Nähtamatu töö ja nähtamatu kahju: Euroopa regulatsioonid ei tsenseeri, vaid kaitsevad inimesi kahju eest

MARJU HIMMA
RiTo peatoimetaja

Platvormide kahju ei piirdu üksnes ebaseadusliku sisuga, vaid hõlmab ka nende disaini ja ärimudeleid, mis võimendavad viha ja polariseerumist. Inimeste kaitsmine selle kahju eest ei ole tsensuur, vaid demokraatlike süsteemide kohustus, leiab Berni Rakendusteaduste Ülikooli sotsiaalse disaini vanemteadur Anna Antonakis. RiTole antud intervjuus õhutab ta läbipaistvuse ja kodanikuühiskonna kaasamise vajalikkust ning näeb kahju vähendamise peamiste teguritena kriitilise digipädevuse arendamist ja moderaatorite vaimse tervise hoidmist.

MARJU HIMMA: Alustame väga laia sissejuhatusega, eriti nende jaoks, kes ei pruugi olla tuttavad kahjuliku veebisisu ja selle modereerimise probleemidega. Kas Te saate tuua näiteid, millist kahjulikku sisu on vaja modereerida ning mida modereerimine sellises kontekstis tähendab?
ANNA ANTONAKIS: Olen interdistsiplinaarne teadlane. Minu uurimistee algas 2010–2011 Tuneesia sotsiaalsete liikumiste ülestõusudest, mida hiljem hakati nimetama Araabia kevadeks. Tollal lähtusin üsna optimistlikust vaatenurgast ja empiirilisest uurimistööst, mis näitas, kuidas inimesed said sotsiaalmeediat kasutada internetitsensuurist möödahiilimiseks ja politseivägivalla paljastamiseks.

Sellest on nüüd möödas peaaegu 15 aastat – olukord on väga palju muutunud. See ei tähenda, et oleksin liikunud täielikult optimismilt pessimismile, pigem olen kogu aeg uurinud sotsiaalmeedia ja infotehnoloogiate kahepalgelisust.

Anna Antonakis meediahariduse konverentsil 2025. aasta oktoobris.

Foto: Aron Urb

Vaatlen neid sotsiaal-tehnoloogilisest vaatenurgast.

Teie küsimuse juurde tulles: mis on kahjulik sisu? Kirjanduses, platvormidel ja ka poliitikakujundajate (nt ELi digiteenuste määrus, DSA) tasandil eristatakse erinevaid kahju liike. Kahju võib olla ebaseaduslik – näiteks vägivalla eksponeerimine, seksuaalvägivald, alaealiste eksplitsiitsed kujutised. Samuti relvade või illegaalsete ainete näitamine. Sellistes olukordades on modereerimine vältimatu ja näeme, et platvormid uuendavad pidevalt oma reegleid – näiteks, kuidas käsitleda videoid, mis õpetavad relvade turvamehhanisme eemaldama.

Kuid kahjulik ei tähenda ainult ebaseaduslikku. Kahju võib olla ka diskrimineeriv või halvustav sisu – naljad või mõnitamised kindlate rühmade ja vähemuste aadressil. Kahju defineerida on keeruline. Vihakõne on juba pikalt olnud uurimisobjekt ning selle mõju inimeste igapäevaelule on märkimisväärne. Üha enam räägitakse ka sisust, mis soodustab radikaliseerumist.

Minu uurimistöö keskmes on eriti sooline vaatenurk. Näiteks meeste ja maskuliinsuse radikaliseerumine. Kuid mind ei huvita ainult sisu ise, vaid ka platvormide disain, mis juhib kasutajat samm-sammult radikaalsema sisuni, kasutades järjestamis- ja soovitusmehhanisme. Küsimus on läbipaistvuses: millised mehhanismid tegelikult töötavad?

Teie teine küsimus: millised mehhanismid on töös? Suur probleem seisneb selles, et suurte tehnoloogiaettevõtete (ingl *big tech*, nt Facebook, X jt) platvormid ei avalikusta täielikult, milliseid modereerimisviise nad kasutavad. See on probleem, sest takistab meil teadlastena mõista sisu, platvormide ja ühiskondlike protsesside, näiteks sotsiaalsete liikumiste vahelisi dünaamikaid.

