HAMK
Hämeen ammatti-
korkeakoulu

KYBERRIKOLLISUUS

Deepfake-huijaukset ja massarikollisuus

Ammattikorkeakoulututkinnon opinnäytetyö

Tieto- ja viestintätekniikka, insinööri (AMK)

Syksy 2024

Sami Rintala

HAMK
Hämeen ammatti-korkeakoulu

Tieto- ja viestintätekniikka, biotalous
Tekijä Sami Rintala
Tiivistelmä
Vuosi 2024
Työn nimi Kyberrikollisuus – Deepfake-huijaukset ja massarikollisuus
Ohjaaja Ismo Turve

Opinnäytetyön tarkoitus oli selvittää, minkälainen on kyberrikollisuuden ja erityisesti deepfake-huijausten tilanne vuonna 2024 sekä miten ne näyttäytyvät massarikollisuuden yhtenä ilmentymänä. Työn tarkoitus oli selvittää suojautumis- ja puolustautumiskeinoja yleisesti kyberrikoksia sekä erityisesti deepfake-huijauksia vastaan. Työssä arvioitiin miten hyvin yritykset ovat varautuneet kyberrikosten torjuntaan sekä miten Euroopan tasoinen NIS 2 -direktiivi ja ISO 27001 -standardi antavat yrityksille linjauksia deepfakeja vastaan. Työssä tutkittiin, miten helppoa on oman deepfaken tekeminen ja kuinka helppo sellainen on tunnistaa tekoälyn tekemäksi. Toimeksiantaja työlle oli Suomen Kyberturvallisuuskeskus.

Opinnäytetyön tietopohja koostuu kyberrikollisuuden historiasta, eri kyberrikollisuuden keskeisimmistä rikostyypeistä ja deepfakejen luomiseen ja suojautumiseen käytetyistä teknologioista. Valittu menetelmä on tutkimuksellinen ja työ pitää sisällään myös suppean toiminnallisen osuuden. Toiminallisen osuuden testiosuudet kerättiin empiirisen tutkimuksen avulla, jossa testattavat henkilöt pyrkivät tunnistamaan aidot mediat deepfakeista. Tämä kerätty aineisto analysoitiin käyttämällä tilastollisen analysoinnin perusmenetelmiä. Toinen toiminallinen osuus työssä koostuu Kyberturvallisuuskeskuksen asiantuntijoiden haastattelusta.

Empiirisen kokeen tuloksena havaittiin pieniä eroja sukupuolten välillä. Miesten kohdalla koulutus korreloi kuvien tunnistamiseen: mitä korkeampi koulutus miehillä oli, sen paremmin aidot kuvat tunnistettiin tekoälyn tekemistä. Vastaavaa ei löytynyt naisten otoksessa. Lisäksi havaittiin, että miehet tunnistivat deepfake-videot hieman naisia paremmin. Erot olivat kuitenkin niin pieniä, että tuloksista ei voida tehdä johtopäätöksiä.

Tutkimuksessa selvisi, että Euroopan laajuisen NIS 2 -direktiivin voimaantulosta ja ISO 27001 -standardista huolimatta yritykset eivät ole riittävän hyvin valmistautuneet kyberhyökkäyksiin ja erityisesti deepfake-huijauksiin. Vastaavasti yksittäiset ihmiset ovat luottavaisia eri medioissa tuleviin tallenteisiin eivätkä osaa epäillä niiden aitoutta. Tutkimuksen perusteella suositellaan, että yritykset nostaisivat valmiuksiaan deepfake-huijauksia vastaan panostamalla sekä teknologisiin ratkaisuihin että kouluttamalla henkilöstöään. Vastaavasti tavallisten kansalaisten ymmärrystä ylipäätään kyberrikoksiin ja erityisesti deepfake-huijauksiin tulee nostaa merkittävästi. Deepfake-huijaukset aiheuttavat merkittäviä rahallisia tappioita yrityksille jo nyt ja menetyksien suuruus tulee kasvamaan tulevina vuosina huimasti.

Toimeksiantajan näkökulmasta työ onnistui nostamaan esiin useita konkreettisia ja tärkeitä havaintoja, kuten uusien teknologioiden luomien muutosten jatkuvan huomioimisen turvallisuuskoulutusten kehittämisessä. Kokonaisuutena tarkastellen toimeksiantajan mielestä työ onnistui vastaamaan jokaiseen asetettuun tutkimuskysymykseen hyvin.

Avainsanat Kyberrikollisuus, deepfake, massarikollisuus
Sivut 75 sivua ja liitteitä 15 sivua

HAMK
Hämeen ammatti-korkeakoulu

Information and Communication Technology, Bio and Circular Economy
Author Sami Rintala
Subject Cybercriminality – Deepfake Scams and Mass Crime
Supervisor Ismo Turve

Abstract
Year 2024

The purpose of this thesis was to picture the situation of cybercrime and especially deepfake scams in 2024, and how they represent one form of mass crime. The purpose was also to find out the protection and defense methods against cybercrimes in general and deepfake scams in particular. The thesis assessed how well companies are prepared to defend themselves against cybercrimes and how well the European level NIS 2 directive and ISO 27001 standard define policies for companies against deepfakes. The thesis also examined how easy it is to make deepfake and how easy it is to recognize one made by artificial intelligence. The client of the thesis was the National Cyber Security Center.

The knowledge base of the thesis consists of the history of cybercrime, the most important types of cybercrime and the technologies used to create and protect companies and individuals against deepfakes. The thesis is mainly theoretical, and it also includes a concise functional part. The functional part includes an empirical test, where randomly selected people tried to identify genuine media from deepfakes. The collected data was analyzed using basic statistical methods and data analysis. The other functional part is expert interviews of cybercrimes.

The empirical experiment showed minor differences between the genders. In males, the higher the education, the better they recognized the real pictures compared to the ones made by artificial intelligence. There was no similar finding among females. In addition, it was found that in general males recognized deepfake videos slightly better than females. However, the differences are so minor, that no overall conclusions can be made.

The research found that despite of the coming European-wide NIS 2 directive and the ISO 27001 standard, companies are not sufficiently prepared for cyberattacks and especially deepfake scams. At the same time, individual people are confident that they recognize from media whether the content is real or fake. Based on this thesis, it is recommended that companies increase their capabilities against deepfake scams by investing in both technological solutions and training their personnel. Accordingly, ordinary citizens' understanding in general of cybercrimes and in particular deepfake scams, must be increased significantly. Deepfake scams are already causing major financial losses to companies, and the size of the losses will grow exponentially in the coming years.

From the client's point of view, the work managed to highlight several concrete and important findings, such as the constant consideration of changes created by new technologies in the development of safety training. As whole, in the opinion of the client, the work managed to answer every research question well.

Keywords Cybercrime, deepfake, mass criminality
Pages 75 pages and appendices 15 pages

Sanasto

AI	Artificial Intelligence eli tekoäly tarkoittaa älykkäästi toimivaa ohjelmaa tai järjestelmää.
Deepfake	Deepfake eli syväväärennös. Deepfakella tarkoitetaan ääntä, kuvaa tai videota, jota on muokattu tekoälyä käyttäen.
GAN	Generative Adversal Networks eli generatiivinen kilpaileva verkosto on neuroverkkoarkkitehtuuri, jota käytetään syväoppimiseen pienellä datamäärällä.
Kyberrikollisuus	Kyberrikollisuudella tarkoitetaan tietoverkossa tapahtuvia rikoksia.
Kyberturvallisuus	Kyberturvallisuudella tarkoitetaan tietojen, laitteiden ja ihmisten suojaamista tietoverkon kautta tulevilta vaaroilta.
Massarikollisuus	Massarikollisuudella tarkoitetaan sellaisia rikoksia, jotka kohdistuvat useisiin ihmisiin saman aikaisesti ja tapahtuvat usein.
MFA	Multi-factor Authentication eli monivaiheinen tunnistautuminen tarkoittaa sitä, että henkilön identiteetti varmistetaan useammalla eri tunnistautumismenetelmällä.
NIS 1	EU:n kyberturvallisuusdirektiivi vuodelta 2016.
NIS 2	EU:n kyberturvallisuusdirektiivi, joka tulee voimaan lokakuussa 2024.
OSINT	Open Source Intelligence tarkoittaa tyypillisesti avoimesta lähteestä, esimerkiksi sosiaalisesta mediasta, etsittyä tietoa.
VPN	Virtual Private Network eli Virtuaalinen erillisverkko on tietoverkko, jolla kaksi tai useampi tietoverkko yhdistetään julkisen verkon yli näennäisesti yksityiseksi verkoksi.

Sisällys

1 Johdanto ... 1

2 Kyberturvallisuus ... 2

2.1 Kyberturvallisuuden historia ... 2

2.2 Massarikollisuus kyberympäristössä ... 6

2.2.1 Kyberrikollisuus ... 7

2.2.2 Tietojenkalastelu eli phishing ... 7

2.2.3 Identiteettivarkaus ... 8

2.2.4 Informaatiovaikuttaminen ... 9

2.2.5 Kiristyshaittaohjelma eli ransomware ... 10

2.2.6 Kuvilla ja videoilla kiristäminen ... 11

2.2.7 Spoofing-hyökkäykset ... 12

2.2.8 Social engineering ... 13

2.2.9 WhatsApp-huijaukset ... 14

2.3 Kyberrikollisuuden rikostutkinta ... 15

2.4 Deepfake-huijaukset ... 16

2.4.1 Autoenkooderi ... 19

2.4.2 Deepfake-huijauksien taustaa ... 20

2.5 Esimerkkejä Deepfake-huijauksista ... 21

2.6 Kyberrikoksissa deepfake-huijaukset kasvamassa ... 23

3 Puolustautumis- ja suojautumismenetelmät ... 25

3.1 NIS – Network and Information System ja ISO 27001 -standardi ... 25

3.1.1 NIS 2 -tietoturvadirektiivi tulee voimaan lokakuussa 2024 ... 26

3.1.2 ISO 27001 -standardi ... 30

3.1.3 ISO 27001 ja NIS 2 turvaamassa yrityksiä deepfake-huijauksia vastaan ... 30

3.2 Cybercrime Exit ... 33

3.3 The Joint Cybercrime Action Taskforce (J-CAT) ... 33

3.4 Suojautuminen kyberrikollisuutta vastaan ... 34

3.4.1 Tietomurtoja vastaan suojautuminen ... 35

3.4.2 Haittaohjelmia vastaan suojautuminen ... 37

3.4.3 Tietojenkalasteluviestejä vastaan suojautuminen ... 37

3.4.4 Spoofing-hyökkäyksiä vastaan suojautuminen ... 39

3.4.5 Informaatiovaikuttamista vastaan suojautuminen ja sen tunnistaminen ... 41

3.4.6 Kiristyshaittaohjelmia vastaan suojautuminen 42
3.4.7 Social engineering hyökkäyksiä vastaan suojautuminen ja puolustautuminen 43
3.4.8 WhatsApp-huijauksia vastaan suojautuminen 44
3.4.9 Deepfake-väärennösten tunnistaminen ja niitä vastaan suojautuminen 45
3.4.10 Työkaluja Deepfake-videoiden tunnistamiseen 47
3.5 Puolustautumismenetelmien toimivuus ja luotettavuus 49
4 Deepfake-huijauksen tekeminen ja tunnistaminen 51
4.1 Deepfake-videon ja äänitallenteen luominen 51
4.1.1 Heygen - www.heygen.com 51
4.1.2 Deepbrain AI – www.deepbrain.io 52
4.2 Deepfake-videoiden ja äänien tunnistamisen testaaminen 54
4.2.1 Testiaineiston luominen empiiriseen kokeeseen 55
4.2.2 Koetilanne ja testiaineiston kerääminen 56
4.3 Testitulosten analysointi 57
4.3.1 Deepfake-käsitteen ja kuvaparien tunnistaminen 57
4.3.2 Kuvaparien tunnistaminen 58
4.3.3 Deepfake ääninäytteet ja niiden tunnistaminen 60
4.3.4 Deepfake-videot ja niiden tunnistaminen 62
4.3.5 Yhteenveto tuloksista 64
5 Tulokset 64
6 Johtopäätökset ja pohdinta 68
7 Yhteenveto 69
Lähteet 70

Kuvat ja taulukot

Kuva 1. Kyberturvallisuuden historia kahdella aikajanalla. (Šimonėlytė, 2023) 6

Kuva 2. Identiteettivarkauden yritys henkilöstöhallintoon. 9

Kuva 3. Autoenkooderin toiminta (Lanham, 2021, luku 2) 19

Kuva 4. Ilmaiseksi saatavilla olevat ohjeet reaaliaikaiseen deepfake-videoon. 24

Kuva 5. NIS 2 -direktiivi verrattuna NIS 1 -direktiiviin. (NIS 2 -Directive, n.d.)..... 28

Kuva 6. Outlookiin rakennettu kuvake tietojenkalastelusta ilmoittamiseen. 32

Kuva 7. Heygenin AI Studio -videoeditori. 52

Kuva 8. Deepbrain AI Studios -videoeditori. 53

Kuva 9. Testattujen sukupuoli- ja ikäjakauma. 58

Kuva 10. Testikuvien tunnistaminen eri ikäluokissa. 59

Kuva 11. Koulutustaso sukupuolittain testikuvien tunnistamisessa..... 60

Kuva 12. Sukupuolten välinen ero ikäluokittain ääninäytteiden tunnistamisessa. 61

Kuva 13. Koulutustaustan vaikutus ääninäytteiden tunnistamiseen..... 62

Kuva 14. Sukupuolten välinen ero ikäluokittain videonäytteiden tunnistamisessa. 63

Kuva 15. Koulutustason vaikutus sukupuolittain videoiden tunnistamisessa. 63

Taulukko 1. Esimerkkejä maailman ja Suomen suurimmista tietovuodoista. (Šimonėlytė, 2023) 4

Taulukko 2. Historian tunnetuimmat kiristyshaittaohjelmat. (F-Secure, n.d.-c)..... 10

Taulukko 3. Koneoppimisen ryhmät ja niiden eroavaisuudet. (Lanham, 2012, luku 1). 18

Taulukko 4. NIS-direktiivejä koskevat sektorit sekä niiden sisältö. (NIS 2 -Directive, n.d.) 28

Taulukko 5. Yrityksiltä vaaditut perusturvatoimenpiteet NIS 2:ssa (NIS 2 -Directive, n.d.). 29

Liitteet

Liite 1. Aineistonhallintasuunnitelma

Liite 2. Testiaineiston kuvaparit

Liite 3. Äänitallenteiden tekstit

Liite 4. Videotallenteiden tekstit

Liite 5. Testien tulokset

Liite 6. Kyberturvallisuuskeskuksen asiantuntijoiden haastattelu litteroituna

Liite 7. Keskusrikospoliisin asiantuntijoiden haastattelu litteroituna

1 Johdanto

Euroopan komission määrittely kyberrikollisuudesta kattaa rikokset, jotka tehdään käyttäen elektronisia viestintäverkkoja ja tietojärjestelmiä verkossa (European Comission, n.d.). Koska olet viimeksi lukenut pankkiryöstöstä, joka on tehty näyttävästi ryntäämällä pankkiin ja ryöstämällä tiskiltä rahaa? Pankkiryöstöt eivät ole kadonneet, mutta ne tehdään nykyään pääosin tietoverkon välityksellä. Rikokset, joissa kohdistetaan hyökkäys yritysten tietoverkkoihin, väärennettyjen linkkien lähettäminen salasanojen saamiseksi tai identiteettivarkaudet, joissa käytetään teknologiaa apuna, ovat pakottaneet yritykset ja valtiot arvioimaan puolustautumis- ja suojautumiskeinojaan aivan uudella tavalla.

