SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 3050

STROJNO GENERIRANJE IZVJEŠĆA SPECIJALISTA TIJEKOM RUKOVANJA SIGURNOSNIM INCIDENTOM

Robert Benić

Zagreb, lipanj 2022.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 11. ožujka 2022.

DIPLOMSKI ZADATAK br. 3050

Pristupnik:	**Robert Benić (0036508373)**
Studij:	Računarstvo
Profil:	Računarska znanost
Mentor:	doc. dr. sc. Stjepan Groš

Zadatak: **Strojno generiranje izvješća specijalista tijekom rukovanja sigurnosnim incidentom**

Opis zadatka:

Tijekom rješavanja incidenta zadaju se zadaće pojedinim specijalistima, primjerice, forenzičarima da provjere neko računalo ili administratorima da pregledaju sistemske i operativne zapise. Kada obave posao, specijalisti izvješća upućuju nadređenima koji na temelju njih donose odluke o daljnjem postupanju. Prilikom simulacije upravljanja incidentnom situacijom, nema na raspolaganju specijalista te se sva izvješća moraju generirati strojnim putem. Način na koji se to radi je da postoje predlošci koji se koriste, ali ti predlošci umanjuju realističnost simulacije. Iz tog razloga bilo bi dobro imati odgovarajuće modele temeljene na strojnom učenju koji bi generirali sintetička izvješća. U sklopu diplomskog rada potrebno je složiti metodologiju treniranja modela koji generiraju izvješća specijalista. Nadalje, potrebno je istražiti potencijalne primjere takvih izvješća te ih dohvatiti za potrebe treniranja modela. Korištenjem definirane metodologije i dohvaćenih primjera, treba istražiti mogućnost generiranja izvješća u simulatoru Cyber Conflict Simulator. Radu priložiti izvorni kod razvijenih i korištenih programa. Citirati korištenu literaturu i navesti dobivenu pomoć.

Rok za predaju rada: 27. lipnja 2022.

Srdačno zahvaljujem svom mentoru, doc. dr. sc. Stjepanu Grošu, na iskazanoj pomoći i podršci u izradi ovog diplomskog rada.

Sadržaj

1. Uvod

Unatoč svim razvijenim postupcima za prevenciju incidenata u informacijskoj i kibernetičkoj sigurnosti, nikad nije moguće u potpunosti jamčiti da se incident neće dogoditi. Prema tome, nužno je uvježbavati sigurnosno osoblje u postupcima rješavanja incidenata nakon što su se oni dogodili. Za potrebe toga je u razvoju niz projekata, među kojima se ističe *Cyber Conflict Simulator* (CCS) [1]. CCS provodi simulirane sigurnosne incidente pri čemu djelovanje napadača, branitelja, pružatelja vanjskih usluga i drugih sudionika u incidentnoj situaciji može biti pod kontrolom korisnika simulacije (stvarne osobe) ili pod kontrolom računala.

Prilikom rješavanja sigurnosnih incidenata se redovito zahtijevaju specijalističke vanjske usluge, za kakve odgovorne osobe u sustavu nisu obučene, npr. analiza zapisa operacijskog sustava ili forenzička analiza računala. Zadatak specijalista je da po završetku obavljanja tražene usluge sastavi i pošalje izvješće nadređenim osobama, odnosno korisnicima usluge.

U simuliranju obavljanja specijalističkih usluga unutar CCS-a se pojavljuje potreba za računalnim generiranjem izvješća koja će čitati korisnici simulacije. Postojeći način kojim se generiraju izvješća u CCS-u se zasniva na parametriziranim predlošcima, što nije idealno jer je na tako generiranim izvješćima lako vidljivo da nisu djelo stvarne osobe, već su sastavljena automatski.

Cilj ovog rada je unaprijediti postupak strojnog generiranja izvješća za potrebe sustava kao što je CCS tako da generirana izvješća više odgovaraju onakvima kakva bi napisao stvarni specijalist. Na taj način bi CCS ili sličan sustav za uvježbavanje sigurnosnog osoblja mogao uključivati i uvježbavanje u tumačenju stvarnih specijalističkih izvješća, što je važno kod rada sa stvarnim specijalistima.

Ovaj diplomski rad predlaže postupak ostvarivanja navedenog cilja pomoću modela prirodnog jezika koji su zasnovani na strojnom učenju. Postupak se dijeli na tri osnovne faze: sastavljanje skupa podataka za učenje, učenje modela i generiranje teksta iz naučenog modela. U trećem poglavlju su opisani modeli prirodnog jezika općenito, odabrano je

nekoliko prikladnih modela za potrebe ovog rada i opisan je način postavljanja okoline za rad s istim modelima. Četvrto poglavlje je posvećeno prikupljanju i obradi podataka za učenje odabranih modela prirodnog jezika. Peto poglavlje opisuje faze učenja i generiranja teksta iz modela prirodnog jezika te navodi primjere generiranih izvješća kao rezultate ovog rada, na temelju kojih je donesen konačni zaključak.

2. Odabir modela i postavljanje okoline

Primjer 2.1 prikazuje nekoliko strojno generiranih izvješća uporabom prethodno navedenih predložaka iz CCS-a. Nasuprot tomu, primjer 2.2 prikazuje tekst iz jednog stvarnog specijalističkog izvješća, s kakvim sigurnosno osoblje mora raditi u stvarnim situacijama i kakav bi valjalo generirati u sklopu CCS-a kako bi se postigla što veća vjernost simulacije.

```
'Bill Due to past contains a link
'https://genom.mefst.hr/webmail/src/login.php' to a website 'Webmail
SquirrelMail'.

'Bitcoin Sale contains a link
'http://webmail.forumofthemall.hr/mail/loging.php' to a website 'Webmail
SquirrelMail Popular Forum'.

'New Dogecoin Crypto Sale contains a link
'http://webmail.forumofthemall.hr/mail/loging.php' to a website 'Webmail
SquirrelMail Popular Forum'.

Log entry found: Firewall (Type: Firewall) detected. [Allowed network
traffic protocol 'smtp:25' from 'PCSRPI08' to'Mail server EP'. Rule
'LAN_SRPI_to_Mail_Server'.]

Log entry found: Firewall (Type: Firewall) detected. [Allowed network
traffic protocol 'smtp:25' from 'server74.aws.com' to'Mail server EP'.
Rule 'Internet_to_Mail_Server'.]

Log entry found: Firewall (Type: Firewall) detected. [Allowed network
traffic protocol 'https:443' from 'Proxy server' to'server74.aws.com'.
Rule 'Proxy_to_Internet, https:443'.]
```

Primjer 2.1 Nekoliko automatski generiranih „izvješća” iz jedne CCS simulacije

```
Stroz Friedberg found direct and compelling digital forensic evidence
that the documents relied upon by Mr. Ceglia to support his claim are
forged. Stroz Friedberg also found what it believes to be the authentic
contract between Mr. Ceglia and Mr. Zuckerberg. That contract contains no
```

references to Facebook. As described more fully in this report, Stroz Friedberg made the following findings bearing on the authenticity of the Work for Hire Document and the Purported Emails:

- Stroz Friedberg did not find any exact copies of the Work for Hire Document on the hundreds of pieces of media produced by Mr. Ceglia, including three computers, three hard drives, 174 floppy disks, and 1,087 CDs (hereinafter, the "Ceglia Media").
- Stroz Friedberg did find a signed copy of an April 28, 2003 contract between Mr. Ceglia and Mr. Zuckerberg, though it concerns only Mr. Zuckerberg's work on the StreetFax project and includes no references to Facebook.

