2024 – 2025 oooZNANOST!
EVROPSKA NOČ RAZISKOVALCEV

Univerza v Mariboru

TRANSFORM 4EUROPE

FACTUMEVENT

Financira Evropska unija

»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 [illegible] odgovorna zanje.«

Resnica v Digitalni Dobi: Zgodbe iz projekta SOLARIS

Glavni naslov

Podnaslov

Od groženj do priložnosti – kako lahko umetna inteligenca krepi demokracijo

Glavni naslov

2024 – 2025 ZNANOST! EVROPSKA NOČ RAZISKOVALCEV

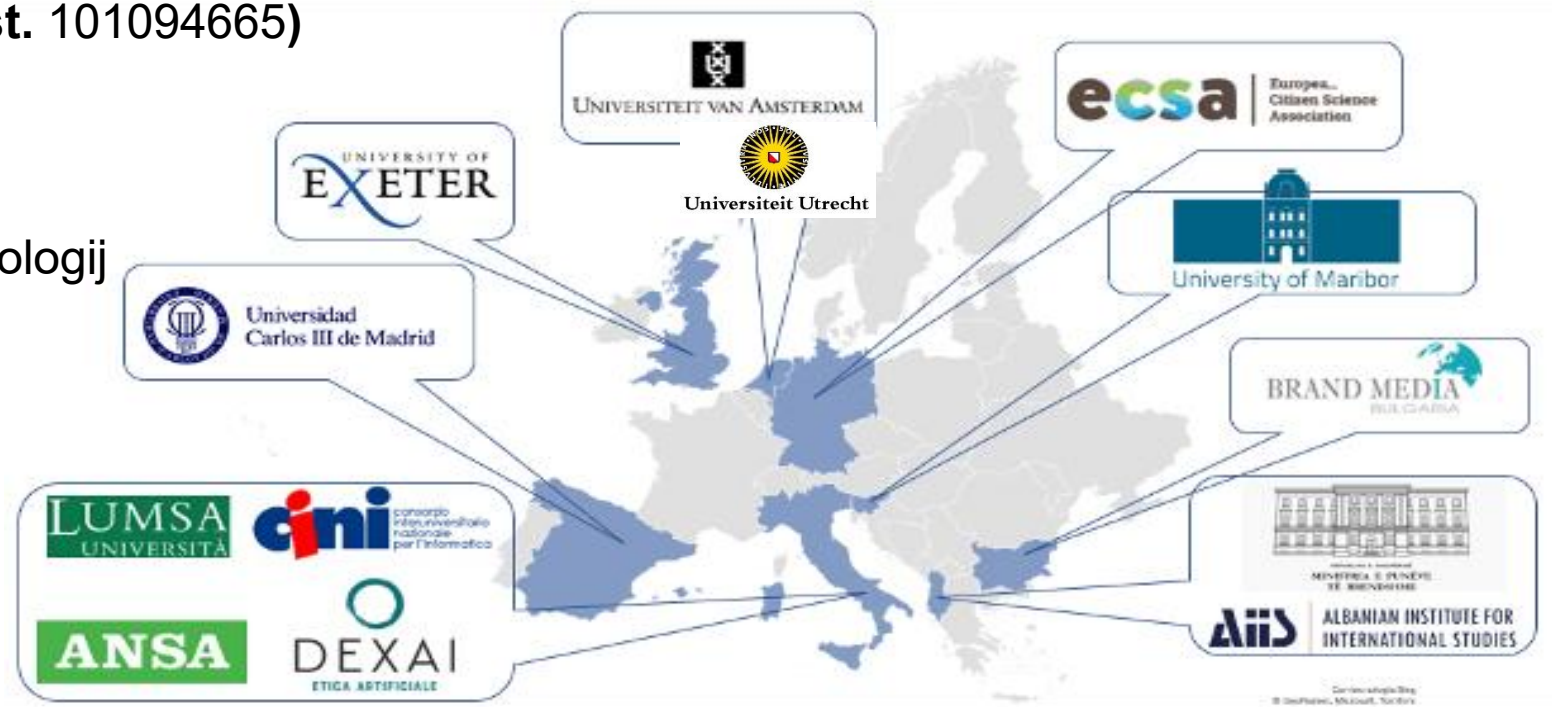
Kaj je projekt SOLARIS?

