Tēma “Dziļviltojumi”

Kas ir dziļviltojumi un kā tie rodas?

Tēzaurs definīcija – *dziļviltojums ir ar mākslīgā intelekta (turpmāk – MI) vai specializētas programmatūras palīdzību veidots vai mainīts attēls, video vai audioieraksts, kurā attēlotie notikumi vai personu rīcība neatbilst patiesībai; bieži tiek izmantots dezinformācijai vai nomelnošanai.*

Svarīgi saprast, ka **ne katrs ar MI radīts attēls vai video ir dziļviltojums**. Piemēram, ja cilvēks izveido izdomātu ainavu un skaidri norāda, ka tā radīta ar MI, tas pats par sevi nav dziļviltojums. Problēma rodas tad, kad mākslīgi radītais saturs tiek pasniegts kā īsts vai tiek izmantots, lai radītu nepatiesu priekšstatu par notikumu vai personu.

Piemērs: Video var radīt iespaidu, ka kāds politiķis ir pateicis konkrētu frāzi, lai gan viņš to nekad nav teicis. Balss var būt ģenerēta ar MI, bet sejas kustības digitāli pārveidotas, tādejādi rezultāts var izskatīties pilnīgi ticams.

Dziļviltojumu pamatā ir ģeneratīvā mākslīgā intelekta tehnoloģijas, kas var analizēt lielu daudzumu materiālu un pēc tiem radīt jaunus attēlus, skaņas vai video.

Vienkāršoti procesu var atspoguļot šādi:

1. Dati → 2. MI apgūst personas izskatu/balsi → 3. tiek radīts vai pārveidots saturs → 4. cilvēks to publicē un/vai izplata.

Lai atdarinātu cilvēka seju vai balsi, sistēma var izmantot jau publiski pieejamus fotoattēlus, video vai audioierakstus. Jo kvalitatīvāki un daudzveidīgāki dati ir pieejami, jo pārliecinošāka un ticamāka var kļūt mākslīgā imitācija.

National Institute of Standards and Technology (*lat. – Nacionālais standartu un tehnoloģiju institūts (NIST)*) norāda, ka mākslīgā jeb sintētiskā satura radīšana ir kļuvusi arvien pieejamāka, bet vienlaikus arī tā noteikšana – sarežģītāka. Detektoru darbība var būt atkarīga no konkrētā ģenerētāja, izmantotā satura un tā, kā saturs vēlāk ticis pārveidots. Turklāt starp satura radītājiem un detektoru izstrādātājiem notiek sava veida “kaķa un peles spēle”: uzlabojoties noteikšanas metodēm, uzlabojas arī ģenerēšanas paņēmieni. (https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-4.pdf?utm_source=chatgpt.com)

Kādi ir galvenie dziļviltojumu veidi?

Sejas un attēli

Personas seja var tikt ievietota cita cilvēka ķermenī vai digitāli pārveidota. Var arī izveidot pilnībā sintētisku seju, kas pieder personai, kura dzīvē nemaz nepastāv.

Video

Video var likt cilvēkam izskatīties tā, it kā viņš būtu teicis vai darījis ko tādu, kas patiesībā nav noticis. Šādās manipulācijās var tikt mainīta seja, lūpu kustības, ķermeņa kustības vai visa situācijas vide.

Balss un audio

MI var izmantot, lai imitētu cilvēka balsi. Tas nozīmē, ka var tikt radīts ieraksts, kurā cilvēks it kā saka kaut ko tādu, ko viņš realitātē nav teicis.

Multimodāls saturs

Mūsdienās iespējams kombinēt vairākus paņēmienus vienlaikus, piemēram, mainīt seju video, ģenerēt balsi un pievienot mākslīgi radītu tekstu. Tādā gadījumā dziļviltojuma problēma vairs nav tikai "neīsts video"; tas var būt vesels mākslīgi ģenerēta stāsta komplekts.

