Nedelo

Država: Slovenija

Stran: 4

Površina: 1.332 cm^2

03.02.2024
Sobota

1 / 4

GLOBOKI PONAREDKI

Samo pomislite, da nič od tega ni res

Včeraj Taylor Swift, danes Nataša Pirc Musar, jutri ...? Zadnje čase je mogoče vsak teden brati o javnih osebnostih, ki so se jih anonimneži lotili s pomočjo globokih ponaredkov, a v senci odmevnih primerov se skriva ogromno zlorab medijsko povsem neizpostavljenih ljudi: včeraj on, danes ti, jutri .. »Globoki ponaredki so zgolj še eno orodje, ki ga imamo po novem na voljo. Včasih smo uporabljali fotošop za slike, potem pa je tehnologija napredovala,« globoke ponaredke v kontekst postavlja komunikator znanosti in raziskovalec na Kemijskem inštitutu dr. Matej Huš.

Lucijan Zalokar

Zamislite si situacijo. Stari ste dvanajst let. Učitelj vas krivično oceni. Vaša simpatija se zagleda v drugega. Sošolci vas zasmehujejo. Radi bi se maščevali. Ampak v analognem svetu ne premorete pravih vzvodov. In nato pridete domov. Sedete za računalnik. Odprete spletni iskalnik. Od maščevanja ste oddaljeni le nekaj klikov. Le nekaj klikov in svoje sovražnike lahko anonimno osramotite pred vsem svetom. Pedofilija? Umor? Posilstvo? Incest? Domišljija ne pozna meja. Tehnologija prav tako ne. Kaj boste storili? Naslednji dan nastavili še drugo lice ali udarili nazaj?

Pred dobrim tednom se je večina medijev z globalnim dosegom znova razpisala o globokih ponaredkih (angl. deepfake), torej o fotografijah in videoposnetkih, ki so s pomočjo algoritmov spremenjeni v tako prepričljiv ponaredek, da ga mnogi ne prepoznajo. Vzrok so bile lažne pornografske fotografije ameriške zvezdnice Taylor Swift, ki si jih je na družbenih omrežjih ogledalo na desetine milijonov uporabnikov. Pozneje so jih odstranili, a svetovni splet je kot tuba zobne paste: ko snov enkrat spraviš na plan, je zlepa ne moreš stlačiti nazaj.

Od Jima Carreyja do samega papeža

Politiki in oboževalci ameriške popzvezdnice so takoj stopili na njeno stran in pozvali k zaostritvi zakonodaje na tem področju. A dejstvo je, da je tehnologija, ki omogoča globoke ponaredke, že več let prosto dostopna sleherniku s kančkom smisla za uporabo spletnih orodij in da je 34-letna glasbenica zgolj naslednja v vrsti slavnih in neslavnih osebnosti, katerih podobe so s pomočjo umetne inteligence predelali v lažne fotografije in videoposnetke. Spomnimo se zgolj videoposnetka papeža Frančiška v beli puhovki, ki so ga ustvarili s pomočjo orodja Midjourney in ki je po spletu zakrožil marca lani, Jima Carreyja, s katerim so v *Izžarevanju* nadomestili Jacka Nicholsona, pa Nicki Minaj in Toma Hollanda, ki sta se brez njune vednosti znašla v skupnem prizoru, ter seveda Nicholasa Cagea, čigar obraz so v enem izmed prvih globokih ponaredkov vstavili v različne filme.

»Globoki ponaredki so obstajali že pred letom 2022, sta pa takrat ChatGPT na področju besedil in Midjourney na področju slik širši javnosti prvikrat pokazala, da za uporabo ne potrebujemo nobenega posebnega znanja,« je dejal kemik in komunikator znanosti

dr. Matej Huš ter izpostavil brezplačno aplikacijo FakeApp, ki je že leta 2018 omogočila spreminjanje obraza osebe na posnetku ali fotografiji. »Takrat smo potrebovali obstoječi posnetek, pa tudi rezultat je bil očitno umeten, zato smo se tedaj izdelkom bolj ali manj zgolj smejali. V naslednjih petih letih pa je razvoj omogočil, da lahko katerokoli osebo posadimo v kakršenkoli položaj in ji v usta položimo poljubne besede,« je pristavil sogovornik, ki redno spremlja razvoj sodobnih tehnologij in z njim v različnih medijih seznanja slovensko javnost.