SOLARIS ne raziskuje samo groženj – raziskuje tudi, kako lahko umetna inteligenca koristi družbi

Projekt EU Horizon EU (pogodba št. 101094665)

Trajanje: 2023-2026

Namen: Raziskovanje deepfake tehnologij in njihovega vpliva na demokracijo



Dvojni pristop:

- Analiza političnih tveganj in razvoj strategij za preprečevanje dezinformacij
- Raziskovanje pozitivnih možnosti umetne inteligence za krepitev demokratične vključenosti državljanov

Partnerstvo: Mednarodna raziskovalna mreža po Evropi (Univerza v Utrechtu, Univerza v Amsterdamu, Univerza v Mariboru, ANSA, ECSA, in drugi)

Univerza v Mariboru
TRANSFORM 4EUROPE
FACTUMEVENT
Financira Evropska unija

Glavni naslov

2024 – 2025 ZNANOST! EVROPSKA NOČ RAZISKOVALCEV

Tri zgodbe, trije pristopi, ena resnica

UC1: Psihologija zaupanja

- Kako ljudje zaznavajo deepfake posnetke?
- Kateri dejavniki vplivajo na zaznavanje verodostojnosti?
- Vključuje: eksperimente, vprašalnike, analize zaupanja v sintetične vsebine
- Teme: klimatske spremembe, migracije

UC2: Novinarji na prvi liniji

- Kako novinarji zaznavajo in preverjajo deepfake vsebine?
- Simulacija "breaking news" situacij z deepfake posnetki
- Analiza uredniških protokolov in strategij preverjanja
- Cilj: Izboljšanje medijske pismenosti

UC3: Državljani kot soustvarjalci

- Sodelovanje državljanov pri ustvarjanju "dobrih" deepfake vsebin
- Delavnice s fokusnimi temami (SDG 3, 5, 13)
- Vključitev javnosti v oblikovanje etične uporabe AI
- Od konceptov do konkretnih vsebin

Univerza v Mariboru
TRANSFORM 4EUROPE
FACTUMEVENT
Financira Evropska unija

Glavni naslov

2024 – 2025 ZNANOST! EVROPSKA NOČ RAZISKOVALCEV

UC1 - Zgodba o zaupanju: Ko naše oči ne zadostujejo

"Predstavljajte si, da vidite posnetke znanega znanstvenika, ki govori o klimatskih spremembah. Ali lahko prepoznate, če je resnični ali ne?"

Raziskovalno vprašanje: Kako kakovost deepfake posnetkov, medijska pismenost in osebni pogledi vplivajo na našo sposobnost prepoznavanja lažnih vsebin?

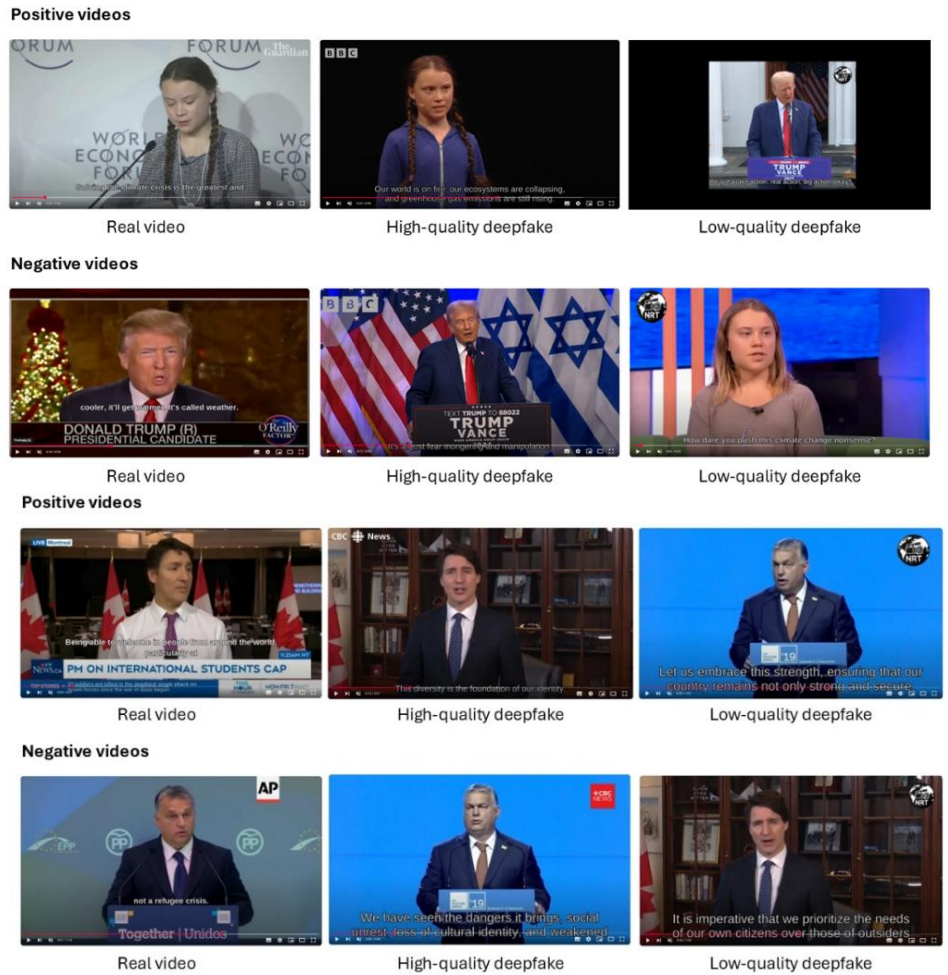
Metodologija: Udeleženci gledajo resnične in deepfake posnetke na polarizirajoče teme.