Kāpēc dziļviltojumi ir medijpratības jautājums?

Medijpratība nav tikai prasme saprast, vai saturs ir īsts vai viltots. Tā ietver spēju izvērtēt informācijas avotu, kontekstu, nolūku un pierādījumus.

Dziļviltojuma gadījumā jāuzdod ne tikai jautājums:

"Vai šis video ir īsts?"

bet arī:

"Kas to publicēja?"
"Kāpēc tas tika publicēts?"
"Vai citi uzticami avoti apstiprina šo informāciju?"
"Vai ir iespējams pārbaudīt satura izcelsmi?"

Tas ir būtiski, jo arī **īsts attēls vai video var tikt izmantots maldinošā kontekstā**. Pētījumos par vizuālo dezinformāciju uzsvērts, ka maldināšanai ne vienmēr nepieciešama attēla viltošana – patiess attēls var tikt parādīts saistībā ar citu notikumu vai izmantots, lai radītu nepatiesu iespaidu. ScienceDirect

Tātad medijpratība nozīmē pārbaudīt ne tikai **satura autentiskumu**, bet arī **kontekstu un apgalvojumu**. (https://www.sciencedirect.com/science/article/pii/S2352250X23002233?utm_source=chatgpt.com)

Kāpēc reizēm var būt grūtības atpazīt dziļviltojumu?

Pastāv mīts, ka dziļviltojumu var vienkārši "ieraudzīt", pievēršot uzmanību sejai, acīm vai lūpām, bet tas nav tik vienkārši un droši.

2024. gadā publicētā sistemātiskā pārskatā un Meta analīzē, kurā apkopoti 56 pētījumi, secināts, ka cilvēku spēja atšķirt sintētisku saturu no īsta daudzos gadījumos ir tuva nejaušas minēšanas līmenim. Tajā pašā laikā pētnieki konstatējuši, ka noteiktas stratēģijas var palīdzēt uzlabot rezultātus. (ScienceDirect)
Vēl jaunākā, 2026. gada pētījumā pētnieki salīdzināja cilvēku un mašīnu spēju noteikt dziļviltojuma attēlus un video. Pētījumā cilvēkiem bija grūti atšķirt statiskus dziļviltojumu attēlus, kamēr rezultāti dinamiskos video bija atšķirīgi. Pētnieki arī konstatēja saistību starp labāku video atpazīšanu, analītiskāku domāšanu un labākām interneta prasmēm. (PubMed)

Tas nozīmē:

“Izskatās īsti” ≠ “ir īsti”.

Un arī:

“Izskatās dīvaini” ≠ “ir pierādīts, ka tas ir viltojums”.

Kādas pazīmes var radīt aizdomas?

Atsevišķos dziļviltojumos var parādīties vizuālas vai audiālas neatbilstības, piemēram, nedabiska lūpu un skaņas sinhronizācija, dīvainas kustības, neierastas sejas izteiksmes vai citas tehniskas anomālijas, taču neviena no šīm pazīmēm nav pilnīgi drošs pierādījums.

Office of Communications (lat. – *Apvienotās Karalistes Komunikāciju birojs*) savā pētījumā uzsver, ka lietotājus nevajadzētu atstāt vienus pašus dziļviltojumu identificēšanas uzdevumā. Turklāt tādi tehniskie risinājumi kā ūdenszīmes var tikt noņemti vai izmainīti pēc vienkāršas satura rediģēšanas. (www.ofcom.org.uk) Daudz labāka pieeja ir izmantot vairākus pārbaudes līmeņus.

Jautājumi, kas palīdz pārbaudīt aizdomīgu saturu

1. KAS to publicēja?

Pārbaudi profilu, mājaslapu vai personu, kura pirmā izplatīja materiālu!

Vai tas ir oficiāls avots? Vai profils ir pazīstams? Vai avotam ir iepriekšēja reputācija?

