B. PROJEKT

B.1. POPIS PROJEKTU (bez uvedenia akýchkoľvek identifikačných údajov žiadateľa)

1. Názov projektu *Umelá inteligencia ako nástroj na eliminovanie prejavov extrémizmu na internete*
2. Cieľ projektu *(definuje sa cieľ alebo čiastkové ciele, čo sa má projektom dosiahnuť)* Cieľom projektu je zníženie výskytu prejavov extrémizmu na internete prostredníctvom zapojenia širokej verejnosti do zberu nenávistných prejavov (overeným prístupom tvz. gamifikáciou) a vybudovaním neurónovej siete na inteligentné automatické vyhľadávanie nezákonných prejavov na internete.
3. Prioritná oblasť výzvy, na ktorú sa projekt zameriava *(definovanie témy, konkrétnej oblasti)* Projekt sa zameriava na zber a filtrovanie nenávistných prejavov na internete, ktoré môžu byť vyhodnotené ako **trestné činy extrémizmu, napĺňajúc tak prioritu č. 1. Znižovanie kriminality a inej protispoločenskej činnosti.**
4. Stručný popis projektu *(uvedie sa stručný popis súčasného stavu a problém, ktorý sa má projektom vyriešiť)* Podľa eurokomisárky pre spravodlivosť, spotrebiteľov a rodovú rovnosť Věry Jourovej sú sociálne médiá jedným z nástrojov, ktoré „rasisti využívajú na šírenie násilia a nenávisti". Množstvo nenávistných prejavov na internete narastá, zároveň však rastie aj odpoveď občianskej spoločnosti vo zvýšených aktivitách smerom k nahlasovaniu nenávistných prejavov na internete a rýchlosť reakcií sociálnych sietí po preskúmaní platných oznámení so žiadosťou o odstránenie nezákonných nenávistných prejavov za menej ako 24 hodín. Keďže len polícia SR môže získať zákonný dôkaz, ktorý môže byť použitý v trestnom stíhaní, je nevyhnutné, aby sa extrémistický prejav dostal do pozornosti orgánov činných v trestnom konaní pred tým, ako aktívni členovia občianskej spoločnosti nahlásia nezákonný prejav prevádzkovateľovi sociálnej siete. Denne je na YouTube nahratých 4 milióny nového obsahu, na Facebooku sa každú minútu objaví 510 000 komentárov, 293 000 statusov a 136 000 fotiek. Preto na dosiahnutie zníženie extrémistických prejavov na sociálnych sieťach je nevyhnutné použiť to, čo ich šírenie umožňuje, teda moderné technológie. Umelá inteligencia je služba postavená na štatistických metódach. Pôjde o tvz. supervised learning, teda umelej inteligencii dodáme vzory, systém sa ich naučí a potom bude schopný s nimi pracovať a vyhodnocovať rádovo niekoľkonásobne rýchlejšie a efektívnejšie ako človek. Umelá inteligencia tak bude zapojená do riešenia globálneho spoločenského problému. Dáta, extrémistické prejavy budú získané uplatnením princípu gamifikácie. Široká verejnosť bude zapojená do nahlasovania extrémistických prejavov cez hranie hry online. Dáta po ich kategorizácii budú následne učiť neurónovú sieť, ktorá bude sama vyhľadávať na základe vložených údajov, extrémistické prejavy, a vytvárať tak databázu pre orgány činné v trestnom konaní. Projekt tak prispeje k šíreniu povedomia o vytváraní kultúry znášanlivosti, a tým znižovaniu kriminality, ako aj k efektívnemu využitiu neurónovej siete pre potreby orgánov činných v trestnom konaní.
5. Východisková situácia *(stav, štruktúra a dynamika kriminality – údaje Policajného zboru, prípadne iných inštitúcií, z vlastných informačných zdrojov, zistené príčiny a podmienky páchania kriminality a inej protispoločenskej činnosti, opatrenia, alebo projekty, ktoré ste doteraz realizovali, dôvod prečo ste sa projekt rozhodli zrealizovať)*

