

2024 – 2025



EVROPSKA NOČ RAZISKOVALCEV



»Financira Evropska unija. Za izražena stališča in mnenja odgovarja samo avtor (ali avtorji) in ne odražajo nujno stališč Evropske unije ali Evropske izvajalske agencije za raziskave. Niti Evropska unija niti Evropska izvajalska agencija za raziskave ne moreta biti odgovorna zanje.«

Psihologija globokih ponaredkov

Doc. dr. Nejc Plohl

Oddelek za psihologijo FF UM

Katedra za psihologijo MF UM

PRAVNO OBVESTILO:

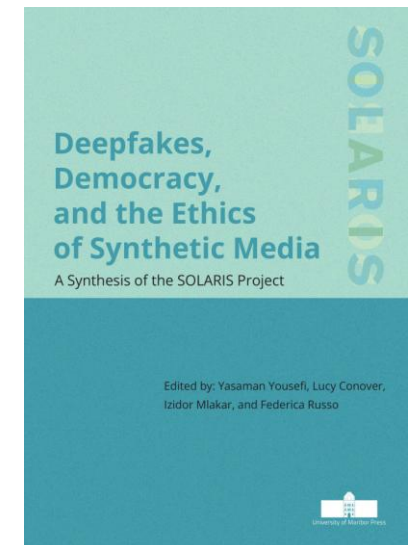
Na vseh dogodkih projekta oooZnanost! poteka snemanje in fotografiranje za namen promocije in poročanja o dogodkih. Če vstopite na lokacijo (spletnega) dogodka, boste lahko posneti in fotografirani. Z vstopom na to lokacijo, dajete dovoljenje organizatorjem in Evropski komisiji, da vas lahko snemajo, fotografirajo, zvočno snemajo in uporabijo vaše posnetke po lastni presoji. Obiskovalci zato ne boste uveljavljali nobene odgovornosti proti organizatorjem in Evropski komisiji v zvezi z zgoraj navedenim.

V kolikor se z zgoraj navedenim ne strinjate, vljudno prosimo, da s tem seznanite organizatorje na: ern@um.si. E-sporočilu obvezno priložite visokokakovostni sken fotografije z osebnega dokumenta, da vas lahko organizator izloči iz vseh posnetkov in fotografij skupaj z navedbo, na kateri lokaciji in katerega dne bi lahko bili posneti s strani organizatorjev. Pooblaščen oseba za varstvo podatkov Univerze v Mariboru je izr. prof. dr. Miha Dvojmoč (dpo@um.si).

Psihologija globokih ponaredkov

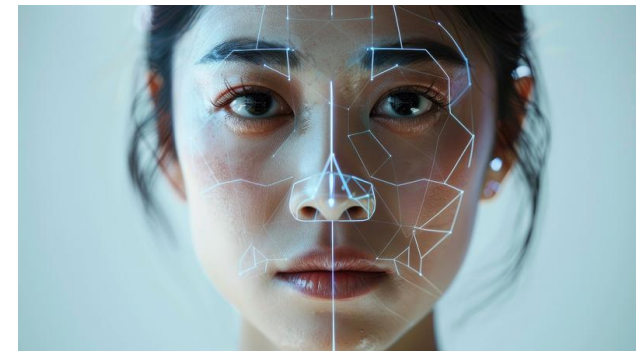
- 1. Vloga psihologije pri razumevanju problematike globokih ponaredkov
- 2. Ali sploh verjamemo globokim ponaredkom?
- 3. Podrobnejši vpogled v zaznavanje globokih ponaredkov
- 4. Individualne lastnosti, ki se povezujejo z zaznavami globokih ponaredkov
- 5. Zakaj so zaznave globokih ponaredkov pomembne?
- 6. Zaključek

Mnoga spoznanja, ki so del današnje predstavitve, so plod projekta „Strengthening democratic engagement through value-based generative adversarial networks“ (SOLARIS; financer: Obzorje Evropa; vodja projekta na UM: dr. Izidor Mlakar)