Primjer 2.2 Nekoliko odlomaka teksta iz forenzičkog izvješća koje je pisao stvarni specijalist [2]

Kao naprednije rješenje za potrebe generiranja teksta od jednostavnih predložaka mogu poslužiti već spomenuti modeli prirodnog jezika.

Model prirodnog jezika je vjerojatnosna razdioba nad skupom svih mogućih tekstova na nekom prirodnom jeziku. Strojno generiranje teksta je samo jedna od mnogobrojnih primjena modela prirodnog jezika, a neke druge primjene su kod prepoznavanja rukopisa, prepoznavanja govora, strojnog prevođenja, odgovaranja na pitanja itd.

Danas su najnapredniji modeli prirodnog jezika u uporabi zasnovani na strojnom učenju, i to uglavnom modeli Transformer arhitekture, koja se zasniva na mehanizmu pozornosti (engl. *attention*). Takvi se modeli uglavnom uče metodama nenadziranog učenja (engl. *unsupervised learning*). Dva najpoznatija modela prirodnog jezika Transformer arhitekture su *Generative Pre-trained Transformer* (GPT) [3] i *Bidirectional Encoder Representations from Transformers* (BERT) [4].

Učenje prethodno nenaučenog modela prirodnog jezika je dugotrajan i resursno zahtjevan postupak. Budući da se isti prirodni jezici, osobito engleski jezik, primjenjuju u ljudskoj komunikaciji u mnogo različitih domena, isti podaci za učenje i naučeni modeli se redovito mogu ponovo upotrijebiti. Stoga i ovaj rad umjesto učenja modela prirodnog jezika s

potpuno novim podacima upotrebljava nekoliko unaprijed naučenih modela koji se mogu brže i jednostavnije prilagoditi danom zadatku.

Za potrebe ovog zadatka model prirodnog jezika treba biti prilagođen domeni informacijske i kibernetičke sigurnosti, što modeli prirodnih jezika općenito nisu. Uz tu činjenicu je lako predvidjeti da bez dodatnih postupaka za prilagođavanje domeni, generirani izlazi iz gotovih modela kao što je GPT neće biti zadovoljavajući. Primjeri 2.3 i 2.4 prikazuju tako generirane izlaze iz dva različita modela, GPT-2 [5] i GPT-Neo [6].

Digital Forensics Analysis Report

The Forensic Investigation Team is conducting its first forensic analysis of the missing persons case at the National Forensic Laboratory (NFL), as part of a collaborative investigation.

While investigators believe the case was stolen, the investigation is continuing.

"The investigation has a timeline and is ongoing," said Dr. Mounir N'Umin.

"We have started to identify and follow up on the suspects in the case, but as we do not know how they could have been affected, we are going to have to determine what the potential impact is on our public safety."

N'Umin also said that after conducting a series of interviews, a team of researchers from the FBL will have an opportunity to find out what the suspects could have done and their motives for leaving their home.

In an interview with The Daily Telegraph in January, N'Umin said the investigators will have more information on the case after the discovery of the phone.

N'Umin also said the investigation will continue.10 pt

The National Forensic Laboratory (NFL) was established to investigate and protect the rights of its citizens as part of the "NHS of Canada Act".

Primjer 2.3 Generirani tekst iz GPT-2 124M modela bez dodatnog podešavanja za ulazni redak teksta „Digital Forensics Analysis Report”

Digital Forensics Analysis Report

When we created our Forensics Report, it didn’t really fit in with what we were looking for. The report we had included in our analysis project was written in two areas – legal and criminal. Both areas were highly detailed and detailed, and the criminal section was highly detailed, too. We had no issues finding the correct information in the first place.

We had several problems with this report. This wasn’t a police report but a criminal one. The criminal section is much larger in size than the police report, and the criminal section is much bigger. The criminal section contains both legal and criminal law. For example, if a suspect is accused of sexual assault, the crime is defined as rape of a child, and the police report can look up the information and search the child’s computer for a suspect who may have been raped by a police officer. There is also a criminal section that does the same thing. You could have a crime report that looks for information that the suspect’s parents said about a rape or assault. This could be the most useful section of the report.

The criminal section doesn’t have to look at a child’s computer, or even the criminal section itself, but it should look at the criminal section. For example, if a suspect is charged with child molestation, the criminal section could look up the charges, the evidence from a child’s computer, and search the suspect’s computer for a suspect that may have been molested by a police officer. This could be the most useful part of the criminal section.

Primjer 2.4 Generirani tekst iz GPT-Neo 125M modela bez dodatnog podešavanja za ulazni redak teksta „Digital Forensics Analysis Report”

Ovakav strojno generirani tekst se može usporediti s izvještajem neke osobe koja nije specijalist u domeni koja se ovdje proučava, već je samo površno upoznata s tom domenom i s problemom koji specijalist rješava, npr. novinara. Iz ovoga slijedi da je za uporabu modela u domeni kao što je kibernetička sigurnost potrebno nadograditi model

dodavanjem informacija koje odgovaraju specijalističkom znanju u toj domeni. Postupak kojime se to izvodi je vrlo sličan prvotnom učenju modela, a obično se naziva fino podešavanje (engl. *fine-tuning*). Primjer 2.5 prikazuje izlazni tekst koji generira fino podešen model za rad u domeni naslova članaka.

```
Nobody Really Likes the Browser

A bike ride through one human DNA

LastPass and Apple users disappoint inactivewire posts to 'best global

Show HN: Pagerbot - A lightweight C++ framework for building HTML5 games
on Windows

Julian Assange further online issue due to patent-up is behind Uber case

Julia: Python for the future, by Cautoti (2002)

The Best Thing Your iPhone Is Child-Trafficking Agents Will Cost $1.25
Million

Hasbro lost its way to Google Plus?

Apple's F-1 Computer

Medicine’s Hidden Message: The Hash Function [pdf]
```

Primjer 2.5 Nekoliko tekstova generiranih modelom prirodnog jezika koji je fino podešen za rad s naslovima članaka (aitextgen [7], Hacker News primjer)

Lako se zaključuje da ako je moguće isti ili sličan postupak provesti nad modelom dvaput, moguće ga je provoditi i tri ili još više puta. Na ovaj način se uz uporabu različitih skupova podataka za učenje (podešavanje) u različitim fazama može dodatno specijalizirati model na užem području unutar neke domene.