On olemas automaatne sisu modereerimine, kuid see on sageli kallutatud. Sellest on üha rohkem teadustöid, kuigi andmetele on raske ligi pääseda. Kallutatus automaatses modereerimises on üks suur probleem. Teine on inimmodereerimine, nn kummitustöö või andmetöö. Need on inimesed, kes töötavad väga keerulistes tingimustes, sageli ekspluateerituna.

Selle aasta konverentsi teema on eriti ka vaimne tervis. Kui vaatame kriitiliselt meediapädevust, peame küsima, kuidas see, mida me sotsiaalmeedias näeme, on tegelikult modereeritud? Ja kelle vaimne tervis selle käigus ohvriks langeb?

Tulles tagasi DSA juurde – milliseid muutusi toob see sisu modereerimisse? Kas see paneb piisava vastutuse tehnoloogiaettevõtetele või peaksid EL ja liikmesriigid rohkem sekkuma?

See on väga huvitav küsimus. DSA on uus projekt, uus seadus, ja on põnev näha, kuidas seda erinevates riikides tõlgendatakse. Mina näen seda nii: poliitikute loodud reeglid (välised regulatsioonid) ja platvormide enda reeglid (sisemised regulatsioonid) peavad toimima koos. Küsimus on, kes neid poliitikaid päriselt kasutada saab?

Oluline on vaadata ühiskonda tervikuna ja kasvatada teadmisi meedia, modereerimise ja digitaalse keskkonna kohta. Poliitika on vaid üks osa. Ka USAs peetakse selliseid regulatsioone tsensuuriks, mistõttu on tähtis, et EL hoiaks kindlalt oma positsiooni ja selgitaks, et see ei ole tsensuur, vaid inimeste kaitsmine kahju eest. Kuid lisaks poliitikale vajame ka haridust, ja kriitilist haridust.

Platvormid on globaalsed, riigid ja EL aga piiridega. Kas üldse saab öelda, kes mille eest vastutab? Mina usun, et platvormid vastutavad oma disaini eest. Platvormide kujundus on suunatud kasumi teenimisele. Ja kahjuks tähendab see, et enim tähelepanu saab polariseeriv, viha ja ärevust tekitav sisu. See on suur probleem ja siin peavad platvormid vastutama.

Milliseid uusi väljakutseid näete lähiaastatel seoses tehisintellekti ja uute manipulatsioonivõimalustega?
AI muutub järjest paremaks piltide, videote ja heli manipuleerimises. Küsimus on, kuidas varustada ühiskonda tööriistadega, et eristada ehtsat võltsist?

See probleem ei ole tegelikult uus – juba varem oleme uurinud, kuidas kaamerad ja meedia raamivad reaalsust. Kuid nüüd on väljakutse palju keerulisem. Näeme, et eriti paremäärmuslikud parteid kasutavad AI loodud pilte emotsioonide õhutamiseks ja valeinfo levitamiseks. Varem oli võltsuudis tekstina, nüüd lisanduvad video ja pildid. See on tohutu probleem, eriti globaalsete konfliktide kontekstis, kus üks pilt võib teatud kohas tekitada rahutusi või viha.

Seetõttu peame varustama nii noori kui ka vanemaid inimesi kriitilise pilguga, et nad oskaksid selliseid manipulatsioone ära tunda.

Aga kas AI loodud valeinfot ei saaks modereerida AI abil?
AI modereerimine on nii hea kui läbipaistev ja väärtuspõhine on selle loomiseks kasutatud infrastruktuur. On olemas „pro-sotsiaalseid algoritme“, mille eesmärk on rahu edendada ja valeinfot tuvastada. Küsimus on, kus ja kuidas neid rakendatakse? Kes saab neid rakendada? Mina ei usu, et kurja peaks alati sama tööriistaga tagasi lööma. Eelistan püsivamaid lahendusi – haridust.

USAs peetakse selliseid regulatsioone tsensuuriks, mistõttu on tähtis, et EL hoiaks kindlalt oma positsiooni ja selgitaks, et see ei ole tsensuur, vaid inimeste kaitsmine kahju eest.

Kas inimmodereerimine asendatakse AIga?
See otsus on suurte platvormide teha. Praegu jäävad inimmoderaatorid siiski hädavajalikuks, et otsustada piiripealse sisu üle – kas midagi on kahjulik või mitte. Ma arvan, et AI ja inimmodereerimise kombineeritud mudel jääb püsima. Ka seetõttu, et inimtöö on odav – ehkki see on ühtlasi ärakasutatud töö. Täielikku asendamist niipea ei tule.