Deepfake-huijaus eli syvähuijaus on yksi esimerkki kasvavasta kyberrikollisuustrendistä. Tässä työssä käytetään termiä deepfake, koska se on vakiintunut termi myös suomenkielisessä aineistossa. Työssä on avattu useampia kyberrikollisuuden lajeja, tehty katsaus kyberrikollisuuden historiaan ja tilanteeseen vuonna 2024. Erityisesti on keskitytty deepfake-rikoksiin sekä pohdittu, mitä ne tarkoittavat massarikollisuudessa. Työssä on avattu suojautumis- ja puolustautumiskeinoja sekä yksityisen henkilön että yrityksen näkökulmasta. Tästä työssä on rajattu sekä psykologinen että sosiaalinen lähestymistapa deepfakeistä pois ja keskitytty pääosin teknologiseen kehitykseen. Toisena henkilönä esiintyminen tavoiteltaessa rikollista hyötyä ei ole uusi asia, mutta koneoppiminen, tekoäly, kasvontunnistusalgoritmit ja keinotekoiset neuroverkot ovat luoneet mahdollisuuden deepfakejen kehittymiselle. Työssä on kokeellinen osuus, jossa on testattu deepfake-huijauksien tekoa ja luotu tekoälyn avulla ääni- ja videotallenteita. Näiden tunnistamista on testattu pienellä kokeellisella joukolla.

Työssä on arvioitu deepfake-teknologian tulevaisuutta osana kyberrikollisuutta sekä kuvattu mahdollisia skenaarioita, mihin se tulee kehittymään. Lisäksi on pohdittu myös tulevaisuuden suojautumis- ja puolustautumiskeinoja yksityiselle ja yrityssektorille. Keskeisinä tutkimuskysymyksinä tässä työssä ovat:

- Miltä tämän hetken yleiskuva näyttää kyberrikollisuudessa?
- Mitkä ovat suojautumis- ja puolustautumiskeinoja deepfake-huijauksia vastaan?
- Mitkä ovat deepfake-huijausten tulevaisuuden näkymät eli miten ne mahdollisesti tulevat kehittymään tulevaisuudessa?

Opinnäytetyön toimeksiantaja on Suomen Kyberturvallisuuskeskus.

2 Kyberturvallisuus

Tässä kappaleessa käsitellään kyberturvallisuutta ja luodaan katsaus sen historiaan. Lisäksi käsitellään massarikollisuutta ja erilaisia kyberrikollisuuden huijausmuotoja. Kappaleessa paneudutaan deepfake-huijauksien tekniikkaan niiden tekemisen osalta ja luodaan katsaus deepfake-huijauksien taustaan. Lopuksi katsotaan vielä esimerkkejä deepfake-huijauksista.

Kyberturvallisuudella tarkoitetaan tietojen, laitteiden ja ihmisten suojaamista tietoverkon kautta tulevilta vaaroilta. Tällaisia vaaroja ovat esimerkiksi kyberhyökkäykset. Kyberhyökkäysten tyypeistä yleisin on palvelunestohyökkäys, jossa hyökkääjä pyrkii kuormittamaan palvelinta keinotekoisella liikenteellä. Kuormitettaessa palvelinta enemmän kuin mihin se pystyy vastaamaan, palvelin tai verkkosivusto saadaan ylikuormittumien vaikutuksesta kaatumaan. Toinen tyypillinen esimerkki turvallisuusvaaroista on hakkerit, jotka pyrkivät ohittamaan tietoturvasuojaukset varastaakseen yrityksen tai yksityisen henkilön arkaluontoista materiaalia. Hakkerien pyrkimyksenä saattaa myös olla haittaohjelmien, virusten tai matojen levittäminen kohteen laitteelle. Haittaohjelmien tarkoitus on saada levitettyä virus, asentaa vakoiluohjelma, kaapata selain tai selaimen tietoja hyökkääjän rikolliseen käyttöön. Nämä tapahtuvat usein ilman kohteen tietoista suostumusta ja tietämättä. (Šimonėlytė, 2023)

2.1 Kyberturvallisuuden historia

Vaikka terminä kybertuvallisuus on suhteellisen uusi, voidaan sen alkusysäys nähdä 1950-luvulla, kun ensimmäiset tietokoneet yhdistettiin toisiinsa tietokoneverkoiksi. Tällöin kantamat olivat kuitenkin niin pieniä, että tietokoneiden oli oltava samassa tilassa toimiakseen verkkona. 1960-luvulla Pentagonin Advanced Research Project Agency – tutkimustoimisto (ARPA) kehitti pakettikytkentäverkon, joka mahdollisti tietokoneiden välisen viestinnän pitkien matkojen päästä. Näin syntyi nykyaikainen internet. (Šimonėlytė, 2023)

Vuonna 1971 tutkija Bob Thomas, joka oli ollut mukana ARPA-projektissa, kehitti viestin, joka pystyi liikkumaan tietokoneelta toiselle ja näyttämään ruudulla tekstin: ”Olen hiipijä. Ota kiinni, jos saat.” (”I am the creeper. Catch me if you can.”). Eli hän tavallaan loi ensimmäisen Creeper-nimisen ”haittaohjelman” eli ”viruksen”. Thomas ei ollut kyberrikollinen, mutta hänen kehittämänsä itsenäisesti toimivan ohjelma, joka monistaa itseään, löi näin alkutahdit haittaohjelmien kehittämiselle. Toinen ARPA-projektissa mukana ollut tutkija Ray Tomlinson, joka tunnetaan myös sähköpostin keksijänä, loi ohjelman, jonka tarkoitus oli etsiä ja tuhota ”virus”. Ohjelman nimi oli Reaper. Tästä voidaan katsoa alkaneen kilpajuoksu haittaohjelmien ja virustorjuntaohjelmien välillä, joka jatkuu yhä nykypäivänäkin. Tietotekniikan kehitys on

muuttanut haittaohjelmien syntyä sekä niiden torjuntaa. Tekninen kehittyminen sekä ohjelmissa, tietoverkoissa että käyttäjissä on luonnollisesti muuttanut tätä perusidean ollessa edelleen sama. (Šimonėlytė, 2023)

1980-luvulla nähtiin ensimmäiset isot kyberiskut, kun suuria yhdysvaltalaisia yrityksiä, esimerkiksi AT&T ja National CSS, vastaan iskettiin tietoverkkoja pitkin. Vienna-virus oli 1980-luvulla tunnetuin tuhoa tehnyt virus. Viennaan reagoi saksalainen kyberturvallisuusasiantuntija Bernt Fix, kun hän huomasi Viennan saastuttaneen laitteensa. Hän loi ensimmäisen varsinaisen virustorjuntaohjelman koodatessaan ohjelman, joka paikallisti ja poisti Viennan. Tätä voidaan pitää ensimmäisenä nykyaikaisena virustorjuntaohjelmana. Ensimmäinen kaupallinen virustorjuntaohjelma julkaistiin vuonna 1988. 1990-luvulla Microsoft julkaisi useita paranneltuja versioita Windows-käyttöjärjestelmästään. Ehkä suurin paranneltu versio oli Windows 95, jonka mukana julkaistu Internet Explorer selain oli hyvin suosittu useamman vuosikymmenen. Windows 95 mukana tuli myös Microsoft Outlook ja tämä antoi monille ihmisille mahdollisuuden nopeaan viestittelyyn sähköpostin muodossa. Valitettavasti myös kyberrikolliset saivat tämän saman mahdollisuuden ja niinpä 1999 iskikin Outlookin liitetiedostona liikkunut Melissa-virus. (Šimonėlytė, 2023)

2000-luvun alussa kyberrikolliset huomasivat, että ihmiset suhtautuvat varoen sähköpostien liitetiedostoihin eivätkä välttämättä enää avaa niitä. Tämän myötä kehiteltiin uusi huijausmuoto, jota käytetään edelleen tänäkin päivänä. Mikäpä saisi tavallisen tallaajan innostumaan enemmän kuin mahdollinen suuri rahapalkinto ja mahdollisuus nopeaan rikastumiseen. Rikolliset huijasivat sähköpostein ihmiset uskomaan, että suuri palkkio tai rahansiirto odottaa, kun vierailee rikollisten perustamilla huijaussivustoilla. Todellisuudessa sivustolta sai vain haittaohjelman tai mikä vielä pahempaa, sivustolla on kirjautuminen jonnekin, jolloin rikolliset saavat uhrin käyttäjätunnuksen ja salasanan. (Šimonėlytė, 2023)

Ensimmäiset VPN-verkot näkivät päivänvalon 2000-luvun puolivälissä. VPN-verkkojen avulla ihmiset pystyivät salaamaan käyttämänsä liikenteen. VPN-ratkaisut veivät aluksi verraten paljon tilaa ja sen aikaisilla tietokoneilla oli huomattavasti nykyistä rajallisempi kapasiteetti. Kapasiteettirajoitteen vuoksi ne eivät vielä juurikaan saaneet suosiota ennen vuotta 2007, kunnes Panda Security ja McAfee julkaisivat ensimmäiset pilvipohjaiset turvallisuusratkaisut ja tilaongelma oli ratkaistu. (Šimonėlytė, 2023)

2010-luvulla voidaan katsoa alkaneen kansainvälisen kybersodankäynnin aikakausi, kun kyberiskuja käytettiin aseena hallitusten iskiessä vihollisiaan vastaan salaa. Tämän seurauksena kyberturvallisuus muuttui rikosten estämisestä ja tietojen suojaamisesta

kansalliseksi turvallisuudeksi. Myös yksityisyydensuoja nousi otsikoihin 2010-luvulla, kun esimerkiksi Facebook ja Google keräsivät suuret määrät tietoa käyttäjistään mainosten kohdentamiseksi tai tietojen myymiseksi ulkopuolisille mainostajille ja muut yhtiöt seurasivat perässä. Sittemmin tähän on puututtu erilaisin laein ympäri maailmaa. Myöhemmin myös käyttäjät ovat pystyneet ostamaan itse ohjelmistoja, jotka suojaavat heidän yksityisyyttään internetissä. Myös VPN-verkkojen suosio kasvoi tämän seurauksena. (Šimonėlytė, 2023)

Tietovuodot nostivat päätään 2010-luvulla aivan uudella tavalla. Tietovuoto tarkoittaa tietojen luvatonta vuotamista verkkoon tai toisille käyttäjille. Vuoto voi tapahtua vahingossa tai hakkerin toimien seurauksena tietoja varastettaessa (Šimonėlytė, 2023). Jos kyseessä on hakkerin toimien seurauksena tapahtunut tietovuoto, puhutaan tietomurrosta. Tietomurto siis tarkoittaa tietojärjestelmään tunkeutumista oikeudetta. Rikoslaissa määritellään tietomurto toiminnaksi, jossa käytetään käyttäjätunnusta luvattomasti tai turva- ja suojausjärjestelyt ohitetaan järjestelmään tai tietoihin käsiksi pääsemiseksi. Tietomurto tai sen yritys ovat rikoksia, joista tulee aina tehdä rikosilmoitus poliisille. Tietomurtoja tehdään yleisimmin siten, että varastetaan kirjautumistiedot uhrilta. Yleisimmin tähän käytetään tietojenkalastelua, jossa uhri ohjataan aidolta näyttävälle kirjautumissivulle, jossa sitten kerätään uhrin tiedot rikollista toimintaa varten. Tietomurto voidaan tehdä myös siten, että tunkeudutaan eri menetelmiä käyttäen turvajärjestelmien ohi. Tietomurron tarkoitus on yleensä hyödyntää järjestelmästä löytyviä tietoja esimerkiksi petoksiin tai muuten varastaa tietoja niiden arkaluontoisuuden vuoksi (Poliisi, n.d.-a). Taulukko 1 on sekä maailman että Suomen suurimpia tietovuotoja ja tietomurtoja 2010 ja 2020-luvuilta.

Taulukko 1. Esimerkkejä maailman ja Suomen suurimmista tietovuodoista. (Šimonėlytė, 2023)

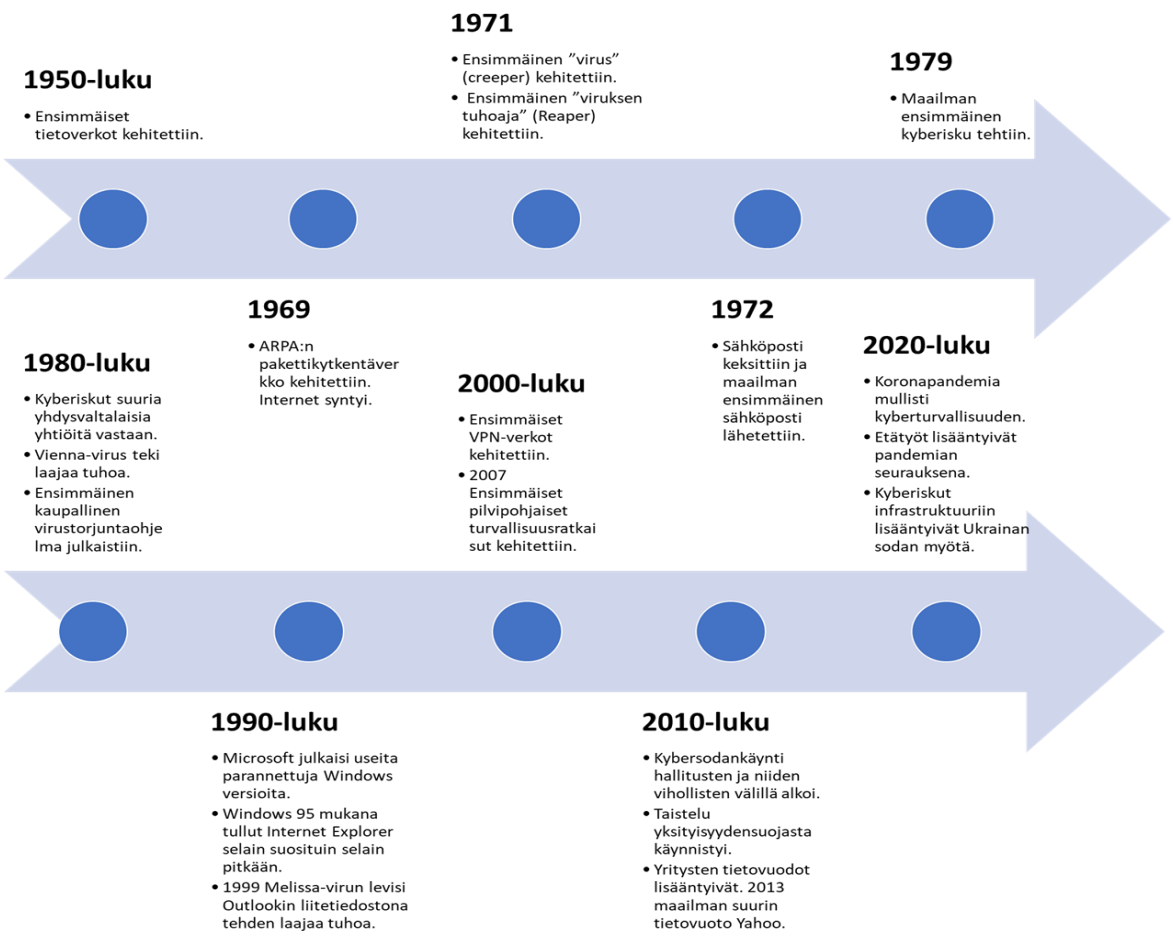
Vuosi	Tietovuoto	Tietovuoto ja sitä koskeneiden ihmisten määrä
2013	Yahoo tietovuoto	Historian suurin tietovuoto, jossa paljastui 3 miljardin ihmisen tietoja. Yritys kertoi tapahtuneesta julkisuuteen vasta vuonna 2016.
2019	First American tietovuoto	Tietovuodossa vuodettiin 850 miljoonaa arkaluontoista asiakirjaa, jotka sisälsivät esimerkiksi sosiaaliturvatunnuksia.
2019	Facebook tietovuoto	Tietovuodossa paljastui 500 miljoonan Facebook käyttäjän tiedot.
2024	Helsingin kaupungin tietomurto	Suomen suurin tietomurto. Tässä tietomurrossa Helsingin kaupungin kasvatuksen ja koulutuksen palveluja käyttäneiden henkilöiden tietojen epäillään paljastuneen jopa 150 000 oppijan sekä 38 000 kaupungin työntekijän osalta.
2019	Psykoterapiakeskus Vastaamon tietomurto	Tässä tietomurrossa 33 086 Vastaamon asiakkaan henkilö- ja potilastiedot varastettiin ja ainakin osa julkaistiin pimeässä Tor-verkossa.