Za rad s modelom prirodnog jezika u sklopu ovog diplomskog rada je upotrijebljena biblioteka aitextgen [7], pomoću koje su generirani i prethodno navedeni primjeri. aitextgen omogućuje rad s unaprijed naučenim GPT-2 [5] i GPT-Neo [6] modelima.

GPT-2 je druga, unaprijeđena inačica modela GPT tvrtke OpenAI, koja je besplatna, slobodna i otvorenog koda. Treća inačica, GPT-3 [8], vlasništvo je tvrtke Microsoft i može mu se pristupiti samo putem Microsoftovog plaćenog programskog sučelja; aitextgen kao zamjenu za GPT-3 podržava GPT-Neo, besplatan i slobodan model koji se zasniva na GPT-3 specifikaciji te uključuje neke dodatne značajke u odnosu na GPT-3. Za vrijeme pisanja ovog rada se razvijaju i dodatno unaprijeđeni modeli u odnosu na GPT-3 i GPT-Neo, npr. GPT-4, GPT-J, GPT-X, GPT-NeoX. Takve modele koji su još u postupku razvoja nije bilo moguće upotrijebiti u sklopu ovog rada. Razlozi tomu su što je za potrebe ovog rada važno da model bude unaprijed naučen, budući da je početni postupak učenja znatno resursno zahtjevniji od naknadnog podešavanja za specifičnu domenu primjene, kao i da se može kasnije podešavati i upotrebljavati za generiranje teksta uz ograničene količine sklopovskih resursa – centralnih procesora, radne memorije, grafičkih procesora i grafičke memorije – koje su dostupne autoru ovog rada.

Za pristup resursima je upotrijebljen servis Colaboratory, skraćeno Colab [9], tvrtke Google, koji omogućuje izvođenje Python programa pisanih u obliku Jupyter bilježnice na Googleovom udaljenom sklopovlju. Google Colab je za korisnike u Hrvatskoj dostupan samo u besplatnoj inačici, koja uključuje dvije Intel Xeon procesorske jezgre, 12 GB radne memorije i Nvidia K80 grafički procesor s 12 GB grafičke memorije ili Nvidia T4 grafički procesor s 16 GB grafičke memorije. Uporaba grafičkog procesora u besplatnoj inačici Colaba je ograničena na 4 sata dnevno.

Uz biblioteku aitextgen je dostupna gotova Jupyter bilježnica za uporabu u Google Colabu, koju je autor ovog rada izmijenio za potrebe rada. Najjednostavniji način učitvanja bilježnice u Google Colab je putem Google Drive servisa za pohranu. Preuzeta bilježnica u `.ipynb` formatu se prenese na Drive, odakle se može izravno pristupiti Colab okruženju. Alternativno, u Colab okruženju se mogu otvoriti i bilježnice pohranjene na drugim vanjskim servisima kao što je GitHub.

Dodatna prednost biblioteke aitextgen je što se može izvoditi i bez grafičkog procesora, što je praktički neupotrebljivo za intenzivne radnje kao što su učenje (podešavanje) i generiranje teksta iz GPT modela, no korisno je za manje intenzivne radnje kao što su pretprocesiranje i kompresija podataka za učenje, čime se smanjuje potreba za uporabom grafičkog procesora u cijelom postupku.

aitextgen također ima mogućnosti pohrane i ponovnog učitavanja naučenog (podešenog) modela iz lokalnog direktorija, ili iz Google Drive direktorija ako se izvodi u Colab okruženju, što omogućuje između ostaloga provođenje više iteracija podešavanja i generiranja teksta uz pomoć grafičke kartice nego što je izvedivo unutar Colabovog ograničenja od 4 sata.

GPT-2 i GPT-Neo modeli su dostupni u nekoliko inačica koje se razlikuju po broju parametara. Veći broj parametara omogućuje modelu veću ekspresivnost, tj. veću mogućnost prilagođavanja različitim skupovima podataka za učenje, što u pravilu dovodi do boljih performansi kod generiranja izlaza. Međutim, veći broj parametara zahtijeva i veću količinu sklopovskih resursa pri uporabi, pa su uz resursna ograničenja Google Colaba autoru ovog rada bile dostupne samo inačice s najmanjim brojem parametara: GPT-2 124M, koji ima 124 milijuna parametara i zauzima 486,7 MiB prostora na disku, odnosno GPT-Neo 125M, koji ima 125 milijuna parametara i zauzima 525,6 MiB prostora na disku.

3. Priprema podataka za učenje

Kod primjena koje zahtijevaju fino podešavanje modela, kao i kod onih gdje se model prirodnog jezika uči iznova na podacima iz konkretne domene, prirodno se postavlja problem prikupljanja tekstualnih podataka iz domene za učenje (podešavanje) modela.

Zajednička karakteristika svih modela strojnog učenja, pa tako i modela Transformer arhitekture kao što je GPT, jest to što uče na način da njihovi izlazi budu što vjerniji ulaznim podacima za učenje, uz ograničenja koja su propisana arhitekturom samog modela. Stoga je za postizanje izlaza u određenom obliku potrebno prikupiti bazu ulaznih podataka za učenje (podešavanje) koji što bolje odgovaraju istom obliku.

Modelima strojnog učenja je također zajedničko da učenjem (podešavanjem) na konačnom skupu podataka nije moguće poboljšavati performanse modela u nedogled, već se optimalne performanse postižu nakon konačnog broja iteracija koji ovisi o količini korisnih informacija sadržanoj u ulaznim podacima i o stopi učenja. Nastavkom postupka učenja (podešavanja) s istim podacima nakon što je postignuta optimalna naučenost se umanjuje mogućnost generalizacije modela. Ova pojava se naziva prenaučenost (engl. *overfitting*).

Uobičajeni način izbjegavanja prenaučenosti je periodičkim prekidanjem postupka učenja (podešavanja) i ispitivanjem performansi modela na različitom skupu podataka od skupa za učenje. Ovaj postupak se naziva unakrsna provjera (engl. *cross-validation*).

Za potrebe finog podešavanja modela prirodnog jezika u domeni informacijske i kibernetičke sigurnosti je potrebno sastaviti novi skup tekstualnih podataka iz iste domene. Korisni tekstualni podaci za tu svrhu uključuju vijesti, članke, blogove, objave s društvenih mreža i sl. iz područja sigurnosti, kakvi se mogu besplatno čitati i preuzimati s weba.