2. NO KURIENES nāk saturs?

Mēģini atrast oriģinālo ierakstu, nevis tikai pārpublicētu video vai ekrānuzņēmumu! Jo vairāk reižu saturs ir pārkopēts, apgriezts un atkārtoti augšupielādēts, jo grūtāk var būt noteikt tā sākotnējo izcelsmi.

3. KO saka citi avoti?

Īpaši nozīmīgi apgalvojumi jāmeklē citos neatkarīgos avotos.

Piemēram, ja video apgalvo, ka kāds politiķis pēkšņi atkāpies no amata, jāmeklē informācija arī uzticamos ziņu medijos vai attiecīgās institūcijas oficiālajos kanālos.

4. KĀDS ir konteksts?

Vai video tiešām attiecas uz norādīto notikumu?

Kad tas uzņemts? Kur? Kas noticis pirms un pēc fragmenta?

5. VAI var pārbaudīt satura izcelsmi?

Daļa tehnoloģiju ļauj saglabāt informāciju par digitālā satura izcelsmi un rediģēšanas vēsturi. Viens no piemēriem ir **C2PA / Content Credentials** pieeja, kurā var tikt saglabāta pārbaudāma informācija par to, kā saturs radīts un kādas izmaiņas tajā veiktas. (C2PA)

Svarīgi, ka arī šāds rīks pats par sevi nenozīmē, ka viss saturs ar vai bez šādām pazīmēm automātiski ir "īsts" vai "neīsts". C2PA galvenais mērķis ir nodrošināt **izcelsmes un izmaiņu informāciju**, nevis piešķirt saturam vispārēju "patiesības sertifikātu". (https://c2pa.org/principles/?utm_source=chatgpt.com)

Kāpēc nevar paļauties tikai uz "MI detektoru"?

Internetā pieejami rīki, kas sola noteikt, vai attēls, video vai audio ir radīts ar MI. Šādi rīki var būt noderīgi, taču to rezultāts jāuztver kā viens informācijas avots, nevis galīgais spriedums.

NIST norāda, ka automātiskās noteikšanas metodes saskaras ar būtiskām problēmām – detektoru darbība var pasliktināties, kad saturs atšķiras no tā, uz kā detektors apmācīts, un ģenerēšanas tehnoloģijai attīstoties, noteikšanas metodes kļūst arvien grūtāk uzturamas. (nvlpubs.nist.gov)

Laba medijpratības prakse ir sekojoša:

detektors + avota pārbaude + konteksts + neatkarīgi avoti.

Nevis:

"Detektors pateica 73 %, tātad tas noteikti ir viltojums."

Kā dziļviltojumi var tikt izmantoti kaitējuma radīšanai?

Dziļviltojumiem ir arī pilnīgi leģitīmi izmantošanas veidi, piemēram, kino, radošajā industrijā, izglītībā un citās jomās, ja auditorija netiek maldināta. Tas nenoliedz to, ka tehnoloģiju var izmantot arī kaitnieciskiem nolūkiem.

Ofcom min vairākus riskus, tostarp krāpšanu, intīma rakstura viltotu materiālu izplatīšanu un dezinformāciju par politiķiem. (www.ofcom.org.uk)

Krāpšana

Personas balss vai tēls var tikt izmantots, lai radītu iespaidu, ka piezvana priekšnieks, radinieks vai cita uzticama un/vai zināma persona.

Reputācijas graušana

Personu var attēlot situācijā, kurā tā nekad nav bijusi.

Dezinformācija

Viltots video vai audio var tikt izmantots, lai ietekmētu sabiedrības priekšstatus par politiskiem, sociāliem vai citiem nozīmīgiem jautājumiem.

"Pierādījumu" viltošana

Dziļviltojums var radīt nepatiesu priekšstatu, ka ir pieejams audio vai video "pierādījums", lai gan pats materiāls ir viltojums.