V dokumente Prezídia Policajného zboru s názvom „Stav vyšetrovania v roku 2016 na úseku extrémizmu v Slovenskej republike", bol identifikovaný presun spôsobu páchania tohto druhu trestnej činnosti na internet a sociálne siete. Ďalej sa uvádza, že trestná činnosť páchaná vo virtuálnom prostredí sa vyznačovala predovšetkým nenávistnými prejavmi, útočnými vyjadreniami, ako aj prejavmi sympatií k rôznym nebezpečným skupinám. Do budúcnosti sa v dokumente explicitne predpokladá naďalej zvyšujúci sa podiel páchania tejto trestnej činnosti v kybernetickom prostredí. Podľa štatistík Policajného zboru Slovenskej republiky, od začiatku roka 2017 do konca septembra 2017 evidovali 202 prípadov extrémizmu, z ktorých bolo v 127 trestných veciach začaté trestné stíhanie. Z celkového počtu trestne stíhaných prípadov extrémizmu (t. j. 127 trestných stíhaní) bolo až 48 skutkov spáchaných prostredníctvom internetu a z nich takmer 90% bolo spáchaných na sociálnej sieti Facebook. Vyššie uvedenými faktami je objektívne opísaný neustále sa zväčšujúci problém s nenávistnými prejavmi na sociálnych sieťach, v slovenskom prostredí najmä na Facebooku.

V súčasnosti existuje IT nástroj, ktorý umožňuje polícii SR monitorovať extrémizmus v prostredí internetu, avšak bez dosahu na skupiny a stránky na sociálnych sieťach, pričom sa odhaduje sa, že sociálnu sieť Facebook navštevuje až 68% slovenskej populácie (GfK, 2016). Preto je nevyhnutné jednak zapojiť návštevníkov sociálnych sietí do boja proti nenávistným prejavom online, ktoré napĺňajú znaky trestného činu extrémizmu, ako aj vystavať IT nástroj, ktorý obsiahne práve sociálne siete. Zároveň je nevyhnutné brať do úvahy fakt, že až 90% dát na internete bolo vytvorených od roku 2016. Ak dnes sa odhaduje množstvo dát vo výške 2,5 kvintiliónov bajtov, je zrejmé, že extrémizmus na internete nie je možné postihnúť štandardným spôsobom. Je nevyhnutné zapojiť umelú inteligenciu, ktorá dokáže porozumieť prirodzenému jazyku a rýchlo a jednoducho vyhľadať relevantný obsah v množstve dát. Podľa odborníkov zmení nástup inteligentných počítačových systémov najmä prístup k verejnej bezpečnosti. Keďže už dnes existujú technológie schopné spracovávať obrovské množstvo dát v reálnom čase a zároveň techniky učenia strojov, umelá inteligencia bude môcť hľadať súvislosti a prepojenia, ktoré by človek nebol schopný nájsť.

Monitorovanie a nahlasovanie nezákonných prejavov na internete je jednou zo základných činností žiadateľa. Od svojho vzniku v roku 2017 sa venuje tejto aktivite, vďaka ktorej sa ako **jediná slovenská organizácia zapojila do 3. monitorovacieho cvičenia**, ktoré bolo realizované v rámci spoločných aktivít **Európskej komisie a spoločností z oblasti informačných technológií** v rámci kódexu správania týkajúceho sa nezákonných nenávistných prejavov na internete (máj 2016). Monitorovacie cvičenie slúži ako kontrola kvality a rýchlosti odstraňovania nahlásených nenávistných prejavov sociálnymi sieťami, ktorú vykonávajú dôveryhodní nahlasovatelia. Tým je žiadateľ pre spoločnosti Facebook a YouTube. Napriek nízkemu počtu užívateľov siete Twitter na Slovensku (3% denne) pokrýva aj Twitter. Je možné skonštatovať, že prístup sociálnych sietí k odstraňovaniu nahlásených prípadov sa markantne zlepšil. Kľúčovým však zostáva vyhľadanie extrémistického prejavu, ktoré by sa malo udiať v momente, kedy je publikovaný na internete. Práve zo skúseností a každodennej praxe vychádza poznatok, že proti prejavom extrémizmu v internetovom prostredí nie je možné postupovať štandardným spôsobom, preto sa žiadateľ rozhodol zapojiť umelú inteligenciu ako nástroj 21. storočia.

6. Nadväznosť cieľov projektu na prioritnú oblasť

Svojím cieľom **zníženie výskytu prejavov extrémizmu na internete** prostredníctvom zapojenia širokej verejnosti do zberu nenávistných prejavov (overeným prístupom tvz. gamifikáciou) a vybudovaním neurónovej siete, ktorá bude následne sama vyhľadávať nezákonné prejavy na internete projekt prispieva k plneniu prioritnej oblasti **Znižovanie**

kriminality a inej protispoločenskej činnosti.