Ručno pregledavanje velikog broja web-stranica i izvlačenje korisnih podataka iz njih bi bilo dugotrajan i mukotrpan posao, pa su osmišljeni računalni programi za istu svrhu. Takav program se obično naziva puzavac (engl. *crawler*) ili pauk (engl. *spider*), a pomoću

sličnih programa su sastavljeni i postojeći skupovi podataka za učenje modela prirodnog jezika, npr. skup WebText kojim je naučen model GPT-2 [6].

Web-puzavci imaju razne primjene, koje nisu ograničene samo na područje strojnog učenja, pa su razvijeni i razni programski alati posebno za razvoj novih web-puzavaca. Za potrebe ovog rada je odabran radni okvir Scrapy [10], koji je napisan u programskom jeziku Python. Scrapy omogućuje dohvat, pohranu i obradu raznih vrsta resursa na webu uz pomoć HTTP protokola.

U postupcima prikupljanja i obrade podataka grafički procesor nije koristan, pa je autor izveo ovu fazu rada u potpunosti na lokalnom računalu. Za izvođenje Scrapy web-puzavca i drugih programa upotrebljenih u ovoj fazi je dovoljno instalirati programski jezik Python 3, Scrapy i druge navedene biblioteke, npr. iz Python Package Index (PyPI) [11] repozitorija.

Scrapy web-puzavac razvijen u sklopu ovog rada radi tako da počevši od indeksa odabranih web-sjedišta slijedi veze do stranica koje sadrže koristan tekst (npr. članaka). Iz teksta se uklanjanju formatne značajke HTML-a pomoću biblioteke BeautifulSoup [12]. Puzavac zaustavlja slanje zahtjeva za dohvatom novih stranica prema poslužitelju domene (web-sjedišta) ako iscrpi sve stranice indeksa, što poslužitelj signalizira praznom sljedećom stranicom ili HTTP statusom `404 Not Found`, i sve pronađene veze prema istoj domeni ili prekorači dopušten broj zahtjeva prema domeni u jedinici vremena, što poslužitelj signalizira HTTP statusom `429 Too Many Requests`.

BeautifulSoup je Python biblioteka za rad s HTML i XML datotekama, što je prikladno za procesiranje teksta s web-stranica koje prikuplja Scrapy puzavac, a koje su uglavnom u HTML formatu. U sklopu ovog rada BeautifulSoup je upotrijebljen samo za uklanjanje značajki formatiranja iz HTML-a, kao što prikazuju primjeri 3.1, 3.2 i 3.3.

```
<p>On April 13, Microsoft said it executed <a
href="https://blogs.microsoft.com/on-the-issues/2022/04/13/zloader-
botnet-disrupted-malware-ukraine/" target="_blank" rel="noopener">a legal
sneak attack</a> against <strong>Zloader</strong>, a remote access trojan
```

```
and malware platform that multiple ransomware groups have used to deploy
their malware inside victim networks. More specifically, Microsoft <a
href="https://noticeofpleadings.com/Zloader/" target="_blank"
rel="noopener">obtained a court order</a> that allowed it to seize 65
domain names that were used to maintain the Zloader botnet.</p>
```

Primjer 3.1 Odlomak iz izvornog teksta članka pisanog na jeziku HTML [13]

```
On April 13, Microsoft said it executed a legal sneak attack against
Zloader, a remote access trojan and malware platform that multiple
ransomware groups have used to deploy their malware inside victim
networks. More specifically, Microsoft obtained a court order that
allowed it to seize 65 domain names that were used to maintain the
Zloader botnet.
```

Primjer 3.2 Odlomak teksta [13] kako se prikazuje u web-pregledniku

```
On April 13, Microsoft said it executed a legal sneak attack against
Zloader, a remote access trojan and malware platform that multiple
ransomware groups have used to deploy their malware inside victim
networks. More specifically, Microsoft obtained a court order that
allowed it to seize 65 domain names that were used to maintain the
Zloader botnet.
```

Primjer 3.3 Čisti tekst odlomka [13] dobiven pomoću biblioteke BeautifulSoup

Konačna verzija skupa podataka prikupljenih Scrapy puzavcem do trenutka objave ovog rada sadrži 37.580 web-stranica preuzetih s 10 domena i 88,1 MiB izdvojenog čistog teksta. Zbog memorijskih ograničenja opisanih u prethodnom poglavlju, skup podataka je razdvojen na 9 dijelova od kojih je svaki velik približno 10 MiB.

Međutim, prethodno opisani skup tekstualnih podataka nije idealan za primjenu u sklopu ovog rada, jer nije u obliku izvješća kao što bi trebao biti izlaz iz modela. Iz toga slijedi da se fino podešeni model mora dodatno, još finije podesiti s tekstualnim podacima koji odgovaraju točnom obliku u kojemu se traži izlaz iz modela, tj. s izvješćima koja su pisali stvarni specijalisti. Takva izvješća nisu javno dostupna u obliku koji bi omogućavao jednostavno preuzimanje i obradu pomoću računalnih programa kao što je slučaj s prethodno prikupljenim podacima iz vijesti, članaka, blogova i dr.

Prema tome, pri sastavljanju ovog drugog skupa podataka za učenje (podešavanje) se autor morao zadovoljiti malom količinom teksta koja je prikupljena ručnim pretraživanjem weba i obrađena također ručno. Za pomoć pri obradi su poslužile autorove Python skripte, koje uključuju ekstrakciju teksta pomoću biblioteke textract [14] i pretvorbu kodiranja pomoću biblioteke chardet [15].

textract omogućuje izdvajanje čistog teksta iz datoteka na sličan način kao što to čini prethodno opisana biblioteka BeautifulSoup, ali podržava veći broj formata ulaznih datoteka, uključujući PDF i Microsoft Word, u kojima se nalaze neka od prikupljenih izvješća za ovu svrhu.

chardet omogućuje automatsko pronalaženje kodiranja teksta koje je u uporabi i pretvorbu teksta na neko drugo kodiranje, u ovom slučaju, UTF-8, što je podrazumijevano kodiranje u programskom jeziku Python. Ovako obrađen tekst još uvijek nije u potpunosti spreman za uporabu u učenju (podešavanju) GPT-2 ili GPT-Neo modela jer sadrži neke suvišne dijelove poput praznih redaka i brojeva stranica, koje je potrebno ukloniti ručno.

Nakon što je prikupljen i obrađen skup tekstualnih podataka, bilo pomoću puzavca ili drugačije, za uporabu s bibliotekom aitextgen je dovoljno pohraniti ga u običnu tekstualnu datoteku s UTF-8 kodiranjem. Ukoliko je datoteka prevelika u odnosu na dostupnu količinu radne memorije, moguće je razdvojiti tu datoteku na nekoliko manjih datoteka i upotrebljavati svaku od tih datoteka kao zaseban skup podataka, kao što je autor ovog rada učinio s podacima prikupljenima puzavcem.