Kakovost vara: Visokokakovostni deepfake posnetki so težje prepoznavni in bolj všečni

Zaupanje povečuje tveganje: Ko ljudje deepfake zaznajo kot verodostojnega, je tveganje deljenja večje

Medijska pismenost pomaga: Osebe z višjo medijsko pismenostjo so bolj previdne pri deljenju

Enkratna izpostavljenost lahko spremeni stališča: Deepfake posnetki lahko vplivajo na mnenja o klimatskih spremembah in migracijah



Univerza v Mariboru · TRANSFORM 4EUROPE · FACTUMEVENT · Financira Evropska unija

Glavni naslov

Ko osebne lastnosti določijo, ali verjamemo svojim očem

Zaupanje (*Perceived Trustworthiness*) kot ključni moderator

Za oblikovanje naših stališč je pomembnejše to, kako verodostojno ocenjujemo vsebino 'deepfake' posnetkov, kot pa njihova tehnična kakovost.

Primer podnebnih sprememb: Ko so udeleženci vsebino ocenili kot verodostojno, so pozitivni 'deepfake' posnetki zmanjšali zanikanje podnebnih sprememb. Ko pa se je vsebina zdela neverodostojna, je isti pozitivni posnetek zanikanje povečal.

Politična usmeritev NE deluje kot moderator

Nasprotno pričakovanjem, politična ideologija **ni vplivala** na to, kako deepfake-i oblikujejo stališča

To pomeni: **Kratka izpostavljenost deepfake-u lahko vpliva na široko javnost**, ne le na tiste z določenimi političnimi prepričanji.

"Deepfake-i manipulirajo ne le z vizualnimi informacijami, temveč izkoriščajo naše psihološke ranljivosti – starost, politična prepričanja, medijska nepismenost in predvsem nagnjenost verjeti sporočilom, ki se 'ujemajo' z našimi pričakovanji, ne glede na tehnično kakovost videa."

Univerza v Mariboru
TRANSFORM 4EUROPE
FACTUMEVENT
Financira Evropska unija

Kdo je bolj ranljiv?

Demografski dejavniki:

- **Starost** (+ tveganje): Starejši posamezniki ocenjujejo deepfake-e kot bolj zaupanja vredne.
- **Uporaba družbenih medijev** (+ tveganje): Pogostejša uporaba medijev za novice povezana z večjim zaupanjem v *vsebino*, ne pa v *prezentacijo*.

Motivacijski dejavniki:

- **Konservativizem** (+ tveganje): Povezan z večjim zaupanjem v vsebino, ne vpliva na zaznavo prezentacije.
- **Zaupanje v medije** (+ tveganje): Dvostranski meč – olajša komunikacijo, a poveča ranljivost za zavajanje (*misplaced trust*).

Kognitivni dejavniki:

- **Medijska pismenost** (– zaščita): Višja pismenost zmanjšuje ranljivost.
- **Reflektivnost** (– zaščita): Sposobnost zavirati intuicijo in kritično razmišljati zmanjšuje tveganje.
- **„Bullshit“ receptivity** (+ tveganje): Nagnjenost, da pseudo-profundnim izjavam pripišemo pomen, je **najmočnejši dejavnik** ranljivosti.
- **Specifično znanje o deepfake-ih** (– zaščita): Edini kognitivni dejavnik, povezan tako z vsebino kot prezentacijo.

Glavni naslov

Od zaznave do delovanja

Pot od neprepoznavanja do širjenja

Če deepfake-a **ne prepoznamo**, ga bolj všečkamo → **povečana namera za deljenje** (indirektni učinek).