Problematika globokih ponaredkov in psihologija

- **Globoki ponaredki (angl. *deepfakes; deep learning + fake*)** so sintetični medijski materiali, ustvarjeni s pomočjo umetne inteligence, ki lahko zelo prepričljivo posnemajo resnične ljudi ali dogodke (Westerlund, 2019)
- Pojavljajo se lahko v različnih oblikah (Diel idr., 2024)
 - **Avdio:** umetno generiran govor, ki posnema glas določene osebe (uporaba za zabavo, vendar tudi finančne prevare, ...)
 - **Slike:** umetno ustvarjeni ali manipulirani obrazi in fotografije (ustvarjanje lažnih identitet ali fotografskih „dokazov“)
 - **Videoposnetki:** kombinacija manipulirane slike in zvoka (lahko prikazujejo osebo, ki govori ali počne nekaj, česar v resnici ni); v sodobnih raziskavah in na današnji predstavitvi bo poudarek predvsem na tej obliki globokih ponaredkov



Problematika globokih ponaredkov in psihologija

- Do nedavnega so se lažne informacije širile predvsem v obliki besedil (npr. članki, objave na spletnih socialnih omrežjih)
- Globoki ponaredki predstavljajo novo generacijo lažnih informacij, saj **združujejo več elementov, ki povečujejo njihovo prepričljivost** (npr. Lee in Shin, 2022), kot so:

Multimodalnost

Močan občutek avtentičnosti

Visoka realističnost

Večja emocionalna angažiranost

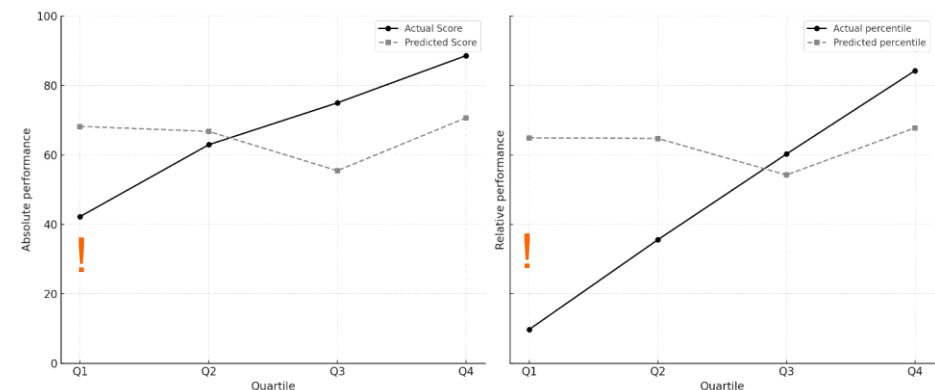
- **Možne družbene posledice?** Globoki ponaredki bi, tako kot druge oblike lažnih informacij, lahko prispevali k širjenju propagande, vplivanju na javno mnenje in politične procese, eroziji zaupanja v institucije, povečanju družbene polarizacije in ustvarjanju splošnega dvoma v informacije (tudi avtentične posnetke; angl. *liar's dividend*; npr. Hameleers idr., 2023)

Ali sploh verjamemo globokim ponaredkom?

- Ljudje radi verjamemo, da sami dobro prepoznavamo globoke ponaredke ...
- ... Vendar se to prepričanje pogosto izkaže za zmotno; svoje sposobnosti pogosto precenjujemo, to pa še posebej velja za tiste, ki so v resnici v tem najslabši (Dunning-Krugerjev učinek; Kruger in Dunning, 1999)