2020-luvulla koronapandemia sai aikaan suuren muutoksen kyberturvallisuuteen ja tietosuojaan yksilön kannalta. Pandemia nopeutti tietokoneiden ja internetin saamista ihmisten koteihin ennennäkemättömällä tavalla, koska pandemian seurauksena suuri osa ihmisistä siirtyi tekemään etätöitä entisen toimistotyön sijaan. Lisääntynyt etätyö mahdollisti myös hakkereille helpomman pääsyn ihmisten tietoihin, kun ihmiset ottivat kotoaan omilla laitteillaan yhteyden työpaikan verkkoon ja hakkereiden oli huomattavasti helpompi murtautua ihmisten omille laitteille kuin suojatuille työpaikkojen laitteille. (Šimonėlytė, 2023)

Brittiläinen turvallisuusohjelmayritys Sophos Group kertoo, että yli puolet yrityksistä joutui kiristysohjelmien uhreiksi vuonna 2020. Myös tietojenkalasteluiskut lisääntyivät koronapandemian aikana, kun ihmiset tilasivat tuotteita kotiin internetin kautta ja hyökkääjä pystyi tekemään kuljetuspalveluhuijauksia kuljetuspalvelun nimissä. Myös tekstiviestihuijaukset kasvoivat pandemian myötä. Ihmisille tarjottiin lääkkeitä ja rokotuksia koronaa vastaan tekstiviestein, jossa on linkki huijaussivustolle. (Šimonėlytė, 2023)

Kriittiseen infrastruktuuriin on kohdistunut paljon iskuja 2020-luvulla. Esimerkkinä tästä on vuoden 2021 toukokuussa tapahtunut isku Colonial Pipeline- yritykseen, joka vastaa polttoainepumppauksesta Yhdysvaltojen itärannikolle. Hakkerit varastivat yli 100 gigabittiä yrityksen dataa ja lukitsivat yrityksen verkot kiristysohjelmalla. Colonial Pipeline päätyi maksamaan rikollisille lunnaat saadakseen verkot takaisin käyttöönsä ja USA:n kriittisen polttoainekuljetuksen jälleen toimintaan. Ennen vuoden 2022 helmikuuta, kun Venäjä hyökkäsi Ukrainaan, nähtiin Ukrainassa jo kyberhyökkäyksiä, kun Ukrainan hallituksen laitteille levitettiin haittaohjelmia ja virallisille verkkosivuille uhkaavia viestejä. Liettuan johdolla perustettu Cyber Rapid Response Team on auttanut Ukrainaa suojautumaan verkkohyökkäyksiltä. Kuva 1 on vielä tiivistettynä kyberturvallisuuden historia kahdella aikajanalla. (Šimonėlytė, 2023)

Kuva 1. Kyberturvallisuuden historia kahdella aikajanalla. (Šimonėlytė, 2023)



2.2 Massarikollisuus kyberympäristössä

Massarikollisuudella tarkoitetaan sellaisia rikoksia, jotka kohdistuvat useisiin ihmisiin saman aikaisesti ja tapahtuvat usein. Kyberrikollisuudessa hyvä esimerkki massarikollisuudesta on tietojenkalastelu eli phishing. Rikolliset pyrkivät lähettämään saman tietojenkalasteluviestin useille eri ihmisille ja tätä tapahtuu usein useiden eri toimijoiden toimesta. Sähköpostitse tällainen toiminta on hyvin yleistä ja helppoa rikollisille, kuten meistä useimmat ovat huomanneetkin. (Polisen, n.d.)

Massarikosten tutkinnalle on valitettavan tyypillistä, että niiden suuren määrän vuoksi, poliisi ei ehdi tutkia niitä kaikkia. Samoin tietoverkkoyhteyksien avulla monien kyberrikosten tekijät ovat toisessa valtiossa, jolloin tutkinta vaikeutuu. Kansallisesti poliisien voimavarojen kohdentamisesta on ollut keskustelua mediassa nimenomaan massarikosten tutkinnassa, ja

eduskunta on muun muassa 2019 antanut kirjallisen vastauksen esitettyyn kysymykseen massarikosten tutkinnasta ja niihin liittyvistä resursseista. Vastauksessa viitataan 2019 annettuun tehtävään Poliisihallitukselle suunnitella keinoja, joilla se kykenee parantamaan ja edelleen kehittämään massarikostutkintaa. (Eduskunta, 2019)

Deepfake-huijaukset ovat yksi uusimmista kyberrikosten ilmenemismuodoista. Ennen deepfake-rikoksen yritystä ja onnistumista, ovat rikolliset tehneet paljon töitä perinteisempien kyberrikostyyppien avulla luodessaan pohjaa onnistuneelle deepfakelle. Siksi on välttämätöntä ymmärtää ja tunnistaa laajasti muita kyberrikollisuuden tyyppejä ennen deepfake-huijausten analyysiä.

2.2.1 Kyberrikollisuus

Kyberrikollisuus on yleisnimitys, jolla tarkoitetaan tietoverkkorikoksia. Suomen Poliisi jaottelee nämä rikokset tietotekniikkaan ja tietoverkkoihin kohdistuviksi rikoksiksi sekä tietotekniikkaa ja tietoverkkoja hyväksi käyttäen tehtyihin rikoksiin. Tietoverkkoavusteiset rikokset ovat rikoksia, jotka eivät kohdistu tietoverkkoihin tai tietojärjestelmään vaan se mahdollistaa jonkin muun rikoksen. Tällaisia rikoksia ovat esimerkiksi huumausainerikokset, rahanpesut sekä petokset. Tietoverkkoihin kohdistuvia rikoksia ovat esimerkiksi tietomurrot, datavahingon teot ja palvelunestohyökkäykset. (Limnell, 2019)

Sisäministeriön mukaan kyberrikollisuus yleisesti tarkoittaa tietotekniikkaan tai tietoverkkoihin kohdistuvia rikoksia tai niitä hyväksi käyttäen tehtyjä rikoksia. Monet rikollisuuden muodot ovatkin siirtyneet internettiin tai niissä käytetään internettiä hyväksi. 2020-luvulla poliisin tietoon tulleista rikoksista tehdäänkin yhä suurempi osa tietoverkoissa tai tietojärjestelmissä. (Sisäministeriö, n.d.)

2.2.2 Tietojenkalastelu eli phishing

Tietojenkalastelu (engl. phishing) tarkoittaa tilannetta, jossa hyökkääjä huijaa kohteen avaamaan haitallisen linkin. Linkki saattaa tulla sähköpostin liitteenä ja se on usein naamioitu sellaiseksi, ettei se herätä epäilyjä; esimerkiksi kiinnostava uutinen tai video. On myös mahdollista, että tällaisen viestin lähettäjäksi on naamioitu kohde johon ihmiset luottavat: poliisi, pankki tai muu virallinen taho. Viestien tarkoituksena voi olla joko saada saastutettua kohteen laite jollakin haittaohjelmalla tai sitten kyseessä voi olla tarkoitus vain saada varastettua henkilötietoja, käyttäjätunnuksia ja salasanoja. Tällaisessa tapauksessa sähköpostin liitteenä yleensä on jonkinlainen kirjautumissivu, jonne syötetyt tiedot päätyvät hyökkääjien käsiin. Tietojenkalastelusta on lisäksi myös muita muotoja kuten kohdennettu

tietojenkalastelu (engl. spear phishing), tekstiviestihuijaukset (engl. smishing) ja huijauspuhelut (engl. vishing). (F-Secure, n.d.-a)

Kohdennettu tietojenkalastelu eli spear phishing voi olla yksityishenkilön lisäksi kohdennettu johonkin yritykseen tai organisaatioon sekä niiden johtohenkilöihin. Tällaista huijausta voi olla vaikea tunnistaa, koska kyseessä on kohdetta varten räätälöity kohdennettu tietojenkalastelu. Kohdennettua tietojenkalastelua voidaan myös tehdä yrityksen tiettyyn henkilöryhmään, joiden luottamus halutaan voittaa ja tätä käyttää hyväksi. (F-Secure, n.d.-a)

Tekstiviestihuijaukset eli smishing huijauksessa hyökkääjä ujuttaa haitallisia viestejä tai linkkejä olemassa oleviin viestiketjuihin hyödyntäen teksti- tai pikaviestipalveluita. Tällaisen viestin tarkoitus on myös saada kohde avaamaan linkki, joka vie huijaussivustolle. Huijaussivustot keräävät sitten henkilökohtaisia tietoja, jotka päätyvät rikollisten käsiin. (F-Secure, n.d.-a)

Huijauspuheluiden eli vishing tarkoituksena on urkkia tietoja puhelimitse esiintyen jonain luottavana tahona, kuten poliisi tai pankki. Myös yrityksen IT-tukena esiintyminen on mahdollista ja tällöin kohde on tarkoitus saada asentamaan esimerkiksi etähallintaohjelma tietokoneelleen, jolla hyökkääjä pääsee käsiksi uhrin koneeseen. (F-Secure, n.d.-a)

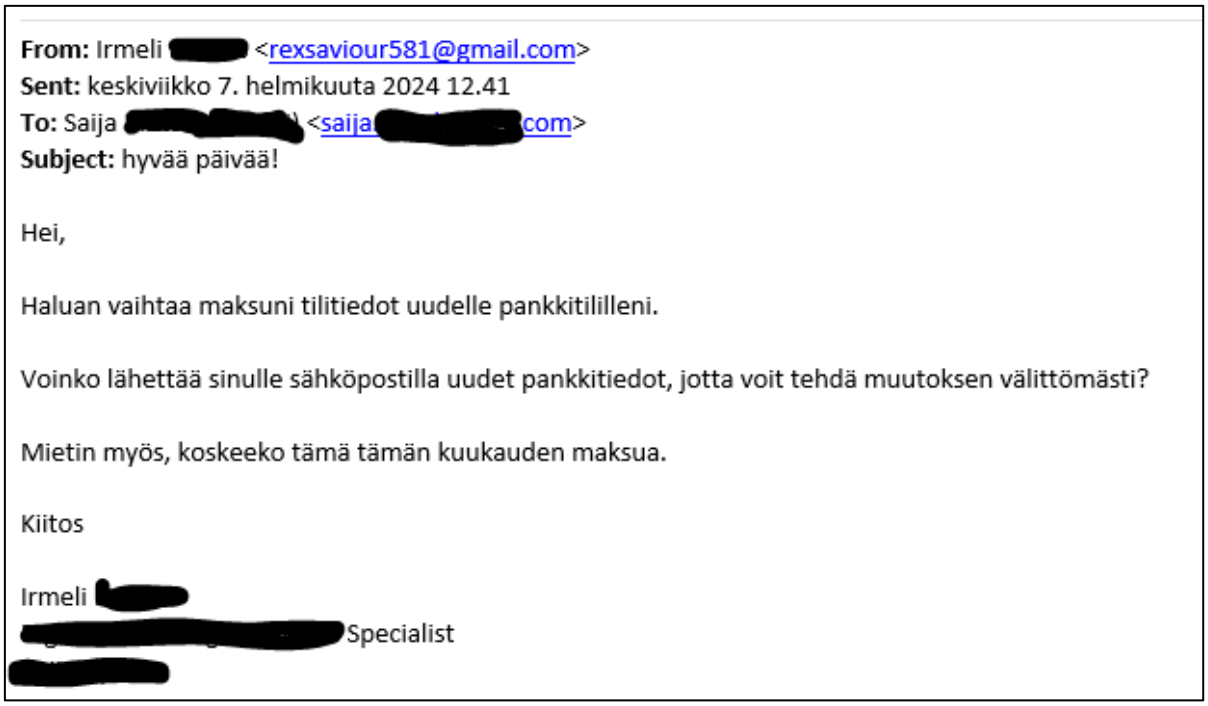
2.2.3 Identiteettivarkaus

Identiteettivarkaudella tarkoitetaan tapahtumaa, jossa esiinnytään toisena ihmisenä käyttäen uhrin henkilötietoja tai muuta vastaavaa yksilöivää henkilötietoa ja näin harhautetaan tarkoituksellisesti kolmatta osapuolta. Tällaisia tilanteita ovat esimerkiksi esiintyminen toisena ihmisenä sosiaalisessa mediassa tai tekemällä uhrin nimissä kauppoja tai tilauksia. (Kyberturvallisuuskeskus, n.d.-a)

Vuonna 2024 ihmiset viettävät aikaa keskimäärin kahdeksan tuntia päivässä verkossa jakaen henkilökohtaisia tietoja useiden eri käyttäjätilien välillä. Koska suurin osa näistä tiedoista on internetissä, tietomurron sattuessa riski niiden paljastumisesta kasvaa. Usein tämä on identiteettivarkauden ensimmäinen askel. Suomen rikoslain mukaan identiteettivarkaus on rangaistavaa vasta kun se aiheuttaa taloudellista tai muuta vahinkoa uhrille. Tästä syystä identiteettivarkaudesta on tullut Suomessa ja monissa muissa maissa suuri kiistanalainenkysymys. Tämän seurauksena myös identiteettivarkauden ennaltaehkäisemisestä on tullut ensiarvoisen tärkeää. (F-Secure, n.d.-b)

Kuva 2 on tyypillinen avoimesta lähteestä etsitty tieto (OSINT, Open source Intelligence), jota käytetään pohjana kohdistettaessa yritykseen tietojenkalastelua ja identiteettivarkauden yritystä. Kuva on otettu globaalin yrityksen sähköpostijärjestelmästä, mutta yritys ei halua työssä näkyvän nimeään. Vastaavia löytyy usealta yritykseltä. Sosiaalinen alusta LinkedIn tarjoaa tarkat tiedot tittelistä sekä nopealla haulla löytyy vastaavan yrityksen henkilöstöhallinon tai talousosaston vastaava, johon kohdistetaan huijausviesti. Massaviesteinä tehtaillut yritykset saattavat mennä läpi erityisesti, jos kyseessä on yrityksen päässä uusi henkilö tai kesälomasijainen.