Dodatan korak kod rada s bibliotekom aitextgen, koji se može izvesti prije samog postupka učenja, čini predobrada (tokenizacija) i sažimanje podataka za učenje. Ovaj korak ne zahtijeva uporabu grafičkog procesora pa je prikladan za izvođenje na lokalnom računalu, tako da se predobrađeni (tokenizirani) podaci pohrane u arhivu formata `.tar.gz` koja se u postupku učenja (podešavanja) može upotrebljavati jednako kao tekstualna datoteka. U suprotnom se podaci za učenje moraju tokenizirati nakon svakog učitavanja iz tekstualne datoteke.

Kako bi se izbjegla potreba za ponovnim slanjem podataka na udaljeni poslužitelj pri svakom pokretanju Google Colab bilježnice, Colab podržava i učitavanje podataka sa servisa Google Drive. Prije pokretanja bilježnice, podaci za učenje se učitaju u odabrani direktorij na Google Driveu te se u bilježnici podesi lokacija s koje se podaci učitavaju. Pri pokretanju je onda potrebno samo odobriti pristup odgovarajućem Drive direktoriju.

4. Učenje i generiranje teksta

Nakon što su svi potrebni podaci učitani u odgovarajuće direktorije, u Jupyter bilježnici je još potrebno podesiti odgovarajuće parametre: broj iteracija i stopu učenja kod učenja (podešavanja) modela, ulazne nizove, broj izlaza i maksimalnu duljinu izlaza kod generiranja teksta iz modela.

Nakon izvršavanja bilježnice u Google Colabu, konačni fino podešeni model i generirani tekstovi će biti pohranjeni u odgovarajuće direktorije na Google Driveu. Umjesto Drivea moguće je konfigurirati da se upotrebljava privremena pohrana Colab sesije, ili lokalna pohrana ako se bilježnica izvršava na lokalnom računalu. Za potrebe eksperimenata koji su izvođeni u sklopu ovog rada su također ugrađene mogućnosti pohrane i generiranja teksta iz djelomično podešenog modela nakon proizvoljnog broja iteracija te mogućnost ponovnog učitavanja pohranjenog modela.

Primjeri 4.1, 4.2 i 4.3 prikazuju generirani tekst iz GPT-2 124M modela nakon finog podešavanja na skupu podataka prikupljenih web-puzavcem za tri različita broja iteracija.

Digital Forensics Analysis Report

Some media reports have a number of headlines that Google’s recent search of “Google”. For example, the one which has been reported on Google’s Google News Feed, Google explained:

"The Google News Feed account is still updated. If it’s removed, it is going to take a look at a Google News feed and then you’ll see a link which claims that the Google News feed has no intention to be accessed from your own Google News feed."

Google’s News Feed, with its data-sharing policy “. Users, and a small businesses, are being forced to report a crime that will have their data-sharing policy turned up to a “Anyone else”

And in this case, Google is reporting that a rogue Google says it has been removed from Google Chrome's services and removed Google.

Primjer 4.1 Generirani tekst iz podnaučenog GPT-2 124M modela nakon prve faze finog podešavanja (1800 iteracija, stopa učenja 0.001) za ulazni redak teksta „Digital Forensics Analysis Report"

Digital Forensics Analysis Report

Last week the hackers used to acquire more files was also detected by the ability to read them to a remote server in question.

After having previously received any spam messages, the scammers would have been using them to trick victims into downloading and installing malware.

So they have already taken advantage of the problem of spamming out unsolicited email addresses of unsuspecting victims.

SophosLabs analysts have managed to detect a number of the emails we have received on the Google web. Other malicious emails have already been spammed out of Google's website and have recently been detected by the Sophos Security Intelligence Report.

It is quite common with this spam and the spammers have sent messages to Google asking for the order (and we have seen some unusual cases of spam spam using Google Chrome).

We also need to remember that the message claims that a YouTube video of a young boy who has died was posted to Google users – they are now finding more people in that way and have brought up a chance to share messages with the social network.

Primjer 4.2 Generirani tekst iz optimalno naučenog GPT-2 124M modela nakon prve faze finog podešavanja (5400 iteracija, stopa učenja 0.001) za ulazni redak teksta „Digital Forensics Analysis Report"

Digital Forensics Analysis Report

While you can read more about our colleagues in a blog I decided to update our website as well as to have a number of personal information (which is used by "news and not limited," it is likely that this information isn't merely about keeping your password security private:

There are some interesting thing to have chosen – in the way that your credentials were stolen, whether they were a free copy or an unauthorised email or a legitimate email.

If you lose control of your passwords then this is certainly a good idea. If you're interested in protecting your passwords from internet cybercriminals, you can find out more here.

SophosLabs have been seeing more of the details of our website – and our website is still at risk of that incident.

A recent survey scam was revealed by my colleague Sarah Palin and I took the survey in a couple of weeks.

While we've previously seen some interesting emails, some of them were in the online world, and all of the sites in question were still in the hands of hackers.

What about a million users of our customers (which I know not to be "t") who had their account stolen from this scam? In all cases, they could have been exploited by a fake support that claims to be about to have been sent out via email.

Primjer 4.3 Generirani tekst iz prenaučenog GPT-2 124M modela nakon prve faze finog podešavanja (9000 iteracija, stopa učenja 0.001) za ulazni redak teksta „Digital Forensics Analysis Report"

Na prethodno navedenim primjerima se pokazuje da je model uspješno preuzeo određene informacije iz domene informacijske i kibernetičke sigurnosti sadržane u podacima za učenje te ih može upotrijebiti prilikom generiranja teksta. Međutim, izlazni tekstovi ne odgovaraju obliku izvješća koji se konačno traži, iz razloga što takvi tekstovi nisu sadržani u upotrebljenim podacima za učenje. Drugim riječima, u ovako podešenom GPT-2 modelu

nedostaju informacije koje bi odgovarale specijalističkom znanju kakvo je potrebno za pisanje izvješća u obliku u kojemu ih piše stvarni specijalist.

Ovdje također valja obratiti pažnju na ponašanje modela za različit broj iteracija i stopu učenja, odnosno primijeniti prethodno spomenutu unakrsnu provjeru. Ako je model podnaučen, tj. izveden je premalen broj iteracija za istu stopu učenja (primjer 4.1), to se manifestira kao nedostatak domenskog znanja u modelu, slično kao i kod uporabe gotovog modela bez naknadnog učenja (finog podešavanja). Ako je model pak prenaučen, tj. izveden je prevelik broj iteracija za istu stopu učenja (primjer 4.3), u njegovom izlazu se pojavljuju naučene beskorisne informacije, što je najočitije u slučajevima gdje su cijele rečenice ili odlomci doslovno prepisani iz podataka za učenje. Autor ovog rada je postigao najbolje rezultate za 5400 iteracija finog podešavanja uz stopu učenja 0.001 (primjer 4.2).

