 April 2024].
- Verhelst, H.M. at all. (2020). "Machine Learning against Terrorism: How Big Data Collection and Analysis Influences the Privacy-Security Dilemma," *Science and Engineering Ethics* 26.
- Wagner, A. (2007). "Intelligence for Counter-Terrorism: Technology and Methods," *Journal of International Affairs* 2:2 (2007): pp. 48-61.
- Ware, J. (2023). "The Violent Far-Right Terrorist Threat to the U.S. Military." *Council on Foreign Relations*. Available at: https://www.cfr.org/blog/violent-far-right-terrorist-threat-us-military. [Accessed 5 April 2024].
- Weimann, G. (2005). *Terror on the Internet, The New Arena, the New Challenges*, Washington, D.C.: United States Institute of Peace Press.
- Weimann, G. (2015). *Terror in Cyberspace: The Next Generation*, New York: Columbia University Press.
- Wells, D. (2022). "Why outsourcing counter-terrorism online won't work in future," Lowy Institute. Available at: https://www.lowyinstitute.org/the-interpreter/why-outsourcing-counter-terrorism-online-won-t-work-future. [Accessed 5 April 2024].
- What is Generative AI? (2023). Goaltide.