Kuva 2. Identiteettivarkauden yritys henkilöstöhallintoon.



From: Irmeli [redacted] <rexsaviour581@gmail.com>
Sent: keskiviikko 7. helmikuuta 2024 12.41
To: Saija [redacted] <saija[redacted]com>
Subject: hyvää päivää!

Hei,

Haluan vaihtaa maksuni tilitiedot uudelle pankkitililleni.

Voinko lähettää sinulle sähköpostilla uudet pankkitiedot, jotta voit tehdä muutoksen välittömästi?

Mietin myös, koskeeko tämä tämän kuukauden maksua.

Kiitos

Irmeli [redacted]
[redacted] Specialist
[redacted]

2.2.4 Informaatiovaikuttaminen

Valtionhallinnon kanslian viestintäsuosituksessa ja tehostetun viestinnän ohjeessa informaatiovaikuttamisella tarkoitetaan toimintaa, jolla pyritään järjestelmällisesti vaikuttamaan yleiseen mielipiteeseen, ihmisten käyttäytymiseen ja päätöksentekijöihin sekä sitä kautta yhteiskunnan toimintakykyyn. Vaikuttamisen keinoja ovat esimerkiksi väärien tai harhaanjohtavien tietojen levittäminen ja painostaminen sekä sinänsä oikean tiedon tarkoitushakuinen käyttö. Usein kyse voi olla strategisesta toiminnasta, jossa tavoitellaan kohteen toiminnan muuttamista itselleen haitalliseksi tai vastoin omaa etuaan.

Voidaan siis todeta informaationvaikuttamisen olevan sellaista haitallista viestintää, jolla pyritään ohjaamaan yksilön tai ryhmän toimintaa. Vaikuttajataho saattaa olla valtiollinen

toimija, joka toimii tietoisesti tai ei-valtiollinen, joka toimii tiedostamattaan jonkun tahon lukuun. Myös yksilöt, erilaiset organisaatiot, järjestöt tai yhteenliittymät voivat toimia vaikuttajana tai olla vaikuttamisen kohteena. (Valtioneuvoston kanslia, 2019)

2.2.5 Kiristyshaittaohjelma eli ransomware

Kiristyshaittaohjelma on haittaohjelma, jonka tarkoituksena on salata osa uhrin laitteen tiedostoista tai lukita koko laite ja pyytää uhrilta lunnasrahat tiedostojen tai laitteen uudelleen avaamiseksi. Tällaisen ohjelman levitys tapahtuu samoin kuin minkä tahansa haittaohjelman eli sähköpostin liitteenä tai rikollisten huijaussivuston kautta. Kiristyshaittaohjelma on yksi eniten häiriötä aiheuttavista haittaohjelmista. Rikolliset pyytävät lunnaat yleensä bitcoinkryptovaluuttana, koska tämä tekee rikollisten jäljittämisen ja kiinni jäämisen vaikeammaksi. Lunnaiden suuruus voi vaihdella muutamista sadoista euroista tähtitieteellisiin summiin. (F-Secure, n.d.-c)

Kiristyshaittaohjelmat voidaan jakaa kahteen ryhmään: kiristysohjelmiin, jotka salaavat uhrin tiedostoja sekä kiristyshaittaohjelmiin, jotka lukitsevat uhrin laitteen. Jos kyseessä on tiedostoja salaava kiristyshaittaohjelma, uhrin laite on normaalisti käytettävissä muuten paitsi salattuihin tiedostoihin ei pääse käsiksi. Kiristyshaittaohjelma voidaan myös liittää tietojenkalasteluhyökkäyksen sähköpostin liitteeksi. Valitettavasti lunnaiden maksaminen ei välttämättä takaa, että uhri saa laitteensa tai tiedostonsa takaisin käyttöön. Tämän vuoksi esimerkiksi Yhdysvaltain liittovaltion poliisi eli FBI ei suosittele lunnaiden maksamista, vaikka se esimerkiksi yritysten kohdalla voi tuntua ainoalta vaihtoehdolta. Lisäksi lunnaiden maksaminen rahoittaa verkkorikollisuutta. Joutuessaan tällaisen uhriksi poliisi suosittelee aina tekemään rikosilmoituksen viranomaisille (F-Secure, n.d.-c). Taulukko 2 on esimerkkejä tunnetuimmista kiristyshaittaohjelmista.

Taulukko 2. Historian tunnetuimmat kiristyshaittaohjelmat. (F-Secure, n.d.-c)

Vuosi	Kiristyshaittaohjelma	Tietoja kiristyshaittaohjelmasta
2016	Petya	Petya tartutti maailmanlaajuisesti miljoonia tietokoneita. Petya lukitsi uhrien laitteet ja vaati lunnaita Bitcoin-kryptovaluuttana. Myöhemmin Petyasta tuli variantti NotPetya.
2017	WannaCry	WannaCry tartutti yli 200 000 tietokonetta ympäri maailmaa. Uhreilta vaadittiin lunnaita Bitcoin-kryptovaluuttana. WannaCry oli tietokoneesta toiseen leviävä mato, jonka kohteina oli myös organisaatioita ja muita isoja kohteita. WannaCry jäljitettiin lopulta Pohjois-Koreaan.
2018	Ryuk	Ryuk lukitsi uhriensa tiedostoja ja ne sai takaisin maksamalla lunnaat Bitcoin-kryptovaluuttana. Ryuk oli suunnattu erityisesti suuriin organisaatioihin, joilla oli mahdollisuus maksaa isommat lunnaat. Ryukin arvioidaan saaneen aikaan yli 61miljoonan dollarin vahingot.

Kiristyshaittaohjelma palveluna eli ransomware-as-a-service (RaaS) konseptilla tarkoitetaan, että kiristyshaittaohjelman kehittäjä myy kehittämäänsä kiristyshaittaohjelmaa ja ostaja voi

käyttää ohjelmaa rahan kiristämiseen uhrilta. Tämän palvelun ansioista tahot, joilla ei ole tietoteknistä osaamista tehdä kiristyshaittaohjelmia, voivat tehdä kiristyshaittaohjelma-hyökkäyksiä. (F-Secure, n.d.-c)

2.2.6 Kuvilla ja videoilla kiristäminen

Kuvilla ja videoilla kiristäminen eli sextortion tarkoittaa tapahtumaa, jossa huijari väittää kuvanneensa salaa uhria, kun tämä on käynyt aikuisviihdesivustoilla. Huijari väittää kuvaamisen tapahtuneen laitteeseen asennetun haittaohjelman kautta. Tarkoitus on saada uhri häpeän tunteeseen ja sen kautta maksamaan rahaa huijarille. Mahdollista toki on, että huijari on saanutkin kuvattua uhria, mutta voi myös olla, ettei kuvia tai videoita todellisuudessa olekaan. Uhri ei kuitenkaan voi olla tästä varma. Kyberturvallisuuskeskus ei ole saanut näyttöä kuvista tai videoista ja neuvookin olemaan maksamatta lunnaita. (Traficom, 2023a)

Suomeenkin on rantautunut vuosia sitten ilmiö, jossa sosiaalisen median alustoilla tai treffipalveluissa jaetuilla intiimeillä kuvilla tai videoilla on ryhdytty kiristämään julkaisijaa. Julkaisija ja kiristäjä ovat tutustuneet sosiaalisen media alustalla tai treffipalvelussa ja tutustumisen edetessä on lähetetty vastapuolelle intiimejä kuvia tai videoita. Tämä on tallentanut tai nauhoittanut videot ja ryhtynyt kiristämään uhria haltuunsa saamalla materiaalilla, vaatien uhria maksamaan rahaa tai ostokortteja. Jos uhri ei suostu maksamaan vaadittuja lunnaita, kiristäjä on uhannut julkaisevansa materiaalit uhrin sosiaalisen median kanavilla. Näistä on saatettu tehdä myös syvähuijaus eli deepfake-video siten, että uhrin kasvot on liitetty johonkin aiemmin kuvattuun videoon. Tällä tuotoksella on sitten ryhdytty kiristämään uhrilta rahaa ja uhattu videon julkaisemisella, jos uhri ei maksa lunnaita. Suomessa kiristyssummat ovat olleet muutamasta kympistä yli tuhanteen euroon. Poliisi varoittaa, että nettituttavuus saattaa myöhemmin osoittautua joksikin ihan muuksi henkilöksi, jona hän on ensin esiintynyt. Poliisi neuvoo tekemään rikosilmoituksen asiasta ja kertoo ettei kiristäjälle kannata maksaa, koska se ei takaa sitä, että kiristäjä ei julkaise hallussaan olevaa materiaali tai lopeta kiristämistä. Poliisin rikostutkintaa vaikeuttaa näissä jutuissa se, että kiristäjät eivät toimi omalla nimellään tai kuvallaan ja toimivat usein ulkomailta käsin. (Poliisi, 2024).

2.2.7 Spoofing-hyökkäykset

Spoofing-hyökkäys tarkoittaa tapahtumaa, jossa hyökkääjä väärentää henkilöllisyytensä ja esiintyy toisena henkilönä tai antaa valheellisia tietoja verkossa huijaten uhria. Spoofing-huijauksille on tyypillistä myös huijaussivustot sekä hyökkääjän GPS-sijainnin väärentäminen. Spoofing-hyökkäyksien tavoite on yleensä esiintyä jonain luotettavana tahona huijaten uhria uskomaan, että hän on tekemisissä jonkun luotettavan tahon kanssa. Tavoitteena näille hyökkäyksille on saada uhrilta varastettua arkaluontoista tietoa, rahaa tai salasanoja. Myös haittaohjelmien levitys kuuluu spoofing-hyökkäyksiin. Spoofing-hyökkäykset muistuttavat tietojenkalasteluhyökkäyksiä. Ne eroavat kuitenkin siinä, että spoofing-hyökkäysten tavoite on esiintyä jonakin toisena henkilönä ja huijata uhria väärennetyillä tiedoilla, kun taas tietojenkalasteluun kuuluun sosiaalinen manipulointi ja uhrin tietojen varastaminen. Seuraavassa käydään läpi erilaisia spoofing-hyökkäysen muotoja. (F-Secure, n.d.-d)

Sähköposti-spoofingissa hyökkääjä väärentää lähettäjän henkilöllisyyden ja sähköpostiosoitteen. Tämä ilmenee siten, että lähettäjän osoitteesta puuttuu kirjaimia tai lähettäjää ei ole edes olemassa. Myös viestin otsikosta tai sisällöstä saattaa myös löytyä kirjoitusvirheitä, koska ne on usein tehty automaattikääntäjällä. (F-Secure, n.d.-d)

IP-spoofingissa hyökkääjä väärentää IP-osoitteensa. IP-osoitetta käytetään internetissä laitteiden yksilöimiseen. Tässä hyökkäyksessä hyökkääjä väärentää todellisen IP-osoitteensa sijaintinsa salaamiseksi. Tämä mahdollistaa palomuurien ohittamisen ja murtautumisen heikosti suojattuihin verkkoihin. (F-Secure, n.d.-d)

URL-spoofingissa hyökkääjä luo huijaussivuston, jonka URL-osoite näyttää aidolta ja jonka ulkoasu on mahdollisimman aidonnäköinen ja sivusto muutenkin mahdollisimman identtinen oikean sivuston kanssa. Tavoitteena hyökkäyksellä on saada uhrilta käyttäjätunnus ja salasana oikealle sivustolle. (F-Secure, n.d.-d)

Caller ID-spoofingissa hyökkääjä yrittää väärentää soittajatunnuksensa eli puhelinnumeron, josta hän soittaa, koska ihmiset eivät vastaa niin helposti tuntemattomasta numerosta tulleisiin puheluihin. Tämä onnistuu VoIP-teknologian (Voice over Internet Protocol) avulla. Esimerkiksi pankit tai muut viralliset tahot eivät koskaan soita ja kysy arkaluontoista tietoa puhelimessa. Samaista toimintaa rikolliset voivat harrastaa myös tekstiviestien välityksellä. (F-Secure, n.d.-d)

DNS-spoofingissa hyökkääjä muokkaa nimipalvelujärjestelmää (Domain Name System). DNS-järjestelmä ohjaa käyttäjät oikealla sivulle URL-osoitteen perusteella. DNS-järjestelmän

muokkaaminen on vaikeaa, mutta sitä on myös vaikea havaita onnistuessaan. (F-Secure, n.d.-d)

2.2.8 Social engineering

Social engineering termille ei ole yhtä vakiintunutta suomennosta. Siitä käytetään usein suomennoksia sosiaalinen manipulointi, vaikuttaminen, käyttäjän manipulointi tai hämääminen. Puhuttaessa social engineeringistä tarkoitetaan sellaista toimintaa, jolla uhrilta huijataan rahaa tai häntä huijataan paljastamaan arkaluonteisia asioita, kuten käyttäjätunnuksia, salasanoja tai muita vastaavia tietoja tai muutoin toimimaan hyökkääjän tahdon mukaisesti. Huijaaminen tapahtuu yleensä pitämällä uhriin yhteyttä sähköpostiviestein tai puhelimella ja käyttämällä hyväksi uhrin hyväntahtoisuutta. Huijarit ottavat yleensä kohteeksi jonkin suuren yrityksen työntekijän ja pyrkivät näin saamaan uhrin antamaan pääsyn yrityksen tietoihin, tietojärjestelmiin tai muuhun omaisuuteen. (F-Secure, n.d.-d)

Käyttäjän manipulointiin perustuvat hyökkäykset toteutetaan yleensä siten, että ensin tunnistetaan kohde ja kerätään kohteesta tietoa. Tämän jälkeen siirrytään itse manipulointiin eli lähestytään kohdetta sähköpostilla tai puhelimella väärennettyä henkilöllisyyttä tai keksittyä tarinaa käyttäen siten, että uhri luulee olevansa tekemisissä luotettavan tahon kanssa. Keinona voi olla myös deepfake-teknologialla tehty videopuhelu tai soitto. Seuraavaksi pyritään saavuttamaan uhrin luottamus ja toteutetaan samalla itse hyökkäys. Kun hyökkäys on toteutettu, siivotaan jäljet, jotta hyökkääjän jäljille on mahdollisimman vaikea päästä. Yleensä näissä hyökkäyksissä hyökkääjä vetoaa kiireeseen, jotta uhri ei ehdi ajatella tekojensa seurauksia. Social engineering hyökkäys on yleensä tarkasti suunniteltu etukäteen ja silloin on jo kerätty tarkkoja tietoja uhrista, hänen työstään ja yrityksestään. (F-Secure, n.d.-d)