- **Moderatorji te poti:**
  - **Pozitiven odnos do osebe v videu**: Močnejši padec všečkanja ob zaznavi manipulacije → večje zmanjšanje namere za deljenje.
  - **Višja medijska pismenost**: Okrepljena negativna povezava med zaznavo in všečkanjem (posebej za podnebne teme).

Sprememba stališč

- Ena sama kratka izpostavljenost deepfake-u **lahko spremeni stališča** o polariziranih temah (podnebje, migracije), če je vsebina zaznana kot zaupanja vredna.

- **Tehnična kvaliteta videa** (visoka vs. nizka) **ni neodvisno vplivala** na prepričljivost – pomembnejša je **vsebinska verodostojnost**.

Glavni naslov

Resnični primeri, resnične posledice

Joe Biden - Lažna najava nacionalne mobilizacije

Primer:

- Deepfake video prikazuje Bidna, ki poziva k mobilizaciji za vojno v Ukrajini
- Prvotno označen kot AI, nato ponovno deljen kot "resnična novica"
- Zbral preko 8 milijonov ogledov, kljub očitnemu opozorilu

Lekcija:

- Kontekst in besedilo lahko spremenita pomen vsebine
- Strah (vojna, mobilizacija) poveča viralnost
- Preveri uradne vire (Bela hiša ni objavila tega videa)

Nekima Levy-Armstrong - Spremenjena fotografija Bele hiše

Primer:

- Bela hiša je objavila spremenjeno fotografijo aktivistke z AI-dodanimi solzami
- Manipulacija je spremenila sporočilo iz protesta v čustveno prizadetost
- Celo uradni vladni računi lahko širijo manipulirane vsebine

Lekcija:

- Preverjaj originalne vire, tudi pri "uradnih" virih
- Čustvene manipulacije (solze) močno vplivajo na percepcijo
- Ne zaupaj slepo, ker je vir "avtoriteta"

Glavni naslov

2024 – 2025 ZNANOST! EVROPSKA NOČ RAZISKOVALCEV

Resnični primeri, resnične posledice

Donald Trump - Lažna podpora Taylor Swift

Primer:

- Trump objavi AI-generirane slike "Swifties for Trump" kampanje
- Taylor Swift ni podprla Trumpa (ga je celo javno kritizirala)
- Lažne fotografije, naslovi člankov in podporniki

Lekcija:

- Zvezdniki povečajo verodostojnost lažnih novic
- AI lahko ustvari celotne scenarije podpore
- Preveri dejanske izjave znanih osebnosti

Vladimir Putin - Pogovor z AI dvojnikom

Primer:

- Ruska TV prikaže Putina v "pogovoru" z njegovim deepfake dvojnikom
- Namen: izobraževanje o deepfake-ih + propaganda proti zahodnim medijem
- Državno sponzorirana demonstracija tehnologije

Lekcija:

- Deepfake-i lahko služijo propagandnim namenom
- Paradoks: uporaba dezinformacije za opozarjanje na dezinformacije
- Razmisli: Kdo je ustvaril vsebino in zakaj?

UC2 - Novinarji med nami: Zgodba uredništva ANSA

Simulacija v Rimu (maj 2025)

- Lokacija: Italijanska tiskovna agencija ANSA
- Metoda: Tri deepfake scenariji predani novinarjem kot "breaking news"
- Novinarji so bili obveščeni, da so posnetki umetni, a niso vedeli, kateri

Tri deepfake scenariji:

- **"Scandal v Vatikanu"** – Kardinal uporablja telefon med Konklavom
- **"Jedrska nesreča"** – Eksplozija v jedrski elektrarni
- **"Olaf Scholz anti-imigracijski govor"** – Nemški kancler v kontroverznem govoru

"Predstavljajte si, da kot urednik prejmete ekskluzivni posnetek jederske nesreče. Imate 15 minut za odločitev. Kaj storite?"

Kako novinarji aplicirajo interne protokole?
Katere metode preverjanja uporabljajo?
Koliko časa potrebujejo za odločitev?
Kako poteka eskalacija do višjih urednikov?

