ŠOLSKI CENTER VELENJE
ELEKTRO IN RAČUNALNIŠKA ŠOLA
TRG MLADOSTI 3, 3320 VELENJE

MLADI RAZISKOVALCI ZA RAZVOJ ŠALEŠKE DOLINE

RAZISKOVALNA NALOGA

PREPOZNAVA GLOBOKIH PONAREDKOV

Tematsko področje: računalništvo

Avtorja:
Tim Povše, 4. TRA
Tim Jevšenak, 4. TRA

Mentorja:
Islam Mušić, prof.
Rok Urbanc, dipl. inž. rač. in inf. tehnol. (UN)

Velenje, 2021

Raziskovalna naloga je bila opravljena na Elektro in računalniški šoli Velenje.

Mentorja: Islam Mušić, prof.
Rok Urbanc, dipl. inž. rač. in inf. tehnol. (UN)

Datum predstavitve: april 2021

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

ŠD ŠC Velenje, Elektro in računalniška šola, 2020/2021

KG globoki ponaredek / deepware / detekcija / algoritem

AV POVŠE, Tim / JEVŠENAK, Tim

SA MUŠIĆ, Islam / URBANC, Rok

KZ 3320 Velenje, SLO, Trg mladosti 3

ZA ŠC Velenje, Elektro in računalniška šola

LI 2021

IN **PREPOZNAVA GLOBOKIH PONAREDKOV**

TD Raziskovalna naloga

OP IX, 49 s., 11 graf., 32 sl.

IJ SL

JI sl / en

AI Globoki ponaredki, znani pod izrazom ‘deepfake’, so oblika medijev, v katerih je človeški obraz spremenjen na tak način, da ponaredek izgleda verodostojen. Raziskovalci so se lotili procesa prepoznave globokih ponaredkov. Zanimale so jih že obstoječe rešitve in njihova učinkovitost. Pri izdelavi lastnih globokih ponaredkov so ugotovila, da je ta tehnologija enostavna za uporabo in dostopna vsakomur. Veliko ponaredkov je tako dobro izdelanih, da jih večina ljudi ne prepozna. Raziskovalci so izdelali anketo, da bi ugotovili, kako dobro prepoznajo globoke ponaredke njihovi vrstniki. Rezultati ankete so pokazali, da jih v večini primerov prepoznajo. Raziskovalci so izdelali tudi aplikacijo, ki z algoritmom za prepoznavo globokih ponaredkov preverja prisotnost posnetkov. V času, ko vedno več komunikacije poteka preko spleta, so postali globoki ponaredki vse bolj popularni. Hkrati predstavljajo veliko grožnjo, saj ne moremo več realno oceniti verodostojnosti videoposnetkov, zato je prepoznavanje globokih ponaredkov izredno pomembno.

KEY WORDS DOCUMENTATION

ND ŠC Velenje, Elektro in računalniška šola, 2020/2021

CX deepfake / deepware / detection / algorithm

AU POVŠE, Tim / JEVŠENAK, Tim

AA MUŠIĆ, Islam / URBANC, Rok

PP 3320 Velenje, SLO, Trg mladosti 3

PB ŠC Velenje, Elektro in računalniška šola

PY 2021

TI **DEEPFAKE DETECTION**

DT RESEARCH WORK

NO IX, 49 p., 11 graf, 32 fig.

LA SL

AL sl / en

AB Deepfakes are a form of media in which the human face is altered in such a way that the counterfeit looks authentic. Researchers have embarked on a process to identify deepfakes. They were interested in existing solutions and their effectiveness. In making their own deep fakes, they found that this technology is easy to use and accessible to anyone. Many fakes are so well made that most people don't even recognize them. Researchers conducted a survey to find out how well their peers recognize deepfakes. The results of the survey showed that they are recognizable in most cases. Researchers have also developed an application that uses an algorithm to detect deepfakes to verify the authenticity of videos. At a time when more and more communication is taking place online, deepfakes have become increasingly popular. At the same time, they pose a major threat, as we can no longer realistically assess the authenticity of videos, so identifying deepfakes is extremely important.

KAZALO VSEBINE

KAZALO SLIK

KAZALO GRAFOV

SEZNAM OKRAJŠAV

npr.	na primer
ipd.	in podobno
t.i.	tako imenovani
API	aplikacijski programski vmesnik
oz.	oziroma

1 UVOD

Vsakodnevna uporaba interneta je postala naša stalnica. Verjameva, da med nami ni nikogar, ki ne bi vsaj enkrat na dan pogledal novic na internetu. Le redki pa se vprašajo, ali so vse prebrane novice tudi resnične. V zadnjem času se je na spletu pojavilo veliko število lažnih novic, še sodobnejša izvedba pa so lažne fotografije in videoposnetki. Čemu sploh služijo vse te ponarejene vsebine? Marsikdaj nastanejo zgolj zaradi zabave, z njimi pa lahko tudi zavajamo ljudi in s tem širimo lažne novice ali pa spremenimo mišljenje o stvareh ali ljudeh, ki so v njih vključeni.

Deepfake ali »globoki ponaredek« je tehnologija, ki temelji na strojnem učenju, ki ga imenujemo globoko učenje. S pomočjo videoposnetka osebe se nauči, kakšne so njene značilne poteze, kako premika usta in kako se spreminjajo obrazne poteze med govorjenjem. Vse to nato uporabi tako, da lahko zamenjamo obraz osebe z obrazom nekoga drugega, ali pa da naredimo, kot da oseba govori nekaj, česar nikoli ni izrekla. Še posebej slednja je zelo nevarna, saj je lahko uporabljena za manipulacijo informacij, podanih preko videa.

Tehnologija globokih ponaredkov je v zadnjem času začela postajati vse bolj napredna. To predstavlja veliko grožnjo verodostojnosti podatkov v času, ko je vse več informacij podanih preko spleta ali v obliki videoposnetkov.

Z raziskovalno nalogo sva želela pobližje spoznati globoke ponaredke. V ta namen sva nekaj ponaredkov ustvarila tudi sama in s tem preizkusila, kako zahtevno je narediti dober globok ponaredek. Prav tako naju je zanimalo, kako dobro znajo najini vrstniki razlikovati med pravimi posnetki in globokimi ponaredki. Ker bomo v prihodnosti očitno nujno potrebovali programe za prepoznavanje globokih ponaredkov, je bila največja teža najine raziskave usmerjena k uporabi in primerjavi različnih detektorjev, ki so namenjeni prepoznavi globokih ponaredkov.

1.1 HIPOTEZE

1. Uspelo nama bo izdelati lasten globok ponaredek.

2. Orodja za prepoznavo globokih ponaredkov so sposobna prepoznati večino le-teh.

3. Večina anketirancev ne bo ločila med globokim ponaredkom in pravim posnetkom.

4. Uspelo nama bo izdelati aplikacijo, ki bo prepoznala globoke ponaredke.

2 PREGLED OBJAV

2.1 KAJ JE GLOBOKI PONAREDEK

Globoki ponaredek je ime za videoposnetke ali slike, ki so bili s pomočjo globokega učenja spremenjeni tako, da prikazujejo spremenjeno vsebino, hkrati pa poskušajo izgledati čimbolj naravno in so narejeni tako, da jih je zelo težko prepoznati.

Tehnologija globokih ponaredkov, ustvarjanje lažnih, ponarejenih videoposnetkov, ki pa so videti neverjetno realistično, se pojavlja vse pogosteje. In če so nekateri posnetki ustvarjeni predvsem kot parodija, za zabavo spletnih uporabnikov, ta ista tehnologija predstavlja tudi nevarnost – predvsem z vidika morebitnega manipuliranja javnega mnenja.

Globoke ponaredke je namreč težko prepoznati kot lažne. Spletne manipulacije pa so to pripeljale do točke, ko praktično ne moremo več presoditi, kaj je resnično in kaj lažno. Tovrstna tehnologija se hitro razvija, predvsem pa postaja vse bolj enostavna in dostopna.

Morda enega najbolj znanih globokih ponaredkov so ustvarili raziskovalci z Univerze v Washingtonu, ki so z globokim ponaredkom nekdanjega ameriškega predsednika Baracka Obame poskušali pokazati, kako se tovrstna tehnologija lahko zlorablja. (Elon Musk …, 2020)

Vzpon tehnologije globokih ponaredkov lahko označimo za globalni problem kot podnebne spremembe, ki pa jih za spremembo dejansko lahko merimo in do neke mere predvidimo, saj so dolgotrajen pojav, o katerem imamo že kar precej informacij. Tehnologija globokih ponaredkov pa je tako nova, da preprosto ni mogoče oceniti, kako prevladujoča je in kako hitro se razvija ter kakšne vplive na družbo s seboj prinaša. A zakaj dejansko je problem? Marsikdo bi namreč lahko rekel, da je tehnologija samo še eden izmed načinov zaslužka v medijski industriji oz. kar digitalni ekonomiji. A problem je (lahko) precej resnejši. Tehnologija je v obdobju manj kot dveh let napredovala od algoritma, ki omogoča preprosto, a vizualno prepričljivo zamenjavo obraza in/ali zvoka, do tega, da lahko v točno določenih okoliščinah in stanju ciljne publike že povzroči masovno manipulacijo, in to v realnem času. (Gorenšek, 2019)

2.2 NASTANEK GLOBOKEGA PONAREDKA

Glavna sestavina globokih ponaredkov je strojno učenje, ki je omogočilo hitrejšo izdelavo ponaredkov. Če želite ustvariti globok ponaredek nekoga, bi ustvarjalec najprej več ur treniral nevronsko mrežo na videoposnetkih osebe, da bi računalnik razumel, kako je videti z različnih

zornih kotov in pod različno osvetlitvijo. Nato bi s pomočjo naučene nevronske mreže te obrazne poteze prilepili na obraz, ki se nahaja v nekem drugem videu. (Adee, 2020)

Čeprav uporaba umetne inteligence proces naredi hitrejši, kot je bil kadarkoli prej, pa treniranje nevronskih mrež, še posebej če želimo prepričljive rezultate, traja dolgo časa. Ustvarjalec globokega ponaredka mora pogosto ročno nastaviti določene parametra in s tem pripomore sami prepričljivosti videoposnetka.

2.3 ZGODOVINA GLOBOKIH PONAREDKOV

Vse skupaj se je začelo z objavo članka na spletnem mediju Vice, ki razkriva uporabo algoritma, s katerim je Reddit uporabnik z imenom *deepfakes* zamenjal obraz pornografske igralke z obrazom Gal Gadot, zvezdnice filma Wonder Woman. Da, začelo se je s pornografijo in spolnostjo, kar nas niti ne sme presenetiti, saj bomo verjetno precejšen del tehnološkega napredka na področjih, kot so robotika, umetna inteligenca ipd., dosegli ravno zaradi čedalje intenzivnejših seksualnih oz. spolnih potreb družbe, kar lahko pripišemo razmahu pornografije v zadnjih 50 letih, predvsem na račun razvoja IK-tehnologije. Prav tako ni skrivnost, da gre napredek v smeri večje podobnosti s človekom, tako fizične kot umske. Ravno v tem strmenju za resničnostjo tehnologije se skriva še en izziv, saj bo prepoznava globokih ponaredkov čedalje bolj otežena, medtem ko bodo možnosti za uspešno manipulacijo percepcije čedalje višje.

V januarju 2018 je Vice objavil nov članek na to temo, Reddit, uporabnik *deepfakes,* pa je medtem nabral več kot 15.000 naročnikov, beseda "deepfake" pa je postajala čedalje bolj trendovska. Neki drug Reddit uporabnik je v tem času razvil aplikacijo, s katero so lahko uporabniki na spletu brez računalniškega znanja ustvarjali "deepfake" vsebine. Sledile so bolj napredne aplikacije, temelječe na uporabniški izkušnji, s čimer je bil, če smo iskreni, izpolnjen najbolj pomemben pogoj za kateri koli digitalni pripomoček – to je uporabniška izkušnja, ki morda tudi zaradi pomanjkanja računalniške pismenosti pri uporabnikih, zavzema višje mesto kot varnost. Orodja za masovno manipulacijo in generiranje ponarejenih video vsebin so bila torej nared za povprečne uporabnike interneta pred manj kot dvema letoma. Razvoj tehnologije globokih ponaredkov se je torej začel v internetnih skupnostih, kot je Reddit. A hkrati je razvoj te tehnologije potekal tudi v akademskih institucijah, konkretno na področju računalniškega vida, ki je podpodročje računalniške znanosti.

Leta 1997 je bil ustvarjen program Video Rewrite, ki je omogočal spremembo glasu osebe v videu z glasom osebe z drugega zvočnega posnetka. Takšne primere smo pogosto lahko gledali

v različnih akcijskih filmih iz Hollywooda. Danes v filmih najdemo ravno primere uporabe tehnologije globokih ponaredkov, s katero akterji spreminjajo percepcijo ciljnega občinstva. Zanimivo, kako tovrstna tehnologija avdiovizualne manipulacije ves čas pronica skozi medijsko industrijo, a več o tem v nadaljevanju. Akademska skupnost je ves čas fokus usmerjala v projekte, ki so omogočali ustvarjanje čedalje bolj realističnih videov ter seveda samo izboljšavo tehnologije. Kakor z vso preostalo (digitalno) tehnologijo je bilo tudi v tem primeru samo vprašanje časa, kdaj bo tehnologijo pričela uporabljati poleg filmske/produkcijske še preostala industrija. Številna (digitalna) podjetja za uporabo aplikacij, ki omogočajo ustvarjanje "deepfakov", ne zaračunavajo ničesar, torej ne ustvarjajo direktnega finančnega prihodka, temveč ga neizbežno ustvarjajo z zbiranjem uporabnikovih osebnih podatkov, na čemer vsaj za zdaj temelji velik del digitalne ekonomije. Tovrstne aplikacije, ki so (bile) dostopne brezplačno v spletnih trgovinah, kot sta App store in Google Play, so postale viralne in povzročile takojšen odziv v kontekstu zaščite zasebnosti, v katero uporaba tovrstne tehnologije na nepremišljen ali zlonameren način tako grobo posega. Takšne primere lahko najdemo po vsem svetu – v ZDA, Evropi, tudi na Kitajskem in v Rusiji. (Gorenšek, 2019)

2.4 ALGORITMI ZA USTVARJANJE GLOBOKIH PONAREDKOV

V zadnjem času je opaziti veliko povečanje algoritmov za ustvarjanje globokih ponaredkov. Kljub velikemu številu novih algoritmov pa so zaradi dodelanosti še vedno najbolj uporabljeni določeni algoritmi, ki so že bolj znani v svetu globokih ponaredkov. Najbolj uporabljeni algoritmi so po navadi seveda tudi najbolj dodelani in zaradi tega lahko z njimi kreiramo najbolj prepričljive in neopazne globoke ponaredke. Nekaj izmed bolj uporabljenih bova tudi naštela.

Algoritmi za ustvarjanje globokih ponaredkov trenirajo modele. Za vsako osebo je potrebno kreirati nov model. Model pri globokih ponaredkih je skupek podatkov. V modelu je zapisano vse, kar se s pomočjo umetne inteligence nauči o osebah, ki jih preučuje, na primer, kako premikajo usta, kakšne obrazne poteze imajo, kako mežikajo in podobno. Dlje kot treniramo model z istima osebama, več se bo naučil in bolj podobno jih bo znal oponašati.

2.4.1 DEEPFACELAB (DFL)

DeepFaceLab je orodje za ustvarjanje globokih ponaredkov, ki ga je ustvaril Github uporabnik z imenom Ivan Perov 'iperov'. Poleg njega je projektu pripomoglo še 18 sodelavcev.

Napisano je v skriptnem programskem jeziku Python na osnovi Googlove platforme za strojno učenje Tensor-Flow. Perov se je pri ustvarjanju orodja DeepFaceLab odločil, da bo uporabnike

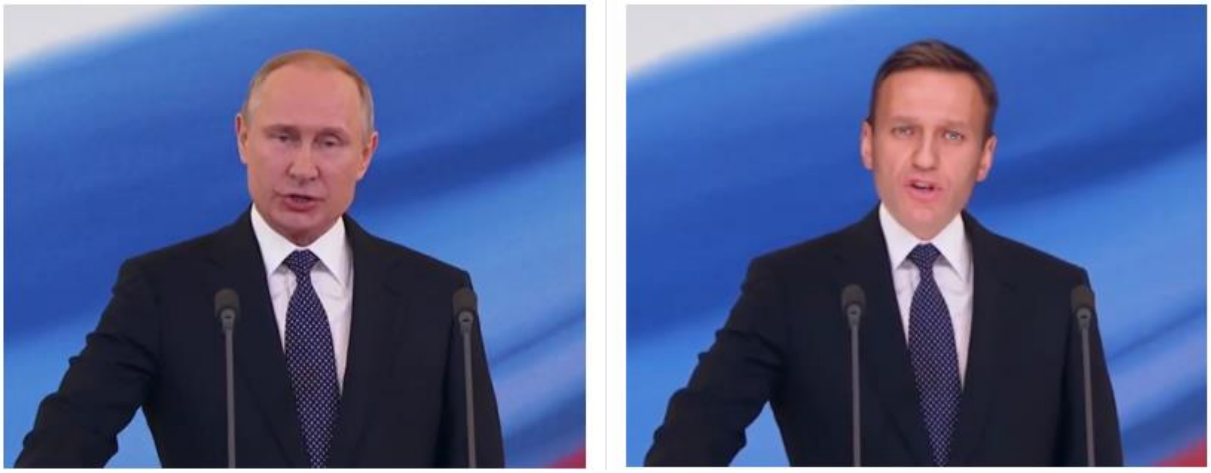
postavil na prvo mesto. Zaradi tega je naredil orodje, ki je lahko za uporabo, hkrati pa je izjemno hitro in ima velik potencial. Sam algoritem si lahko naloži na računalnik vsak in na njegovi spletni strani je podroben opis, kako se ga uporablja ter kaj katera funkcija naredi.

Prav tako je to orodje odprtokodno in modularno. To pomeni, da si vsak lahko na svoj računalnik naloži izvorno kodo in jo spreminja, prav tako pa je koda napisana tako, da je spreminjanje določenih modulov zelo lahko in celo zaželeno.

