

Süvavõltsingute tuvastamine

Võltsitud audio-visuaalne sisu pole mõjutustegevuses üksinda piisav, sinna ümber ehitatakse "usaldusväärne" keht, mille toel võltsinguid paremini tõe pähe müüa - kuidas seda päriselt ka tehtud on võid lugeda sellest [Euronewsi artiklist](#). Süvavõltsitud materjali sellena tuvastamine on aina keerulisem ning 100% abi ei leia ühestki rakendusest.

Kuidas tuvastada süvavõltsinguid

Kiirkontroll: Kui sisu tekitab tugevat emotsiooni või šokki → kontrolli põhjalikumalt.

Visuaalsed märgid:

- Ebaloomulik valgus/varjud
- Moonutused näos või kätes
- Kummalised silmaliigutused

Heli ja kõne:

- Huulte liikumise ja kõne desünkroon
- Ebaloomulik tempo või hingamine

Kontekstuaalsed vihjed:

- Allika usaldusväärsus (kas tegemist võib olla kloonitud või spoof-lehega?)
- Narratiiv „meedia vaikib“ või „ametlikud kanalid kustutasid“

Kuidas kontrollida:

Parim viis kontrollimiseks on vaadata, kas mõni usaldusväärne allikas on sama pilti/videot jaganud. Tihti piisab [Google pildiotsingust](#), mis otsib sarnaseid pilte või teisi allikaid samale pildile. Nii saab veenduda kas tegu on päris pildiga ning kas postitusel/uudisel on mingigi tõepõhi all. Turul on olemas ka eraldi tööriistad süvavõltsingute tuvastamiseks, aga ainult nende tulemusi ei saa valideerimisel usaldada.

Kuidas hinnata mõju

- **Mainekahju:** kas sihtmärgi isiklikku või ametialast reputatsiooni rünnatakse?
- **Institutsioonide õonestamine:** kas väited seavad kahtluse alla valimisi vms?
- **Leviku ulatus ja tempo:** esimesed 4-8h näitavad leviku ulatust.
- **Narratiivide sihtus:** kas rünnak on suunatud teatud poliitika või liitlussuhete vastu?
- **Ajakriitilisus:** vali kas „*vaikne eemaldus*“ või *avalik ümberlükkamine*, sõltuvalt olukorrast.

Mida teha kui märkad?

- Teavita asjakohaseid partnereid
- Salvesta kuvatõmmis, originaalfailid, URL-id, metaandmed
- Kasuta Meta ja Google raporteerimisvõimalusi

Avalik kommunikatsioon: reageeri selgelt ja faktipõhiselt. Väldi valeväidete kordamist pealkirjades ja visuaalides.

Platvormide deepfake-poliitika

Ilmselged süvavõltsingud markeeritakse platvormide poolt automaatselt (kui kasutaja seda teinud pole), aga enamjaolt tuleb toetuda kasutajate-poolsele raporteerimisele ja kaebuste esitamisele. Kokkkuvõtvalt on meil populaarsemate platvormide poliitika järgmised (muutuvad ajas pidevalt):

Meta (Facebook, Instagram, Threads):

- Märgistamise kohustus AI-video ja heli puhul
- Automaatne märgistus piltide korral
- Märgistada võidakse ka sellist sisu, mis võib inimesi olulistel teemadel eksitada.
- Eemaldatakse kogukonnareeglite rikkumise korral (nt valimistesse sekkumine, vägivaldale õhutamine)

Google / YouTube:

- Märgistamiskohustus realistlikult muudetud sisu puhul
- Märgistamata jätmisel lisab YouTube märgise ise või piirab kasutajat
- Võimalus esitada **privaatsuskaebus** väärkasutuse korral