Po njegovem mnenju je povedno, da se je silovit medijski odziv zgodil šele potem, ko so tehnologijo začeli zlorabljati za posnemanje zvezdnikov, a po drugi strani opozarja, da obstajajo dokazi, da sistematično in organizirano širjenje dezinformacij z namenom vplivanja na javno mnenje ali destabilizacije družbe na družbenih omrežjih poteka že celotno desetletje. »Globoki ponaredki omogočajo še hitrejše in cenejše širjenje, saj je treba imeti zelo malo znanja za njihovo ustvarjanje. Čeprav se običajno sorazmerno hitro ugotovi, da gre za lažne posnetke, so posledice še vedno lahko opazne. Do vseh ljudi popravek informacije ne pride, z obiljem dezinformacij se povečuje splošno nezaupanje, včasih pa zadostuje že nekajminutno vladanje neresnice – pomislite samo na dogodke na borzah.«

Preprosto in cenovno dostopno

V senci bolj ali manj škandaloznih primerov iz sveta slavnih se skriva na tisoče zlorab umetne inteligence, ki so prizadele povsem običajne ljudi. Te so v marsikaterem pogledu še bolj travmatične kot manipulacije slavnih osebnosti, saj lahko Taylor Swift s pomočjo vrhunskih pravnikov in strokovnjakov za odnose z javnostjo, svetovno prepoznavnih medijev in navsezadnje s svojim vplivom v nekaj urah seznani ves svet, da so fotografije z njeno podobo lažne, medtem ko slehernik nima teh možnosti.

O tej problematiki je lani spregovorila ameriška spletna vplivnica Gabi Belle, ki je na spletu našla na desetine fotografij, na katerih je bila gola. Jasno ji je bilo, da gre za ponaredke, a trajalo je precej časa, preden je dosegla, da so jih umaknili. In takoj zatem so se pojavile nove. »Kot ženska nisi nikoli varna,« je dejala za *Washington Post*.

Prava zakladnica tovrstnih zlorab so pornografske spletne strani. Podjetje Sensity AI je leta 2019 opravilo raziskavo, v kateri so ugotovili, da se v več kot 90 odstotkih globoki ponaredki uporabljajo za ustvarjanje nesporazumnih pornografskih posnetkov in da so v 99 odstotkih primerov žrtve ženske. S tem se zgolj potrjuje nepisano »pravilo 34«, ki pravi, da vse, kar obstaja na spletu, obstaja tudi kot oblika pornografije. Po mnenju Mateja Huša to niti ni tako nenavadno, saj »so po nekaterih statistikah znatni deleži – ocene nihajo od štiri do 40 odstotkov – interneta namenjeni pornografiji«.

K temu veliko pripomoreta cenovna dostopnost in preprosta uporaba različnih orodij, ki s pomočjo umetne inteligence ustvarijo ponarejeno vsebino. Kar je bil nekoč fotošop, so danes aplikacije za slačenje oblečenih ljudi na fotografijah.

»Mama, poglej, kaj so mi naredili,« je v omenjenem prispevku za *Washington Post* začetek pogovora s svojo 14-letno hčerko opisala Miriam Al Adib Mendiri. »Nato je pokazala fotografijo, na kateri je bila gola.«

Primer je obravnavala policija, a škoda je bila že storjena. Njena hči nikoli ni pozirala brez oblačil, toda skupina lokalnih fantov je na družbenih omrežjih našla povsem običajne fotografije vrstnic in jih s pomočjo aplikacije AI nudifier pretvorila v gole. Fantje niso bili računalniški eksperti, toda tovrstne aplikacije so tako preproste, da je treba zgolj naložiti fotografijo, nato pa algoritmi opravijo svoje. Četudi žrtev doseže, da ponarejene fotografije izbrišejo, nikoli ne more vedeti, kdo jih je shranil na svojih zasebnih napravah.

Enako velja pri ustvarjanju nesporazumnih pornografskih posnetkov. Revija *MIT Technology Review* je že pred tremi leti poročala, da obstaja spletna storitev, ki fotografijo poljubne osebe umesti v vnaprej pripravljene pornografske vsebine.