7. Cieľová skupina *(konkrétna cieľová skupina, počet osôb, pre ktorú bude projekt realizovaný, uvedie sa spôsob zapojenia cieľovej skupiny, vzdelanie cieľovej skupiny, vek, pohlavie, sociálny pôvod napr. bezdomovec, týraná žena, popíšte aj zapojenie širokej verejnosti do projektu)*

Cieľovou skupinou sú užívatelia sociálnych sietí, najmä tí do 35 veku rokov života. Prieskumy ukazujú, že 86% mladých do 24 rokov navštevuje Facebook niekoľkokrát denne, YouTube 59%. Priemerný vek hráča hier online v roku 2017 sa zvýšil na 35 rokov. Štatistiky ukazujú, že hrá viac mužov ako žien (70% v porovnaní s 30%). Do cieľovej demografickej vrstvy patria občania, ktorí sa zaujímajú o spoločenské dianie a sú schopní rozpoznať protizákonné prejavy, prípadne prejavy destabilizujúce demokratické zriadenie Slovenska.

8. Pôsobnosť projektu *(celoštátna, regionálna, miestna)* celoštátna

9. Udržateľnosť projektu *(stručne popíšte aktivity, ktoré budú pokračovať aj po ukončení financovania z dotácie rady)*

Po skončení financovania projektu z dotácie rady bude hra umiestnená na sociálnych sieťach voľne, bez sponzorovania, čo však nevyvolá prekážky pre jej ďalšie šírenie a používanie. Používame voľne dostupné technológie, ktoré používajú vývojári po celom svete. Žiadna časť programu nevyžaduje zakúpenie licencie a nepožaduje poplatky za autorské práva. Databáza ako výsledok práce neurónovej siete bude umiestnená na serveri žiadateľa, čo si vyžiada minimálne finančné náklady.

10. Personálne zabezpečenie projektu *(projektový tím, koľko ľudí sa podieľalo na príprave projektu a koľko sa bude podieľať na realizácii projektu, počet dobrovoľníkov zapojených do prípravy a realizácie projektu)*

Príprava projektu: 3 (senior analytik prejavov extrémizmu, projektový manažér a IT konzultant)

Realizácia projektu: 5 (senior analytik prejavov extrémizmu, junior analytik prejavov extrémizmu, psychológ pre dizajn hry a poradenstvo, projektový manažér a IT konzultant)

Počet dobrovoľníkov: 3

11. Aktivity a časový harmonogram realizácie projektu *(uvedie sa začiatok a ukončenie projektu a ďalšie dôležité udalosti, ak viete, uveďte aj dátum jednotlivých aktivít)*

Projekt prirodzene naviaže na činnosť žiadateľa, ktorej jedným z pilierom aktivít je monitorovanie a reportovanie nenávistných prejavov na internete.

Aktivita 1: Zber dát z vybraných serverov

Kategorizácia dát

Zber dát bude prebiehať kontinuálne počas celého trvania projektu, tj. 9 mesiacov, pričom kategorizácia dát bude realizovaná v prvý mesiac realizácie projektu.

Aktivita 2: Vytvorenie hry a zber dát verejnosťou prostredníctvom hry

Prostredníctvom hry, ktorá sa bude šíriť sociálnymi sieťami, napomôže verejnosť s plnením databázy nenávistných prejavov online. Hra bude technicky vypracovaná otvorenými technológiami. Na zreteľ sa bral aj fakt, aby prevádzkovateľ nemal žiadne dodatčoné náklady na licencie:

- Javascript (ECMAScript 8); PHP 7.1; Serverové prostredie je postavené na

operačnom systéme linux.

Z funkčného hľadiska ide o plugin do internetového prehliadača (Chrome, Firefox), ktorý rozšíri funkčnosť vybraných stránok (Facebooku,YouTube) o možnosť označiť nenávistný príspevok. Za takúto činnosť je používateľ hry ohodnotený bodovým systémom. Bodový systém pri dosiahnutí určitého počtu bodov generuje rôzne ocenenia ako aj možnosť prezentovať svoje výsledky v hre na verejnosti (mezdi priateľmi).

Nový hráč prechádza najskôr "skúšobnou dobou" - určitú dobu administrátor kontroluje, či hráč zámerne neoznačuje komentáre zmätočne za účelom znehodnocovania datasetov. Ak hráč prejde skúšobnou dobou, jeho profil je pridaný do verejného zoznamu hráčov a dáta ktoré pomáha kategorizovať použijeme na učenie umelej inteligencie. Oficiálne sa stáva plnohodnotným hráčom.

Administrátor môže blokovať alebo meniť skóre jednotlivých hráčov, pridávať a upravovať hodnosti prípadne organizovať špeciálne odmeny za extra úlohy. Správnym označovaním nenávistných komentárov získa hráč body ("skóre") do svojho účtu. Po dosiahnutí určitého počtu bodov sa mu zvýši hodnosť. Nová získaná hodnosť bude spojená s pekným grafickým odznakom a možnosťou "pochváliť" sa verejne facebookovým statusom. (podobne ako pri iných FB hrách). Na oficiálnej stránke hry bude vidieť globálny zoznam hráčov s hodnosťami. Svoje skóre bude možno porovnať s inými hráčmi (priateľmi).