Kaj so nas naučili novinarji: Strategije in izzivi

Uporabljene metode preverjanja:

- Preverjanje metapodatkov posnetkov
- Povratno iskanje slik (reverse image search)
- Kontaktiranje neposrednih virov
- Preverjanje konsistentnosti zvokov in slik
- Analiza detajlov (mimika, osvetlitev, sence)
- Uporaba strokovnjakov in zunanjih orodij

Ugotovljeni izzivi:

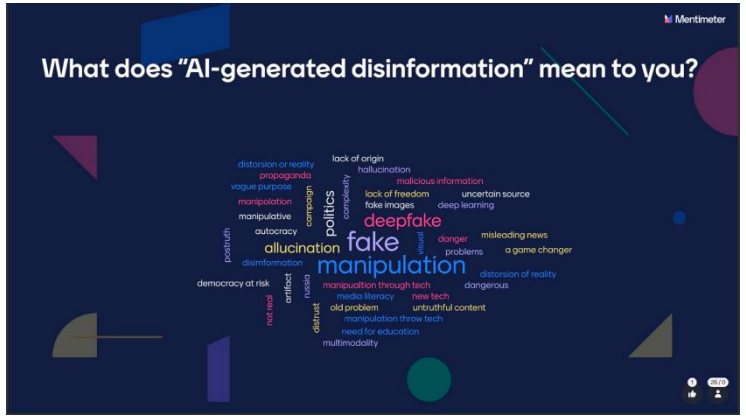
- Časovni pritisk "breaking news" situacij
- Neustreznost obstoječih protokolov za AI-generirano vsebino
- Pomanjkanje dostopnih tehnoloških orodij
- Nezadostna usposobljenost novinarjev
- Pravne in etične nejasnosti

Ključne priporočila:

1.Nujnost AI-pismenostnih programov za novinarje
2.Razvoj hitrejših in dostopnejših orodij za preverjanje
3.Jasnejši pravni okviri in odgovornost
4.Sodelovanje med uredništvi in tehnološkimi podjetji

"Eden od novinarjev je dejal: 'Najtežje je vedeti, ali smo dovolj hitri za novico, a dovolj previdni za resnico.'"

Tehnološka podpora: Kaj novinarjem dejansko pomaga?



Kaj dejansko deluje:

✓ Kombinacija več metod preverjanja
✓ Človeški presoja + tehnološka podpora
✓ Sodelovanje s forenzičnimi strokovnjaki
✓ Blockchain preverjanje avtentičnosti

Kaj NE deluje:

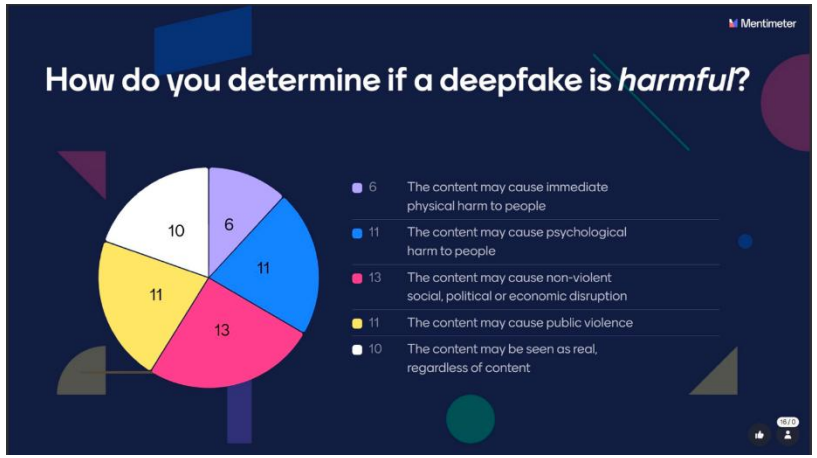
✗ Zanašanje samo na eno orodje
✗ Popolna avtomatizacija brez človeške presoje
✗ Počasna orodja v hitrem novičarskem ciklu

Predstavljena orodja med projektom:

Analiza čustev: Orodja, ki zaznavajo nenormalne čustvene vzorce v videoposnetkih

Analiza metapodatkov: Avtomatizirana preverjanja časovnih žigov in izvora

AI detektorji: Algoritmi za prepoznavanje AI-generiranih vsebin



Glavni naslov

2024 – 2025 ZNANOST! EVROPSKA NOČ RAZISKOVALCEV

UC3 - Državljani kot soustvarjalci

Raziskovalno vprašanje: Ali lahko deepfake tehnologija služi pozitivnim družbenim ciljem?