Roke ne lažejo

Če pogledamo širšo sliko, so globoki ponaredki med drugim ena izmed oblik medijske manipulacije, ki je stara toliko kot mediji sami. Seznam je neskončno dolg: podkupljivi literarni kritiki v 19. stoletju, o katerih se je v *Izgubljenih iluzijah* tako obširno razpisal Honoré de Balzac, brisanje politično kompromitiranih posameznikov s fotografij, značilno predvsem za totalitarizme 20. stoletja, radijska igra *Vojna svetov*, s katero je tedaj mladi in dokaj neznani Orson Welles na vzhodni obali ZDA povzročil pravcato moralno paniko, nagrajevani novinarji, ki so si izmišljali sogovornike v svojih reportažah ... Vsem naštetim primerom je skupno, da so manipulatorji morali pokazati dobršno mero znanja in truda, če so hoteli doseči svoj cilj. Vsakdo pač ne more spisati prepričljive literarne kritike ali novinarske reportaže, propagandni oddelki totalitarnih sistemov so zaposlovali na stotine ljudi, Welles pa velja za enega največjih filmarjev vseh časov.

Kot že omenjeno, globoki ponaredki to perspektivo postavljajo na glavo še precej bolj, kot jo je pred leti fotošop. Ponarejene fotografije in videoposnetki so trenutno že zelo prepričljivi, a na koncu strokovnjaki, včasih pa tudi laiki, še vedno lahko prepoznajo, ali je vsebina lažna ali ne. Pri tem se poraja vprašanje, ali bo ostalo pri tem ali pa se bo tehtnica v prihodnje nagnila v korist ponarejevalcev. »Gre za tekmo med ustvarjalci orodij za generativno umetno inteligenco in pisci orodij za prepoznavanje tovrstnih izdelkov. Pogosto so to iste skupine, saj ustvarjanje globokih ponaredkov ni nujno slabo. Včasih imajo povsem legitimne razloge in ustrezna dovoljenja za uporabo,« meni sogovornik s Kemijskega inštituta.

Na družbenih omrežjih je v zadnjem času mogoče najti ogromno vsebin, ob katerih piše: »Samo pomislite, da nič od tega ni res.« Ponavadi gre za fotografije raznoraznih prizorov iz vsakdanjega življenja: skupina ljudi večerja v restavraciji, jata ptic leti nad jezerom, moški sedi v letalu, ženska v knjižnici ... Kdor podobe preleti zgolj bežno in brez predhodnega opozorila, nikakor ne more vedeti, da jih je ustvarila umetna inteligenca, a številni uporabniki se jim temeljito posvetijo in prej ali slej najdejo kakšno nedoslednost. »The hands, chico, they never lie,« *(Roke, fant, te nikoli ne lažejo, op. p.)* je eden bolj duhovitih komentatorjev parafraziral znani citat Ala Pacina iz filma *Brazgotinec*.

»Res je, prve verzije generativne umetne inteligence za slikanje so imele največ težav z obrazi in rokami. Za prve je razlaga tudi v našem dojemanju slike, saj ljudi prepoznavamo prav po obraznih potezah in imamo nesorazmerno velik del možganskih kapacitet namenjen ravno razpoznavanju obrazov. Že majhne nedoslednosti nas močno zmotijo in jih takoj opazimo, s čimer je povezan tudi termin *uncanny valley* iz robotike,« se je Huš navezal na izraz, ki opisuje srhljiv občutek, ki ga človek občuti, ko sreča robota s človeškimi lastnostmi. »Čim bolj so humanoidni roboti podobni lju-

dem, tem bolj smo jim naklonjeni, dokler niso že zelo podobni ljudem, a ne povsem enaki. Takrat se nam zdijo spačeni, groteskni, odvratni. Če pa so nam povsem podobni, imamo spet pozitivno mnenje.«

Da je verodostojna upodobitev dlani in prstov na roki izjemno zahtevna, so govorili že številni slikarji, a nespretnost umetne inteligence pri risanju okončin ni povezana z njihovimi preglavicami. Kot pravi Huš, so težava vhodni podatki za trening umetne inteligence, saj roke in prsti običajno niso zelo izpostavljeni na slikah, zato je informacij manj. »Umetna inteligenca se uči samo z gledanjem slik, zato nima zavedanja, da dlan sestavlja pet prstov s členki, zapestje ... Skratka, ne pozna biomehanike, dlani pa so na slikah lahko v zelo različnih položajih. Podobno velja, recimo, še za zobe in ušesa.«