DeepFaceLab je najbolj uporabljen izmed vseh orodij, saj je tehnološko dovršen in ima veliko raznih funkcij. Prav tako je modularen in uporabniku prijazen algoritem. Omogoča tri vrste globokih ponaredkov.

Prva vrsta so posnetki, kjer algoritem med danima videoma zamenja obraze. Najprej preuči oba videa, kakšne obrazne poteze imata, kako se ob govorjenju premikajo usta in podobno, nato pa ju zamenja.

Druga vrsta so posnetki, kjer algoritem zamenja celotno glavo. S tem zamenja med osebama vse, kar je od vratu navzgor. Z uporabo te vrste ponaredkov lahko preprosto zamenjamo celoten obraz dvema osebama.



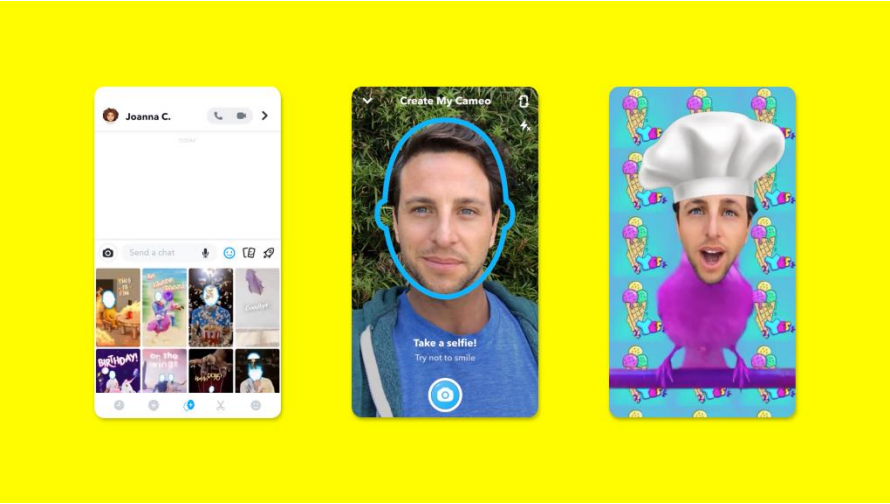
Slika 1 - Na levi vidimo original, na desni pa zamenjano celotno glavo. (Perov, 2020)

Tretja in zadnja vrsta ponaredkov, ki jih DeepFaceLab podpira, pa so globoki ponaredki, s pomočjo katerih lahko pomlajšamo osebo na posnetku. Delovanje te funkcije je lepo predstavljeno na primeru pomlajšanja Arnolda Schwarzeneggerja.

Slika 2 - Na levi Schwarzenegger pred pomlajšavo, na desni pa po pomlajšavi

2.4.2 MOBILNE APLIKACIJE

Poleg resnejših orodji za ustvarjanje globokih ponaredkov lahko za kreiranje le-teh uporabimo tudi katero izmed mnogih mobilnih aplikacij, ki so na voljo na internetu. Že nekatere izmed bolj uporabljenih aplikacij za socialna omrežja, kot so Snapchat ali Instagram, imajo vgrajene algoritme za kreiranje globokih ponaredkov, vendar to niso takšni globoki ponaredki, kot smo jih navajeni. Večinoma so algoritmi v teh aplikacijah uporabljeni za uporabo raznih filtrov, kot na primer na Snapchatu. Tam lahko svoj obraz prilepimo v različne videoposnetke oseb. Temu sami rečejo »cameo«, v bistvu pa je to ena izmed najpreprostejših vrst globokega ponaredka. Vzame naš obraz in ga zamenja z obrazom igralca, ki je v tistem videu dejansko igral. Po potrebi seveda doda tudi mežikanje ali premikanje in odpiranje ust.



Slika 3 - Primer Snapchat Cameo (Snapchat Cameo brings ..., 2019)

Aplikacije, ki so specializirane za kreiranje globokih ponaredkov, lahko po navadi v relativno kratkem času naredijo dober globoki ponaredek. Ti globoki ponaredki večinoma niso visoke kakovosti in so mišljeni le za zabavo. Dober primer takšne aplikacije je Reface. Na aplikaciji Reface lahko že le z eno našo sliko naredijo v nekaj sekundah globoki ponaredek. Na tej aplikaciji lahko svoj obraz prilepimo v enega izmed mnogih že vnaprej pripravljenih videoposnetkov. Po navadi so to kratki smešni videi. Zanimivo je dejstvo, da je aplikacija popolnoma brezplačna, saj porabi ustvarjanje globokih ponaredkov velike količine strežniške moči.

Slika 4 - Primer Reface ponaredka (Reface, 2021)

V zgornjem desnem kotu slike vidimo obraz, iz katerega je aplikacija jemala obrazne poteze. Nato je te obrazne poteze s pomočjo umetne inteligence združila s premikanjem igralke na originalnem videoposnetku in nastal je posnetek, ki ga vidimo na sliki.

2.5 ALGORITMI ZA PREPOZNAVANJE GLOBOKIH PONAREDKOV

Za razliko od algoritmov za ustvarjanje globokih ponaredkov je algoritmov za prepoznavanje globokih ponaredkov veliko manj, prav tako pa jih je večina nedostopnih vsakdanjemu uporabniku. Tako moramo za uporabo le-teh stopiti v kontakt s podjetji, ki se ukvarjajo z razvojem le-teh. Prav tako zaznavanje globokih ponaredkov še ni dovršeno, saj je vsak ponaredek unikaten. Še posebej težko je zaznati ponaredke, ki jih uporabniki še dodatno obdelajo z uporabo raznih orodij, kot je Adobe After Effects.

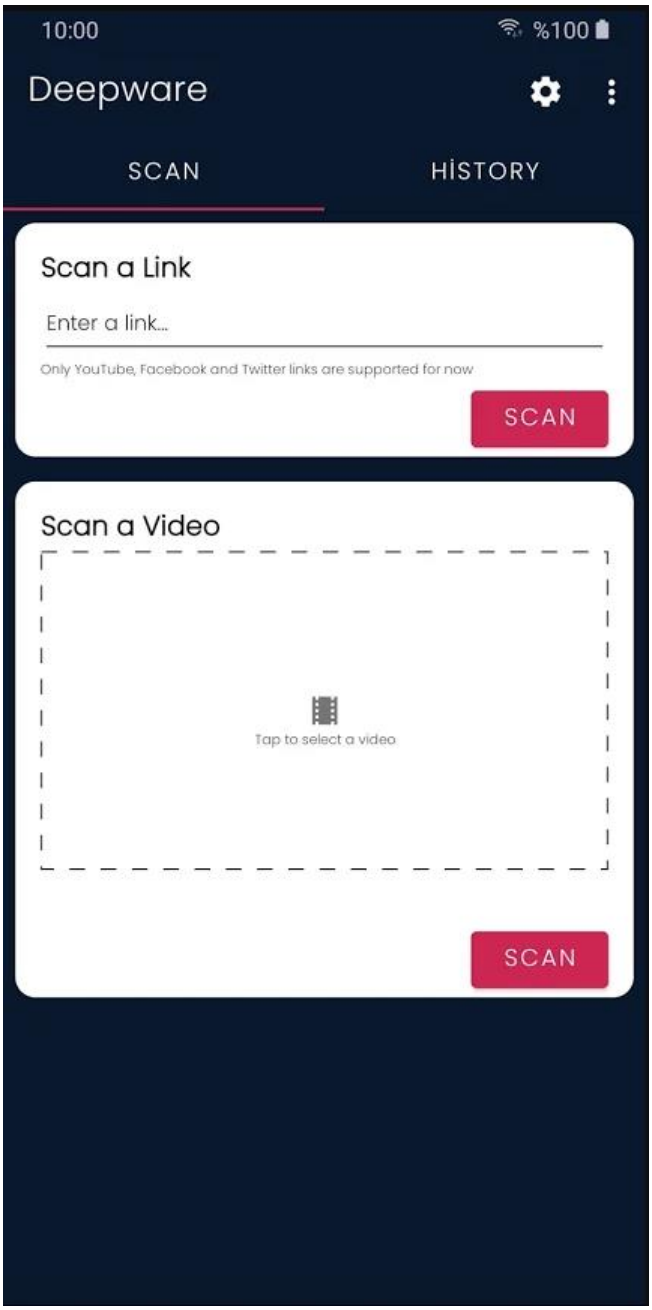
Prepoznavanje globokih ponaredkov je postal velik problem, ki se bo s časom le še povečal, saj se lahko s ponarejenimi videi širijo lažne informacije. V skrajnih primerih se z njimi lahko tudi oškoduje ali prizadene ljudi. Problem zaznavanja globokih ponaredkov je v tem, da je razlika med pravim videom in globokim ponaredkom zelo mala. Prav tako razlike niso konstantne in so odvisne od uporabe orodja. Pri enem orodju je lahko pokazatelj globokega ponaredka ena stvar, pri drugem pa nekaj čisto drugega.

2.5.1 DEEPWARE

Deepware je turško podjetje, ki se ukvarja z zaznavanjem globokih ponaredkov. Z zaznavanjem globokih ponaredkov so se začeli ukvarjati že leta 2018, ko je njihovo starševsko podjetje Zemana razvilo antivirusni program, ki uporablja umetno inteligenco, in tako se je začel razvoj detektorja globokih ponaredkov.

Njihov detektor je odprtokoden, kar pomeni, da si lahko vsak ogleda izvorno kodo in jo spremeni po želji, prav tako pa ga lahko vsak uporablja. Med izdelavo najine raziskovalne naloge so izdali tudi aplikacijo, podobno najini, preko katere lahko uporabniki preverijo, ali so videi globoki ponaredki ali niso.

Aplikacija, ki so jo izdali, je dostopna vsem uporabnikom naprav, ki jih poganja operacijski sistem Android, saj je trenutno na voljo le v Google Play trgovini. V aplikaciji imajo uporabniki neomejeno število skenov, ki jih lahko opravijo in se s tem prepričajo, da njihovi videoposnetki niso na nikakršen način spremenjeni.



Slika 5 - Zaslonski posnetek aplikacije (Deepware, 2020)

2.5.2 QUANTUM INTEGRITY

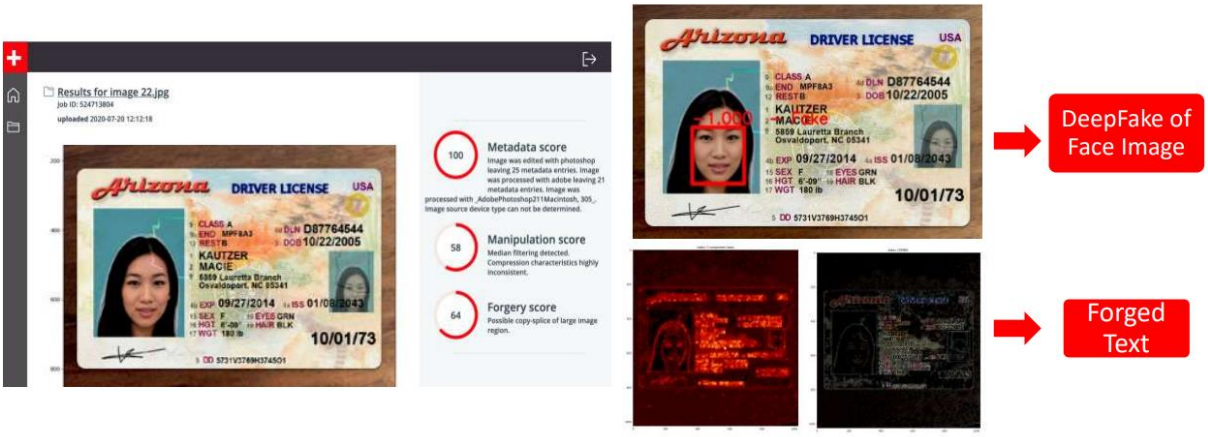
Quantum Integrity je vodilno švicarsko podjetje, ki se ukvarja z zaznavo globokih ponaredkov. Za razliko od podjetja Deepware se Quantum Integrity večinoma ukvarja z zaznavo preurejenih in ponarejenih dokumentov ali slik.

Del njihovega podjetja se ukvarja z zaznavanjem možnih zavarovalniških goljufij, saj so nekatere zavarovalnice prijavljanje škode na avtu začele opravljati digitalno, tako pa lahko vsak uredi sliko avta, da izgleda poškodovan in tako dobi zavarovalniški denar. Pri Quantum

Integrity so ravno v ta namen razvili poseben detektor. Njihov detektor s pomočjo umetne inteligence prepozna slike, ki so bile urejene ali na kakršenkoli način spremenjene in na to opozori uporabnika.

Slika 6 - Primer prepoznanega ponaredka (Quantum Integrity, 2021)

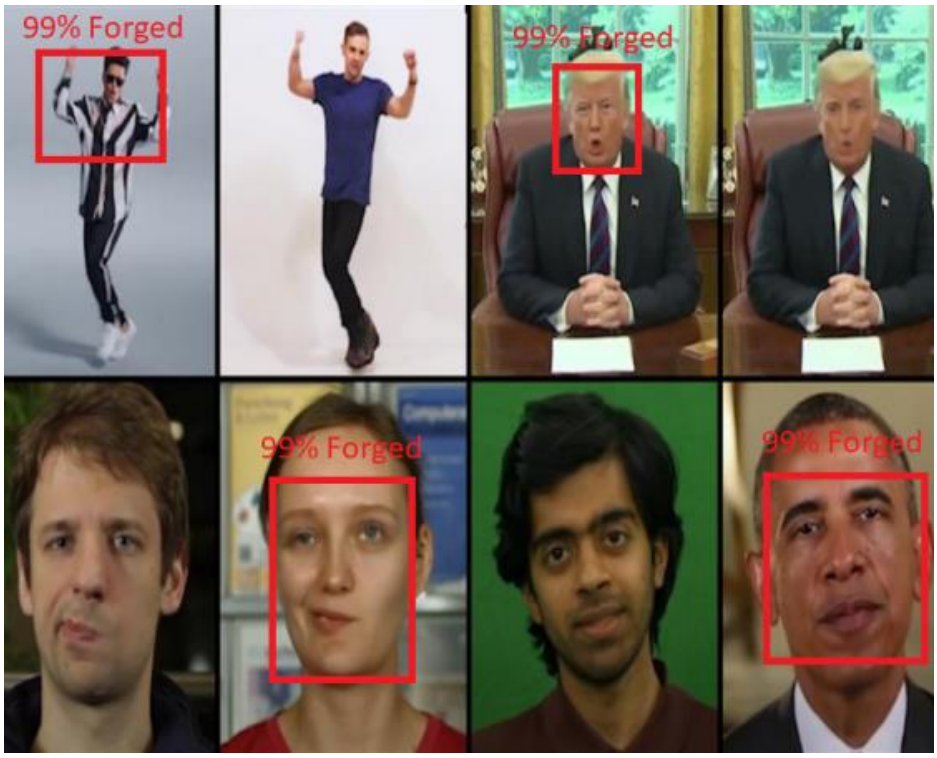
Drugi del njihovega podjetja pa se ukvarja s prepoznavo ponarejenih dokumentov, kot so bančne kartice, potni listi in podobno. Ponarejeni potni listi in osebne izkaznice so še posebej nevarni, saj gre za krajo identitete. Pri Quantum Integrity so razvili detektor tudi za takšno vrsto ponaredkov, kot vidimo na spodnji sliki. S slike je jasno razvidno, da je detektor zaznal ponarejeno vozniško izkaznico in točno prikazal, kaj je ponarejeno in kje.



Slika 7 - Prepoznan ponarejen dokument (DeepFake in KYC, 2021)

Pred kratkim se je podjetje začelo ukvarjati tudi s prepoznavo globoko ponarejenih videoposnetkov, saj so tudi ti še posebej zaradi epidemije korona virusa postali velik problem. Tudi sama sva se z njihovim podjetjem povezala in primerjala njihov detektor z detektorjem podjetja Deepware. Z najinimi podatki bodo svoj algoritem še dopolnili in izboljšali.

Njihov namen je ohraniti integriteto podatkov v modernem svetu. Skrbi jih tudi podajanje informacij preko videoposnetkov in govori znanih politikov, ki bi jih lahko kdo oškodoval, če bi videoposnetke ponaredil.



Slika 8 - Prikaz delovanja detekcije (Quantum Integrity, 2021)

3 METODE DELA

3.1 USTVARJANJE GLOBOKEGA PONAREDKA

V sklopu raziskovalne naloge sva se odločila, da bova poskusila izdelati lasten globoki ponaredek. Prva odločitev je bila, katero orodje bova uporabila. Želela sva uporabiti orodje, ki bi nama omogočalo popolno prilagajanje najinim potrebam. Izdelava ponaredka s pomočjo mobilnih aplikacij se nama ni zdela smiselna, saj so namenjene širši uporabi in ne ustvarijo globokega ponaredka, ki bi ga lahko kdorkoli zamenjal z originalom. Želela sva narediti takšen ponaredek, da bi brez povečane pozornosti zelo težko opazila, da video ni originalen, ter da je globoki ponaredek.

Po daljšem raziskovanju po spletu sva ugotovila, da je najboljši algoritem za najine namene DeepFaceLab, saj omogoča popolno prilagoditev glede na želje uporabnika, prav tako pa ob pravilni uporabi in dovolj dolgem času učenja omogoča rezultate, na katerih je skoraj neopazno, da je video globoki ponaredek.

3.1.1 IZBIRA VIDEOPOSNETKOV

Za najin globoki ponaredek sva morala najprej najti videe, ki bi jih rada ponaredila. Glede na temo raziskovalne naloge sva se odločila, da bova poskusila ustvariti posnetek, kjer se znan košarkar Luka Dončić zahvaljuje mladim raziskovalcem.

Kot videoposnetek Luka sva vzela video, kjer se zahvaljuje ob pridobitvi naziva »Rookie of the year«. Še vedno pa sva potrebovala posnetek, na katerega bova kasneje »prilepila« Lukov obraz. Za ta videoposnetek se je soavtor Tim Jevšenak posnel, kako se zahvaljuje. Ta posnetek sva kasneje uporabila kot podlago, na katero sva s pomočjo umetne inteligence »prilepila« Lukov obraz.