Metodologija: Tri online delavnice (april-junij 2025), 44 udeležencev iz vse Evrope, Nobeden ni AI strokovnjak

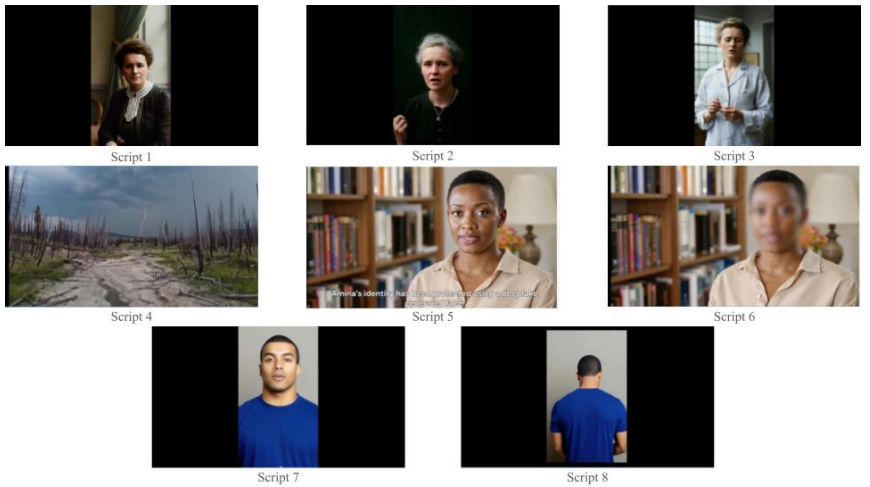
Teme vsebin (Cilji trajnostnega razvoja – SDGs):

SDG 3: Zdravje in dobro počutje (duševno zdravje – Casey)

SDG 5: Enakost spolov (ženske v znanosti – Marie Curie)

SDG 13: Podnebne spremembe (Amina, žrtev dezertifikacije)

8 eksperimentalnih videoposnetkov:
različni formati (zgodovinske osebe, izmišljeni liki, AI avatarji), različni pristopi (monolog, dialog, zabrisane obraze, deepfake obrazi), različne nastavitve (realistične, abstraktne)



Univerza v Mariboru

TRANSFORM 4EUROPE

FACTUMEVENT

Financira Evropska unija

Čustvene in kognitivne reakcije: Kdaj UI deluje?

POZITIVNE MOŽNOSTI, ki so jih prepoznali udeleženci

1. **Izobraževanje o temah brez posnetkov**
   - Zgodovinski dogodki
   1. Prihodnje scenarije
2. **Duševno zdravje in socialna vključenost**
   1. Interaktivni avatarji proti osamljenosti (starejši)
   2. Virtualni vodniki (muzeji, zdravstvene ustanove)
3. **Ozaveščanje in dostopnost**
   1. Prelom jezikovnih in kulturnih ovir
   2. Predstavitev relacijskh vzorov (lokalni akcenti, geste)
4. **Občutljive teme**
   1. Spolna vzgoja, duševno zdravje
   2. Uporaba "varnejših" likov za otroke/mladino

IZZIVI, ki so jih izpostavili:

- **Avtentičnost:** Trenutna kvaliteta glasu in mimike ustvarja čustveno distanco
- **Manipulacija:** Strah pred zlorabo, če gledalci ne vedo, da je AI-generirano
- **Zaupanje:** Potreba po transparentnosti – jasno označiti AI vsebine
- **Odgovornost:** Kdo je odgovoren za vsebino? (primerjava z dokumentarci)

Orodja UI SAMA PO SEBI niso problematična, pomembno je, KAKO in ZAKAJ jih uporabljamo.

Kdaj je deepfake "dober"? Semiotična analiza

TEKSTUALNA TAKSONOMIJA (kako je vsebina strukturirana)

1. **Diskurzivna oblika:**
   - **Evokativna:** Zunanja perspektiva, informativna (Marie Curie o zgodovini)
   - **Pričevanjska:** Osebna izkušnja, intimna (Casey o tesnobnih motnjah)
2. **Identitetna funkcija:**
   - **Pasivna:** Zaščita identitete (zabrisani obrazi – Amina, Casey)
   - **Aktivna:** Afirmacija identitete (zgodovinske osebe – Marie Curie)
3. **Destinacija:**
   - **Javna:** Široka publika, izobraževanje (muzejski posnetki)
   - **Specifična:** Ciljne skupine, personalizirano (terapevtski konteksti)

INTERPRETACIJSKA TAKSONOMIJA (kako ljudje razumejo vsebino):

1. **Plastična interpretacija:** Vizualni/avdio detajli (sinhronizacija ustnic)
2. **Diskurzivna interpretacija:** Koherentnost zgodbe in vrednot
3. **Etično-kognitivna interpretacija:** Primernost uporabe v kontekstu
4. **Pasyonalna interpretacija:** Čustveno ujemanje med obliko in vsebino
5. **Metareflektivna interpretacija:** Kritična refleksija o tehnologiji sami

Verodostojnost in učinkovitost sintetičnih vsebin ne temelji samo na tehnični realističnosti, ampak na kompleksnem usklajevanju oblike, identitete, destinacije, občutljivosti in etičnega zavedanja.