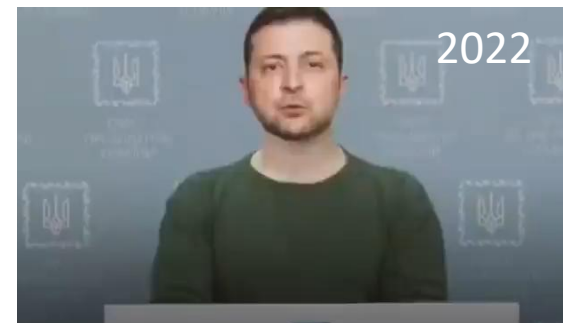
V eni od raziskav (Plohl idr., 2025a) smo 229 udeležencev izpostavili 12 kratkim videom (8 globokih ponaredkov). Udeleženci so med drugim morali presoditi, ali gre za resničen posnetek in oceniti delež pravilno prepoznanih posnetkov (absolutna uspešnost) ter uspešnost glede na druge sodelujoče (relativna uspešnost)



Ali sploh verjamemo globokim ponaredkom?

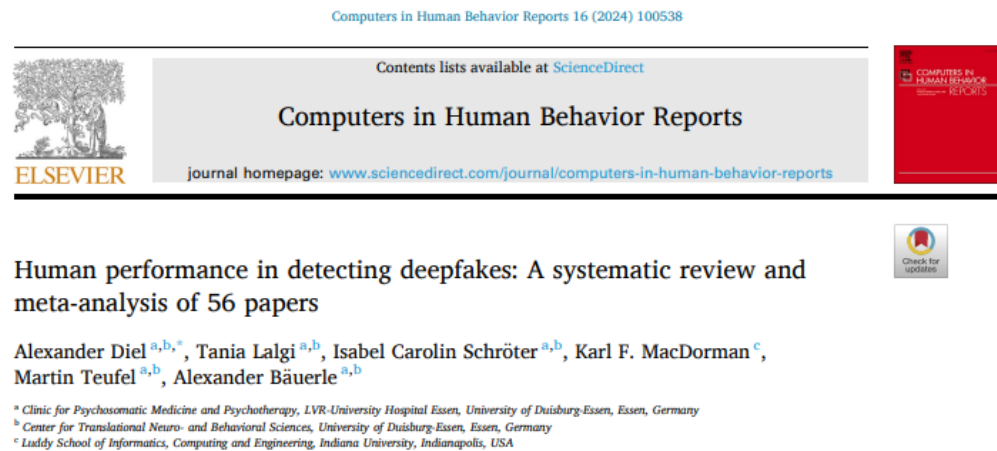
- Podatki, vezani na uspešnost prepoznavanja globokih ponaredkov, se med individualnimi raziskavami precej razlikujejo (različni dražljaji in načini merjenja prepoznave)
- Köbis idr. (2021): 16 kratkih videoposnetkov (~10 sekund); povprečna uspešnost 57,6 %
- Somoray in Miller (2023): 20 kratkih videoposnetkov (~10 sekund); povprečna uspešnost 60,7 %
- Naša nedavna raziskava (Plohl idr., 2025b): 2 videoposnetka (~1 minuta) iz večjega nabora možnih dražljajev, ki so se razlikovali v dovršenosti; povprečna uspešnost se je za nižje kakovostne videoposnetke gibala med 43,5 % in 60,4 %, za višje kakovostne videoposnetke pa med 30,9 % in 36,6 %

↓
Implikacije?



Ali sploh verjamemo globokim ponaredkom?

- Najbolj veljaven odgovor na vprašanje, ali sploh verjamemo globokim ponaredkom, nudi nedavni sistematični teoretični pregled z meta-analizo



- Pregled 56 raziskav z več kot 86.000 udeleženci, ki so bili izpostavljeni globokim ponaredkom
- Povprečna uspešnost prepoznavne globokih ponaredkov v obliki videoposnetkov: 57,3 % (podobno metanju kovanca)

Psihologija globokih ponaredkov

Ali sploh verjamemo globokim ponaredkom?



An AI-generated photo of Tom Holland and Zendaya's wedding has somehow surpassed 10 million likes on Instagram.