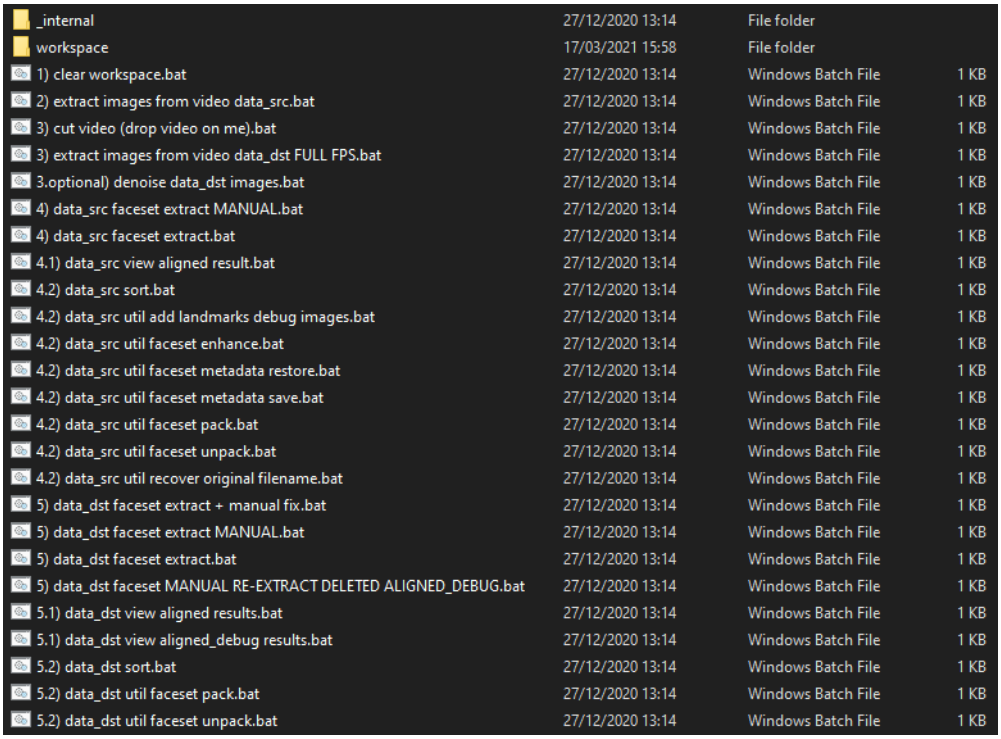
Pri izbiri videoposnetkov je priporočljivo, da sta oba videoposnetka enake resolucije. To lahko spremenimo tudi naknadno, kar sem tudi moral storiti, saj je bil posnetek Luka slabše kakovosti, kot posnetek Tima. Če je temu tako, se to opazi tudi na končnem izdelku.

3.1.2 UPORABA PROGRAMA DEEPFACELAB

Program DeepFaceLab je odprtokoden, kar pomeni, da si ga lahko vsak naloži ter si ogleduje in spreminja njegovo izvorno kodo. Prav tako je pri DeepFaceLabu zelo razširjena skupnost, ki

si med sabo deli različne metode za izdelavo boljših globokih ponaredkov in že vnaprej na trenirane modele, s katerimi lahko ustvarimo še boljši globoki ponaredek v krajšem času.

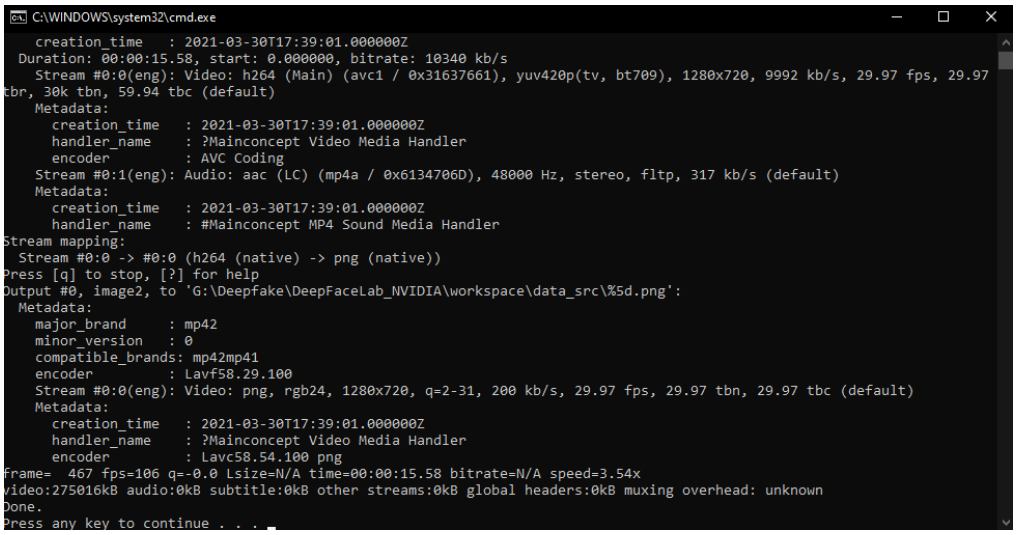
Samo delo v programu poteka tako, da zaganjamo posebne .bat datoteke. Te datoteke vsebujejo skupke ukazov, ki se nato izvedejo na našem računalniku. Ko naložimo DeepFaceLab, ima v mapi več datotek, preko katerih upravljamo celoten program. Z imeni datotek tudi vemo, kaj katera naredi.



Slika 9 - Primer mape DeepFaceLab

Najboljše je začeti ustvarjanje novega globokega ponaredka s tem, da počistimo t. i. delovni prostor. S tem se vse stare datoteke izbrišejo in začnemo znova.

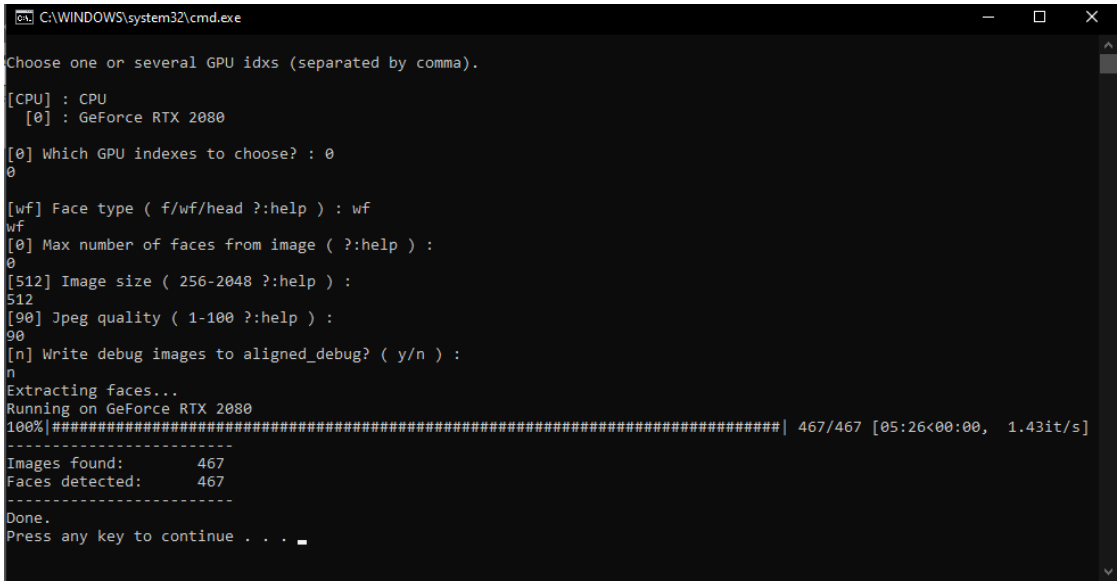
Naslednji korak je izvleči vsako sliko posebej iz videoposnetkov. Te slike DeepFaceLab kasneje uporabi za učenje modela. To najprej storimo s prvim videoposnetkom.



Slika 10 - Ekstrakcija slik iz videoposnetka

Enako ponovimo tudi z drugim videoposnetkom. S tem smo dobili slike obraza iz različnih perspektiv ter med govorjenjem.

Ker nas zanimajo le obrazi, ne pa okolica osebe ali karkoli drugega, je naslednji korak ekstrakcija obrazov s slik. Na srečo ima DeepFaceLab vgrajeno funkcijo, kjer avtomatsko zazna obraze ter slike obreže tako, da ostanejo samo obrazi.

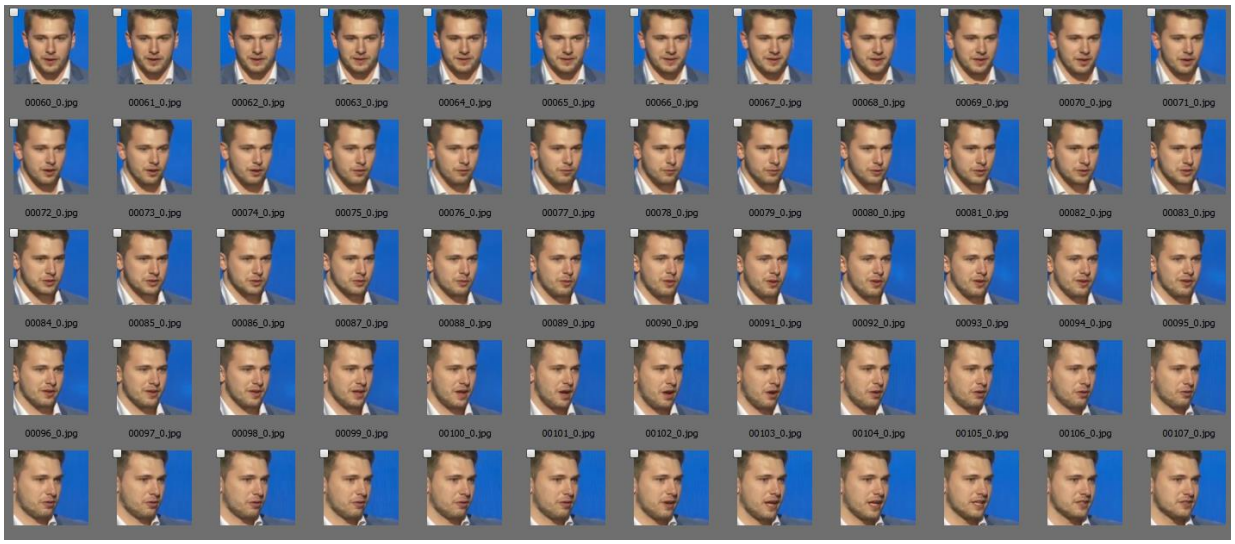


Slika 11 - Ekstrakcija obrazov s slik

Kot vidimo na zgornji sliki, algoritem sam prepozna obraze na slikah in jih izreže. Na začetku nas vpraša nekaj stvari, kot na primer, katero grafično kartico bomo uporabili, kakšen tip izreza želimo, velikost slik in kakovost le teh.

Pri izbiri izreza imamo tri različne možnosti. To so obraz, celoten obraz in pa glava. Razlika pri teh treh je, koliko slike pusti algoritem okoli obraza. Če izberemo možnost obraz, potem algoritem izreže skoraj vse razen obraza. Pri izbiri celotnega obraza algoritem pusti približno 30 % več prostora okoli obraza, medtem ko pri izbiri glave algoritem glavo izreže tako, da ostanejo tudi lasje. Izbira je predvsem pomembna od naših nadaljnjih namenov. Če želimo, kot v najinem primeru, zamenjati le obraz, potem izberemo obraz ali celoten obraz, če pa bi želeli zamenjati celotno glavo, torej vse od vratu naprej, pa bi izbrali glavo.

Kot vidimo, je z 467 slik pri najinem prvem videu algoritem našel 467 obrazov. Torej je na vsaki sliki obraz. Algoritem nam da tudi možnost, da preverimo obraze, ki jih je izvlekel in obrezal. To izgleda tako:



Slika 12 - Izvlečeni in obrezani obrazi

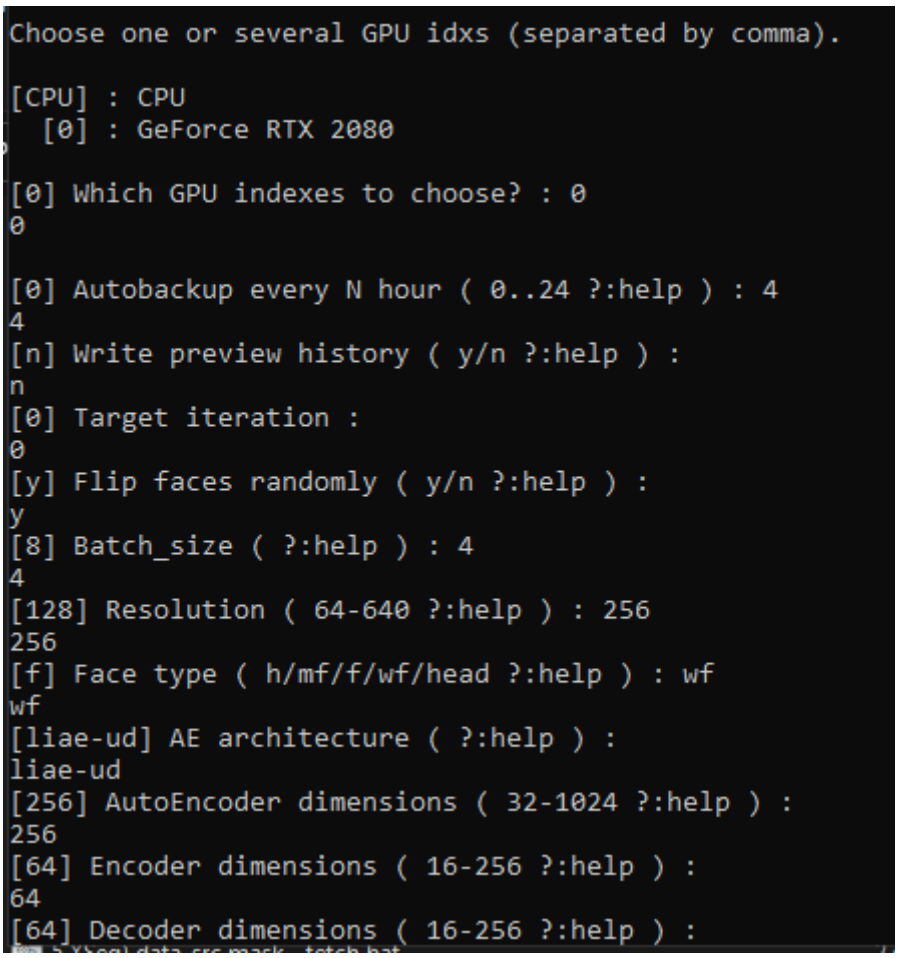
Ko je ekstrakcija obrazov iz obeh videov končana, lahko začnemo s treniranjem modela. Tudi tukaj nam DeepFaceLab ponudi več različnih možnosti izbire. Izbiramo lahko med XSeg, Quick96 in SAEHD.

XSeg podpira popolno modifikacijo in je običajno uporabljen za treniranje modelov za menjavo celotne glave. Prav tako to orodje zahteva veliko ročnega dela, najin cilj pa je s čim manj ročnega dela ustvariti lasten globoki ponaredek. Zato ga nisva izbrala.

Quick96 je, kot že ime pove, orodje, narejeno za hitro treniranje modelov. Zaradi same hitrosti ni tako natančno in so globoki ponaredki slabše kakovosti.

SAEHD je po težavnosti in po kakovosti videa nekje med Quick96 in XSeg. Je počasnejši, kot Quick96, hkrati pa je kakovost primerljiva z XSeg. Ko zaženemo SAEHD-algoritem, moramo najprej nastaviti nekaj parametrov. Spet moramo najprej izbrati, katero grafično kartico bo

algoritem uporabljal za treniranje. Nato nas vpraša, če želimo, da se ob določenem času avtomatsko shrani, da tudi ob nenadni zaustavitvi programa ne bi izgubili podatkov.



Slika 13 - Parametri za treniranje SAEHD

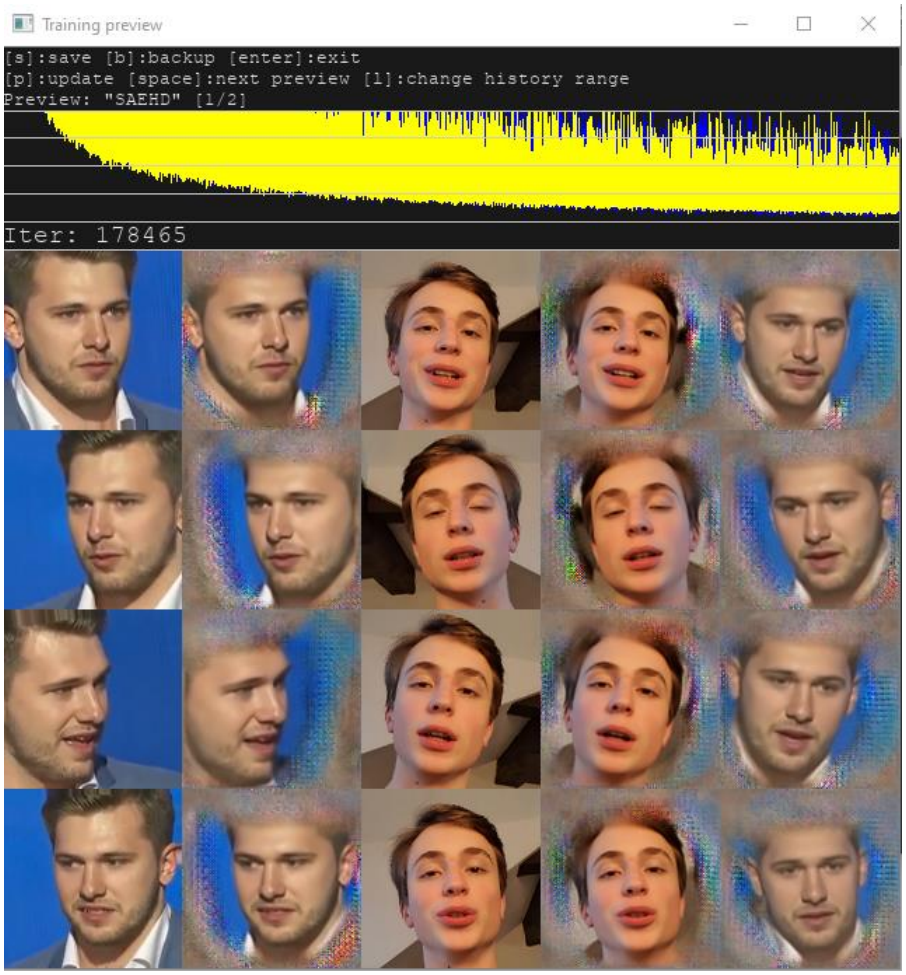
Ko nastavimo vse željene parametre, algoritem najprej zbere vse slike obrazov, nato pa jih primerja med seboj in trenira model, v katerega se vsi ti podatki sprotno shranjujejo. Če želimo narediti globoki ponaredek slavne osebe, je velika možnost, da na internetu že obstaja vnaprej pripravljen model, ki se je treniral po več 100 ur. Preverila sva, če obstaja takšen model za Dončiča, vendar takšnega modela nisva našla zato sva morala ustvariti svojega. S pomočjo takšnih modelov lahko v kratkem času naredimo prepričljiv globoki ponaredek, saj je najbolj zamudno treniranje modela.