Hyökkääjällä on usein keksittynä jokin uskottava tarina huijauksen varmistamiseksi. Tällainen huijaus yleensä edellyttää sitä, että uhri luulee olevansa tekemisissä luotettavan tahon, esimerkiksi työkaverin tai esihenkilön kanssa. Näin hyökkääjällä on paremmat mahdollisuudet saada huijaus onnistumaan ja uhri luovuttamaan tietoja, klikkaamaan linkkiä tai luovuttamaan rahaa. Tällaista huijausta kutsutaan englanniksi nimeltä ”pretexting”. Myös houkuttelu eli englanniksi baiting on yksi keino saada huijaus onnistumaan. Tällaisessa huijauksessa hyökkääjä on toimittanut CD- tai DVD-levyn tai mahdollisesti USB-muistitikun julkiselle paikalle esimerkiksi yrityksen tiloihin, josta se huomataan helposti ja luottaa siihen, että uhri on tarpeeksi utelias tarkistaakseen mitä media sisältää. Hyökkääjä voi lisäksi lisätä mediaan jonkun kuvan tai logon, jotta se ei herätä epäilyjä. Tällaisen median avulla

hyökkääjä yrittää saada toimitettua esimerkiksi jonkin haittaohjelman uhrin tai yrityksen laitteeseen. (F-Secure, n.d.-d)

2.2.9 WhatsApp-huijaukset

WhatsApp on maailman suosituin viestisovellus. Sillä on maailmanlaajuisesti käyttäjiä yli 2 miljardia. Tämän vuoksi WhatsApp soveltuu erinomaisen hyvin huijauksien alustaksi sen saavuttaessa helposti ja nopeasti massoittain ihmisiä. Tämän vuoksi WhatsApp-huijaukset ovatkin tyypillistä massarikollisuutta. Hyvin yleinen huijaus WhatsAppilla on niin kutsuttu Posti-huijausviesti. Huijauksessa Postin nimissä lähetetään johonkin tekaistuun pakettiin tai sen toimitukseen tai toimitusongelmaan liittyvä viesti ja pyydetään käyttäjää avaamaan saapunut linkki. WhatsAppilla huijarit lähettävät myös päivityshuijauksia, joissa kerrotaan, että sovellus pitää päivittää tai että se vanhenee. Todellisuudessa WhatsApp ei lähetä päivityksistä tietoja tekstiviestein eikä niitä ladata mitään linkkiä käyttämällä vaan sovellus päivitetään asetuksista tai sovelluskaupasta. Päivitys kannattaa tehdä aina kun mahdollista laitteen suojaamiseksi (Augustėnė, 2023). Vastaavia huijauksia voidaan tehdä käyttämällä muita vastaavia alustoja kuten Telegramia. Telegram ja WhatsApp viestien lähettäminen on maksutonta, joten massoittain lähetettävistä viesteistä ei aiheudu mitään kuluja.

Identiteettivarkauden pohjana käytettävä WhatsApp-koodihuijaus tapahtuu siten, että uhri saa vahvistusviestin, joka lähetetään normaalisti, kun luodaan uutta tiliä tai yritetään kirjautua uudella laitteella. Seuraavaksi huijari lähettää uhrille tutulta käyttäjältä viestin, jossa hän kertoo pyytäneensä vahvistuskoodin vahingossa uhrin numeroon oman numeronsa sijaan ja pyytään lähettämään viestin hänelle. Tällaisessa tapauksessa huijari on luultavasti kaapannut uhrille tutun käyttäjän tilin ja yrittää nyt tehdä samoin uhrin tilille. (Augustėnė, 2023)

”Hei äiti” -huijauksessa huijari esiintyy uhrin läheisenä, yleensä uhrin lapsena ja kertoo että hänen puhelimensa on rikki ja siksi hän lähettää viestin kaverin numerosta. Yleensä huijari kertoo, että puhelin on rikki, jonka vuoksi hän ei pääse pankkisovellukseen ja pyytää äitiä lähettämään hänelle pikaisesti rahaa johonkin tekaistuun tarkoitukseen (Augustėnė, 2023). Vastaava huijaus voidaan tehdä deepfake-huijauksena äänen avulla.

Deittisovellus-kryptohuijaus on huijaus, joka alkaa yleensä deittisovelluksesta, josta keskustelu siirretään WhatsAppiin. Seuraavaksi huijari kertoo jonkun kaverinsa vaurastuneen kryptovaluutalla ja kertoo, että jos uhria kiinnostaa tämä hänen kannattaa siirtää rahaa huijatulle kryptosivustolle. Tämän jälkeen huijari katoaa ja estää uhrin WhatsAppissa. (Augustėnė, 2023)

WhatsAppin yritystilihuijaus on WhatsAppin huijaus, jossa huijari on luonut tilin, joka vaikuttaa viralliselta WhatsAppin yritystililtä. Tällä tavoin huijari saa lisää uskottavuutta huijauksiinsa. (Augustėnė, 2023)

2.3 Kyberrikollisuuden rikostutkinta

Yleisimpiä tietoverkkoihin kohdistuneita rikoksia ovat tietomurrot. Poliisille ilmoitettiin tietomurtoja ja törkeitä tietomurtoja yhteensä 1939 kappaletta vuonna 2022. Toukokuun loppuun mennessä vuonna 2023 tietomurtoja oli ilmoitettu poliisille 907 kappaletta. Tietomurtojen määrä on kasvanut tasaisesti viimeisten vuosien aikana eikä määrän arvioida kääntyvän laskuun. (Limnell, 2019)

Yle uutisoi jo vuonna 2022, että poliisi saa selvitettyä vain pienen osan tietomurroista. Vuodesta 2012 lähtien tietomurtojen selvitysaste on laskenut tasaisesti. Selvitysprosentti oli ollut vuonna 2021 vain 4 %. Vastaavasti samalla ajanjaksolla oli raportoitujen tietomurtojen määrä kasvanut lähes eksponentiaalisesti. Keskusrikospoliisin kyberrikostorjuntakeskuksen päällikkö, rikostarkastaja Mikko Rauhamaa kertoo Ylen uutisessa selvitysprosentin pienuuden johtuvan kyberturvallisuuden osaajien vähyydestä ja toisaalta monien rikosten juontavan juurensa Suomen ulkopuolelle, jolloin tutkinta monimutkaistuu ja hidastuu. (Mäntysalo, 2022)

Vaikka uutisissa kerrotaan suuriin yrityksiin kohdistuneista tietomurroista, Mikko Rauhamaan mukaan ne ovat vain jäävuoren huippu ja valtaosa tietomurroista kohdistuu tavallisiin kansalaisiin. Sosiaalisten tilien salasanojen päätyminen rikollisten käsiin ja sitä kautta tilin tietojen väärinkäyttö on Rauhamaan mukaan yksi yleisimmistä tietomurtotyypeistä. (Mäntysalo, 2022)

Identiteettivarkaudet ovat yleensä oheis- tai valmistelevaa rikollisuutta ja niitä käytetään monesti tietomurtojen ja petosten toteuttamisessa. Identiteettivarkauksia ilmoitettiin poliisille vuonna 2022 yli 3700 kappaletta. Toukokuun loppuun mennessä vuonna 2023 identiteettivarkauksia oli ilmoitettu yli 1700 kappaletta (Limnell, 2019). Yhdysvaltain kauppakomission mukaan identiteettivarkauden kasvoivat maassa vuosina 2012–2022 270 000:sta 1,4miljoonaan kappaleeseen. Tämä tarkoittaa siis 518 % kasvua. Asiantuntijat kertovat, että identiteettivarkauksista ilmoitetaan harvoin, joko tietämättömyyden, hämmennyksen tai todisteiden puutteen vuoksi. Tämän vuoksi organisaation uskovat, että identiteettivarkauksien todellinen määrä on huomattavasti suurempi kuin viralliset tilastot osoittavat. (F-Secure, n.d.-b)

Viestintä on hyökkäyksen kohteena, kun puhutaan tietoliikenteen häirinnästä. Tällöin puhutaan yleensä palvelunestohyökkäyksistä. Palvelunestohyökkäysten eri tekomuotoja yhdessä tietomurtojen kanssa tuli poliisin tietoon vuonna 2022 yhteensä 2093 kappaletta. Tietojärjestelmä on kohteena, kun puhutaan tietojärjestelmän häirinnästä. Vuonna 2022 tietojärjestelmän häirinnän eri tekomuotoihin liittyviä rikoksia tuli poliisin tietoon 42 kappaletta. (Limnell, 2019)

2.4 Deepfake-huijaukset

Deepfake voidaan kääntää suomeksi ”syväväärennös”, mutta suomen kielessä käytetään niin termiä deepfake kuin syväväärennöskin. Termi deepfake tulee englannin kielen sanoista deep learning ja fake eli syväoppiminen ja väärennös. Näin ollen syväväärennös on osuva käännös. Deepfake-huijauksella tarkoitetaan ääntä, kuvaa tai videota, jota on muokattu tekoälyä eli AI:ta käyttäen. (Zieniūtė, 2022)

Kyberturvallisuuskeskuksen erityisasiantuntija Tretjakovin mukaan suomenkielinen termi syväväärennös ei riittävän selkeästi kuvaa, minkälainen väärennös on kyseessä. Hän suosittelee käyttämään tarkentavasti joko äänivääärennös-, kuvamanipulaatio- tai videovääärennöstermejä selventämään, minkälaisesta deepfakesta on kysymys. Myös suomalaisessa mediassa on uutisia, joissa terminologialla avataan tarkemmin, minkälaisesta väärennöksestä on kysymys. Tätä tapaa Tretjakov pitää onnistuneena. (Autio ym, haastattelu, 30.8.2024)

Ensimäinen deepfake-teknologia ilmestyi 2017 Reddit-alustalla. Tuntematon käyttäjä jakoi algoritmin, jossa tekoälyn avulla pystyi luomaan todentuntuisia videoita (Organization for Social Media Safety, n.d.). Itse teknologia juontaa juurensa kuitenkin paljon kauemmas. GAN eli Generative Adversal Networks teknologialla kyetään syväoppimiseen pienellä datamäärällä. Neuroverkkoteknologia emuloi ihmisaivoja: suuren datamäärän prosessoinnin jälkeen ihmisaivot ennustavat, miltä muut samantyyppiset datat näyttävät. GAN teknologia perustuu kahteen algoritmiin, jossa toinen luo väärennöksen ja toinen pyrkii tunnistamaan sen. Tunnistamisen onnistuessa, algoritmi oppii tunnistamaan väärennöksen ja syöttää tämän tiedon toiselle algoritmille. Näin algoritmit syöttävät toisilleen uutta tietoa, joiden avulla sekä väärennökset että niiden tunnistaminen kehittyvät. Tätä prosessia kutsutaan syväoppimiseksi (Creswell ym., 2017). Kuvia ja videoita olemassa olemattomista ihmisistä voidaan siis luoda neuroverkkoteknologian avulla. GAN prosessi on verrattavissa taiteen tuntijaan ja taiteen väärentäjään, jossa taiteen väärentäjä pyrkii tekemään yhä parempia väärennöksiä ja taiteen tuntija pyrkii tunnistamaan aidot väärennöksistä. Molempien oppiessa samanaikaisesti ja tulemaan yhä paremmiksi. (Lanham, 2021, luku 1)

Deepfake-huijaukset ovat jatkuvasti kasvava huolenaihe, koska tekoäly teknologia kehittyy jatkuvasti. Voidaankin sanoa, kuten muissakin rikostyypeissä, rikolliset tehtailemassa deepfake-teknologialla huijauksia ja toisaalta kansalliset ja yksityiset toimijat kehittämässä suojautumiskeinoja, kilpailevat kiivaasti kumpi on edellä. (Zieniūtė, 2022) Kyberturvallisuuskeskus arvioi kuitenkin, että niin kauan kuin tekoälyn käyttö on kalliimpaa kuin pelkän geneerisen massaviestin lähetys, rikollisten saavuttama hyöty on pienempää ja näin ollen tekoälyllä tuotettujen yksityisiin henkilöidin kohdistuvien yritysten määrä on pieni. Kun tuotto-odotus on riittävän suuri, tekoälyn käyttö deepfakeissa myös yksittäisiin henkilöihin, on kannattavaa. Tästä esimerkkinä luvussa 2.5 kerrottu Arupin tapaus. (Autio ym, haastattelu, 30.8.2024)

Tekoälyllä tarkoitetaan älykkäästi toimivaa ohjelmaa tai järjestelmää. Tekoälyyn kuuluvat koneoppiminen sekä kuvien ja luonnollisen kielen prosessointi. Tekoäly voidaan jakaa heikkoon ja vahvaan tekoälyyn. Heikolla tekoälyllä tarkoitetaan ohjelmaa, jonka on mahdollista toimia rajoitetussa ympäristössä, joka sille on etukäteen määritetty. Vahvalla tekoälyllä taas tarkoitetaan ohjelmaa, joka kykenee kommunikoimaan, oppimaan, tekemään päätöksiä, ajattelemaan ja ratkomaan mahdollisia ongelmia. (Lanham, 2012, luku 1)

Koneoppiminen voidaan jakaa kolmeen ryhmään: ohjattu koneoppiminen, ohjaamaton koneoppiminen ja vahvistusoppiminen. Taulukko 3 on kuvattu nämä ryhmät sekä niiden eroavaisuudet. (Lanham, 2012, luku 1)

Taulukko 3. Koneoppimisen ryhmät ja niiden eroavaisuudet. (Lanham, 2012, luku 1)

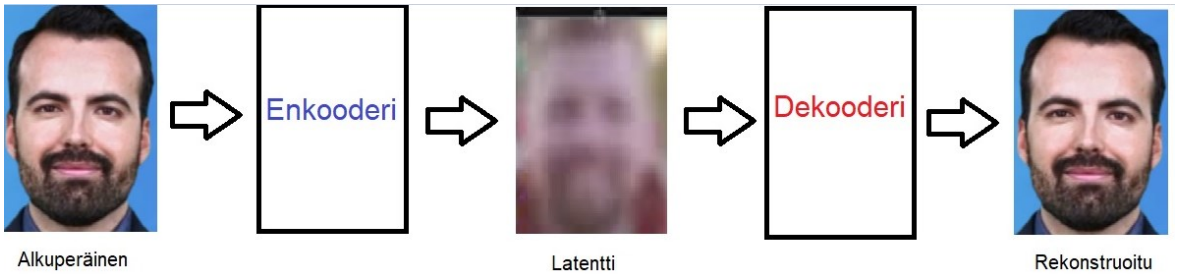
Koneoppimisen ryhmä	**Ryhmän kuvaus**
Ohjattu koneoppiminen	Syötetään koneoppimisohjelmalle valmiiksi luokiteltua materiaalia, jonka tavoitteena on opettaa ohjelma luokittelemaan dataa materiaalin perusteella.
Ohjaamaton koneoppiminen	Syötetään koneoppimisohjelmalle luokittelematonta materiaalia. Ohjelma jakaa dataa eri luokkiin samankaltaisuuksien perusteella ja tunnistaa poikkeamat. Oppimisen etuna on, ettei sitä ohjaa luokittelun painotukset.
Vahvistusoppiminen	Koneoppimisohjelma saa toimintansa perusteella positiivista tai negatiivista palautetta. Ohjelman tavoitteena on saada mahdollisimman paljon positiivista palautetta.