- lynnemagic** • 2d ago
Ai is getting out of hand
1.3K 1 Reply 1 Award 1 Share
- ChemicalNatural5506** • 2d ago
please tell me they can sue or something cause this is so weird
950 1 Reply 1 Award 1 Share
- Available_Cold_7659** • 2d ago
this is getting really bad... i can't even tell that this is Ai!!! Or maybe I'm dumb 🤡
258 1 Reply 1 Award 1 Share



ŠOK! Kardiologa so pretepli v neposrednem prenosu, ker je razkril medicinsko skrivnost!



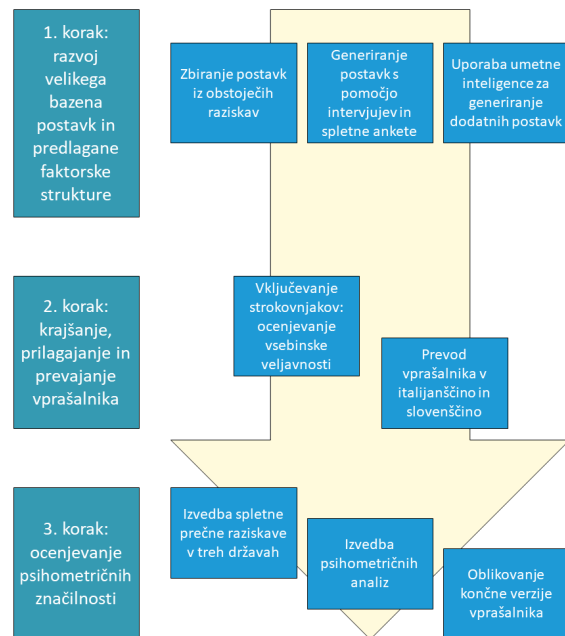
Skandal na TV-showu: znani slovenski kardiolog je dobil „pljunc“



Kardiolog je bil premagan v oddaji PTV SLO - povedal je, kako živeti brez srčnega infarkta!

Podrobnejši vpogled v zaznave

- Zaupanje globokim ponaredkom variira ne le med dražljaji (glede na objektivne lastnosti globokih ponaredkov), temveč tudi med posamezniki, v odvisnosti od **subjektivnih zaznav globokih ponaredkov** in individualnih lastnosti oseb



Kljub temu da gre za kompleksen preplet zaznav različnih vidikov, v preteklosti ni bilo celovitih vprašalnikov, ki bi jih ustrezno zajeli

To vrzel smo naslovili v projektu SOLARIS: razvoj celovitega vprašalnika **zaznane verodostojnosti globokih ponaredkov**

INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION
2025, VOL. 41, NO. 11, 6786–6803
<https://doi.org/10.1080/10447318.2024.2384821>



OPEN ACCESS [Check for updates](#)

Development and Validation of the Perceived Deepfake Trustworthiness Questionnaire (PDTQ) in Three Languages

Nejc Plohl^a, Izidor Mlakar^b, Letizia Aquilino^{c,d}, Piercosma Bisconti^c, and Urška Smrke^b

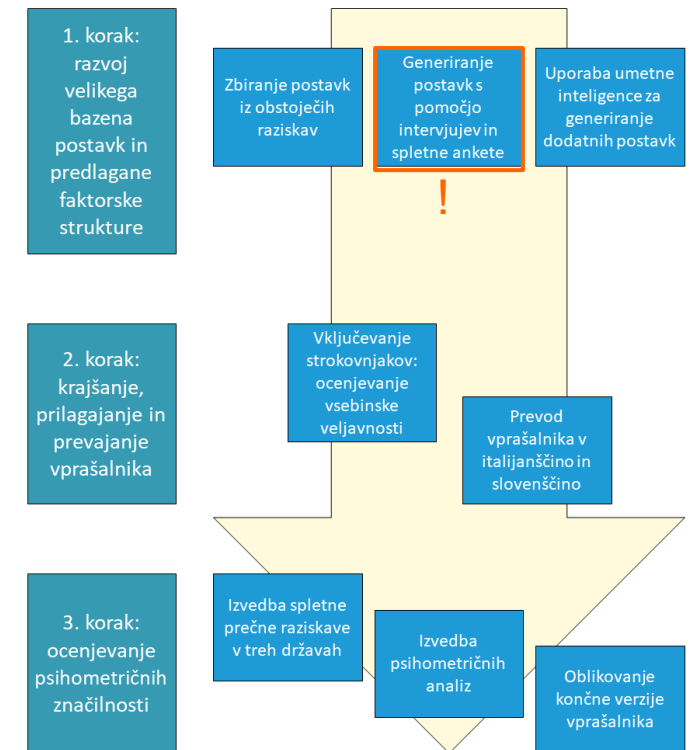
^aDepartment of Psychology, University of Maribor, Maribor, Slovenia; ^bFaculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia; ^cDEXAI – Artificial Ethics, Rome, Italy; ^dResearch Unit on Theory of Mind, Department of Psychology, Università Cattolica Del Sacro Cuore, Milan, Italy