Kod učenja ili podešavanja s podacima iz izvješća, posljedica ograničene količine tih podataka je da uz istu stopu učenja dolazi do prenaučenosti za znatno manji broj iteracija nego na većem skupu podataka prikupljenih puzavcem. Stoga je najprikladniji način uporabe tih podataka u drugoj fazi podešavanja, uz manji broj iteracija i stopu učenja nego u prvoj fazi.

Primjeri 4.4, 4.5 i 4.6 prikazuju generirani tekst iz GPT-2 124M modela nakon finog podešavanja u dvije faze, najprije na skupu podataka prikupljenih web-puzavcem a nakon toga na skupu podataka iz izvješća.

Digital Forensics Analysis Report

```
The report showed that no evidence has been seen in the file and the
evidence has not been seen, but it appears that there is a large amount
of evidence to have been released.

The report showed that that when a file is found the report shows that
the file is not part of the file.

In the report, the files are displayed at a file that contains the
following text:
```

The document contained text:

The file is a hard disk and the hard disk contained files have been created by the user.

Sophos intercepts the contents and other files.

There is no way to stop the file and other files from being seen, as well as some useful tools to deal with the case and the file is displayed.

Primjer 4.4 Generirani tekst iz podnaučenog GPT-2 124M modela nakon druge faze finog podešavanja (100 iteracija, stopa učenja 0.000001) za ulazni redak teksta „Digital Forensics Analysis Report”

Digital Forensics Analysis Report

In this incident, the company claimed that the company had made it sound like a warning that a customer has been exposed to the hackers' activities.

The information was said to be encrypted by the hackers who had written an article on their web pages.

This is difficult to identify and respond to questions regarding the sensitive data was not encrypted in this case.

As far as I can tell, the hackers's online presence was no small business, and when it came to properly securing it.

The only difference from this is that they posted their information through public feeds (even if they were the victim of this attack).

The "dirty" story was the fact that when the data was publicised as an opportunity for users to connect to it via a public forum and share it with their online activities.

It turns out that the information was stolen from the data breach, not from a private direct message. It was only when the victim had sent a malicious code to the accounts of a 35-year-old Canadian Pharmacyster.

Primjer 4.5 Generirani tekst iz optimalno naučenog GPT-2 124M modela nakon druge faze finog podešavanja (600 iteracija, stopa učenja 0.000001) za ulazni redak teksta „Digital Forensics Analysis Report"

Digital Forensics Analysis Report

On April 11, 2011, Paul Ceglia filed an Amended Complaint seeking a share of Facebook. Mr. Ceglia based his claim on a purported contract between Mr. Ceglia and Mr. Stroz Friedberg (the "Work for Hire Document"). In addition, the Amended Complaint included excerpts of purported emails between Mr. Ceglia and Mr. Zuckerberg (the "Full Footage video recordings made available by Mr. Ceglia. Stroz Friedberg (the "Supplemental Document" lacking pertinent conversation.
Furthermore, four of the five raw audio recordings, which also contained audio captured from the video recording device, are accompanied by a raw audio recording captured from a separate audio-only recording device. The raw audioonly recordings last for the duration of their associated raw videos. These raw audio recordings support the completeness and authenticity of the raw video recordings since they depict the same events within the same duration as captured from the two separate video recorders.

Evidence Acquisition Processing Procedures
Coalfire employed industry standard tools and techniques throughout handling, processing, and analysis. A sealed FedEx Express envelope was received into Coalfire Labs via FedEx Expressalfire Labs via FedEx Expressalfire Labs in parts, that contains not included an International official training requirement, as contribution to the availability of a new video recording service.
Coalfire employed industry standard tools and techniques throughout handling, processing and analysis of the evidence. A Chain of Custody was established upon opening the package. The package contained one USB flash drive sealed in a FedEx label pouch. Details about the enclosed media are included below.

Primjer 4.6 Generirani tekst iz prenaučenog GPT-2 124M modela nakon druge faze finog podešavanja (1000 iteracija, stopa učenja 0.001) za ulazni redak teksta „`Digital Forensics Analysis Report`”

Generirani tekstovi nakon druge faze podešavanja sadrže informacije, kao i stilske karakteristike koje bolje odgovaraju stvarnim specijalističkim izvješćima nego nakon prve faze. Dakle, potvrđuje se da je GPT-2 model usvojio informacije koje odgovaraju specijalističkom znanju za potrebe pisanja izvješća. Autor ovog rada je postigao najbolje rezultate u drugoj fazi za 600 iteracija uz stopu učenja 0.000001 (primjer 4.5), što potvrđuje prethodno iznesenu hipotezu o odnosu ovih parametara prema količini podataka za učenje (podešavanje).

Generiranje teksta iz GPT-2 ili GPT-Neo modela, kao što je vidljivo i na nekima od prethodno navedenih primjera, moguće je bez ulaza ili s ulaznim nizom znakova (dijelom teksta). Generiranjem bez ulaza se dobiva tekst slučajnog informacijskog sadržaja, koji nije prikladan za uporabu u CCS-u jer je za CCS bitno da se ne promijeni ulazno značenje, tj. da se ne promijene zaključci specijalista koje generira sam CCS, npr. „na analiziranom računalu/disku/datoteci je/nije pronađen zlonamjerni softver”. Stoga valja upotrebljavati generiranje s ulazom, pa je za potrebe ovog rada autor odabrao nekoliko ulaznih nizova znakova preuzetih iz zapisa jedne CCS simulacije ili iz iste baze izvješća koja je upotrebljena u posljednjoj fazi podešavanja, iz koje je preuzet i prethodno upotrebljavani ulazni redak „`Digital Forensics Analysis Report`”.

Primjer 4.7 prikazuje generirani tekst iz optimalno naučenog GPT-2 124M modela bez ulaznog niza znakova.

```
In the first place, in this case, you look at the message from the
sender/downloader. I didn’t know how much I’m going to do for a premium
rate service. The message then directs me to an email which appears to be
from an American newspaper based in Turkey.

I’ve tried to get my attention after a short podcast, with one of my
friends who claims to come from the United Arab Emirates.
```

Here’s the video I have posted at the end. It is very easy to receive from people you might like to give an early warning to your friends.

This is the latest in the computer security saga on Facebook. I’ll be going to be working on my behalf as a free premium rate service, and I’ll do my bit of the work this month.

I’ll be doing a couple of questions about this year, but my guess is that there’s nothing that needs to be done with it, and I’ll be doing it for you.

But what should this case be? In this case I would love to have a news story, and I will be giving a very rare time to answer my questions (and, I’m a bit sure to pick up the answer if they’re right… I’ve got a good job at my pocket, but I can’t tell you just how to do it to me or my friend in total love.

I will also be giving away the “News” of the week and bringing together more information about this scam. Over a week and a week I’ll be discussing the latest scams, which will try to spread across Facebook and make a difference to your email account.