Teie uurimistöö on käsitlenud nähtamatuid moderaatoreid. Miks see on nii nähtamatu töö ja millised on selle tagajärjed?
See on nähtamatu vägivalla vorm. Näeme, kuidas vägivaldsed struktuurid taastoodetakse just nähtamatuse kaudu. Näiteks *shadow banning* (ingl varjatult ärakeelamine – toim). Sisu ei kustutata, vaid selle nähtavus viiakse madalale ja see muutub

tähelepandamatuks. Tihti puudutab see feministlikku või vaimse tervisega seotud sisu.

Samuti on nähtamatu inimtöö, mis peitub iga meie infovoo või „kureeritud *feed*'i" taga. Enamik kasutajaid ei tea, et inimesed – sageli madala palgaga töötajad vaestes riikides – puhastavad meie sotsiaalmeedia voogu traumeerivast sisust. Sellel on neile tõsised psühholoogilised tagajärjed. Seda nähtamatust hoitakse üleval, et ärimudel püsiks. Kui inimesed teaksid, mis nende tarbitava sisu taga on, võib-olla nad loobuksid.

Inimeste kaitsmine kahju eest ei ole tsensuur, vaid demokraatlike süsteemide kohustus.

Kuidas seda teadlikkust ühiskonnas suurendada?
Alustada tuleb juba koolist. Õpilased peavad mõistma, kes sisu toodab ja kes seda kujundab. Sisu modereerimine on alati inimeste ja algoritmide koostöö, seega tuleb mõista ka platvormide disaini.

Praegu keskendutakse liigselt tehnoloogilistele oskustele. Me vajame kriitilist digipädevust, mis hõlmab ka teadmisi meedia tootmise kohta. Ja lisaks humanitaarteadused, kirjandus, kunst – need aitavad inimestel maailma eri viisidel näha. See on oluline osa meediakirjaoskusest.

Te mainisite intervjuu alguses, et meedia- ja infopädevus on digitaalse kahjuliku sisu kontekstis vananenud. Mida täpsemalt mõtlesite?
Pidasin silmas, et meediapädevus peab hõlmama ka modereerimise mehhanismide mõistmist – eriti algoritmilise, automaatse modereerimise kallutatust.

Lisaks ei tohi meediapädevus piirduda oskustega. Vajame rohkem kriitilist mõtlemist, uudishimu ja loovust. Samuti tuleb õpetada mitte ainult õpilasi, vaid ka õpetajaid, psühholooge, tervishoiutöötajaid – kõiki, kes töötavad inimestega. Näiteks, kui inimesed arutavad oma vaimse tervise muresid ChatGPTga, peaksid psühholoogid oskama selgitada, milleks see tööriist sobib ja milleks mitte.

Millised soovitused on Teil poliitikakujundajatele?
Esiteks võtta kallutatust tõsiselt, eriti sisu automaatses modereerimises. AI ei saa olla alati lahendus. Kaasata tuleb kodaniku-ühiskonna organisatsioone – nemad on sageli esimesed, kes platvormipoliitikate probleeme välja toovad.

Teiseks, seadusandjad ei tohiks järele anda. DSA, DMA, AI Act – need on olulised instrumendid ja neid tuleb kaitsta. Inimeste kaitsmine kahju eest ei ole tsensuur, vaid demokraatlike süsteemide kohustus.

Kolmandaks rahastada kriitilist, interdistsiplinaarset teadust. Mina olen politoloogia taustaga, seejärel liikusin kommunikatsiooniteadusesse, nüüd uurin platvormide disaini. Üksi ei saa ma seda teha – vajame meeskondi, kus on AI- ja IT-eksperte. Ja see peab olema avalikult rahastatud, muidu jääb teadmiste ja andmete monopol eraettevõtetele.

Anna Antonakis esines oktoobri lõpus Eesti Rahvusraamatukogu korraldatud meediahariduse konverentsil „Ekraanid ja hinged – elu digimaailmas", mida rahastab Šveitsi-Eesti koostööprogrammi projekt MeediaRadar. Projekt toimub aastatel 2024–2028 RaRa ja Kultuuriministeeriumi koostöös ning seda on kaasrahastanud Šveitsi riik majanduslike ja sotsiaalsete erinevuste vähendamiseks Euroopa Liidus.