Podrobnejši vpogled v zaznave

- 26 intervjujev z različnimi relevantnimi deležniki; ogled različnih videoposnetkov, razmišljanje na glas ob odločanju glede avtentičnosti posnetkov; skupno 305 relevantnih trditev

„Vsebina ni prepričljiva in deluje izmišljeno“
„To ni resnica, zdi se, kot da me nekdo želi prenesti naokoli“
„Ustnice niso bile popolnoma usklajene z govorom“
„Njegove geste so bile nenavadne; zdi se zelo nenaravno, da bi nekdo to razlagal na tako teatralen način“
„Video je bil slabe kakovosti, z nizko ločljivostjo“
„Video je objavila neka nepoznana oseba“

- Po 1. koraku : 419 trditev → 123 trditev in prva struktura (vsebina, oseba na posnetku, vir, tehnični aspekti, kontekst)
- Po 2. koraku: 31 trditev z ustrezno veljavnostjo in jasnostjo (kontekst izločen)



Podrobnejši vpogled v zaznave

- V zadnjem koraku smo izvedli večjo raziskavo na treh vzorci (ZK, Italija, Slovenija): $N = 733$
- Rezultati faktorских analiz so pokazali 2-faktorsko strukturo 22-postavčnega vprašalnika: **zaznana verodostojnost informacij in vira** (11 postavk) in **zaznana verodostojnost predstavitve** (11 postavk)

Spodnji vprašalnik vsebuje trditve, ki se nanašajo na vaše zaznavanje ravnokar ogledanega videoposnetka. Pozorno preberite vsako trditev in označite svoje strinjanje s pomočjo 7-stopenjske lestvice od "Sploh se ne strinjam" do "Povsem se strinjam". Če menite, da na določeno trditev ne morete odgovoriti, prosimo izberite odgovor "Neutrarno".

	Sploh se ne strinjam	Ne strinjam se	Nekoliko se ne strinjam	Neutrarno	Nekoliko se strinjam	Strinjam se	Povsem se strinjam
1. Predstavljene informacije so se zdele prepričljive.	1	2	3	4	5	6	7
2. Premiki ust osebe na videoposnetku se niso popolnoma ujemali z zvokom.	1	2	3	4	5	6	7
3. Ozadje na videoposnetku je vsebovalo nerelevantne predmete ali predmete, za katere se zdi, da tja ne sodijo.	1	2	3	4	5	6	7
4. Predstavljene informacije so se zdele verjetne.	1	2	3	4	5	6	7
5. Glas osebe na videoposnetku se mi je zdel nenaraven.	1	2	3	4	5	6	7

- Dimenziji sta relativno neodvisni, visoko zanesljivi in konsistentni skozi vse tri vzorce

Individualne lastnosti, povezane z zaznavami

- Na vprašanje, zakaj ljudje verjamejo globokim ponaredkom, ne moremo odgovoriti samo z analizami videa; pomembne so tudi **splošnejše individualne razlike med ljudmi**
- Raziskave kažejo, da so ti dejavniki **pogosto pomembnejši od samih lastnosti sporočila**, saj umerjajo način, kako informacije interpretiramo
- Pri presojanju globokih ponaredkov imajo pomembno vlogo sociodemografski, motivacijski in kognitivni dejavniki