Treniranje modela lahko kadarkoli prekinemo. Več iteracij kot algoritem naredi, boljše se nauči obraznih potez.

Slika 14 - Izgled po 200 iteracijah

Na zgornji sliki vidimo, koliko se je algoritem naučil po 200 iteracijah. Naenkrat se lahko uči iz dvanajstih slik. V mojem primeru sem omejil na osem slik hkrati, saj je algoritem zelo zahteven in že ob osmih slikah deluje moj računalnik na skoraj največji obremenitvi. Zelo zanimivo je to, da lahko sproti vidimo, kako se model uči in se obrazi posodabljajo. Zraven vsakega obraza vidimo, kako ga model poskuša rekonstruirati, na skrajni desni pa vidimo oba obraza združena.

Ko smo zadovoljni s številom ponovitev in izgledom obrazov, lahko video izvozimo. To naredimo tako, da najprej prekinemo treniranje algoritma in nato zaženemo datoteko, ki bo s pomočjo modela, ki smo ga na trenirali, in videoposnetka, ki smo ga na začetku izbrali, izvozila združen videoposnetek.



Slika 15 - izgled po 178 tisoč iteracijah

3.2 ODKRIVANJE GLOBOKIH PONAREDKOV Z ALGORITMI

Po tem ko sva se prepričala, da lahko sama ustvariva prepričljiv globoki ponaredek, sva začela raziskovati možnosti prepoznave le-teh. Z razvojem tehnologije postaja prepoznavanje globokih ponaredkov s človeškim očesom vedno težje. Zanimalo naju je, če že obstajajo kakšni algoritmi, ki so sposobni prepoznati globoke ponaredke. Želela sva preizkusiti točnost teh algoritmov. Prav tako naju je zanimalo, kako dolgo ti algoritmi porabijo za procesiranje podatkov. Menila sva namreč, da mora biti optimalen algoritem za prepoznavo globokih ponaredkov zanesljiv, hiter in predvsem zelo natančen.

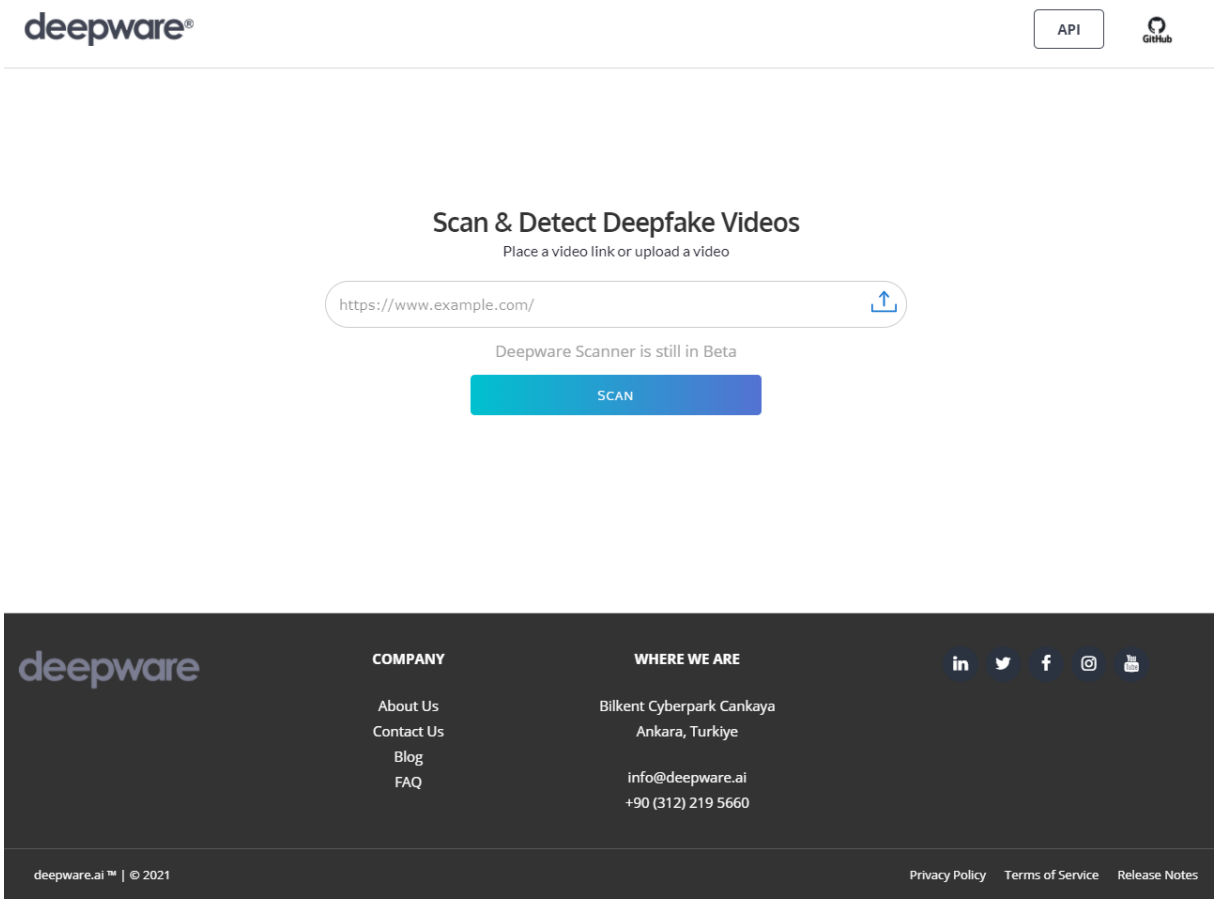
3.2.1 ISKANJE IN IZBIRA ALGORITMOV

Želela sva preizkusiti čim več algoritmov, zato sva poskusila uporabiti čisto vsak algoritem, ki sva ga našla na spletu. Vsi algoritmi, ki sva jih našla, so napisani v skriptnem programskem jeziku Python. Ugotovila sva, da so nekateri zastareli in ne delujejo. Nekaj algoritmov pa nima ustrezne dokumentacije, ki bi razjasnila njihovo uporabo. Nasploh ni veliko razvitih algoritmov, ki bi bili dostopni širši javnosti. Ugotovila sva, da se je na področju globokih

ponaredkov ustvarila skupnost ljudi, ki sodelujejo pri razvoju tega področja. Uspelo nama je najti dva dobra in dostopna algoritma. Eden od njiju je Deepware-ov algoritem za skeniranje in zaznavanje globokih ponaredkov. Drugi pa je bil zaseben algoritem švicarskega podjetja Quantum Integrity. Z njimi sva vzpostavila kontakt preko elektornske pošte. V zameno za povratne informacije so nama omogočili dostop do njihovega algoritma za prepoznavo globokih ponaredkov.

3.2.2 REZULTATI DEEPWARE ALGORITMA

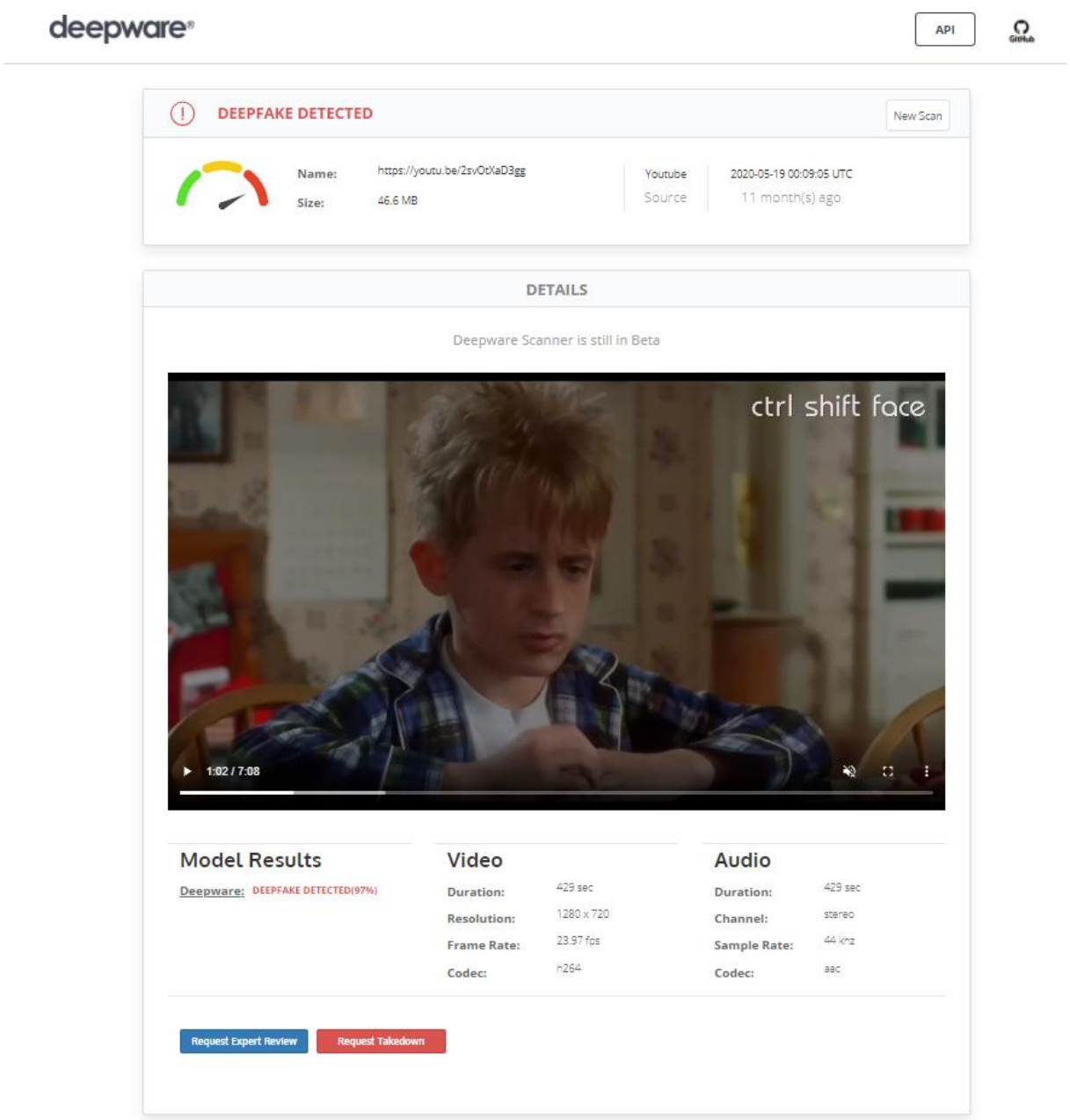
Deepware omogoča pregledovanje globokih ponaredkov na njihovi spletni strani. Stran je zelo pregledna in enostavna za uporabo. Skener deluje tako, da v dano polje preprosto kopiraš povezavo do videoposnetka, za katerega sumiš, da je globoki ponaredek. Druga možnost je, da videoposnetek naložiš iz svojega računalnika. Nato klikneš na tipko »Scan«. Nato si preusmerjen na posebno stran z rezultati. S tem postopkom sva preverila kar veliko videoposnetkov. Izbrala pa sva jih deset, ki sva jih shranila in testirala tudi na algoritmu Quantum Integrity. Izbrani videoposnetki so bili različno dolgi. Najkrajši je bil 10 sekund, najdaljši pa čez 8 minut. S različnimi dolžinami posnetkov sva želela preizkusiti, če ima dolžina videoposnetka kakšen vpliv na rezultate algoritma. Ugotovila sva, da je imel ta algoritem težave pri preverjanju krajših videoposnetkov. Pri daljših posnetkih je v večini primerov natančno prepoznal globoki ponaredek. Pri krajših posnetih je algoritem vrnil, da ni zaznal globokega ponaredka. Rezultati so se izpisali v odstotkih. Največji možen rezultat je 100 %, najnižji pa 0 %. Če program vrne 100 %, pomeni, da je algoritem prepričan, da gre za globoki ponaredek. Če pa program vrne 0 %, to pomeni, da videoposnetek po mnenju algoritma ni globoki ponaredek. Videoposnetki, ki sva jih testirala midva, so vsi bili globoki ponaredki. Deepware-ov algoritem ja pri krajših videoposnetkih vračal nizke rezultate (najnižji je bil 1 %). Ko pa sva preverjala daljše videoposnetke, pa sva dobila bolj točne rezultate. Najvišji rezultat je algoritem vrnil za sedmi videoposnetek.



Slika 16 - Deepware algoritem (Deepware Scanner, 2021)

Ugotovila sva, da je dolžina posnetka imela takšen vpliv na rezultate algoritma. Ta podatek je zaskrbljujoč, saj je dolžina posnetkov, ki krožijo po socialnih omrežjih, večinoma kratka. Deepware-ov algoritem je narejen tako, da analizira vsako sliko v videoposnetku. To je precej zanimivo, saj gre pri tem za obdelavo velike količine podatkov, algoritem pa rezultate vrne v relativno hitrem času. Slabo je, da se algoritem zanaša na veliko količino podatkov za analizo in v primeru, da je videoposnetek krajši, ne vrne ustreznega rezultata.

Ocenjujeva, da je ta algoritem dober in ima veliko procesorsko moč. Ni pa še sposoben točno prepoznati globoke ponaredke v krajših videoposnetkih. Zaradi tega ne moreva reči, da so njegovi rezultati zanesljivi.



Slika 17 - Rezultati Deepware algoritma (Deepware Scanner, 2021)

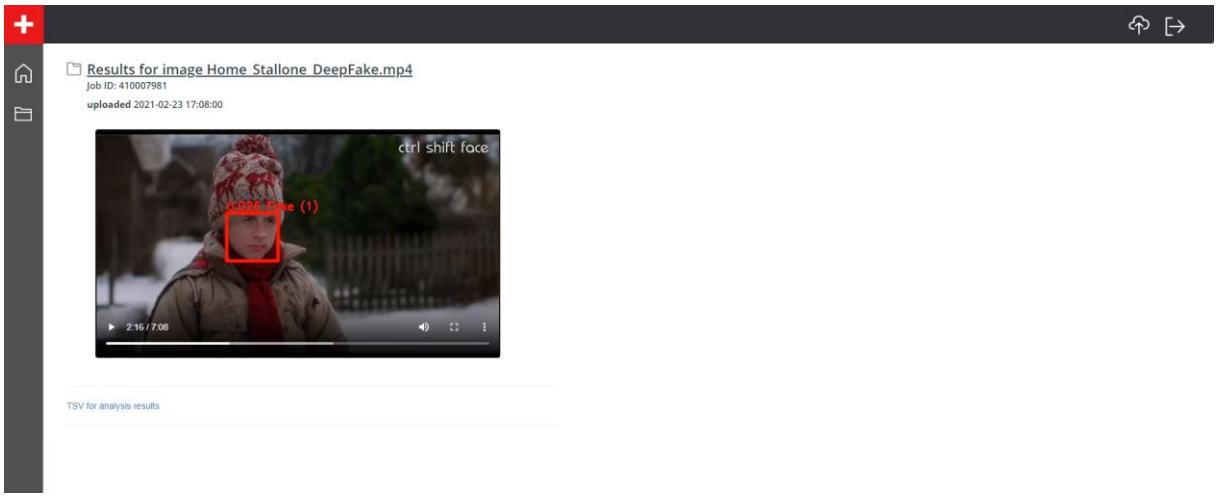
3.2.3 REZULTATI ALGORITMA PODJETJA QUANTUM INTEGRITY

Podjetje Quantum Integrity se v prvi vrsti ukvarja s prepoznavo lažnih prijav škode zavarovalnicam. Za ta namen imajo poseben algoritem. Ukvarjajo pa se tudi s prepoznavo globokih ponaredkov. Sami so rekli, da je njihov algoritem še v razvoju. Kljub temu pa so nama omogočili dostop do njihovega algoritma za prepoznavanje globokih ponaredkov preko njihove spletne aplikacije. Dodelili so nama uporabniško ime in geslo za prijavo v storitev. Spletna aplikacija nama je omogočala, da nanjo naloživa videoposnetke iz svojega računalnika. Te je nato algoritem sam avtomatsko pregledal. Ko se je skeniranje zaključilo, se je videoposnetek

obarval zeleno in ob kliku nanj ti je odprlo novo stran z rezultati. Tudi s tem algoritmom sva preverila kar nekaj videoposnetkov. Za primerjavo z Deepware-ovim algoritmom sva uporabila istih deset posnetkov kot prej. Podobno kot prej naju je zanimalo, ali dolžina posnetka vpliva na rezultate. Pri tem algoritmu dolžina videoposnetka ni bila izrazit problem. Algoritem je bil dokaj uspešen pri prepoznavi globokih ponaredkov. Za polovico videoposnetkov je pravilno prepoznal, da gre za globoki ponaredek. Le pri dveh videoposnetkih je ocenil narobe. Za tri videoposnetke pa so rezultati bili nekje v sredini (približno 50 %). Rezultate je algoritem vrnil za vsak izsek s posnetka posebej. Najnižji možni rezultat je 0 in najvišji 1. Če je rezultat 0, to pomeni, da je algoritem prepričan, da gre za globoki ponaredek. V nasprotnem primeru, torej če je rezultat 1, pa je obraz prepoznan kot resničen oz. nespremenjen. Rezultat je izpisan na tri decimalna mesta natančno (npr. pri prvem posnetku je rezultat 0.492). Ta rezultat sva nato pretvorila v odstotke in odštela od 100 %. Tako sva dobila podatke, ki so bili primerljivi s tistimi, ki jih je vrnil Deepware-ov algoritem (npr. pri prvem posnetku je tako končni rezultat 50.8 %). Algoritem podjetja Quantum Integrity je pri krajših videoposnetkih analiziral vsako sliko posnetka posebej ter vrnil rezultat. Pri daljših videoposnetkih pa je algoritem pregledal samo ključne izseke s posnetkov (namesto vsake slike samo 1 od 20). Tako se je skrajšal tudi čas procesiranja, a je ta še vedno bil dosti daljši kot pri Deepware-ovem algoritmu. Algoritem švicarskega podjetja je bil počasnejši in manj prepričljiv, kar je bilo pričakovati, saj podjetje za enkrat še ni dalo prevelik poudarek na zaznavanje globokih ponaredkov.