Syväoppiminen prosessoi ihmisaivojen tapaa prosessoida tietoa. Keinotekoisia neuroverkkoja käytetään tässä apuna. Tällaisessa tapauksessa kasvatetaan kerros kerrokselta järjestelmän informaation määrää opittavasta aiheesta. Syväoppiminen yhdistetään aivoihin ja tämä saattaa aiheuttaa epävarmuutta. Aivojen toiminta kuitenkin on paljon syväoppimista monimutkaisempaa. Ihmisaivoista voidaan kuitenkin ottaa syväoppimiselle malli yhteyksien luomisesta. (Lanham, 2021, luku 1)

2.4.1 Autoenkooderi

Kun deepfake-videota tehdään, koneoppiva neuroverkko muodostaa aidosta materiaalista mallin, joka sitten upotetaan toiseen videoon. Tällaista järjestelmää kutsutaan autoenkooderiksi. Autoenkooderin toisessa päässä on dekooderi ja toisessa päässä enkooderi ja näiden välissä latentti tila. Kun autoenkooderia käytetään, on se ohjaamatonta oppimista. Tässä mallissa kone purkaa datan alempaan muotoon ja rakentaa sen tämän jälkeen uudelleen. Kuva 3 on esitelty autoenkooderin toiminta. (Lanham, 2021, luku 2)

Kuva 3. Autoenkooderin toiminta (Lanham, 2021, luku 2)



Prosessi alkaa, kun enkooderille syötetään analysoitavaksi massoittain kasvokuvia. Enkooderin tehtävä on etsiä ja oppia kuvista samankaltaisuudet ja pelkistää kuvat. Se siis varastoi mahdollisimman paljon informaatiota kasvoista rajoitettuun tilaan. Tätä tilaa kutsutaan latentiksi tilaksi. Tärkeintä onkin, että enkooderi tunnistaa vaihtuvat ominaisuudet kuvista ja varastoi ne latenttiin tilaan, esimerkiksi katseen suunnan. Rajoitetun tilan vuoksi latenttiin tilaan ei ole tarkoitus varastoida kuvista samoina pysyviä ominaisuuksia, kuten silmien väriä. (Lanham, 2021, luku 2)

Toisen algoritmin eli dekooderin tehtävä on päinvastainen kuin enkooderin, eli sen on tarkoitus opetella löytämään enkooderin läpikäymästä datasta pysyvät kasvojen piirteet ja rekonstruoida ne takaisin. Jos käytetään kahta eri autoenkooderia kaksista eri kasvoista ja halutaan vaihtaa ne keskenään, tämä vaihto tehdään latentissa tilassa eli annetaankin toiselle dekooderille, joka on opetettu hakemaan eri henkilön kasvojen piirteet, rekonstruoitavaksi eri enkooderin latentti. Lopputuloksena on, että eri henkilön kasvot ja liikkeet rekonstruoidaan toisen alkuperäisen ihmisen kasvojen päälle. (Lanham, 2021, luku 2)

Autoenkooderit käyttävät valvomatonta koneoppimista rekonstruoidun kuvan lopputuloksen arviointiin, koska tavoitteena on saada mahdollisimman identtinen jäljennös alkuperäisestä materiaalista. Autoenkooderi siis vertaa jokaista pikseliä jäljennöksen ja alkuperäisen välillä. Neuronit, jotka ovat tuottaneet huonolaatuisia pikseleitä jäljitetään ja autoenkooderi muokkaa parametreja hieman seuraavalle iteraatiokierrokselle. (Lanham, 2021, luku 2)

Yksinkertaistettuna deepfake-videon luominen siis tapahtuu seuraavasti: teknologiana käytetään syväoppimisprosessia median analysointiin ja uudelleen luomiseen tekoälyjärjestelmissä. Jotta saadaan mahdollisimman uskottava lopputulos, on käytössä kaksi eri tekoälyverkostoa, jotka kilpailevat toisiaan vastaan. Näin lopputulos paranee ja oppiminen on syvempää. Käytännössä tekoälyllä luodun tai muokatun videon lähtökohta on usein tavallinen video, jossa esiintyy näyttelijä. Näyttelijä korvataan sitten toisella henkilöllä siten, että tekoäly katsoo lukemattomia kuvia toisesta henkilöstä tuottaen hänen piirteensä uskottavasti videoon. Kuvat henkilöstä yhdistetään videolla alun perin esiintyneen näyttelijän liikkeisiin ja eleisiin siten, että ilmeet ja huulten liikkeet synkronoidaan. Tässä käytetty teknologia ei ole science fictionin tarinoita vaan syväväärennettyjä videoita ja kuvia voi tehdä kuka tahansa. Freeware-sovelluksia on netistä saatavana useita. Näiden sovellusten tuotokset ovat usein huonolaatuisia, koska ne ovat ilmaisia mutta maksullisilla sovelluksilla jälki on huomattavasti vakuuttavampaa. (Zieniūtė, 2022)

2.4.2 Deepfake-huijauksien taustaa

Deepfake-huijaukset eivät koske pelkästään yksityishenkilöitä vaan enenevässä määrin myös kansallisia viranomaisia ja maiden turvallisuutta, maalle tärkeää infrastruktuuria sekä maiden hallituksia. Tekoälyn kehittyessä äänet, kuvat ja videot ovat erittäin aidon näköisiä ja vaativat edistyksellistä teknologiaa tunnistaa väärennökset. Deepfake-teknologiaa on käytetty useisiin eri tarkoituksiin. Esimerkkeinä sitä on käytetty mm. Hollywoodissa näyttelijöiden nuorentamiseen sekä toisessa ääripäässä propagandasodissa poliittisia vastustajia vastaan. Deepfake-teknologian avulla voidaan luoda aidolta näyttäviä, väärennettyjä videoita ja saada siinä esiintyvät ihmiset sanomaan tai tekemään asioita, joita ei he eivät todellisuudessa ole tehneet tai sanoneet. (Zieniūtė, 2022)

Deepfaken kohteena voi olla yksittäinen henkilö, kansalaisjoukko tai yritys; toistaiseksi yrityksiin kohdistuneet deepfaket ovat vielä vähemmistössä. Yksittäiseen henkilöön kohdistuvassa huijauksessa pyritään usein saamaan rahallista hyötyä nopeasti ennen kuin henkilö saa selville kyseessä olevan huijauksesta tai ehtii reagoimaan huijaukseen. Henkilöä erehdytetään viestin tulevan hänelle tutulta henkilöltä, jolloin ääni, kuva tai jopa video näyttää hänelle tutulta, joka tarvitsee apua. Kansalaisjoukon kohdalla on kyse mielipidevaikuttamisesta ja ohjaamisesta tiettyyn käytökseen. Usein näissä deepfakessa käytetään julkisuuden henkilöä tai poliitikkoa, joka nähdään mielipidevaikuttajana. Yrityksiin kohdistuvassa deepfakessa on lähes poikkeuksetta kyse tavoitteesta saada rahallista hyötyä. (Zieniūtė, 2022)

Deepfake-huijauksista tekee erityisen tehokkaita ja vaarallisia tekoälyjärjestelmien kyvykkyys luoda yksittäisten ihmisten ääniä vain muutaman sekunnin näytteestä. Näitä voidaan käyttää useassa väärässä tarkoituksessa: valeuutisissa, propagandassa kuin myös kostotarkoituksissa. Niillä voidaan häpäistä esimerkiksi poliitikkoja tai julkisuuden henkilöitä tai saada heidän avullaan uskoteltua ihmisjoukolle heidän tukevan tiettyä aatetta tai toimintaa. Väärennetty video leviää tehokkaasti, kun asiaan uskotaan ja jaetaan eteenpäin. (Zieniūtė, 2022)

Kaikki deepfakeilla tehdyt väärennykset eivät kuitenkaan täytä rikollisen toiminnan merkkejä. Hollywoodin studiossa vanhempia näyttelijöitä nähdään nuorennettuina versiona ja jo edesmenneitä näyttelijöitä on herätetty valkokankaalle henkiin. Tämä on laillinen tapa käyttää deepfake-teknologiaa eikä mitenkään rikollista toimintaa. (Zieniūtė, 2022)

Deepfake-video on yksi uusimpia ja pelottavimpia kyberrikollisuuden muotoja. Rikollisen videon tarkoituksena on saada kohdehenkilö toimimaan epärationaalisesti tai pakottaa toimimaan vastoin tahtoaan. Rikollisen tavoitteen mukaan tarkoituksena voi olla valheellisen tiedon levittäminen tai kohdehenkilön maineen tuhoaminen. Deepfake-videon tarkoituksena voi olla myös harhauttaa päättävissä asemissa olevia henkilöitä toimimaan yritykselle tai heille itselleen epäedullisella tavalla. Laittomaan ja rikolliseen toimintaan liittyen deepfake-videoita onkin käytetty myös esimerkiksi väärennetyn pornografisen sisällön tuottamiseen ja levittämiseen. Usein tämä on tehty siten, että näyttelijän tilalle videoon on rakennettu esiintyjäksi joku kuuluisuuden henkilö. Vaikka henkilö kieltää olevansa videolla, on hän joutunut epäedulliseen valoon tai kiristyksen kohteeksi. Deepfake-videoita voivat myös käyttää hallitukset esimerkiksi vihollisina pitämiensä tahojen mustamaalaamiseen tai propagandan levittämiseen. Deepfake-videoita voidaan siis käyttää täysin laillisesti, mutta myös laittomiin ja rikollisiin tarkoituksiin. (Zieniūtė, 2022)

2.5 Esimerkkejä Deepfake-huijauksista

Toistaiseksi maailman suurin julkisuuteen tullut deepfake-huijaus tuli julki helmikuussa 2024, kun Hongkongin poliisi kertoi, että lontoolaisen insinööritoimisto Arupin Hongkongin toimipiste on joutunut deepfake-huijauksen kohteeksi. Arup on suunnitellut muun muassa Sydneyn oopperatalon sekä Manchesterissa sijaitsevan Etihad-stadionin. Yritys työllistää yli 18 000 työntekijää ja sillä on toimisteitä ympäri maailmaa yhteensä 34. Rikolliset järjestivät videopuhelun, johon osallistui vain yksi Arupin työntekijä. Videopuhelussa onnistuttiin keinotekoisesti luomaan tapaaminen, johon oli väärennetty Arupin johtajien osallistuminen deepfake-teknologialla. Tämä näyttäytyi aidolle osallistujalle siten, että rikolliset olivat luoneet videon ja äänen avulla johtajat tapaamiseen. Näin työntekijä saatiin huijattua uskomaan, että

yrityksen johtajat ovat oikeasti mukana tapaamisessa. Tapaamisesta oli myös etukäteen kerrottu huijatuksi tulleelle työntekijälle siten, että tapaaminen on salainen ja siitä ei saa kertoa kenellekään tapaamisen aiheen ollessa suuret yrityskaupat. Työntekijä uskoi tämän, koska ei tunnistanut johtajien videoita väärennöksiksi. Näin huijarit manipuloivat työntekijän siirtämään suuria summia rahaa useassa erässä rikollisille. Kaikkiaan rahaa siirtyi rikollisille 23 miljoonaa euroa 15 eri rahan siirrolla, viidelle eri tilille. Arup on kertonut, että kyseessä oli puhdas rahanhuijaus eikä sen sisäiset järjestelmät vaarantuneet. Ennen onnistunutta huijausta huijarit olivat yrittäneet samaa taktiikkaa useaan muuhun Arupin työntekijään kuitenkaan siinä onnistumatta. Hongkongin poliisi pidätti kuusi ihmistä tapaukseen liittyen. (Tekniikan Maailma, 2024)

Myös Suomessa tuli julkisuuteen samoihin aikoihin deepfake-huijaus, kun silloisesta presidenttiehdokkaasta ja nykyisestä Suomen presidentistä Alexander Stubbista tehtiin deepfake-video. Videolla luotu Stubb-hahmo kehuu automaattisen kaupankäynnin sovellusta, joka on yhdistetty sijoitushuijauksiin. Oikeasti presidentti Stubb ei ole mukana videolla eivätkä mitkään videon sanomiset ole peräisin hänen suustaan. Tämän videon oikeita tarkoitusperiä voidaan vain arvailla. Uskooko videon tekijä suomalaisten uskovan siihen, että Stubb olisi rikastunut kyseisellä sovelluksella ja saisi näin huijattua ihmisiä käyttämään sovellusta? Vai onko kyseessä yritys tehdä lovi presidentin ja sitä kautta koko Suomen valtion maineeseen ja uskottavuuteen? Vai oliko kenties tarkoitus vaikuttaa kansalaisten vaalikäyttäytymiseen ja presidentinvaalien tulokseen? Video on pyörinyt Facebookissa ja Presidentinkanslia onkin vaatinut Facebookia poistamaan videon siinä kuitenkaan onnistumatta. (Korhonen, 2024)

Mediaväärennösten käyttö juuri vaalivaikuttamisessa ja disinformaation jakamisessa ovat tunnistettuja ilmiöitä, joita on tarpeen seurata myös Suomessa herkällä korvalla. Erityisesti vaalien yhteydessä tuotettujen deepfake-medioiden osalta Kyberturvallisuuskeskus tekeekin yhdessä Oikeusministeriön kanssa kansainvälistä yhteistyötä. Yhteistyössä vaihdetaan kokemuksia ja tietoja muiden maiden viranomaisten kanssa sekä keskustellaan laajasti myös median kanssa valemedian käytöstä osana informaatiovaikuttamista (Autio ym, haastattelu, 30.8.2024). Myös keskusrikospoliisi kertoo, että keskeisenä osana heidänkin työssään kyberrikosten selvittämisessä on kansainvälinen yhteistyö. (Keskusrikospoliisin kyberrikostorjuntayksikkö, 2024)

Myös Suomen entistä presidenttiä Sauli Niinistöä on käytetty deepfake-videon lähteenä. Tämä video toteutettiin vuonna 2019, kun YLE testasi silloisia ohjelmistoja deepfake-videoiden luomiseen. Deepfake-videota varten Jarkko Tamminen imitoi presidentti Niinistön uudenvuodenpuhetta ja Tammisen kasvot korvattiin presidentti Niinistön kasvoilla. Video on

suhteellisen huonolaatuinen ja koska siinä ei käytetty Niinistön ääntä, myös äänestä voi videon tunnistaa deepfake-väärennökseksi. (Yle, 2019)

Kesällä 2024 tehtiin myös toisesta entisestä Suomen presidentistä Urho Kekkosesta deepfake-video. Videolla Kekkonen kertoo, kuinka tekoälyä voidaan käyttää väärin. Videon onkin tarkoitus olla varoittava esimerkki siitä, mitä tekoälyllä luodun materiaalin mahdollisuudet ja vaarat ovat ja kuinka kehittyneitä deepfake-videot nykyään ovat. (Järveläinen, 2024)

Tämän kaltaiset deepfake-huijaukset ovat lisääntyneet räjähdysmäisesti vuonna 2023. Financial Times uutisoi, että yksin Britanniassa varastettiin noin 50 miljoonaa puntaa alkuvuonna 2023 esittämällä yritysjohtajia, pankkivirkailijoita ja poliiseja. Varmennuspalveluyhtiö Regula kertoo, että äänihuijauksista on raportoinut 37 % ja videohuijauksista 29 % yrityksistä vuonna 2023. (Haajanen & Tuominen, 2024)

Kotimaan mediassa on ollut useitakin uutisia, jossa tekoälyn avulla on luotu ihmishahmoja ja näille sosiaalisen median tilejä. Monessa tilissä kerrotaan suoraan hahmon olevan tekoälyn luoma eikä henkilö ole oikeasti olemassa. Tästä yksi esimerkki on Ylen vuonna 2023 julkaisema artikkeli 24-vuotiaasta Milla-Sofiasta, joka oli täysin tekoälyn luoma hahmo. Taustaselityksestä huolimatta monet Milla-Sofian tilin seuraajista halusivat puhua hänen kanssaan yksin tai jopa tavata. Milla-Sofian takana oli 44-vuotias mies, joka oli luonut hänet yrityksensä mainoskasvoksi (Kymäläinen, 2023). Voi vain arvailla kuinka monta vastaavaa tekoälyllä luotua tiliä on olemassa, jossa seuraajat kuvittelevat henkilön olevan oikea ihminen ja orastavan suhteen olevan aluillaan.