Glavni naslov

Od groženj k priložnostim – kaj lahko AI naredi za demokracijo?

Večja angažiranost in akcijskost

- Čustveno bogate AI vsebine lahko spodbudijo k akciji
- Posebej pri družbenih in okoljskih temah

Učinkovitost, stroški in skalabilnost

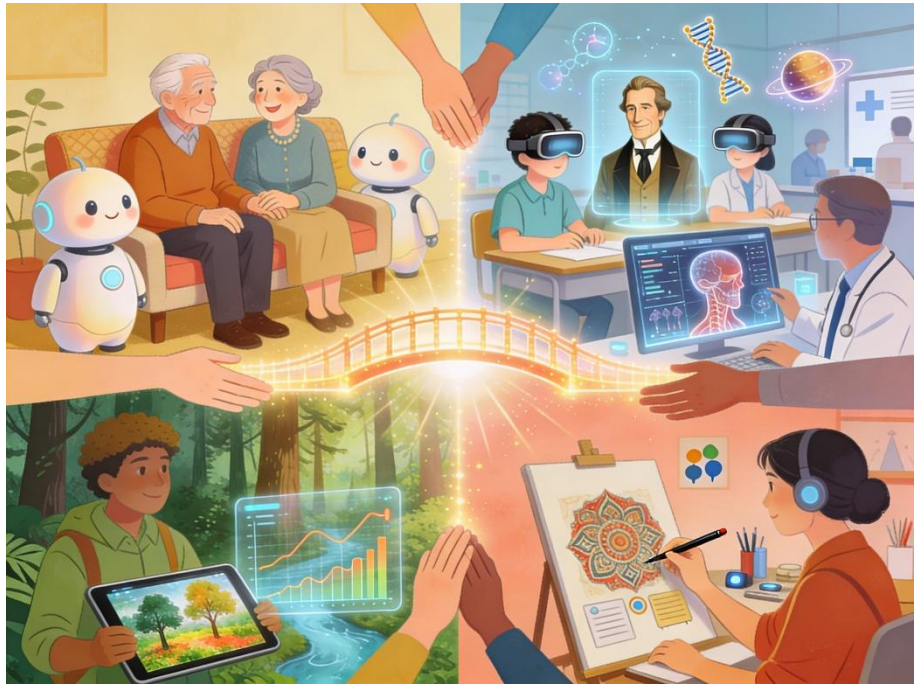
- Hitrejša in cenejša produkcija
- Več verzij sporočila za različne ciljne skupine
- Prilagodljivost za različne kontekste

Prelom jezikovnih in kulturnih ovir

- Lokalni akcenti in kulturno primerni znaki
- Večja relacija in vključenost

Prilagodljivost različnim potrebam

- Prilagajanje za različne starostne skupine
- Različne ravni znanja
- Personalizirani izobraževalni pristopi



POMEMBNO: vsebine UI morajo biti:

✓ Čustveno izrazne (naravni glasovi, toni, geste)

✓ Kulturno in jezikovno vključujoče

✓ Podpora (NE NADOMESTITEV) resnični človeški interakciji

Kaj smo se naučili v projektu SOLARIS

Deepfake niso samo grožnja:
Lahko služijo izobraževanju, ozaveščanju, vključevanju
Odvisno od transparentnosti, konteksta, etičnih načel

Zaupanje je ključno:
Zaznana verodostojnost poveča tveganje
Potreba po transparentnem označevanju AI vsebin
Odgovornost ustvarjalcev in distributerjev

Tudi novinarji potrebujejo podporo:
AI-pismenostni programi nujni
Boljša orodja, jasnejši protokoli
Sodelovanje med uredništvi

Tehnologija sama ni rešitev:
Človeška presoja ostaja ključna
Kombinacija orodij + izkušenj + etike
AI kot podpora, ne nadomestilo

Državljani morajo biti vključeni:
Participatorna demokracija v dobi AI
Soustvarjanje etičnih standardov
Empowerment, ne panika

Glavni naslov

Morda v razmislek?

Kako bi prepoznali deepfake posnetek?

Kdaj ste nazadnje delili vsebino brez preverjanja?

Komu zaupate v digitalnem prostoru in zakaj?