Sodelovanje s podjetjem Quantum Integrity je bilo za naju koristno in tudi zanimivo. V zameno za dostop do njihovega algoritma sva jima beležila rezultate za vsak videoposnetek, ki sva ga preverila z algoritmom. O tem, kako bo naše sodelovanje potekalo, smo se zmenili na klicu preko Skype-a. Za sprotna vprašanja pa smo si dopisovali preko e-pošte.

Tudi ta algoritem še ni najbolje optimiziran za prepoznavo globokih ponaredkov. Čeprav je veliko globokih ponaredkov prepoznal, so bili rezultati še vedno premalo natančni, da bi lahko govorili o optimalni rešitvi za prepoznavo globokih ponaredkov.



Slika 18 - Rezultati algoritma podjetja Quantum Integrity (Quantum Integrity Scanner, 2021)

3.3 ANKETIRANJE

Ker naju je zanimalo, kako dobro lahko človeško oko zazna novejše globoke ponaredke, sva se odločila, da bova izdelala kratko anketo ter jo poslala vsem dijakom Elektro in računalniške šole. Pri tem nama je pomagal mentor Islam Mušić, ki je ankete poslal vsem dijakinjam in dijakom na njihove šolske elektronske naslove. Anketiranci so imeli podanih osem videoposnetkov, ki sva jih skrajšala ter uredila, da so bili primerni za anketo. Med gledanjem teh videoposnetkov so se morali odločiti, ali gre za globoke ponaredke ali ne. Med videoposnetki je bilo pet globokih ponaredkov in trije nespremenjeni posnetki.

3.3.1 IZBIRA VIDEOPOSNETKOV ZA ANKETO

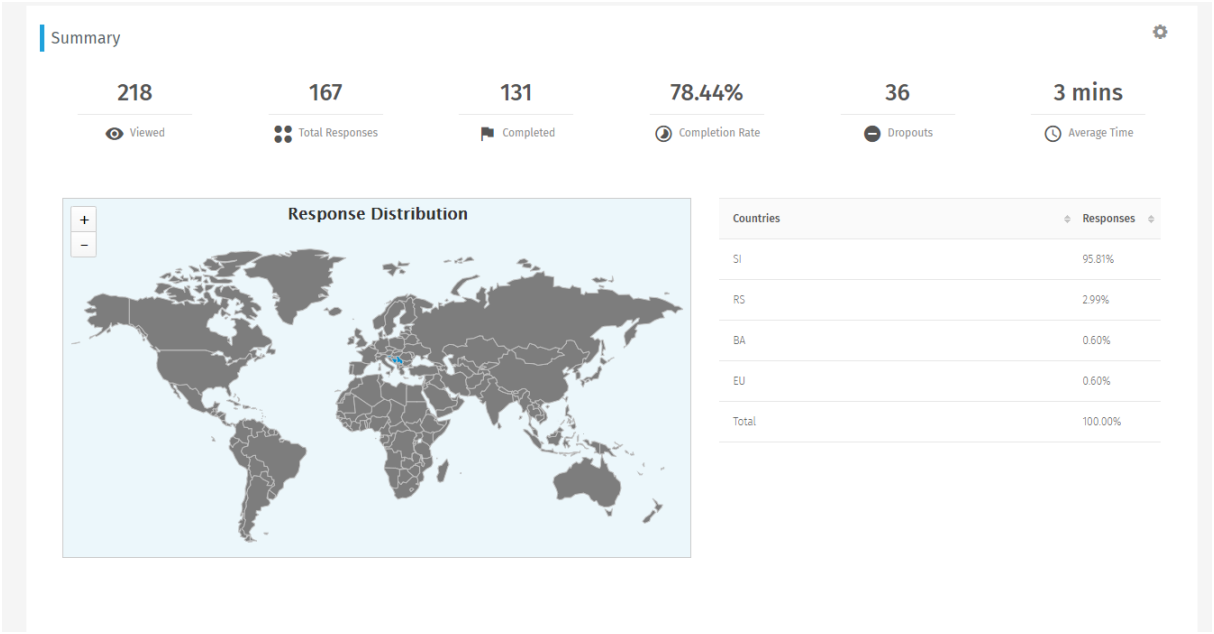
Že preden sva se odločila za raziskovalno nalogo, sva videla veliko globokih ponaredkov. Nekateri so bili precej slabi in se je dalo takoj videti, da niso resnični. Nekateri pa so bili tako dobro izdelani, da če avtor ne bi navedel, da gre za globoki ponaredek, tega sama ne bi ugotovila. Zato naju je zanimalo, kako dobri pa so drugi v prepoznavanju globokih ponaredkov. To je bil tudi glavni razlog, da sva izvedla anketo. Vse videoposnetke za vprašanja sva našla na socialnih omrežjih. Pregledala sva veliko posnetkov, dokler nisva našla osem primernih za anketo. Odločila sva se, da bo osem posnetkov dovolj, saj nisva hotela imeti predolge ankete in s tem zmanjšati zanimanje. Anketirancem bi lahko podala še več posnetkov, vendar je bilo po najini oceni osem videoposnetkov dovolj, da se lahko anketiranec zmede. Večina posnetkov je iz Youtube-a, eden pa je s portala Twitter. Videoposnetki vsi vsebujejo znane osebe. Ker so v originalu posnetki dolgi nekje od 3 do 8 minut, sva jih za uporabo v anketi skrajšala. Iz vsakega videoposnetka sva vzela tisti del, v katerem znana oseba največ govori. Skrajšani posnetki so

bili dolgi okoli 10 sekund. Ker sva želela anketirance zmesti, sva v anketo vključila tudi posnetke, ki niso globoki ponaredki. Tako sva imela izbranih pet videoposnetkov, ki so bili globoki ponaredki. Trije posnetki pa niso bili ponarejeni. Urejene videoposnetke sva nato objavila na svoj zaseben Youtube kanalu. V anketo sva jih dodala s spletno povezavo do videoposnetka.

3.3.2 IZBIRA ORODJA ZA IZDELAVO ANKETE

Ko sva imela izbrane videoposnetke, je bilo potrebno izdelati anketo. Sprva sva mislila uporabiti spletno orodje 1KA, vendar sva naletela na težave pri vključevanju videoposnetkov. Ti se niso hoteli predvajati sami. Prav tako so se po enem predvajanju zaključili, midva pa sva hotela, da se neprestano ponavljajo. Za to sva poiskala drugo orodje. Izbrala sva spletno orodje QuestionPro, ki je prav tako kot 1KA brezplačno. To orodje je imelo možnost ponavljanja posameznega videoposnetka. Tako da so si anketiranci posamezen videoposnetek lahko ogledali večkrat, ne da bi morali klikniti na gumb za ponavljanje. Posnetek začel predvajati sam, ko je anketiranec prišel do posameznega vprašanja. Na uvodni strani ankete sva anketirancem na hitro razložila, kaj so globoki ponaredki. Nato pa sva jih pozvala, naj poskušajo ugotoviti, kateri videoposnetki so ponarejeni.

Sam postopek izdelave ankete je bil s tem spletnim orodjem enostaven. Enostavno nama je bilo preverjati, ali je posamezen anketiranec oddal samo en odgovor ali ne, saj ima spletno orodje to funkcijo že vgrajeno. Tudi kasnejše preverjanje rezultatov je bilo precej olajšano, saj spletno orodje samo analizira podatke in pripravi organizirano poročilo.



Slika 19 - Poročilo o anketi (QuestionPro, 2021)

3.3.3 SEZNAM ANKETIRANCEV

Anketo sva želela poslati čim večjemu številu ljudi. Meniva, da bodo globoki ponaredki v prihodnosti vedno večji problem. S širjenjem lažnih informacij je možno manipuliranje ljudi za najrazličnejše namene. Videoposnetki, pri katerih gre za globoki ponaredek, se najhitreje širijo po spletu. Zanimali so naju odgovori vrstnikov, saj sva del generacije, ki večino svojih novic dobi na spletu preko različnih socialnih omrežij. Anketo sva s pomočjo najinega mentorja poslala vsem dijakom Elektro in računalniške šole.

3.4 IZDELAVA MOBILNE APLIKACIJE

Po raziskovanju raznih algoritmov za detekcijo globokih ponaredkov sva ugotovila, da jih velika večina ni dostopnih širši javnosti. Ker sva to želela spremeniti, sva stopila v kontakt s podjetjem Deepware, ki je eno izmed vodilnih podjetij na področju globokih ponaredkov in jih prosila, če bi lahko dobila dostop do njihovega API. Ker so bili tudi oni enakega mnenja, so nama zelo hitro odgovorili in nama ponudili dostop do njihovega API. Dostop je bil sicer omejen na 100 skeniranj, vendar so rekli, da lahko brez težav dobiva še nadaljnji dostop, če bi to potrebovala.

Ko sva stopila v kontakt s podjetjem Quantum Integrity iz Švice, so tudi oni imeli željo po aplikaciji. Aplikacijo sva tudi naredila in jim jo poslala, vendar zaenkrat še ni v uporabi.

3.4.1 IZBIRA ORODJA ZA IZDELAVO APLIKACIJE

Pri izdelavi aplikacije sva se odločala med več različnimi orodji. Želela sva preprostost, saj je bil namen le čisto preprosta aplikacija, ne pa kaj bolj zahtevnega. Želela sva tudi, da je aplikacija lahko na voljo na operacijskih sistemih iOS in Android.

Pomembna pa je tudi preprosta povezava z API. Na koncu sva ugotovila, da je ogrodje Ionic kot nalašč za najino aplikacijo, saj omogoča tudi kasnejšo razširitev na spletne aplikacije, ker temelji na Javascript in HTML.

3.4.2 APLIKACIJA

Pri izdelavi aplikacije sva se najprej lotila povezave z API. Vsakič, ko uporabnik vnese nek URL do videoposnetka, ga aplikacija posreduje naprej in pošlje zahtevo na API. API nato posnetek procesira in dokler se posnetek ne procesira do konca, vrača zahtevek na API vrednost

0 oziroma false, ko pa se posnetek dokončno procesira, pa nam zahtevek na API vrne vrednost 1 oziroma true in pa verjetnost tega, da je videoposnetek ponaredek.

```
async presentLoading() {

  this.dataService.startScan(this.inputValue).subscribe(data =>
    {
      console.log(data);
      this.reportID = data["report-id"]
      console.log(this.reportID);

      const requestReport = () => {
        setTimeout(() => {
          this.dataService.getReport(this.reportID).subscribe(result =>
            {
              this.result = result;
              console.log(result["completed"] + " - je stanje");
              console.log(result);
              this.completed = result["completed"];
              if(this.completed == false){
                requestReport();
              }
              else{
                loading.dismiss();
                console.log(this.result.results.deepware["score"]);
                this.score = this.result.results.deepware["score"];
                this.scoreSmall = this.score/100;
                if(this.score < 40){
                  this.barva = "success";
                  this.isItFake = "Your video probably isn't fake!";
                }
                else if(this.score > 40 && this.score < 70){
                  this.barva = "warning";
                  this.isItFake = "Your video might be deepfaked!";
                }
                else{
                  this.barva = "danger";
                  this.isItFake = "Your video is probably a deepfake!";
                }
              }
```

Slika 20 - izsek kode, ki skrbi za komunikacijo z API

Kot vidimo, je koda narejena tako, da po tem, ko pošlje URL do videoposnetka, vsake 2 sekundi pošlje zahtevek in preveri, kakšna je vrednost. Če videoposnetek še ni procesiran, spet počaka 2 sekundi in pošlje zahtevek. To dela dokler, ne dobi potrditve, da je videoposnetek dokončno procesiran. Ko dobi to potrditev, vzame podatke, ki jih je API vrnil in jih prikaže na aplikaciji.

Aplikacija je narejena tako, da jo lahko hitro prilagodiva za potrebe kakršnegakoli drugega podjetja ali drugega API.

Report {

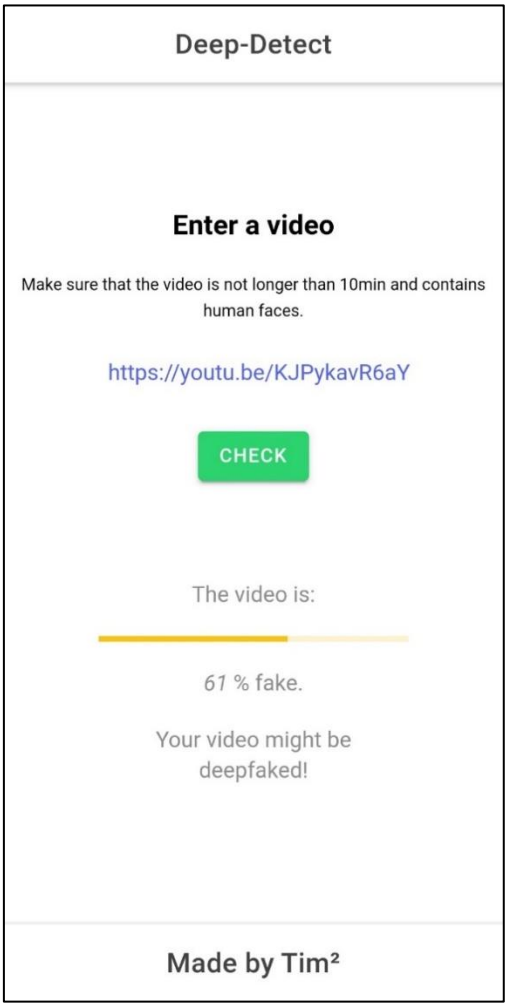
report-id	string
video-id	string
type	string
size	integer($int64)
names	[string]
total	integer($int32)
positive	integer($int32)
complete	boolean
results	{
< * >:	**Result** {
	detected boolean
	score integer($int32)
	}
}	

}

Slika 21 - podatki, ki jih aplikacija pridobi iz API (Deepware, 2021)

Kot vidimo na zgornji sliki (slika 21) aplikacija iz Deepware API-ja pridobi več podatkov. Najbolj pomemben podatek za naju je podatek o rezultatu in o točkah, ki jih je posnetek dosegel. Več točk kot posnetek doseže, večja je verjetnost, da je globok ponaredek. Največje število možnih točk je 100, najnižje pa 0.

API vrne tudi velikost in podatkovno vrsto posnetka, vendar naju ti podatki ne zanimajo. Pomemben podatek je tudi podatek o zaključenosti, ki ga preverjama vsake 2 sekundi.



Slika 22 – Aplikacija

Kot vidimo, je aplikacija videoposnetek, ki sva ga vstavila, zaznala kot potencialno ponarejenega. Z 61 % verjetnostjo trdi, da je videoposnetek mogoče ponarejen. Kot vidimo, nam aplikacija to lepo prikaže najprej z odstotki, zgoraj vidimo oranžno črto. Barva te črte se spreminja v odvisnosti od rezultata. Če algoritem meni, da video ni ponarejen, bo črta zelena, če pa je siguren, da je ponarejen, pa bo črta rdeča. Spodaj pa je tudi z besedo razloženo, kaj približno algoritem meni.