Časový harmonogram: máj – september 2018

Aktivita 3: Učenie neurónovej siete

Pomocou datasetov získaných v predchádzajúcej kapitole natrénujeme neurónovú sieť, ktorá sama rozpozná a označí nezákonný obsah. Táto časť projektu je závislá na kvalite vstupov a dát získaných v predchádzajúcej fáze od čo najväčšieho počtu užívateľov sociálnych sietí.

Časový harmonogram: september – december 2018

Aktivita 4: Nahlasovanie nezákonných prejavov

Analytik žiadateľa vyhodnotí výstupy získané vďaka umelej inteligencii a tie, ktoré budú niesť znaky skutkovej podstaty trestného činu, nahlási národnej kriminálnej agentúry Prezídia Policajného zboru.

Časový harmonogram: počas trvania projektu, najintenzívnejšie september – december 2018

12. Výsledky a výstupy projektu ***(uveďte kvalitatívne a kvantitatívne ukazovatele dosahovania cieľov projektu, spôsob vyhodnotenia úspešnosti projektu.*** *Napríklad: počet ľudí zúčastnených na projekte, počet, množstvo realizovaných aktivít a akcií, seminárov, školení, počet ľudí zúčastnených na jednotlivých aktivitách, počet vydaných publikácií, metodických príručiek, zborníkov, v prípade kultúrnych akcií počet dokumentačných materiálov (CD nosiče, obrazové publikácie. Za kvalitatívne ukazovatele môžeme považovať napríklad: zefektívnenie spolupráce medzi relevantnými inštitúciami).*

Aktivita 1:

Výsledkom je:

1. webová aplikácia so zoznamom vybraných nenávistných komentárov.

2. robot, ktorý zberá komentáre a príspevky z vybraných sociálnych sietí (FB stránky, YouTube kanály)

3. administrácia zberu dát, administrácia uverejňovania príspevkov a manažment kategórií nenávistných príspevkov.

Aktivita 2:

Výsledkom je:

1. plugin do webového prehliadača, ktorý rozširuje funkcionalitu stránok facebook.com a youtube.com o možnosť nahlasovať nenávistné a dezinformačné prejavy.

2. odmeňovací systém pre používateľov hry.

3. kumulácia datasetov pre učenie neurónovej siete v nasledujúcej aktivite.

4. ako kvantitatívny ukazovateľ je možné brať počet používateľov tejto hry, „hráčov" z radov verejnosti.

Aktivita 3:

Výsledkom je:

1. automatická kategorizácia nenávistných prejavov v aplikácii, ktorá vznikla počas 1. aktivity.

2. automatická detekcia nenávistných prejavov a dezinformačných aktivít v prostredí sociálnych sietí.

Aktivita 4:

Výsledkom je:

1. filter extrémistických prípadov a prípadov, ktoré nenesú znaky trestného činu extrémizmu

2. postúpená informácia Policajnému zboru o prípadoch vykazujúcich znaky trestného činu extrémizmu.

Kvalitatívne ukazovatele cieľov projektu	**Spôsob vyhodnotenia úspešnosti projektu**
Väčšia digitálna občianska angažovanosť	Počet nahlásených prípadov extrémizmu online prostredníctvom hry
Zvýšená miera dôvery v právny štát	Spoločenský diskurz na sociálnych sieťach
Efektívnejšie odhaľovanie prejavov extrémizmu online	Nahlásený počet prípadov extrémizmu online Policajnému zboru SR
Kvantitatívne ukazovatele cieľov projektu	**Spôsob vyhodnotenia úspešnosti projektu**
Počet nahlásených prípadov prostredníctvom hry	Databáza v administrátorskom rozhraní
Počet odohraných minút v rámci hry (angažovanosť pre dobrú vec)	Počítadlo nastavené v administrátorskom rozhraní
Počet nahlásených prípadov Policajnému zboru SR	Databáza žiadateľa

13. Publicita projektu *(uvádzať informáciu na dostupných tlačených a elektronických materiáloch, v mediálnych výstupoch uskutočnených v súvislosti s projektom v znení:*

„Projekt bol finančne podporený Radou vlády Slovenskej republiky pre prevenciu kriminality".