Regulacija in učenje

Kljub temu že zdaj živimo v dobi, ko sta pojma resnica in resničnost ogrožena kot nikoli doslej. Večina zakonov, ki urejajo naša življenja, je nastala v časih, ko je ideja o globokih ponaredkih obstajala zgolj v znanstvenofantastičnih romanih ali pa še tam ne, zato ne presenča, da bi večina ljudi rada videla, da se to področje čim prej pravno regulira. »Pojavljajo se pobude, da bi vsaj široko dostopna komercialna orodja v izdelke vgrajevala vodne žige, iz katerih bi bilo jasno, da gre za izdelke umetne inteligence. Lani so se OpenAI, Google, Meta in nekaj drugih velikanov na načelni ravni zavezali, da bodo sprejeli ustrezne ukrepe za varno uporabo, kamor sodi tudi označevanje izdelkov umetne inteligence. A številna orodja so odprtokodna in jih lahko prenese, predrugači in uporabi kdorkoli. Tam je to precej težje in tudi pravne omejitve bi bilo zelo težko izvajati,« pravi Huš, ki je pred dobrim letom objavil knjigo *Od kvarkov do galaksij*, v kateri so zbrani njegovi najboljši poljudni in strokovni članki o znanosti in tehnologiji.

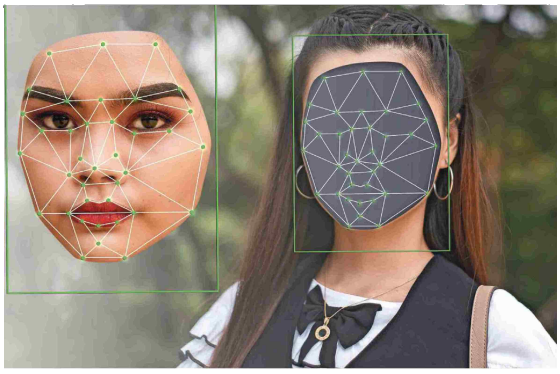
Medtem ko se zakonodajni organi ukvarjajo z vprašanji, kako regulirati umetno inteligenco in vse njene implikacije, lahko slehernik v Kennedyjevi maniri vprašanje postavi na glavo: ne gre samo za to, kaj nam bodo prinesli globoki ponaredki, ampak tudi za to, kako bomo ravnali z njimi.

»Po mojem mnenju se človeška narava in tudi fiziologija v zadnjih 20.000 letih nista bistveno spremenili, drugačni so zgolj okolje in orodja. Kakor smo se morali naučiti, da ne gre slepo verjeti vsemu, kar preberemo na internetu, se bomo očitno morali naučiti, da ne gre slepo verjeti vsakemu posnetku, ki se pojavi. Zato pa je pomembno vedeti, kaj tehnologija omogoča, da se lahko na to pripravimo,« je dejal sogovornik in za konec izpostavil, da je s tega vidika dobro, da so globoki ponaredki in drugi produkti generativne umetne inteligence pobegnili v javnost in so na voljo vsem. »Samo pomislite, če bi imela tehnologijo le vojska ali kakšne obveščevalne službe, mi pa sploh ne bi vedeli, da je to mogoče.«

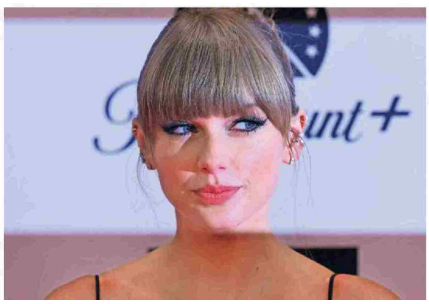
Kakor smo se morali naučiti, da ne gre slepo verjeti vsemu, kar preberemo na internetu, se bomo očitno morali naučiti, da ne gre slepo verjeti vsakemu posnetku, ki se pojavi.

dr. Matej Huš

Taylor Swift je zgolj naslednja v vrsti slavnih in neslavnih osebnosti, katerih podobe so s pomočjo umetne inteligence predelali v lažne fotografije in videoposnetke.



Ponarejene fotografije in videoposnetki so že zelo prepričljivi, a na koncu strokovnjaki, včasih pa tudi laiki, še vedno lahko prepoznajo, ali je vsebina lažna ali ne. FOTO SHUTTERSTOCK



Ena izmed ponarejenih fotografij zvezdnice je v samo 17 urah, preden so jo odstranili, na omrežju X imela kar 47 milijonov ogledov. FOTO WOLFGANG RATTAY/REUTERS

Argentinec Santiago Barros je s pomočjo aplikacije Midjourney ustvaril podobo moškega, ki je kot deček izginil pred več kot štiridesetimi leti in ga odtlej niso več videli. FOTO AGUSTIN MARCARIAN/REUTERS