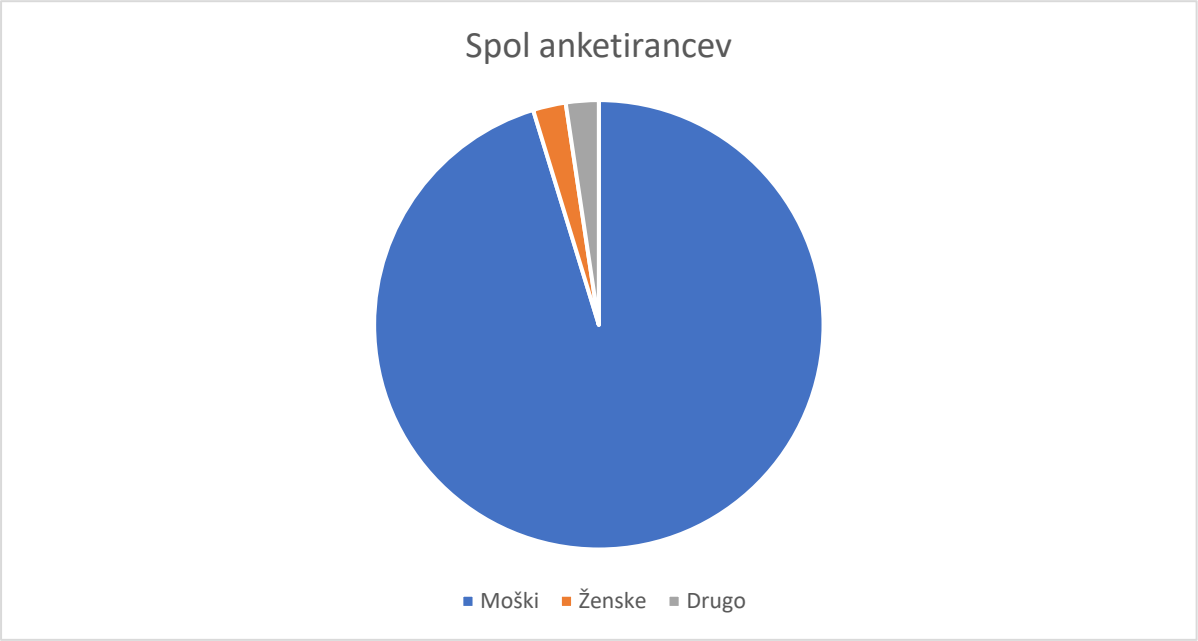
4 REZULTATI

4.1 ANALIZA ANKETE

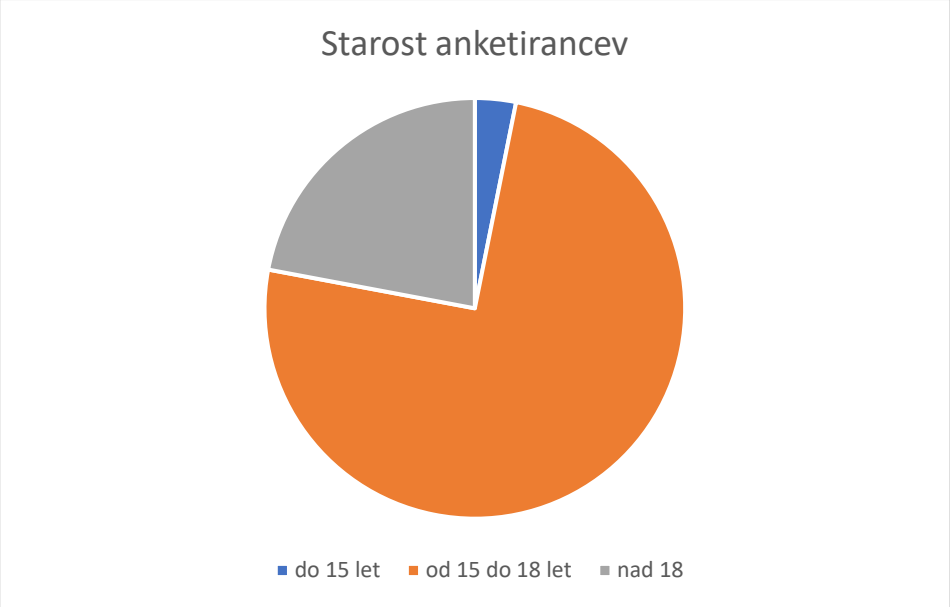
Dobila sva 131 odgovorov. Vse skupaj si je anketo ogledalo 167 ljudi. V povprečju so anketiranci za reševanje ankete porabili 3 minute. Število anketirancev naju je pozitivno presenetilo, saj nisva pričakovala tako velik odziv. Glede na rezultate sklepava, da so anketiranci v resnici poskušali ugotoviti, ali gre za globoki ponaredek ali ne, torej sva dobila realne rezultate.

4.1.1 PODATKI O ANKETIRANCIH

Večina najinih anketirancev je bila moškega spola (95,28 %), anketirank je bilo za 2,36 %, 2,36 % anketirancev pa ni navedel spola. Večinoma so anketiranci bili stari med 15 in 18 let (74,2 %), 3,15 jih je bilo mlajših od 15 let, 22,05 % pa starejših od 18 let.

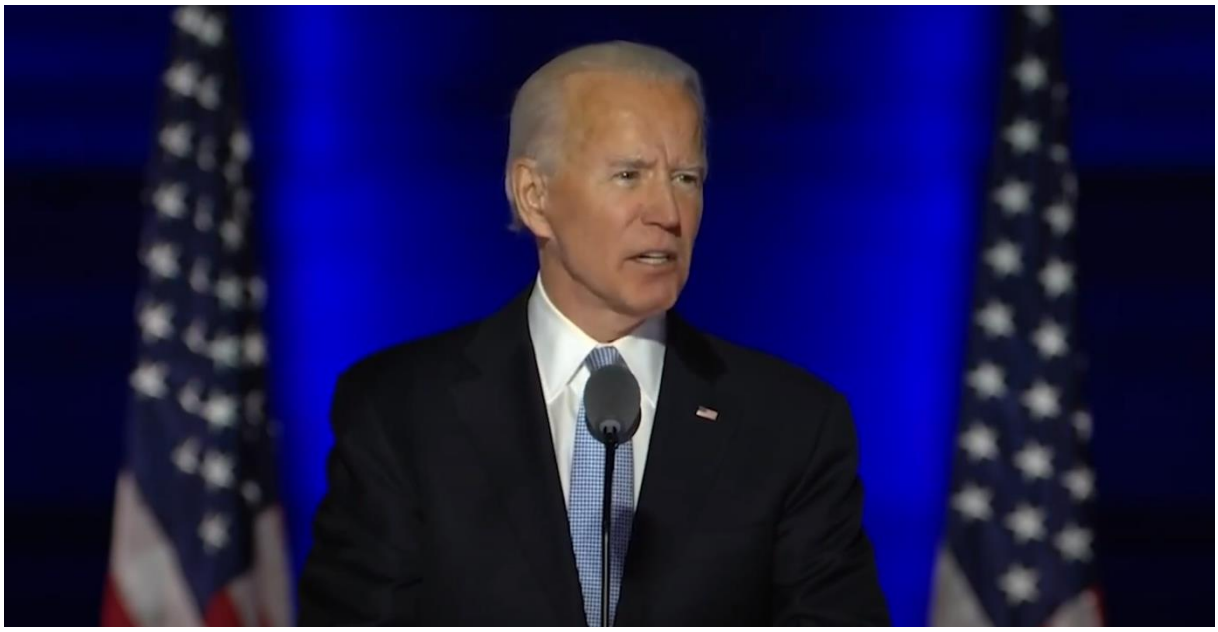


Graf 1 - Spol anketirancev



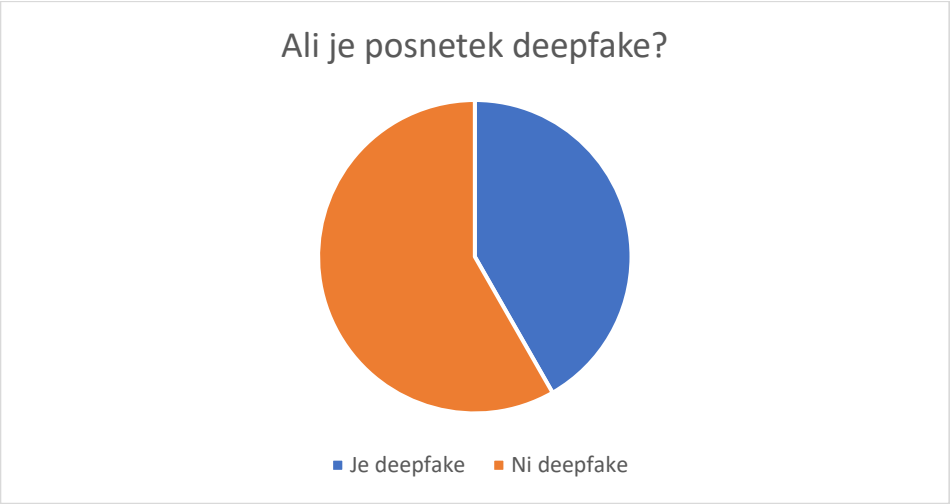
Graf 2 - Starost anketirancev

4.1.2 1. VIDEOPOSNETEK



Slika 23 - Izsek iz videoposnetka pri prvem vprašanju

Ta posnetek ni globoki ponaredek. Video prikazuje kratek odsek govor ameriškega predsednika Joea Bidena. Statistika odgovorov naju je pri tem prvem posnetku presenetila, saj sva bila prepričana, da bo večina anketirancev ugotovila, da ne gre za globoki ponaredek, kar pa ni bilo res.



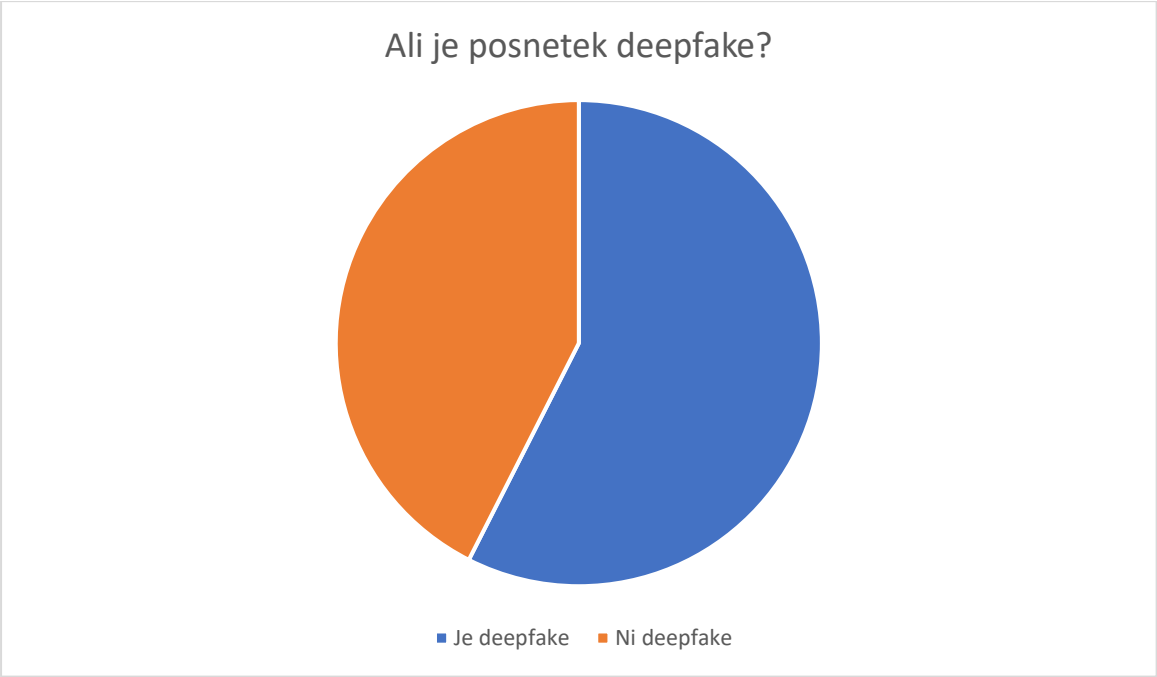
Graf 3 - Rezultati 1. vprašanja

Rezultati so zelo razdeljeni. 41,73 % vprašanih je reklo, da gre za globoki ponaredek, 58,27 % vprašanih pa je pravilno odgovorilo, da obraz v posnetku ni spremenjen.

4.1.3 2. VIDEOPOSNETEK

Slika 24 - Izsek iz videoposnetka pri drugem vprašanju

Ta posnetek je globoki ponaredek. V posnetku je nekdanji predsednik ZDA Barack Obama, vendar pa, to kar govori, ni izrekel on, ampak je njegov obraz spremenjen. To daje izgled, kot da govori nekaj, kar v resnici ni nikoli izrekel. Tudi za ta video so bili odgovori precej razdvojeni.

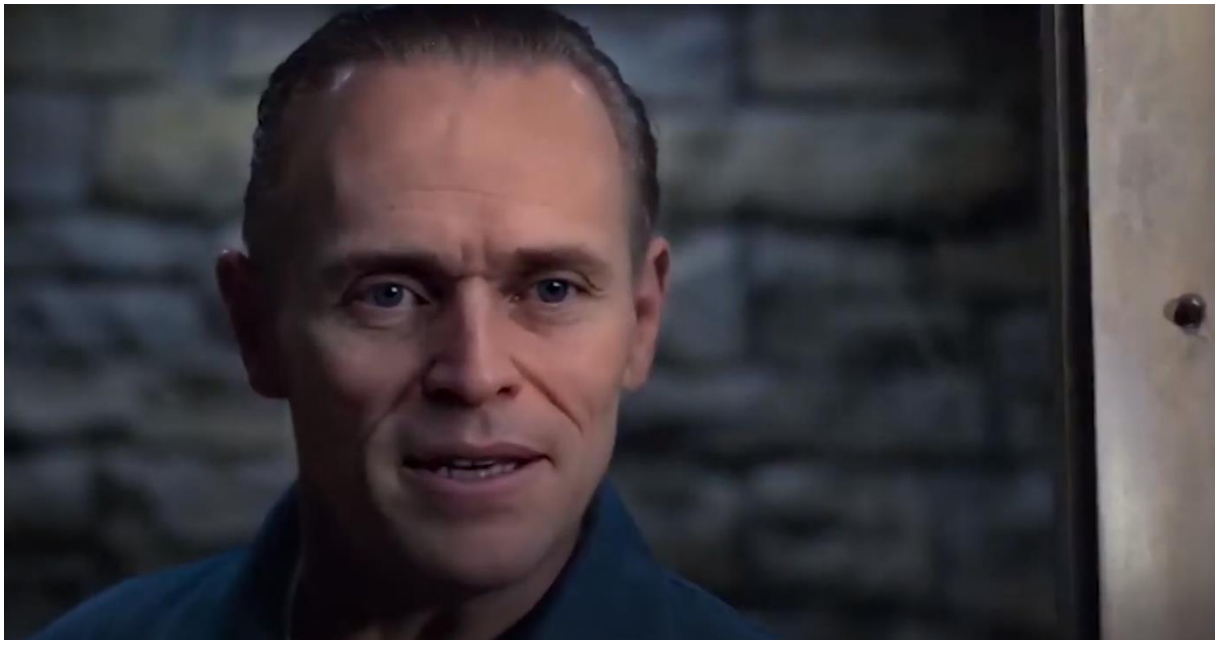


Graf 4 - Rezultati 2. vprašanja

Kot je prikazano na sliki, je 57,48 % anketirancev pravilno odgovorilo, da gre za globoki ponaredek. Ostalih 42,52 % vprašanih pa se je zmotilo.

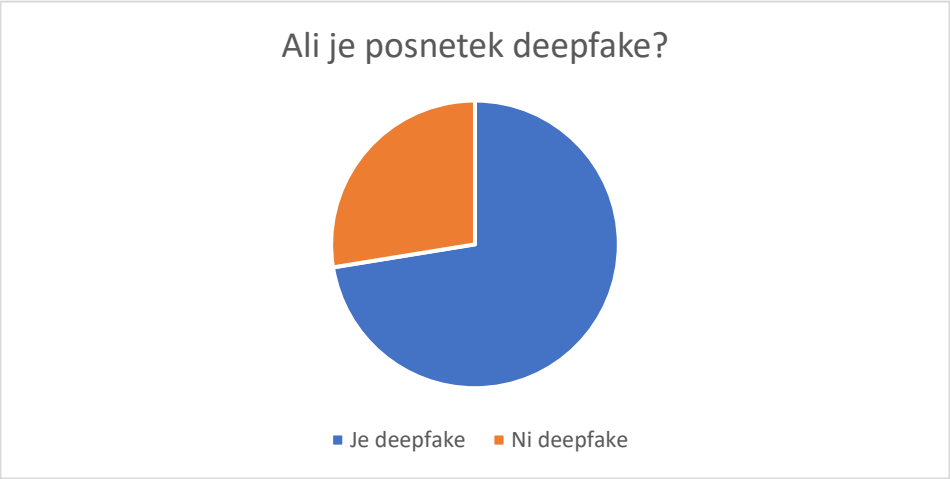
Zanimivo je, da so v primerjavi s prvim vprašanje pri tem vprašanju odgovori ravno nasprotno.

4.1.4 3. VIDEOPOSNETEK



Slika 25 - Izsek iz videoposnetka pri tretjem vprašanju

Ta posnetek je globoki ponaredek. Gre za prizor iz filma Silence of the lambs, v katerem je obraz igralca Anthonya Hopkinsa zamenjan z obrazom igralca Willema Dafoea.



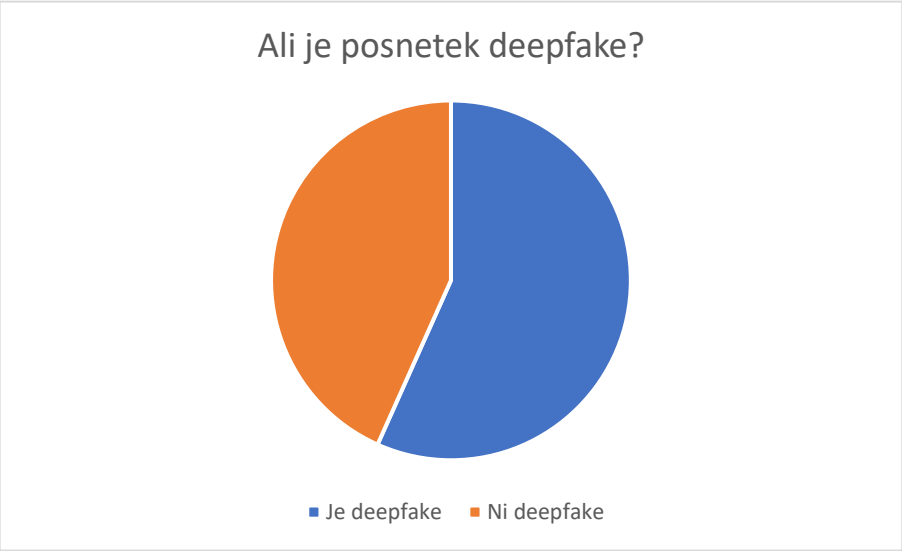
Graf 5 - Rezultati 3. vprašanja

Pri tem vprašanju so anketiranci večinoma ugotovili, da gre za globoki ponaredek. 72,44 % vprašanih je odgovorilo pravilno, 27,56 % vprašanih pa je bilo prepričanih, da je posnetek nespremenjen.

4.1.5 4. VIDEOPOSNETEK

Slika 26 - Izsek iz videoposnetka pri četrtem vprašanju

Ta posnetek je globoki ponaredek. V videu je obraz Marka Zuckerberga spremenjen na takšen način, da izgleda, kot da govori nekaj, kar ni on nikoli rekel.