Sociodemografski dejavniki

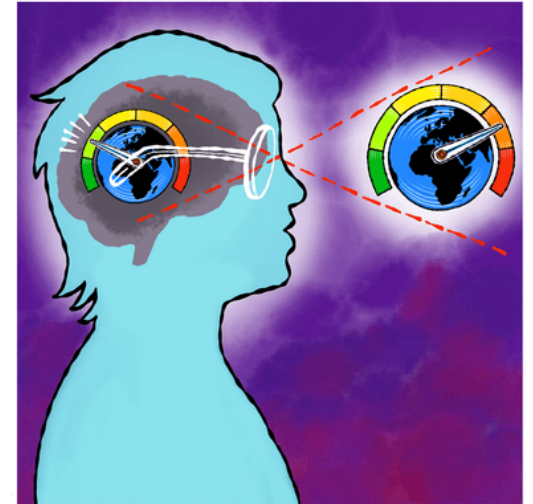
- **Starost:** da so starejši posamezniki bolj nagnjeni k zaupanju globokim ponaredkom, tako z vidika vsebine kot prezentacije videoposnetka
- **Uporaba spletnih socialnih omrežij:** pogostejša uporaba je povezana z večjim zaupanjem vsebini videoposnetka, ne pa tudi s prepoznavanjem manipulirane prezentacije



Individualne lastnosti, povezane z zaznavami

Motivacijski dejavniki

- Velik del raziskav dovzetnosti za lažne informacije temelji na **teoriji motiviranega rezoniranja** (Kunda, 1990): ljudje informacije pogosto presojava tako, da potrdijo naše obstoječe prepričanja ali cilje → pričakovali bi predvsem povezave z zaznano verodostojnostjo vsebine
- **Politična orientacija:** večja politična konservativnost povezana z večjim zaupanjem v vsebino globokih ponaredkov (pomen (ne)skladnosti s predstavljenim sporočilom!), ni pa bila povezana z zaznavanjem prezentacije
- **Zarotniška miselnost:** v naših raziskavah brez jasne povezave z zaznano verodostojnostjo; ta dejavnik verjetno močno odvisen od konkretnega konteksta sporočila
- **Zaupanje v medije:** posamezniki z višjim zaupanjem v medije bolj nagnjeni k temu, da zaupajo vsebini globokih ponaredkov → zaupanje kot dvorezen meč (širjenje zanesljivih informacij vs. ‚misplaced trust‘)



Individualne lastnosti, povezane z zaznavami

Kognitivni dejavniki

- Raziskave lažnih informacij dodatno pogosto izhajajo tudi iz t. i. **hipoteze nepozornosti** (van der Linden, 2022); v okolju informacijske preobremenjenosti ljudje pogosto ne namenimo dovolj pozornosti kritični presoji informacij → pričakovali bi povezave z obema dimenzijama zaznane verodostojnosti
- **Izobrazba:** formalna izobrazba se v naših raziskavah ne povezuje z zaznavami globokih ponaredkov
- **Medijska pismenost in reflektivno mišljenje:** višja izraženost povezana z nižjim zaupanjem v manipulativno vsebino
- **„Dovzetnost za sranje“:** višja izraženost povezana z večjim zaupanjem v vsebino globokih ponaredkov
- **Znanje o globokih ponaredkih:** edini kognitivni dejavnik, ki je povezan tudi z zaznavanjem manipulirane prezentacije videa!



Zakaj so zaznave pomembne?