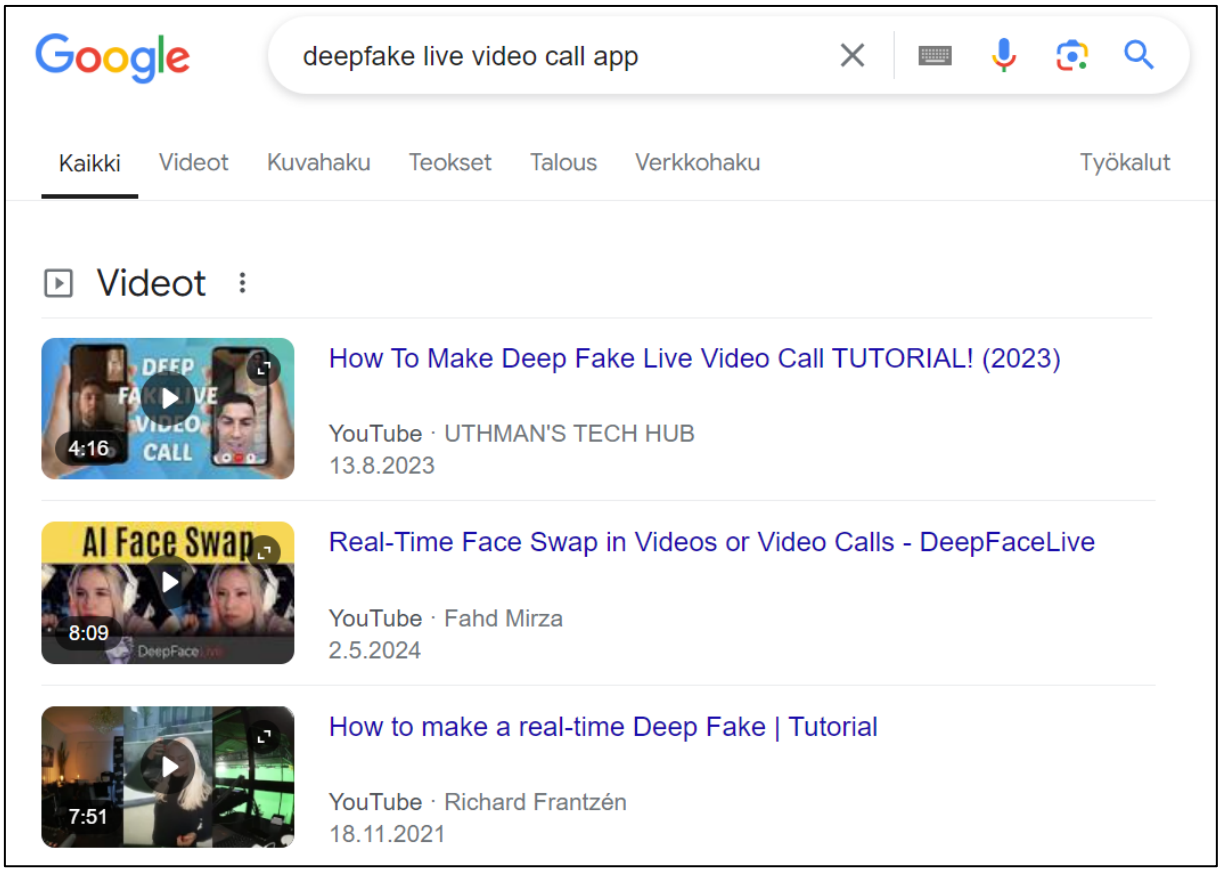
2.6 Kyberrikoksissa deepfake-huijaukset kasvamassa

Mihin sitten yritysten tulisi varautua tulevaisuudessa? Mihin tavalliset kansalaiset voivat luottaa? Kyberrikokset ovat kasvava rikollisuuden laji ja deepfaket kehittyessään entisestään sekä onnistuessaan, ennen näkemättömän tehokkaita. Ennen deefaken tekemistä ja onnistumista pohjalla on joku muu kyberrikostyyppi. Arupin huijausyritykseen oli tehty ennen toteutusta runsaasti työtä. Huijarit olivat todennäköisesti muun muassa etsineet avoimen lähteen tietoa (OSINT), etsineet julkisesti (tai sisäisesti) saatavilla olevaa video- ja tai äänimateriaalia henkilöistä, joista luotiin deepfake-videot Teams-keskusteluun. Koska jälkikäteen kerrottiin rikollisten kokeilleen samaa ensin toiseen Arupin henkilöön, oli työ tehty useampaan kertaan. Ymmärrettävistä syistä Arup ei ole kertonut julkisuuteen kaikkea selvitystyössä löydettyä, jotta viranomaiset ovat saaneet työrauhan. Hyvin todennäköistä on, että ennen onnistumista on tehty social engineeringiä ja luotu luottamusta kohdehenkilöön.

Tätä on mahdollisesti tehty useampaan henkilöön yhtä aikaa, jotta joku tarttuisi syöttiin. Tarkoituksena on ollut luottamuksen saavuttaminen, jotta Teams-kutsu varsinaiseen kokoukseen ei herättäisi kysymyksiä. Deepfaken käyttö on identiteettivarkauden yritys ja onnistuessaan identiteettivarkaus. (Tekniikan Maailma, 2024)

Yksityinen kansalainen on vaarassa joutua deepfaken kohteeksi vuonna 2024 luvussa 2.2.9 mainitun ”Hei, äiti” -huijauksen kautta, jossa käytetään ääntä. Tässäkin tapauksessa on kyseessä pohjalla ensin social engineering, jotta oman lapsen ääni voidaan yhdistää huijausviestiin. Jo tämänhetkisellä teknologialla pystytään luomaan keinotekoisesti reaaliaikainen deepfake-video, joten todennäköisesti ”Hei äiti” -huijauksissa on jo nähty tai tullaan näkemään äänen jälkeen myös mahdollinen video tai videopuhelu. Aiemmat WhatsApp-tekstiviestihuijaukset ovat kehittyneet ensin ääneksi ja seuraavaksi käyttämään videota huijauksen pohjalla. Kuva 4 näkyy Google-haun tulos vapaasti saatavilla olevista ohjeista ja applikaatioista reaaliaikaisten deepfakejen tekemiseen.

Kuva 4. Ilmaiseksi saatavilla olevat ohjeet reaaliaikaiseen deepfake-videoon.



Toinen yksityisiä ihmisiä koskettava huijaustyyppi luvussa 2.2.4 on informaatiovaikuttaminen, jossa deepfaken avulla oli luotu silloisesta presidenttiehdokas Alexander Stubbista videoväärennös. Video on levinnyt suurille massoille sitä jaettaessa erilaisten

tekstiviestipalveluiden ja sosiaalisten alustojen avulla, ja sen pysäyttäminen on lähes mahdotonta, vaikka se alkuperäisestä lähteestä poistettaisiinkin. (Korhonen, 2024)

3 Puolustautumis- ja suojautumismenetelmät

Yritysten johtohenkilöt saattavat ajatella esimerkiksi ”Miksi kukaan meidän yritykseemme hyökkäisi, ei meillä ole mitään varastamisen arvoista” tai vaikka että ”Suomen kieli on niin outo ja maa niin pieni, ettei tänne kukaan hyökkää”. Valitettavan usein yritykset laiminlyövät kyberturvallisuuden ajatellen, ettei yrityksellä ole mitään varastamisen arvoista. Valitettavasti hyökkääjä usein ajattelee, että kaikilla yrityksillä on jotain varastamisen arvoista. Tämä siksi, että kaikilla on jotain dataa ja kaikenlainen data on pimeillä markkinoilla arvokasta.

Miten sitten voidaan suojautua kyberhyökkäyksiä vastaan? Tässä kohtaa voidaan tarkastella keinoja yrityksen ja toisaalta yksityisen henkilön kannalta. Yrityksessä kohteena voivat olla yrityksen järjestelmät, toimintaympäristö ja toisaalta yrityksen henkilöstö. Yksityisen henkilön suojautuminen noudattaa pääasiassa samoja polkuja kuin yrityksessä henkilöstön suojautuminen. Tässä kappaleessa käydään läpi puolustautumis- ja suojautumismenetelmiä, joita on tullut sekä EU:n että Suomen tasolta. Käydään läpi myös niitä menetelmiä, joita yritykset ja yksityiset ihmiset voivat käyttää halutessaan suojautuakseen kyberrikoksia vastaan.

3.1 NIS – Network and Information System ja ISO 27001 -standardi

Tässä luvussa käydään läpi NIS 1 -tietoturvadirektiivin eroja tulevaan NIS 2 -tietoturvadirektiiviin sekä katsotaan vaatimuksia, joita NIS 2 -direktiivi asettaa yrityksille. Luvussa arvioidaan myös ISO 27001 -standardia sekä sen vaikutuksia erityisesti deepfake-huijauksia vastaan. Lisäksi pohditaan, luodaanko NIS 2 -direktiivin ja ISO 27001 -standardin kaltaisilla säädöksillä kattava turva kyberrikollisuutta ja erityisesti deepfakeja vastaan.

Massarikollisuus ja kasvava kyberrikollisuus ovat tuoneet yrityksille kasvaneita kuluja onnistuneiden huijausten kautta. Tässä luvussa katsotaan, minkälainen kehitys näillä kustannuksilla on ollut viime vuosina sekä mihin mittasuhteisiin asiantuntijat arvioivat niiden kasvavan tulevina vuosina.

3.1.1 NIS 2 -tietoturvadirektiivi tulee voimaan lokakuussa 2024

NIS eli Euroopan Unionin verkko- ja tietoturvadirektiivi, NIS (Network and Information System), on koko Euroopan unionin alueen yhteinen säädöstö, jolla pyritään varmistamaan direktiivin mukainen yhtenäinen turvallisuustaso etenkin kyberturvallisuuden alueella. Ensimmäinen NIS direktiivi tuli voimaan 2016 ja vuonna 2024 kansallisen lainsäädännön mukaisesti tulee jäsenvaltioiden implementoida NIS 2 -direktiivi. NIS direktiivi (2016) koskee ensisijaisesti yhteiskunnallisesti kriittisten infrastruktuurien toimijoiden sekä tarjoajien velvollisuuksia ilmoittaa poikkeamista tietoturvahäiriöiden osalta. (Enisa, n.d.)

Vuonna 2024 lokakuussa voimaan tuleva direktiivi tarkoittaa käytännössä monelle suomalaisellekin yritykselle perinpohjaista nykyisen tietoturvallisuusarkkitehtuurin katselmointia, riskipaikkojen selvittämistä ja analysointia sekä varautumista hyökkäyksestä toipumiseen. Suuret ja keskisuuret yritykset, jotka toimivat muun muassa energian, liikenteen, rahoitusmarkkinoiden, terveydenhuollon, vesihuollon sekä muiden yhteiskunnan kriittisillä alueilla toimivat alat, ovat NIS 2 -direktiivin kohteena. (Traficom, 2023b)

Euroopan parlamentin julkaisussa helmikuulta 2023 todettiin kyberhyökkäysten olevan nopeimmin kasvava rikollisuuden lähde maailmanlaajuisesti. Kyberrikosten todettiin kasvavan mittakaavassaan pienistä valtavan kokoisiksi, vaikuttaen merkittävästi kustannuksiin sekä samalla niiden muuttaneen muotoaan entistä korkealaatuisemmiksi. Julkaisussa viitattiin tutkimukseen vuodelta 2021, jossa oli arvioitu hyökkäysten aiheuttamien kustannusten nousevan 57 kertaisiksi vuoteen 2015 verrattuna. Kasvun Yhdysvaltojen dollareissa arvioitiin kasvavan jopa 20 miljardiin vuositasolla. Kiristyshaittaohjelmien (ransomware) hyökkäysten määrän arvioitiin keskimäärin kohdistuvan joka 11. sekunti vuonna 2021 kun viisi vuotta aiemmin määrä oli ollut joka 40. sekunti. Eurobarometrin tutkimuksen mukaan samassa julkaisussa oli 76 % vastaajista arvellut riskin joutua onnistuneen kyberhyökkäyksen uhriksi kasvaneen. (European Parliament, 2023)

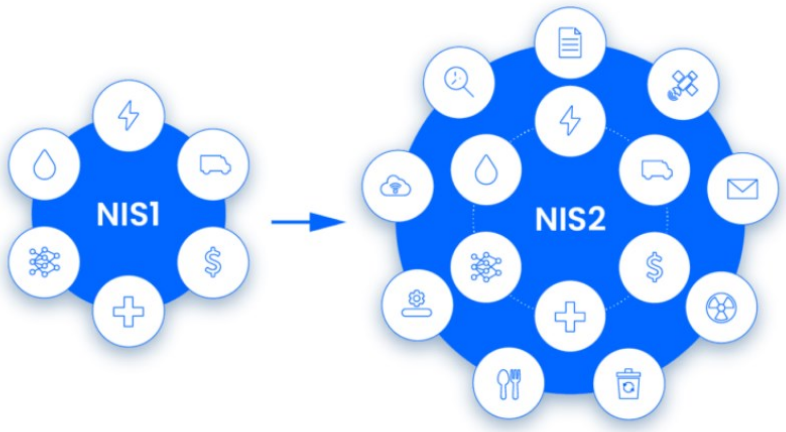
Suomessakin vuosi 2023 oli kiireinen huijareilla. Finanssialan julkaisun mukaan huijauksia yritettiin tehdä yhteensä Suomessa 76,9 miljoonan euron edestä ja pankit pystyivät torjumaan 32,7 miljoonan euron edestä näitä. Pelkästään verkkorikollisille nousua oli edellisestä vuodesta tapahtunut 36 %. Pankkien estämien ja palauttamien maksujen määrä oli noussut 132 %, joten pankkien järjestelmät ovat selkeästi kehittyneet vuoden aikana. Artikkelissa Finanssiala ry:n petos- ja rikostorjunnasta vastaava johtaja kuitenkin kertoo, että vaikka pankkien järjestelmät paranevat, rikolliset kehittävät samaa vauhtia uusia huijauskeinoja (Palmgren, 2024). Huijausten kehittyessä deepfakella tehty puhelu tai muu yhteydenotto pankin nimissä on jo tätä päivää.