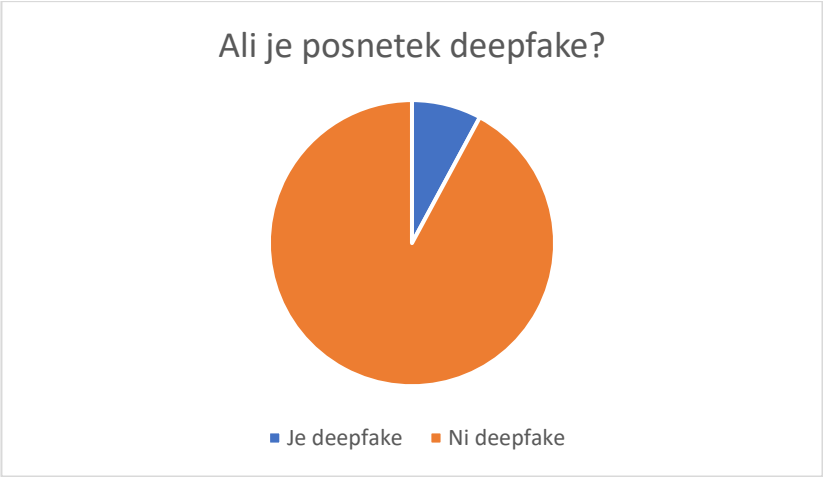
Graf 6 - Rezultati 4. vprašanja

Ponovno so bili odgovori zelo razdvojeni. 56,69 % vprašanih je pravilno odgovorilo, da je posnetek globoki ponaredek, 43,31 % vprašanih pa je verjelo, da gre za resničen posnetek.

4.1.6 5. VIDEOPOSNETEK

Slika 27 - Izsek iz videoposnetka pri petem vprašanju

Ta posnetek ni globoki ponaredek. Gre za pogovor s trenutno najbogatejšim Zemljanom Elonom Muskom.



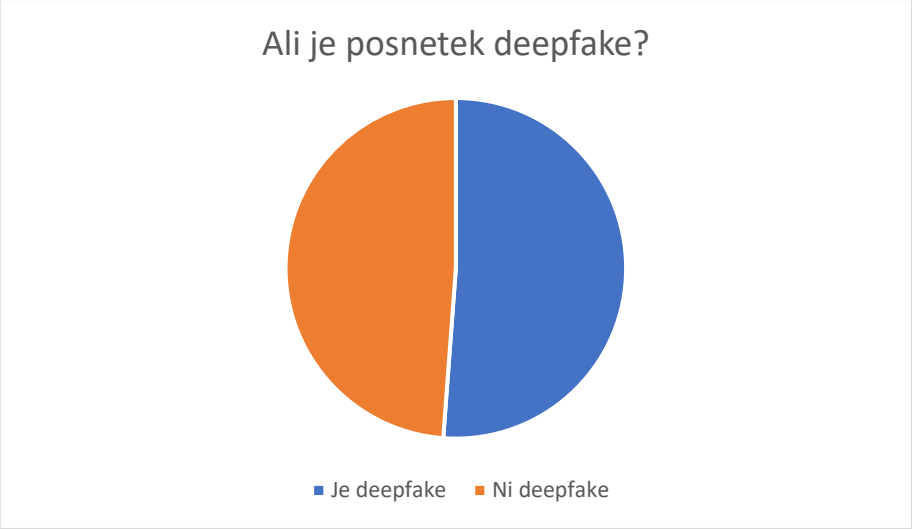
Graf 7 - Rezultati 5. vprašanja

Na to vprašanje je večina anketirancev odgovorila pravilno, in sicer 91,47 % vprašanih je reklo, da ne gre za globoki ponaredek. Le 8,53 % vprašanih je menilo, da je posnetek ponarejen. To naju je presenetilo, saj nisva pričakovala, da bodo ljudje v takem številu ugotovili, da je video resničen.

4.1.7 6. VIDEOPOSNETEK

Slika 28 - Izsek iz videoposnetka pri šestem vprašanju

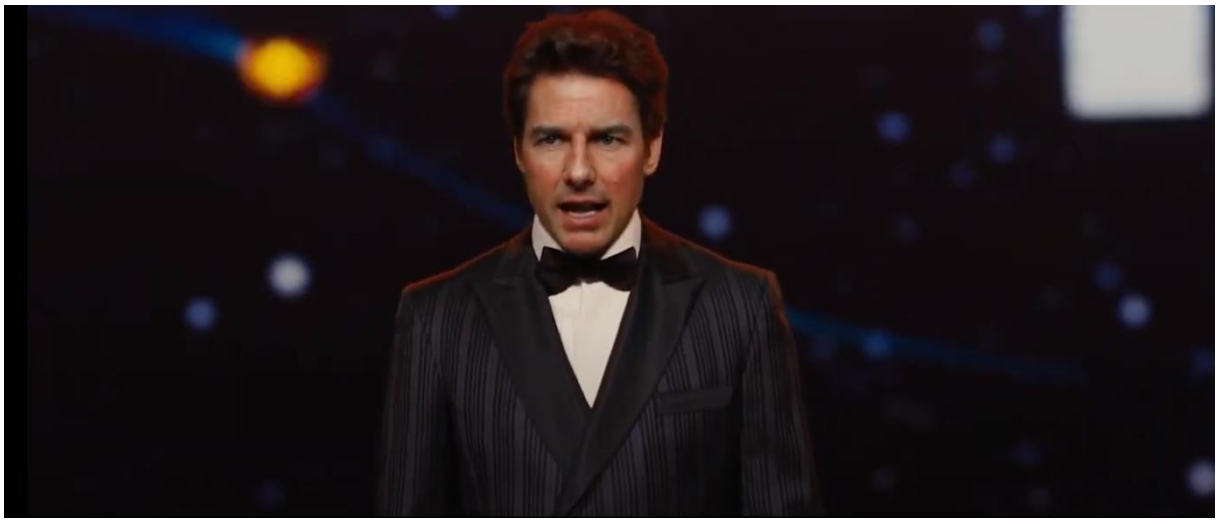
Ta posnetek je globoki ponaredek. Prikazuje igralca Ramija Maleka, ki sprejema svojo nagrado, vendar pa je njegov obraz zamenjan z obrazom pevca Freddieja Mercuryja.



Graf 8 - Rezultati 6. vprašanja

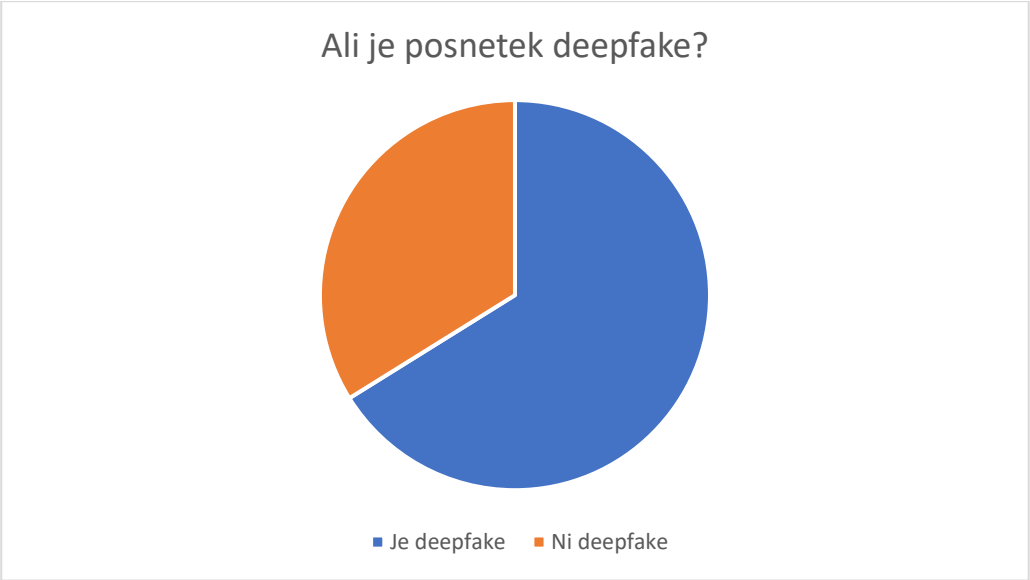
Rezultati so pri tem vprašanju še najbolj razdvojeni. 51,16 % je pravilno uganilo, da gre za globoki ponaredek. 48,84 % vprašanih pa je bilo mnenja, da videoposnetek ni ponarejen.

4.1.8 7. VIDEOPOSNETEK



Slika 29 - Izsek iz videoposnetka pri sedmem vprašanju

Ta posnetek je globoki ponaredek. Gre za sceno iz filma Iron Man, v kateri je obraz glavnega lika zamenjan z obrazom igralca Toma Cruisea.



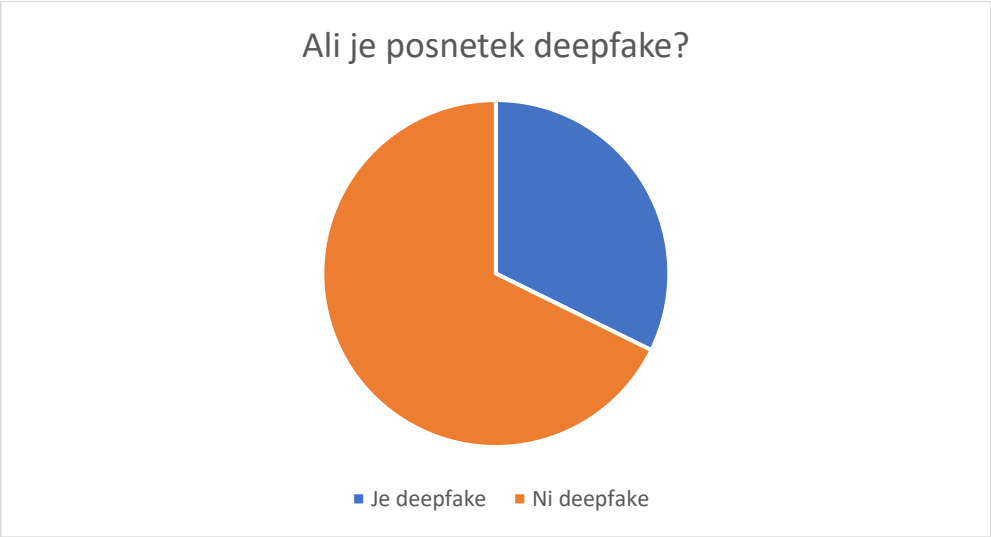
Graf 9 - Rezultati 7. vprašanja

Pri tem vprašanju so bili anketiranci dokaj uspešni pri prepoznavanju globokega ponaredka. 66,67 % vprašanih je pravilno odgovorilo, da gre za ponaredek. 33,33 % vprašanih pa se je zmotilo in so odgovorili, da posnetek ni globoki ponaredek.

4.1.9 8. VIDEOPOSNETEK

Slika 30 - Izsek iz videoposnetka pri osmem vprašanju

Ta posnetek ni globoki ponaredek. Gre za del govora nekdanjega predsednika ZDA Donalda Trumpa.

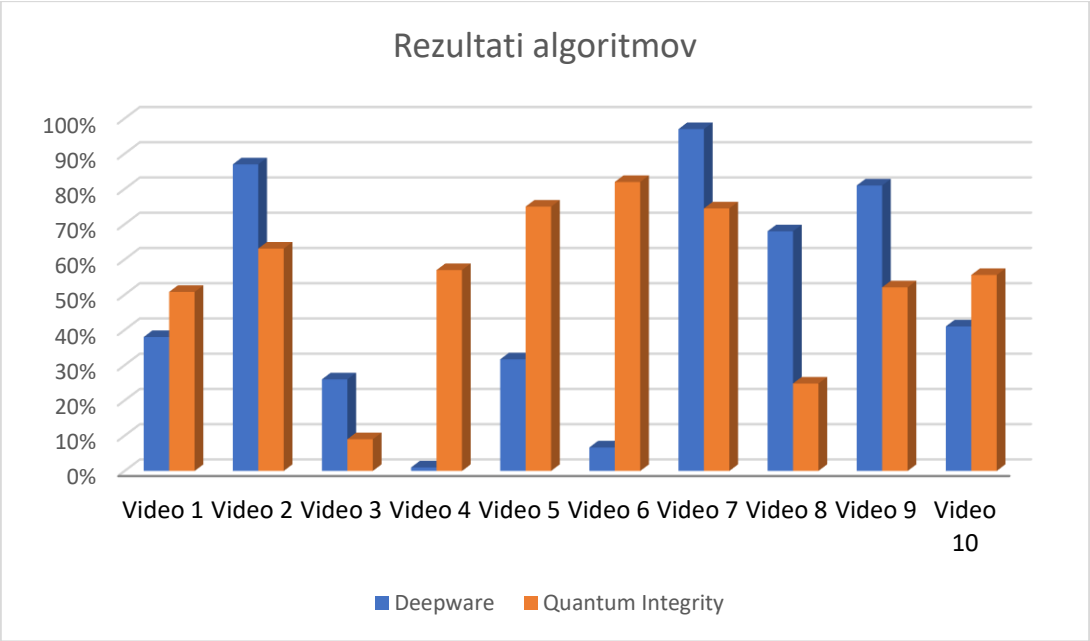


Graf 10 - Rezultati 8. vprašanja

Tudi pri tem vprašanju so imeli anketiranci večinoma prav. 67,44 % vprašanih je vedelo, da posnetek ni spremenjen, medtem ko je bilo 32,56 % vprašanih prepričanih, da gre za globoki ponaredek.

4.2 ANALIZA REZULTATOV ALGORITMOV

Z algoritmoma Deepware in Quantum Integrity sva preverila veliko videoposnetkov in ugotavljala, kako natančna sta. Preverjala sva tudi, če pravilno zazna videoposnetke, ki niso globoki ponaredki. Za primerjavo teh dveh algoritmov sva izbrala deset posnetkov in jih testirala z obema algoritmoma. Pri testiranju obeh algoritmov sva ugotovila, da je Deepware-ov algoritem natančnejši, če je videoposnetek daljši. Pri prepoznavi kratkih videoposnetkov pa ima težave. Algoritem podjetja Quantum Integrity pa uspešno prepozna tudi globoke ponaredke, ki so krajši. Oba algoritma imata prednosti in slabosti. Za nobenega ne moreva reči, da je boljši od drugega. Deepware-ov je hitrejši in preveri čisto vse izseke iz videoposnetka, tudi pri daljših posnetkih. Algoritem švicarskega podjetja pa uspešno prepozna tudi kratke globoke ponaredke, s katerimi ima Deepware težave. Spodnji graf lepo prikaže razliko v natančnosti pri vsakemu od teh desetih videoposnetkov.



Graf 11 - Primerjava med rezultati obeh algoritmov

Na zgornjem grafu (graf 11) je viden rezultat algoritmov. Bližje kot je rezultat 100 %, bolj siguren je algoritem, da je videoposnetek globoki ponaredek, če pa je posnetek bližje 0 %, pa algoritem meni, da posnetek ni preurejen.

Presenetilo naju je predvsem število lažno pozitivno zaznanih obrazov na videoposnetkih. Sicer so bili vsi videoposnetki, ki sva jih testirala, globoki ponaredki, vendar so vsebovali tudi osebe, katerih obrazi niso bili spremenjeni. Algoritma sta večkrat tudi te obraze zaznala kot globoke ponaredke.

Iz rezultatov je razvidno, da ima tehnologija zaznavanja globokih ponaredkov še veliko prostora za izboljšave. Problem je predvsem, da se ob izboljševanju algoritmov za zaznavo globokih ponaredkov izboljšujejo tudi algoritmi za kreacijo globokih ponaredkov. Tako je izdelava algoritma za detekcijo neskončna dirka med izdelovalci globokih ponaredkov in iskanjem rešitev za detekcijo.

Problem je tudi, da so globoki ponaredki vse bolj človeški in vsebujejo vse manj lastnosti, s pomočjo katerih lahko ugotovimo, da gre za spremenjen videoposnetek.

5 DISKUSIJA

Algoritmi za ustvarjanje globokih ponaredkov se med seboj razlikujejo po različnih lastnostih. Delujejo na različnih principih in nam omogočajo različne načine uporabe. Ustvarjanje globokih ponaredkov z njimi je vse lažje in ponaredki so vse bolj prepričljivi.

V prihodnosti bo detekcija globokih ponaredkov vse bolj pomembna. Meniva, da se bo iz algoritmov za detekcijo ponaredkov razvila velika industrija, saj je veliko podjetij, ki so lahko oškodovana z neprepoznanim globokim ponaredkom.

Pred nevarnostjo globokih ponaredkov svari celo FBI. Kot so sami rekli: »Prihajajo sintetični ljudje, pazite, komu verjamete.« Na spletu obstaja celo spletna stran, na kateri lahko vidimo portrete ljudi, ki so bili generirani s pomočjo umetne inteligence, čeprav takšna oseba sploh ne obstaja. S tem vidimo prave nevarnosti te tehnologije.

Slika 31 - Primer umetno generiranega portreta (This Person ..., 2021)

Pri izdelavi raziskovalne naloge sva že od samega začetka pričakovala, da nama bo uspelo izdelati lasten globoki ponaredek, saj je algoritmov in aplikacij za izdelavo le-teh ogromno. Na koncu sva uspela izdelati celo bolj napreden globoki ponaredek z uporabo enega od mnogih algoritmov in ne z uporabo bolj preproste aplikacije.

Najbolj naju je presenetilo, ko sva ugotovila, kako resnični izgledajo naprednejši globoki ponaredki. Še sama nisva ločila večine od neobdelanih videoposnetkov. S tem sva odkrila, kako resen je problem globokih ponaredkov in ugotovila, da sva tudi sama že kdaj najbrž videla globoko ponarejeni videoposnetek, ne da bi se tega sploh zavedala.

Ob začetku izdelave raziskovalne naloge sva si zadala štiri hipoteze, ki sva jih z najinimi rezultatu tudi potrdila ali ovrgla.

Hipoteza 1: Uspelo nama bo izdelati lasten globok ponaredek.

Prvo hipotezo lahko le delno potrdiva, saj nama je s pomočjo preprostih aplikacij za izdelavo globokih ponaredkov na telefonu res uspelo izdelati lasten globoki ponaredek, ko pa sva želela izdelati svoj globoki ponaredek na računalniku z uporabo programa DeepFaceLab pa sva imela nekaj težav.