- Če ljudje globoke ponaredke zaznajo kot verodostojne, so **bolj nagnjeni k njihovemu deljenju na spletnih socialnih omrežjih** (širjenje lažnih informacij!)
- V eni od naših raziskav (Plohl idr., 2025b) smo ugotovili, da je bila slabša prepoznava globokih ponaredkov povezana z bolj pozitivno splošno evalvacijo videoposnetka, ta pa se je naprej povezovala z izrazitejšo namero, da bi udeleženci video delili na spletnih socialnih omrežjih
- Zaznana verodostojnost globokih ponaredkov (Plohl idr., 2024) pomemben napovednik „viralnega“ vedenja (npr. namere všečkati, deliti ali priporočiti videoposnetek) tudi po upoštevanju demografskih dejavnikov in individualnih razlik
- Zanimivo: ljudje vidoe širijo predvsem zato, **ker se jim vsebina zdi verodostojna**, ne zato, ker bi bil video tehnično zelo prepričljiv



Zakaj so zaznave pomembne?

- V eksperimentalni raziskavi (Plohl idr., 2025c) smo dodatno proučevali, **ali lahko že enkratna izpostavljenost globokemu ponaredku vpliva na stališča**
- 1,124 udeležencev iz ZK, Italije in Slovenije si je ogledalo resnični video, visoko kakovosten globok ponaredek ali nizko kakovosten globok ponaredek na temo podnebnih sprememb in imigracij, stališča pa smo merili pred in po izpostavitvi videom
- Globoki ponaredki so bili podobno prepričljivi kot resnični videoposnetki, so pa bile spremembe stališč v splošnem majhne
- V rezultatih izstopa **ključna vloga zaznane verodostojnosti globokih ponaredkov**; ko so udeleženci video dojemali kot verodostojen, so se njihova stališča premaknila v smeri sporočila videa. Pri tem je ponovno prišlo do izrazitejše vloge zaznane verodostojnosti vsebine kot zaznane verodostojnosti prezentacije
- Širša implikacija: **vpliv globokih ponaredkov na javno mnenje je verjetno bolj odvisen od psihološke prepričljivosti sporočila kot od same tehnološke dovršenosti manipulacije**

Positive videos



Real video



High-quality deepfake



Low-quality deepfake

Negative videos



Real video



High-quality deepfake



Low-quality deepfake

Zaključek

- Globoki ponaredki predstavljajo **nov izziv za informacijsko okolje**. Napredek umetne inteligence omogoča ustvarjanje zelo realističnih avdiovizualnih manipulacij, ki lahko otežijo razlikovanje med resničnimi in manipuliranimi vsebinami
- Ljudje globokih ponaredkov **ne prepoznavamo dobro**. Raziskave kažejo, da je povprečna uspešnost prepoznave le malo nad naključjem, svoje sposobnosti pa pogosto precenjujemo
- **Zaznave globokih ponaredkov so kompleksne** (vidiki vsebine in prezentacije) in močno psihološko pogojene. Pri presojanju ne igrajo vloge le lastnosti videa, temveč tudi **individualne razlike**, kot so politična prepričanja, medijska pismenost in predhodno znanje
- Zaznana verodostojnost **lahko vpliva na stališča in vedenje**. Ko posamezniki globoki ponaredek dojemajo kot verodostojen, lahko to vpliva na njihova stališča ter poveča verjetnost vedenj, kot je deljenje videoposnetka na spletnih socialnih omrežjih
- **Razumevanje načinov, kako ljudje zaznavajo in interpretirajo takšne vsebine, je ključno za razvoj učinkovitih pristopov za omejevanje širjenja manipulativnih informacij**

Zaključek

- Plohl idr. (2025a): statistično pomemben učinek **kratke izobraževalne infografike** na uspešnost prepoznavanja globokih ponaredkov (74,4 % uspešnost vs. 63,8 % uspešnost) in nekatere vidike metakognitivnih sodb (večja skromnost pri osebah, ki so bile izpostavljene infografiki)
- Na prepoznavo in potencialne izide, ki sledijo (ne)prepoznavi globokih ponaredkov je možno vplivati → potrebne so bolj celovite intervencije!
- Hvala za pozornost; vprašanja?
- Kontakt: nejc.plohl1@um.si, izidor.mlakar@um.si