Dziļviltojumi un politika

Politikā dziļviltojumi var būt īpaši ietekmīgi, jo cilvēki nereti reaģē ātri, īpaši vēlēšanu laikā vai krīzes situācijās.

UNESCO (*lat. – Apvienoto Nāciju Izglītības, zinātnes un kultūras organizācija*) savā medijpratības materiālā apraksta 2023. gada Slovākijas vēlēšanu gadījumu. Divas dienas pirms vēlēšanām sociālajos tīklos tika izplatīts audioieraksts, kurā it kā politiķis Michal Šimečka un žurnāliste Monika Tódová apsprieda vēlēšanu manipulēšanu. Tik noskaidrots, ka ieraksts bija ar MI radīts dziļviltojums. UNESCO šo piemēru izmanto, lai parādītu, cik nozīmīga ir kritiska informācijas izvērtēšana tieši politiski jutīgos brīžos. (UNESCO)

Šeit ļoti svarīgs ir princips:

Jo nozīmīgāks apgalvojums, jo augstāks pierādījumu standarts.

Dziļviltojuma "stiprā puse" nav tikai tehnoloģija

Dziļviltojumi var būt pārliecinoši ne tikai tāpēc, ka tie izskatās labi – svarīga ir arī mūsu pašu kritiskā domāšana.

Cilvēks biežāk var noticēt saturam, kas apstiprina to, kam viņš jau tic, raisa spēcīgas emocijas vai šķiet īpaši pārsteidzošs. Pētījumi par vizuālo dezinformāciju rāda, ka attēli var palielināt uzmanību un ietekmēt to, kā cilvēki uztver un atceras informāciju.

(ScienceDirect) Tāpēc bīstams saturs bieži vien nav tas, kas ir vistehnoloģiskākais – tas var būt saturs, kas ļoti labi "strādā" jeb ietekmē cilvēku emocionāli.

Piemēram:

"Steidzami! To nedrīkst rādīt televīzijā!"

"Šis video pierāda, ka viņš meloja!"

"Nosūti tālāk, pirms to izdzēš!"

Šādi formulējumi var mudināt cilvēku dalīties ar saturu pirms tā kārtīgas pārbaudīšanas un faktu un avotu izsvēršanas.

"Slikts" dziļviltojums un "labs" dziļviltojums

Medijpratībā ir svarīgi neradīt priekšstatu, ka jebkurš MI ģenerēts saturs ir slikts. Problēma bieži vien nav pati tehnoloģija, bet gan tās izmantošanas veids un caurspīdīgums.

Piemēram, var radīt mākslīgi ģenerētu video izglītībai vai kino un skaidri norādīt, ka tas ir MI ģenerēts saturs.

Partnership on AI savā Synthetic Media Framework īpaši izceļ trīs principus: consent, disclosure un transparency – **piekrišanu, atklātu informāciju par ģenerēta satura izmantošanu un caurspīdīgumu**. (Partnership on AI - Synthetic Media)

Tas nozīmē:

MI izmantošana pati par sevi ≠ maldināšana.

Par maldinošu situācija kļūst tad, kad auditorijai tiek apzināti radīts nepatiesa autentiskuma iespaids.

Ko nosaka Eiropas Savienība?

ES Mākslīgā intelekta akta (AI Act) pārredzamības prasības sāka piemērot 2026. gada 2. augustā. Eiropas Komisija norāda, ka noteiktiem MI radīta vai manipulēta satura veidiem jābūt atpazīstamiem, tostarp dziļviltojuma saturam. Komisijas skaidrojumā teikts, ka dziļviltojumi ir jāmarķē un MI ģenerētam vai pārveidotam saturam jābūt arī mašīnlasāmi identificējamam noteiktajos gadījumos. (Digitālā stratēģija Eiropā)

Vienlaikus ir svarīgi nesarežģīt šo jautājumu skolēniem ar domu, ka "tagad jebkurš MI attēls internetā obligāti būs redzami marķēts". Prasības ir atkarīgas no konkrētā MI sistēmas un izmantošanas scenārija. Eiropas Komisija ir publicējusi atsevišķas vadlīnijas par šo pienākumu piemērošanu, ar kurām var iepazīties – Digitālā stratēģija Eiropā.