Primjer 4.7 Generirani tekst iz optimalno naučenog GPT-2 124M modela nakon druge faze finog podešavanja bez ulaznog teksta

Primjer 4.8 prikazuje generirani tekst iz optimalno naučenog GPT-2 124M modela s ulaznim retkom teksta koji je preuzet iz jedne CCS simulacije.

New Dogecoin Crypto Sale contains a link 'http://webmail.forumofthemall.hr/mail/loging.php' to a website 'Webmail SquirrelMail Popular Forum'.

To gain your account, the site uses an HTTPS connection and then automatically posts the following HTTP proxy request to a remote site in order to send you your traffic to a service called your ISP.

That is a huge task, even if it is a known-yet open HTTPS connection.

If your browser does use SSL for HTTPS, or Firefox 2.3, use some more secure HTTP connection, this attack won't be very surprising.

Many websites rely on HTTPS for their own security, but in order to keep secure online, it isn't always possible to visit.

That's an issue, of course.

As in the last few months of Microsoft and Adobe, they seem to have noticed that in the United States it's important that users don't trust them.

By the way, if you use Internet Explorer 6.5 you should protect against the malicious attack by disabling the use of HTTPS.

If you use Internet Explorer 6.0 you can do a lot worse than upgrade to Internet Explorer 8.0. The attack is now targeting Windows users from Windows XP and Windows 7.1. In this case it's an "Revealed" attack which is being spammed out widely across email.

So it's not surprising for users of Internet Explorer 6.0 are not vulnerable to any attacks – at least in the USA and in Europe.

Primjer 4.8 Generirani tekst iz optimalno naučenog GPT-2 124M modela nakon druge faze finog podešavanja za ulazni redak iz CCS-a: „New Dogecoin Crypto Sale contains a link 'http://webmail.forumofthemall.hr/mail/loging.php' to a website 'Webmail SquirrelMail Popular Forum'."

Rezultati dobiveni uz pomoć GPT-Neo modela su slični prethodno navedenim rezultatima iz GPT-2 modela, s time što je GPT-Neo u 125M inačici resursno zahtjevniji nego GPT-2 u 124M inačici, tj. za podešavanje GPT-Neo 125M modela uz isto sklopovlje, podatke za učenje i parametre je utrošeno više vremena nego za GPT-2 124M.

Primjeri 4.9, 4.10 i 4.11 prikazuju tekst koji generira optimalno naučeni GPT-Neo 125M model uz iste ulazne nizove kao u prethodnim GPT-2 primjerima: bez ulaznog niza, uz ulazni niz iz skupa za učenje, odnosno uz ulazni niz iz CCS-a.

s, hackers have created their own botnets that share spam messages.

A new botnet will be infected as soon as a computer can reach its peak, and the hackers will attempt to steal information about your company and your company that you've been able to log on for some time. It's only possible to be infected on the machine, but even if it has a copy of itself, this could be a significant security risk for your company.

It's certainly an interesting turn for some anti-spam and anti-virus companies, but let's face it, if you've ever encountered a malware attack like this, you'll know that you should be careful about the computer security of its users.

Earlier this week we published another security report about a software zero-day vulnerability in Java which Sophos products were seeing every day.

We are seeing reports of this and other vulnerabilities being found in other versions of Java.

There are several major vulnerabilities currently at the risk, with the largest remote hacker being a sysadmin in your country, and the second being "Bad Antivirus".

Two of the vulnerabilities at the time, both in the Java version of Java, are remote code execution vulnerabilities (RCEs) in Java 2.0 versions, though Java 2.0 versions are vulnerable, and both vulnerabilities could allow malicious code to run and run on the client.

Two of them, CVE-2009-32 and CVE-2009-29, lead you to believe that in one of these attacks you would be able to gain remote control of a Java Java server connected to your web servers – this is one of many ways of exposing your Java computers (even if the Java server is located in the UK).

Fortunately, a reliable web certificate for a Java server is available.

In other words, the Java version of Java is quite legitimate, and we won't be seeing a critical one soon. We will release a patch for the next update in time.

Primjer 4.9 Generirani tekst iz optimalno naučenog GPT-Neo 125M modela nakon druge faze finog podešavanja bez ulaznog teksta

Digital Forensics Analysis Report

The following video, made by Mark Harris, demonstrates how the malware has evolved over recent months and a half. The key points out that Sophos has produced proactive detection for this type of malware for the OSX/iT Trojan on OSX/iTK. It also illustrates how successful these detections are.

This weekend in Sydney SophosLabs, the American lab got infected with Troj/Spy-B. The files on that hard drive were not at all.

The first thing to do, is download a disassembler and run it. It is difficult to tell whether the infected files are actually malicious or not. The following video demonstrates just how these infections have occurred over the past few days.

It appears that a hacker has been able to gain access to a company's servers by planting malware on a website using the "spyware" file format.

The following video of the attack shows what's to be believed, at least on any website.

Somehow the attack can be seen below.

A number of compromised webpages that have the domain name of the web server with the file format file are detected as Troj/Spy-B.

The following chart shows that the infected webpages are blocked:

The good news is that the victim is not the victim of this attack – this was not the result of this hack.

The bad news is that some of the sites that are hacked are related to security software applications. As a result, their sites have been compromised, with several compromised domains having been compromised.

One of the biggest concerns was, of course, whether the site's owners could make up the password being used to post to their own blog feed. It was very unusual to see website users' comments on the situation.

As soon as the victim's PC was updated, the Trojan was able to run itself. On the same day the page was not compromised and the malware was able to spread itself.

What is perhaps of more interest to be found is what's to be believed, rather than the attack being targeted by hackers. In an interview published on BBC Radio Five Live, the hacker claim that he broke into the websites for his criminal activities, and has posted photographs of young women and girls in front of a message message about a child sex video.

As this attack appears to have originated over the past few weeks, the truth will no doubt be that the hackers may have gained access to several other online websites. The latest piece of evidence suggests that the hackers are using the internet forum, which allows them to post pictures of young women in front of other victims.

Primjer 4.10 Generirani tekst iz optimalno naučenog GPT-Neo 125M modela nakon druge faze finog podešavanja za ulazni redak teksta „Digital Forensics Analysis Report"

New Dogecoin Crypto Sale contains a link 'http://webmail.forumofthemall.hr/mail/loging.php' to a website 'Webmail SquirrelMail Popular Forum'.

The site also contains a link, however the page was a compromised website containing a malicious script that was proactively detected as Mal/ObfJS-C. The script contained the file "Spam mail/spam.php" and contained the malicious JavaScript file:

The malicious script also contained two files (Troj/Dloadr-C and Troj/Dloadr-D) the files have been detected as Mal/Dloadr-D. The first file contained a malicious script detected as Mal/EncPk-C.