Publicita a informovanosť bude zabezpečená počas celého obdobia trvania realizácie projektu. Informácia o projekte bude uvedená na webovej stránke žiadateľa, na jeho Facebookovom profile a pri všetkých mediálnych aktivitách, ktoré v čase trvania projektu bude realizovať. Najväčšia mediálna prítomnosť sa predpokladá pri spustení hry na sociálnych sieťach. Vzhľadom na danú cieľovú skupinu a virtuálne prostredie sa bude publicita sústreďovať na online priestor.

14. Uveďte, či ste už v minulosti získali finančnú dotáciu na podobný projekt zameraný na prevenciu kriminality a inej protispoločenskej činnosti *(v ktorom roku, na aké konkrétne aktivity, program a názov subjektu)*

2017- podpora projektu Nahlás.to! pre žiadateľa, teda Digitálnu inteligenciu v rámci dotácie podľa zákona číslo 526/2010 Z. z. o poskytovaní dotácií v pôsobnosti Ministerstva vnútra Slovenskej republiky v znení neskorších predpisov na nasledujúce aktivity:

1. aktivizácia mladých ľudí pri nahlasovaní nenávistných prejavov na internete prostredníctvom funkčnej linky nahlás.to, 2. scitlivovanie na prejavy extrémizmu a neznášanlivého správania v online priestore cez online kampaň, 3. posilnenie odborných kapacít preventistov škôl, školských psychológov, učiteľov občianskej náuky a iných, a to prostredníctvom workshopov vo vybraných regiónoch Slovenska, 4. na základe vyhodnotenia online kampane a druhovej typológie nenávistných príspevkov vytvorenie poznatkovej bázy pre správne nastavenie komunikačnej stratégie zameranú na šírenie myšlienok vzájomného rešpektu a predchádzanie rasizmu a xenofóbii.

15. Stanovisko obce *(priložte stanovisko obce k projektu, v prípade kamerového systému stanovisko príslušného Krajského riaditeľstva Policajného zboru)*

Neaplikuje sa.

16. V prípade inštalácie kamerového systému uveďte nasledovné informácie:

a) akým spôsobom a v akom režime bude (je) zabezpečený stály, teda nepretržitý 24 – hodinový monitoring kvalifikovaným operátorom (príslušníkom Policajného zboru, príslušníkom mestskej alebo obecnej polície, a pod.);

b) akým spôsobom bude (je) zabezpečená obsluha a jej kvalita (spôsob výberu a výcvik operátorov, absolvované školenia, certifikáty a pod.);

c) akým spôsobom bude (je) vykonávaná práca so záznamami (tvorba a evidencia záznamov, doba uložených záznamov, dokumentácie, využívanie informácií, existencia vnútorných smerníc, prístupové práva k záznamom a pod.);

d) akým spôsobom bude (je) garantované právo na súkromie občanov pri fungovaní kamerového systému (zákon č. 122/2013 Z. z. o ochrane osobných údajov a o zmene a doplnení niektorých zákonov v znení zákona č. 84/2014 Z. z., ktorým sa mení a dopĺňa zákon č. 122/2013 Z.z. o ochrane osobných údajov a o zmene a doplnení niektorých zákonov a ktorým sa mení zákon Národnej rady Slovenskej republiky č. 145/1995 Z.z. o správnych poplatkoch v znení neskorších predpisov);

e) akým spôsobom bude (je) zabezpečené informačné alebo iné prepojenie na Policajný zbor, prípadne na ďalšie systémy (mestský rozhlas, integrované bezpečnostné systémy a pod.);

f) akým spôsobom bude (je) realizované vyhodnocovanie účinnosti kamerového systému a zber štatistických údajov o znížení kriminality v lokalite monitorovanej kamerovým systémom (aké kritéria budú (sú) sledované, aké hodnoty vyčíslené, napr. pokuty za priestupky odhalené kamerami, odhalenie škôd na verejnom

majetku, poškodzovanie majetku obce a pod.).

g) v prípade "mobilného kamerového systému" uveďte okrem písm. b), c), d), e), f) či je využitie takéhoto systému účelné, opodstatnené, zmysluplné a efektívne.

Neaplikuje sa.

B.2. ROZPOČET PROJEKTU *(popíšte rozpočet projektu max. 1 strana)*

Komentár k štruktúrovanému rozpočtu:
Riadok 29 Požadovaná dotácia vo výške 10 000,00 €
Služby IT expertov za účelom kategorizácie, programovania hry a jej grafiky, učenie neurónovej siete, testovanie (400 človekohodín x 25/h)

Spolufinancovanie: Riadok 29 vo výške 2 500,00 €
Psychologické poradenstvo pri tvorbe hry (40 človekohodín x 25/h)
Monitorovanie sociálnych sietí (60 človekohodín x 25/h)