Prva težava je bila to, da potrebujejo programi za izdelavo globokih ponaredkov ogromno časa, če model ni že vnaprej treniran. Z najinima videoposnetkoma sva ga trenirala približno 20 ur in rezultat je bil globoki ponaredek, ki bi ga lahko vsak z gotovostjo prepoznal.

Slika 32 - rezultat po 20 urah učenja

Druga težava je bila to, da nisva bila pozorna in sva ob izbiri videoposnetka izbrala posnetek Luke Dončiča na katerem niti enkrat ne pomežikne. Tukaj je problem, ker se brez njegovega pomežika tudi algoritem ne more naučiti kako narediti pomežik z njegovim obrazom. Tako je na končnem videoposnetku bil obraz Luke Dončiča, ki je premikal usta in govoril, vendar ni nikoli pomežiknil.

Če bi globoki ponaredek poskusila izdelati še enkrat in bi si izbrala boljše videoposnetke, ter bi ga procesirala dlje časa meniva, da bi bil rezultat veliko bolj prepričljiv.

Hipoteza 2: Orodja za prepoznavo globokih ponaredkov so sposobna prepoznati večino le-teh.

V najini raziskavi sva testirala dva orodja za prepoznavo globokih ponaredkov, ki sta nama bila na voljo oziroma sva do njih lahko dobila dostop. Velika verjetnost je, da obstaja še kakšen bolj napreden algoritem, ki javnosti žal ni dostopen. Ker je bil celoten cilj raziskovalne naloge pogledati, kako natančno prepoznajo globoke ponaredke algoritmi, ki so javnosti dostopni, ne pa algoritmi, do katerih lahko dostopa le nekaj oseb, sva z naborom algoritmov zadovoljna.

Ob najinem testiranju teh algoritmov sva bila zelo presenečena z rezultati. Pričakovala sva, da bo natančnost večja, ter da algoritmi v večini ne bodo prepoznali videoposnetkov, ki niso globoki ponaredki kot ponaredke. Po testiranju sva ugotovila, da sicer algoritmi res zaznajo nekatere globoke ponaredke, ampak je vse odvisno od načina izdelave posnetka in trajanja videoposnetka.

Med algoritmoma Deepware in Quantum Integrity je bila največja razlika pri krajših videoposnetkih, saj je algoritem podjetja Quantum Integrity bolj uspešno zaznal globoke ponaredke v krajših videoposnetkih, medtem ko je imel algoritem Deepware težave pri zaznavanju krajših videov.

Oba algoritma sta imela težave pri videoposnetkih, kjer je bilo naenkrat več oseb. To je bilo najbrž zaradi tega, ker algoritem ne uspe vedno ločiti obrazov in potem primerja obraz ene osebe z obrazom druge osebe. Ker tam opazi večje neujemanje, obraz zazna kot globoki ponaredek. Rešitev za to bi bil algoritem, ki bi obraze najprej sortiral in primerjal le obraze ene osebe. Algoritma tudi nista bila enotna, saj je bila velikokrat razlika med sigurnostjo velika. Najbolj opazno je to pri videoposnetku 6, kjer je bil algoritem Quantum Integrity 82 % siguren, da je videoposnetek globoki ponaredek, medtem ko je bil algoritem Deepware le 7 %, kar pomeni, da se mu ni zdel kot ponaredek.

Algoritmi za detekcijo so tehnologija, ki jo bomo v prihodnosti nujno potrebovali, zato je treba delati na izboljšanju trenutnih algoritmov in izdelovanju novih, preden bo prepozno. Meniva, da je premalo denarja in časa namenjeno razvoju teh tehnologij, saj se bomo zavedali, kako pomembne so šele, ko jih bomo zares potrebovali. Videli smo, da se lahko v kratkem času vse spremeni, in da je vse več pomembnih informacij podanih preko spleta. Če bi kdo želel, bi lahko

preuredil videoposnetke in koga prepričal, da je bilo povedano nekaj čisto drugega, kot je bilo zares.

Hipotezo sva delno potrdila, saj so algoritmi res zaznali veliko število globokih ponaredkov, vendar potrebujejo še veliko razvoja, preden bodo lahko s sigurnostjo zaznali vse globoke ponaredke, ne da bi zaznali videoposnetke, ki niso globoki ponaredki. Prav tako se morajo algoritmi ves čas izboljševati, saj se tudi globoki ponaredki izboljšujejo in brez izpopolnjevanja algoritmov za zaznavanje jih kmalu več ne bodo sposobni zaznati.

Hipoteza 3: Večina anketirancev ne bo ločila med globokim ponaredkom in pravim posnetkom.

Ko sva anketo poslala drugim dijakom, sva jih izzvala, naj se preizkusijo v reševanju ankete, saj naju je zanimalo, kakšni bodo rezultati in koliko ljudi bo odgovorilo pravilno. Bila sva presenečena, ko sva ugotovila, da je večina anketirancev na vprašanja odgovorila pravilno. Ko sva si postavljala hipoteze, sva bila sigurna, da večina ljudi ne bo znala prepoznati med globokim ponaredkom in originalom. Prepoznava globokih ponaredkov bi bila za anketirance veliko lažja, če bi ob videoposnetkih pustila tudi zvok. Za odstranitev zvoka sva se odločila, ker je velika večina globokih ponaredkov ustvarjena v satirične namene, in stvari, ki jih v ponaredkih govorijo, niso logične.

Zanimivo je bilo videti, kako zelo raznoliki in razdvojeni so bili odgovori. Kot primer bi vzela 6. vprašanje, kjer so bili odgovori razdeljeni skoraj na polovico. Nekaj več kot polovica, torej 51,16 % anketirancev, je pravilno ugotovilo, da je videoposnetek globoki ponaredek, 48,84 % anketirancev pa je pri tem vprašanju zgrešilo in so odgovorili, da videoposnetek ni globok ponaredek.

Tretjo hipotezo lahko ovrževa, saj je čisto pri vsakem vprašanju več kot polovica anketirancev odgovorila pravilno, torej so večinoma uspeli razlikovati med globokimi ponaredki in originali.

Čeprav je večina anketirancev odgovorila pravilno, pa misliva, da niso bili sigurni, ampak so večinoma tudi ugibali. Pomagalo pa je tudi to, da so bile na vseh videoposnetkih znane osebe, kar pomeni, da so jih lažje prepoznali. Anketo bi lahko izboljšala tako, da bi jih vprašala tudi, kako sigurni so s svojim odgovorom, ter bi tako tudi ugotovila, ali le ugibajo ali jih res prepoznajo.

Hipoteza 4: Uspelo nama bo izdelati aplikacijo, ki bo prepoznala globoke ponaredke.

Na svojo aplikacijo sva zelo ponosna. Sicer aplikacija uporablja API podjetja Deepware. Aplikacijo bi lahko nadgradila še s tem, da bi ustvarila svoj algoritem, vendar sva med ustvarjanjem raziskovalne naloge ugotovila, da je za izdelavo takšnega algoritma potrebno več časa in ljudi.

Ker nama je uspelo izdelati aplikacijo, ki si jo uporabniki lahko naložijo na telefon in z njo testirajo svoje videoposnetke, lahko četrto in zadnjo hipotezo popolnoma potrdiva.

Z rezultati raziskovalne naloge sva zelo zadovoljna. Nekateri so naju presenetili, nekatere pa sva pričakovala. Ob izdelavi sva se tudi veliko naučila.

Seveda je še nekaj prostora za izboljšave. Lahko bi na primer poskusila ustvariti lasten algoritem, vendar je to zaradi pomanjkanja časa, znanja in denarja neizvedljivo. Anketo bi lahko razširila tako, da bi od anketirancev dobila še več informacij glede izbire.

Raziskovalna naloga bi bila lahko uporabna pri izboljšavi obstoječih algoritmov, ki sva jih testirala, in tudi pri izdelavi novih algoritmov. Prav tako sva dobila vpogled v poznavanje globokih ponaredkov najinih sovrstnikov in prepoznavanje le-teh.

Aplikacija se lahko uporabi za razširjanje ozaveščenosti o globokih ponaredkih in možne nevarnosti, ki jo ti s sabo prinesejo. Prav tako je priročno orodje za testiranje videov, saj doda plast zaščite, preden nekemu videoposnetku popolnoma zaupamo.

6 ZAKLJUČEK

Področje globokih ponaredkov je zelo obsežno. V kratkem času se je hitro razvilo veliko različnih pripomočkov za ustvarjanje teh ponaredkov. Ti lažni posnetki so postali že tako napredni, da veliko ljudi sploh ne prepozna, ali je posnetek ponarejen ali ne. Preprosto verjamejo, da gre za resnične posnetke brez kakršne koli manipulacije ali spreminjanja. Zelo fascinantno je, da je ustvariti takšen videoposnetek ali sliko precej enostavno. To je tudi razlog, da je na spletu tako veliko globokih ponaredkov. Ti se dajo najti na skoraj vsakem področju in panogi. Nameni za izdelavo takšnih medijev so tudi zelo različni. Nekateri jih izdelujejo za zabavo, drugi imajo slabše namene. Takšni videoposnetki ali slike so lahko za neko osebo, ki je lažno prikazana na njih, zelo škodljivi. Uporaba algoritmov za ustvarjanje globokih ponaredkov ni omejena in kakovost posnetkov se hitro izboljšuje. Za v prihodnje je zelo pomembno, da se začnejo hitro razvijati tudi algoritmi za prepoznavo tovrstnih ponaredkov. Obstoječi namreč še niso dovolj natančni, da bi se lahko v prihodnje na njih zanesli. So pa ti algoritmi zagotovo korak v pravo smer.

Sama sva na začetku bila mnenja, da je globoki ponaredek možno opaziti, če si le dovolj pozoren. Vendar sva hitro ugotovila, da temu ni tako. Posnetki so v nekaterih primerih že tako dobro izdelani, da je nemogoče s samim človeškim očesom zaznati znake manipulacije. Rezultati ankete so pokazali, da imajo tudi drugi ljudje težave s prepoznavanjem globokih ponaredkov.

Oba algoritma za prepoznavo globokih ponaredkov, ki sva ju testirala, sta prepoznala večino ponarejenih posnetkov. Vsak je imel svoje prednosti in slabosti. Pri nekaterih posnetkih sta celo oba imela težave z detekcijo. Meniva, da sta oba kakovostna, ampak še potrebujeta veliko razvoja.

Ker se tehnologija globokih ponaredkov razvija zelo hitro, je pomembno, da to področje pride v širšo javnost. Če se bo veliko ljudi zavedalo, da se da videoposnetke in slike tako prepričljivo spreminjati, bo tudi širjenje lažnih novic težje.

7 POVZETEK

Globoki ponaredki se že pojavljajo v našem vsakdanjem življenju. Ti ponaredki so se v zelo kratkem času hitro razvili. V veliko primerih človeško oko ni več sposobno prepoznati, da gre za spremenjen videoposnetek ali sliko. Skupnost, ki se je razvila okoli globokih ponaredkih, se zaveda, da lahko ta tehnologija povzroči veliko škode, če se uporablja napačno. Na srečo avtorji globokih ponaredkov večinoma nimajo slabih namenov in jim je kreiranje tovrstnih posnetkov le v zabavo. Potrebno je pravočasno razviti algoritme za prepoznavo globokih ponaredkov, da bodo lahko natančno prepoznali, kateri posnetki so resnični in kateri ponarejeni.

Skozi postopek raziskovanje sva se naučila veliko o globokih ponaredkih. Prepričana sva, da je to tehnologija bližnje prihodnosti, ki pa zna biti problematična, če se bo uporabljala v slabe namene.

Obstoječe orodje za izdelavo globokih ponaredkov so enostavni za uporabo in z njimi je možno ustvariti zelo dobre ponaredke. Testirala sva več teh algoritmov. Najboljše se nama je zdelo orodje DeepFaceLab, s katerim sva izdelala kar precej globokih ponaredkov.

Zanimali so naju tudi algoritmi za prepoznavo globokih ponaredkov. Malo sva bila razočarana, da jih na voljo ni bilo tako veliko in da nekateri sploh niso imeli napisane dokumentacije. Za testiranje sva si izbrala dva algoritma, Deepware-ov algoritem in algoritem podjetja Quantum Integrity. Deepware-ov je imel na voljo tudi API, zato sva jih prosila za dostop. Na podlagi njihovega API-ja sva izdelala tudi aplikacijo za prepoznavo globokih ponaredkov. Sodelovala pa sva tudi s podjetjem Quantum Integrity. Z njimi sva se dogovorila za možnost testiranja njihovega algoritma za prepoznavo. To so nama omogočili preko njihove spletne aplikacije, ki je povezana z njihovim algoritmom. Oba algoritma sta prepoznala večino globokih ponaredkov, vendar pa nista bila čisto natančna. Problem se je pojavil tudi pri testiranju neponarejenih videoposnetkov, ki sta jih algoritma občasno označila za globoke ponaredke.

Izdelala sva anketo, ki jo je rešilo 131 dijakov Šolskega centra Velenje. Na podlagi rezultatov sva ugotovila, da so dijaki večinoma pravilno ugotovila, kateri videoposnetek je globoki ponaredek in kateri ne. Opazilo pa se je, da so bili kar precej zmedeni, kar pa sva pričakovala.

8 ABSTRACT

Deepfakes are already becoming a part of our day-to-day lives. Their development has progressed quickly in a short amount of time. In most cases the human eye isn't able to recognize these deepfakes. The deepfake community is aware of the potential danger this technology could create, if it's used for malicious purposes. Luckily most of the people creating deepfakes make them for fun and don't have malicious intent. It's important that proper detection algorithms are created. This will enable us to tell deepfakes apart from genuine videos.

During research we learned a lot about deepfakes. We are sure that deepfakes are a technology that will be evermore present in our day-to-day lives in the future. This could be a problem if they are used for spreading misinformation.

Current tools for creating deepfakes are easy to use and produce very realistic results. We tested a bunch of tools. Our personal favourite was a tool called DeepFaceLab. We used it to create numerus deepfakes of our own.

We were also interested in deepfake detection algorithms. Not finding that many good ones was a disappointment for us. Some algorithms didn't even have proper documentation, so we didn't use them. For our testing we used two algorithms. One was called Deepware. The other was made by a Swiss company called Quantum Integrity. The Deepware algorithm had an API, which we used to develop a mobile app for detecting deepfakes. We also collaborated with Quantum Integrity. They agreed to give us access to their deepfake detection algorithm for test purposes. Both algorithms successfully detected most of the videos we tested, but weren't completely accurate. There was also a problem, where videos that weren't deepfakes occasionally got detected as deepfakes.

We also conducted a survey, which got 131 responses from students. The results showed that most respondents were quite successful at detecting deepfakes. We noticed that the videos confused them a bit.

9 ZAHVALA

Najprej gre zahvala najinima mentorjema za potrpežljivost in strokovno pomoč pri delu.

Zahvalila bi se tudi podjetju Quantum Integrity za vso pomoč in dostop do njihovega detektorja.

Prav tako gre zahvala podjetju Deepware, ki nama je dovolilo brezplačno uporabo njihovega API za izdelavo najine aplikacije.

Profesorici Lidiji Šuster se zahvaljujeva za lektoriranje raziskovalne naloge, profesorici Vlasti Leban pa za lektoriranje angleškega dela.

Hvala tudi najinim staršem za moralno podporo.

10 LITERATURA IN VIRI

~ Elon Musk, ki poje sovjetsko pesem? Gre za deepfake tehnologijo. https://www.24ur.com/novice/znanost-in-tehnologija/elon-musk-deepfake.html (5. 2. 2021)

~ Gorenšek, T. 2019. Izziv prihajajočega desetletja: 'deepfake' tehnologija. https://www.24ur.com/novice/slovenija/izziv-prihajajocega-desetletja-deepfake-tehnologija.html (5. 2. 2021)

~ Adee, S. 2020. What are deepfakes and How Are They Created? https://spectrum.ieee.org/tech-talk/computing/software/what-are-deepfakes-how-are-they-created (6. 2. 2021)

~ Perov, I. 2020. DeepFaceLab: A simple, flexible and extensible face swapping framework. https://arxiv.org/abs/2005.05535 (5. 2. 2021)

~ Snapchat Cameo brings deepfake tool to social media app https://eu.usatoday.com/story/tech/2019/12/08/snapchat-cameo-deepfake-tool-social-media-app/4376790002/ (30. 3. 2021)

~ Reface https://hey.reface.ai/ (30. 3. 2021)

~ Deepware https://deepware.ai/ (30. 3. 2021)

~ Quantum Integrity https://quantumintegrity.ch/ (30. 3. 2021)

~ DeepFake in KYC https://quantumintegrity.ch/wp-content/uploads/2021/03/KYC-04-March-21.pdf (30. 3. 2021)

~ Deepware Scanner https://scanner.deepware.ai/ (3. 4. 2021)

~ Quantum Integrity Scanner http://mvp.quantumintegrity.ch/ (3. 4. 2021)

~ QuestionPro https://www.questionpro.com/ (5. 4. 2021)

~ This Person Does Not Exist https://thispersondoesnotexist.com/ (31. 3. 2021)

VIRI TESTNIH VIDEOPOSNETKOV:

1. Videoposnetek: https://youtu.be/DdpsYojwW5c
2. Videoposnetek: https://www.youtube.com/watch?app=desktop&v=VhFSIR7r7Yo
3. Videoposnetek: https://www.youtube.com/watch?app=desktop&v=vo8MIOPoejM
4. Videoposnetek: https://www.youtube.com/watch?app=desktop&v=2hUkBHtOMg4
5. Videoposnetek: https://www.youtube.com/watch?app=desktop&v=ZUOrNVEQ18Q
6. Videoposnetek: https://www.youtube.com/watch?app=desktop&v=LtVURjxACi4
7. Videoposnetek: https://www.youtube.com/watch?app=desktop&v=2svOtXaD3gg
8. Videoposnetek: https://www.youtube.com/watch?app=desktop&v=aau8qa3xgFs
9. Videoposnetek: https://www.youtube.com/watch?app=desktop&v=H3pV-_iyT4U
10. Videoposnetek: https://www.youtube.com/watch?app=desktop&v=bPhUhypV27w