The second file contained the relevant files, W32/Spam-B and W32/Spam-B. The malicious script detected as Mal/Dloadr-C will point to the malicious site and its malicious DNS records. This is a bit of an overkill of some sites.

The bad news is that the DNS records of the hacked site are not necessarily associated with a malicious script. So if the sites was compromised, it would have been safe to clean up any malicious webpages using this method.

Primjer 4.11 Generirani tekst iz optimalno naučenog GPT-Neo 125M modela nakon druge faze finog podešavanja za ulazni redak iz CCS-a: „New Dogecoin Crypto Sale contains a link 'http://webmail.forumofthemall.hr/mail/loging.php' to a website 'Webmail SquirrelMail Popular Forum'."

5. Zaključak

Rezultati ovog rada su pokazali da se uporabom modela prirodnog jezika zasnovanim na strojnom učenju zaista mogu generirati tekstovi koji bolje odgovaraju stvarnim specijalističkim izvješćima nego uobičajenim generiranjem na temelju predložaka, što je korisno u primjenama kao što je *Cyber Conflict Simulator* za uvježbavanje osoblja u području informacijske i kibernetičke sigurnosti. Model prirodnog jezika se za takve svrhe mora naučiti i fino podesiti uz pomoć odgovarajućih postojećih podataka, na sličan način kao što se izučavaju stvarni specijalisti. Učenje i fino podešavanje modela se mogu izvoditi u više faza, počevši od osnovnog znanja prirodnog jezika, zatim od šireg područja primjene sve do najuže specijalizacije, što odgovara tijeku obrazovanja kroz koji prolazi osoba.

Modeli prirodnog jezika i tehnike strojnog učenja su još uvijek u ranim fazama razvoja u trenutku pisanja ovog rada, pa je njihova upotrebljivost pri rješavanju određenog problema još uvijek bitno ograničena raspoloživim sklopovskim resursima. Ipak, uz daljnji razvitak algoritama, programske podrške i sklopovlja koji omogućuju strojno učenje, u budućnosti se od njega mogu očekivati veća raspoloživost i bolji rezultati. Strojno učenje je samo jedna od brojnih tehnologija koje se aktivno razvijaju i čija važnost u svijetu raste. Kako bi postojeća programska rješenja ostala konkurentna na tržištu, nužna su stalna ulaganja u istraživanje i razvoj uz uporabu novih tehnologija, što je osobito važno u području sigurnosti jer se na brojnim primjerima pokazalo da loša sigurnosna rješenja mogu rezultirati katastrofalnim posljedicama.

6. Literatura

[1] Utilis d.o.o. i Sveučilište u Zagrebu Fakultet elektrotehnike i računarstva. *Cyber Conflict Simulator* (2017). Poveznica: https://ccs.utilis.biz/; pristupljeno 6. lipnja 2022.

[2] Stroz Friedberg LLC. Izvješće digitalne forenzičke analize u sudskom postupku *Paul D. Ceglia v. Mark Elliot Zuckerberg, Individually, and Facebook, Inc., Civil Action No: 1:10-cv-00569-RJA* (2012). Poveznica: https://www.wired.com/images_blogs/threatlevel/2012/03/celiginvestigation.pdf; pristupljeno 6. lipnja 2022.

[3] Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. *Improving language understanding by generative pre-training* (2018).

[4] Devlin, J., Chang, M. W., Lee, K., Toutanova, K. Bert: *Pre-training of deep bidirectional transformers for language understanding* (2018).

[5] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. *Language models are unsupervised multitask learners*. OpenAI blog, 1(8), 9 (2019).

[6] EleutherAI. GPT-Neo (2020). Poveznica: https://github.com/EleutherAI/gpt-neo; pristupljeno 8. lipnja 2022.

[7] Woolf, M. aitextgen (2019), Poveznica: https://github.com/minimaxir/aitextgen; pristupljeno 8. lipnja 2022.

[8] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P. i dr. *Language models are few-shot learners*. *Advances in neural information processing systems* 33 (2020), str. 1877-1901.

[9] Google Colaboratory (2018). Poveznica: https://colab.research.google.com/; pristupljeno 9. lipnja 2022.

[10] Scrapy (2015). Poveznica: https://scrapy.org/; pristupljeno 10. lipnja 2022.

[11] Python Package Index (2003). Poveznica: https://pypi.org/; pristupljeno 25. lipnja 2022.

[12] Richardson, L., BeautifulSoup (2004). Poveznica: https://www.crummy.com/software/BeautifulSoup/; pristupljeno 10. lipnja 2022.

[13] Krebs, B. *Conti's Ransomware Toll on the Healthcare Industry*, *Krebs on Security* (18. travnja 2022.). Poveznica: https://krebsonsecurity.com/2022/04/contis-ransomware-toll-on-the-healthcare-industry/; pristupljeno 10. lipnja 2022.

[14] Malmgren, D. textract (2014). Poveznica: https://github.com/deanmalmgren/textract; pristupljeno 11. lipnja 2022.

[15] Pilgrim, M., Rose, E., Blanchard, D., Cordasco, I. chardet (2011). Poveznica: https://chardet.github.io/; pristupljeno 11. lipnja 2022.

Sažetak

Naslov: Strojno generiranje izvješća specijalista tijekom rukovanja sigurnosnim incidentom

Ovaj diplomski rad iznosi mogući način poboljšanja strojnog generiranja specijalističkih izvješća za potrebe simulacije sigurnosnih incidenata uz pomoć modela prirodnog jezika i pokazuje ga na nekoliko primjera.

Upotrebljeni su GPT-2 i GPT-Neo modeli prirodnog jezika zasnovani na strojnom učenju, koji su unaprijed naučeni na skupu općih tekstualnih podataka. Unaprijed naučeni model se fino podešava u dvije faze, najprije na skupu tekstualnih podataka iz šireg područja informacijske i kibernetičke sigurnosti, a nakon toga na manjem skupu podataka koji točno odgovaraju specijalističkim izvješćima u traženom obliku.

Ključne riječi: informacijska i kibernetička sigurnost, incident, specijalist, izvješće, simulacija, generiranje teksta, model prirodnog jezika, strojno učenje, Transformer, GPT-2, GPT-Neo, Cyber Conflict Simulator

Summary

Title: Computer Generation of Specialist Reports during Security Incident Handling

This master's thesis presents a possible means of improving machine generation of specialist reports for the purpose of simulating security incidents using natural language models and shows it in several examples.

GPT-2 and GPT-Neo natural language models based on machine learning were used, which were pre-trained using a generic text dataset. The pre-trained model is fine-tuned in two phases, first using a dataset in the broader field of information and cyber-security, and then using a smaller dataset that exactly matches specialist reports in the required format.

Keywords: information and cyber-security, incident, specialist, report, simulation, text generation, natural language model, machine learning, Transformer, GPT-2, GPT-Neo, Cyber Conflict Simulator