Kakšno vlogo imajo šole, mediji, civilna družba?

Kako uravnotežiti inovacije in zaščito pred zlorabami?

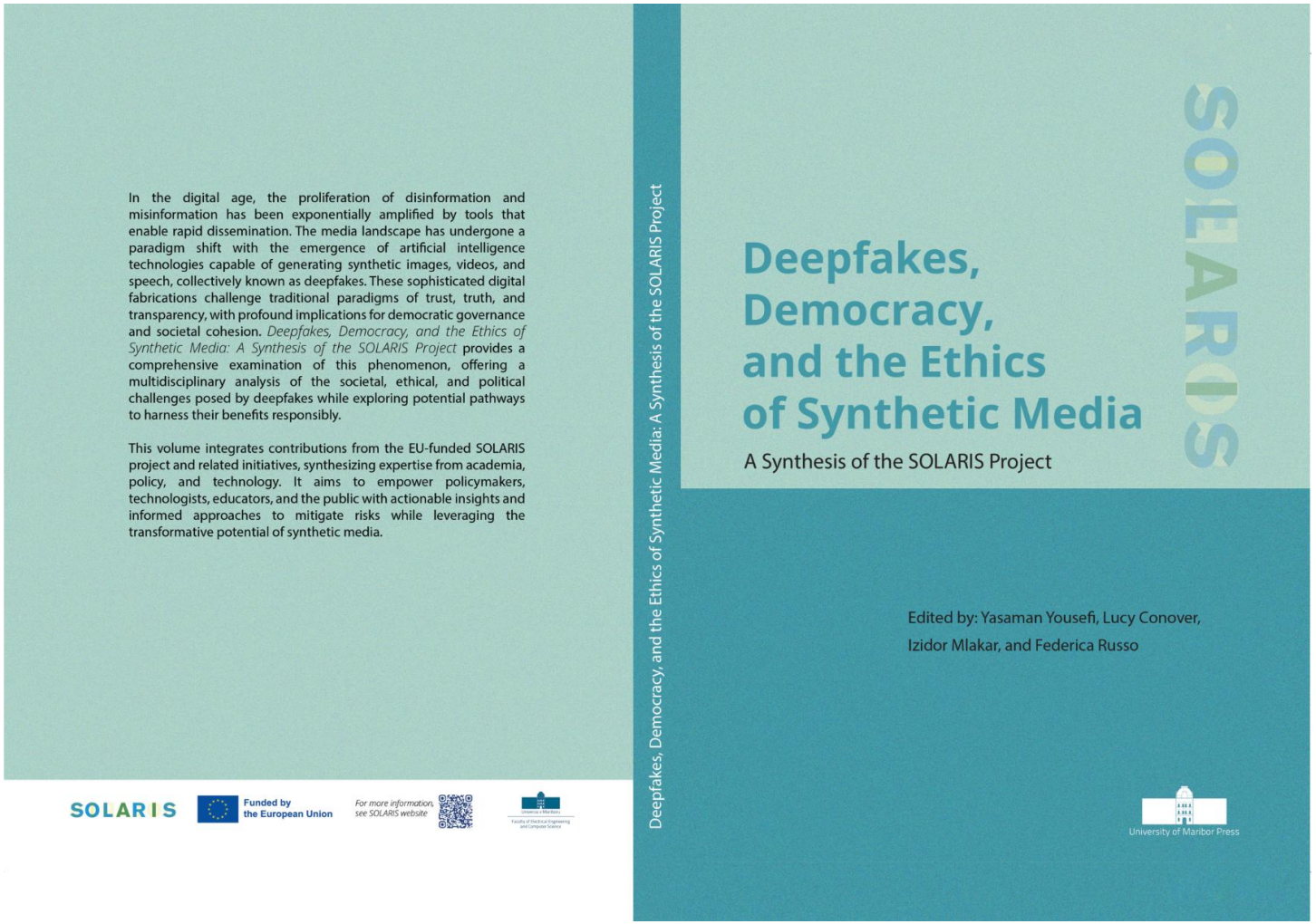
Kakšno demokracijo želimo v digitalni dobi?

Glavni naslov

2024 – 2025 ZNANOST! EVROPSKA NOČ RAZISKOVALCEV

Hvala Za Pozornost

dr. Izidor Mlakar
izidor.mlakar@um.si
dr. Nejc Plohl
nejc.plohl1@um.si



Univerza v Mariboru · TRANSFORM 4EUROPE · FACTUMEVENT · Financira Evropska unija

https://press.um.si/index.php/ump/catalog/book/1089