Kā rīkoties, ja redzu aizdomīgu video?

APSTĀJIES

Nedalies ar saturu uzreiz, īpaši tad, ja tas izraisa ļoti spēcīgas emocijas.

PĀRBAUDI

Noskaidro avotu, publicēšanas laiku, oriģinālo ierakstu un kontekstu.

SALĪDZINI

Meklē neatkarīgus uzticamus avotus.

MEKLĒ IZCELSMI

Pārbaudi, vai pieejama informācija par satura izcelsmi, rediģēšanu vai autentiskumu.

TIKAI TAD SECINI

Ja pierādījumu nepietiek, drošāk ir teikt:

“Es nevaru apstiprināt, ka tas ir īsts.”

nevis:

“Tas noteikti ir dziļviltojums.”

Šī atšķirība ir ļoti svarīga medijpratībā: **šaubas nav tas pats, kas pierādījums**.

Galvenās atziņas, ko vērts atcerēties

1. Dziļviltojums var izskatīties un izklausīties ļoti pārliecinoši.
2. Ar acīm vai ausīm vien ne vienmēr iespējams droši noteikt viltojumu.
3. Viena pazīme vai viena AI detektora rezultāts nav pietiekams pierādījums.
4. Jāpārbauda avots, konteksts un neatkarīgi avoti.
5. Īsts materiāls arī var tikt izmantots maldinošā kontekstā.
6. MI radīts saturs ne vienmēr ir kaitīgs – būtiska ir caurspīdīga un atbildīga tā izmantošana.
7. Jo spēcīgāku emociju saturs izraisa, jo svarīgāk ir apstāties un pārbaudīt saturu pirms dalīšanās ar to.

Praktiskie uzdevumi

1. uzdevums – īsts vai aizdomīgs?

Aplūko doto attēlu/video un atbildi:

- Kas tev tajā šķiet aizdomīgs?
- Nosauc 2 pazīmes, kas rada šaubas.
- Vai ar šīm pazīmēm pietiek, lai pierādītu, ka tas ir dziļviltojums? Kāpēc?

Mērķis: saprast, ka viena vizuāla pazīme nav pietiekams pierādījums.

2. uzdevums – avota pārbaude

Tev tiek parādīts sociālo tīklu ieraksts ar šokējošu video.

Izmanto 5 jautājumus: Kas? No kurienes? Ko saka citi avoti? Kāds ir konteksts? Vai var pārbaudīt izcelsmi?

Uzraksti 3 pārbaudes darbības, ko veiktu pirms dalīšanās ar šo ierakstu.

Mērķis: praktizēt informācijas avota un konteksta pārbaudi.

3. uzdevums – apturi emocijas!

Izlasi apgalvojumu:

“Steidzami! Šis video pierāda, ka politiķis ir izdarījis kaut ko nelikumīgu. Dalies, pirms video tiek izdzēsts!”

Atbildi:

- Kādas emocijas šis ieraksts cenšas izraisīt?
- Kāpēc šāds formulējums var mudināt cilvēku dalīties nepārbaudītā informācijā?
- Uzraksti vienu teikumu, ko tu sev pateiktu pirms dalīšanās.

Mērķis: atpazīt emocionālu manipulāciju un apstāties pirms satura izplatīšanas.

Dziļviltojumu piemēri

1. https://youtu.be/IfUladg3-SU?si=DvDQadgxECEmSeMS
2. https://youtu.be/X17yrEV5sl4?si=637pjhoq1cYUjZc8
3. https://youtu.be/cQ54GDm1eL0?si=j_206GuSEsazZzE5



5. 

6. 