Kyberrikollisuuden muuttuessa massiiviseksi rikollisuuden muodoksi, massarikollisuudeksi, käyttävät yritykset sekä yksityiset henkilöt vuosittain valtavat summat suojautumiseen. Koronapandemia lisäsi merkittävästi digitalisaation käyttöä sekä yrityksissä että yksityisissä talouksissa. Etätöiden myötä yritykset siirsivät työntekijöitään yrityksen oman verkon ulkopuolelle ja tarjosivat erilaisia ratkaisuja liittyä yritysten verkkoon. Yksityiset henkilöt lisäsivät mm. verkko-ostosten määrää sekä digitaaliset kanssakäymiset lisääntyivät erilaisten fyysisten rajoitteiden ollessa voimassa. Tämä on omalta osaltaan lisännyt alustaa kyberrikollisuudelle. (Fortune Business Insights, 2024)

Kyberturvallisuusmarkkinoiden kasvua on vaikea ennustaa. Useat ekonomistit ovat kuitenkin arvioineet markkinoiden erittäin nopeaa kasvua myös tuleville vuosille. Kyberuhkien muuttuessa yhä älykkäämmiksi ja regulaation lisääntyessä, yritysten kustannukset uhkien torjumiseksi nousevat. Fortune Busienss Insights arvioi pelkästään globaaleiden kyberturvallisuus markkinoiden kasvavan 2,5 kertaiseksi vuoden 2023 172 miljardin Yhdysvaltojen dollarin tasosta 423 miljardiin vuoteen 2030 mennessä. Kun markkinoiden vuotuinen kasvuvauhti samalle ajalle on lähes 14 %, on mahdollisesti maltillisempikin kasvu vielä hurjaa. (Fortune Business Insights, 2024)

Alkuperäisessä direktiivissä on määritelty seitsemän eri kriittistä, yhteiskunnan kannalta keskeistä toimijaa energian tuotannosta vesijärjestelmiin ja logistiikasta terveydenhuoltoon (Traficom, 2023b). Tänä vuonna voimaan tuleva NIS 2 -direktiivi korvaa aiemman NIS direktiivin sekä samalla laajentaa aiemman direktiivin kohdeorganisaatioita. Aiempien seitsemän kohteen lisäksi on otettu kahdeksan muuta keskeistä toimijaa nostaen kohteiden määrän 15:sta. NIS 2 -direktiivin myötä tulevat voimaan (neljä vaatimusaluetta) myös tiukemmat vaatimukset riskien hallinnalle, raportoinnille poikkeamien osalta, yritysvastuulle kyberturvallisuuden alueelta sekä liiketoiminnan jatkuvuuden varmistaminen mahdollisen kyberhyökkäyksen sattuessa. Direktiivi määrittelee myös aiempaa kovemmat rangaistukset direktiivin noudattamatta jättämiselle. Direktiivin voimaantulo koskettaa satoja tuhansia yrityksiä Euroopassa (NIS 2 -Directive, n.d.). Kuva 5 on esitetty NIS 1 -direktiivin muutos tulevaan NIS 2 -direktiiviin.

Kuva 5. NIS 2 -direktiivi verrattuna NIS 1 -direktiiviin. (NIS 2 -Directive, n.d.)



Seuraavaksi tarkastellaan Kuva 5 näkyvien merkkien selityksiä.

Taulukko 4 on esitetty Kuva 5 näkyvien NIS 1 ja NIS 2 -direktiivien merkkien selitykset eli sektorit, joita direktiivit koskevat sekä näiden sektorien sisältö.

Taulukko 4. NIS-direktiivejä koskevat sektorit sekä niiden sisältö. (NIS 2 -Directive, n.d.)

NIS1 sektorit	Sisältää	NIS2 sektorit	Sisältää
Energia-ala	Sähkö, öljy, kaasu, kaukolämpö ja vety	**Julkishallinto**	Sosiaalipalvelut, yleinen turvallisuus, taloussääntely sekä poliittinen edustus
Terveydenhuoltoala	Terveyspalvelut sekä lääkinnälliset laitteet	**Digitaalisten palvelujentarjoajat**	Hakukoneet, verkkokaupat ja sosiaaliset mediat
Kuljetus ja liikenne	Ilma-, rautatie-, maantie- ja vesiliikenne	**Postipalvelut**	Posti- ja kuriiripalvelut
Rahoitusala	Pankki- ja rahoitusmarkkinat sekä niiden infrastruktuuri	**Jätehuoltopalvelut**	Jätteiden kerääminen, kuljettaminen, käsittely ja hävittäminen
Vesihuolto	Juoma- ja jätevesi	**Avaruus**	Avaruuteen liittyvä tietoliikenne, navigointi ja kansallinen turvallisuus
Digitaalinen infrastruktuuri	Telecom, DNS, TLD, datakeskukset, luottamuspalvelut ja pilvipalvelut	**Ruokapalvelut**	Viljely, ruuan tuotanto, pakkaus, kuljetus ja vähittäismyynti
		Valmistuspalvelut	Lääkinnällisten laitteiden, tietokoneiden, elektro-niikan, koneiden ja laitteiden, moottoriajoneuvojen ja muiden kuljetusvälineiden valmistus
		Kemikaalit	Petrokemikaalit, polymeerit, perus epäorgaaniset aineet, erikoistuotteet ja kulutus kemikaalit
		Tutkimuspalvelut	Erilaiset tutkimukset, jotka voivat sisältää arkaluontoisia tietoja

Vaikka direktiivin pääasiallinen tarkoitus on kyberturvallisuuden varmistaminen, se sisältää myös fyysiseen ja ympäristöturvallisuuteen liittyviä tekijöitä. Yritysten tulee neljän vaatimusalueen lisäksi toteuttaa määritellyt perusturvatoimenpiteet torjuakseen kybertuvallisuus muodot. Taulukko 5 on lueteltu nämä toimenpiteet.

Taulukko 5. Yrityksiltä vaaditut perusturvatoimenpiteet NIS 2:ssa (NIS 2 -Directive, n.d.).

Minimivaateet toimenpiteille
Riskiarvioinnit ja turvallisuus -politiikat tietojärjestelmille
Politiikat ja menettelytavat, joilla varmistetaan turvatoimien tehokkuus
Kryptografian, salauksen käyttöön liittyvät käytännöt ja menettelytavat
Suunnitelma turvallisuuspoikkeamien käsittelyyn
Tietoturvaan liittyvien hankintojen sekä niihin liittyvän kehittämisen haavoittuvuuksien raportointi ja hallinta
Kyberturvallisuus sekä yleinen tietoturvallisuus koulutus
Tietoturvapolitiikka ja menettely arkaluonteisiin tietoihin pääsevien henkilöiden osalta. Yleiskuva kaikista mahdollisista vaaroista tältä alueelta sekä menettelytapa varmistamaan prosessin toimivuus
Suunnitelma yritystoiminnan jatkuvuuteen sekä mahdollisen turvapoikkeaman aikana että sen jälkeen.
Monivaihetodennuksen, jatkuvan todennuksen, äänen, videon, tekstin salausta sekä salattu hätäviestintä yrityksen sisällä.
Yrityksen toimitusketjujen turvallisuuden varmistaminen sekä tarvittavat varmistukset suoriin yrityksen ja toimittajan välisiin kanssakäymisiin.

NIS 2 luo selkeän linjan tavoitteille EU tasoisesti valmistautuessa kyberturvallisuus uhkiin. Direktiivillä pystytään entistä paremmin kasvattamaan sekä yritysten että yksittäisten kansalaisten ymmärrystä kyberrikollisuudesta sekä antamaan keinoja tunnistaa vaaranpaikkoja sekä ehkäistä kyberrikollisuuden uhriksi joutumista. Toinen direktiivin tuoma edistysaskel on yhteistyö Euroopan unionin sisällä sekä yritysten että kansallisella tasolla. Tieto kulkee nopeammin ja selkeämmin mahdollisten riskien tai hyökkäysten osalta ja tätä kautta myös puolustautuminen ja toiminnan jatkaminen paranevat. Kolmantena positiivisena asiana NIS 2 -direktiivissä on perusteltu syy yrityksille tarkistaa ”oma pesä” ja sekä ymmärtää, että pieni maa ja vaikea kieli eivät ole syitä jättää varautumatta kyberuhkiin. Näitä

kaikkia vauhdittavat NIS 2 -direktiivin sanktiot noudattamatta jättämiselle sekä yrityksen johdon henkilökohtainen vastuu direktiivin implementoinnin laiminlyönnistä. (European Parliament, 2023) Kyberturvallisuuskeskus arvioi, että kansallinen ja sitä myöden myös kansainvälinen tilannekuva parantuu nykyisestä, koska NIS2 piirissä olevien yritysten raportointivelvollisuus laajenee valvovan viranomaisen suuntaan. (Autio ym, haastattelu, 30.8.2024)

3.1.2 ISO 27001 -standardi

Tietoturvallisuuden alueelle on oma ISO standardi; ISO 27001. Standardi on vuodelta 2022 ja sen kohteena on Information Security Management Systems (ISMS) eli tietoturvallisuuden hallintajärjestelmät. Standardi määrittelee minimivaateet, jotka yrityksen tietoturvallisuuden tulee pitää sisällään. (ISO, n.d.)

Kyberriskien hallinta voi tuntua monesta yrityksestä erittäin vaikealta tai jopa mahdottomalta tehtävältä. Rikollisten käyttämä teknologia on kuitenkin edellä puolustautumiskeinoja ja sen vuoksi on tärkeää luoda yritykselle selkeä pohja myös kyberrikollisuutta estämään. ISO/IEC 27001 luo pohjan organisaatiolle riskien tiedostamiseen, tunnistamiseen sekä ennalta tunnistamaan heikkouksia ja korjaamaan niitä. ISO 27001 -standardi lähestyy tietoturvaa kokonaisvaltaisesti: se tuo yhteen ihmiset, käytännöt ja teknologian. Standardi ei kuitenkaan anna turvaa itsessään; se vaatii systemaattista lähestymistä, jatkuvaa päivittämistä sekä koko henkilöstön ymmärrystä standardin ulkopuolelta kokonaiskuvasta. Kuten muutkin ISO standardit myös ISO 27001 antaa peruspohjan tekemiselle. Standardin mukaisen toiminnan ja sertifioinnin myötä ei voida kuitenkaan taata, että yritys olisi resilientti tietoturvauhille.

3.1.3 ISO 27001 ja NIS 2 turvaamassa yrityksiä deepfake-huijauksia vastaan

ISO 27001 ja NIS 2 -direktiivin tavoitetila on yritysten suuntaan sama: yrityksillä on perusedellytykset turvata oma tietoturvallisuusarkkitehtuurinsa, kouluttaa henkilöstönsä riittävästi tunnistamaan ja estämään mahdolliset tietoturvaloukkaukset sekä kyberturvallisuushyökkäyksen sattuessa, jatkaa toimintaansa mahdollisimman pienin vaurioin. Voidaankin sanoa, että NIS 2 -direktiivi ja ISO 27001 sertifiointi ovat kivijalkana suojautumiskeinoja mietittäessä yrityksille.

Suojaavatko NIS 2 -direktiivi implementoituina tai ISO 27001 -standardi sitten deepfake-huijauksia vastaan? Tähän voidaan vastata kyllä ja ei. Direktiivi ja standardi luovat perusteet valtiollisella tasolla ja osana yrityksen tietoturvapolitiikkaa, -teknologiaa ja -prosesseja kattavan kokonaisuuden, jossa teknologia ja osaaminen voidaan rakentaa turvaamaan

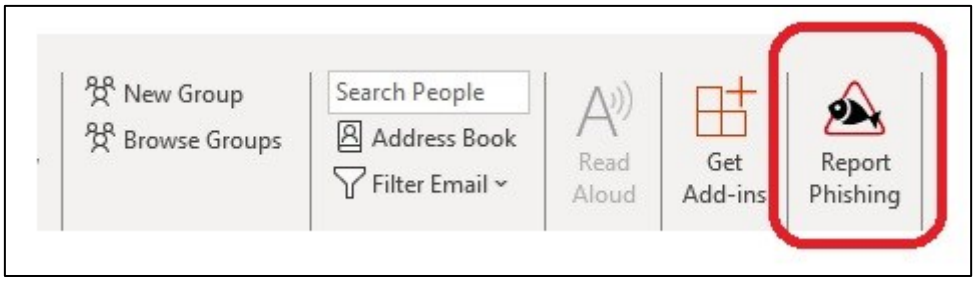
deepfake-huijauksia vastaan. Suoraan ne eivät nykyisessä muodossaan tarjoa riittävän yksityiskohtaisia vaateita tai perusteita rakentaa kyberturvallisuuden osalta järjestelmää tai prosesseja, joilla estetään deepfake-huijaukset kohdistuen yrityksiin tai yksityisiin kansalaisiin.

Myös Kyberturvallisuuskeskuksen asiantuntijat toteavat sääntelyn kehittyvän hieman jälkijunassa. Kyberturvallisuuden alueella olevaa lainsäädäntöä on jo olemassa, mutta maltillisesti. Heidän arvionsa mukaan EU:n tekoälyasetus ja mahdollisesti myös digipalvelusäädös lisäävät säätelyä sekä vahvistavat läpinäkyvyyttä. Vaikka mediamanipulaatioiden osalta on materiaali selkeästi merkittävä keinotekoisesti luoduksi, jää nähtäväksi, kuinka hyvin palvelun tarjoajat tätä tulevat noudattamaan. (Autio ym, haastattelu, 30.8.2024)

Euroopan parlamentin direktiivissä (EU) 2022/2555 artiklassa 21 luetellaan kymmenen kyberturvallisuuden riskien hallinnan osa-aluetta, joiden on oltava rakennettaessa kyberturvallisuuskokonaisuutta fokuksessa. Taulukko 5 oli nämä kokonaisuudet listattu. Näistä lähes jokaiseen kohtaan voidaan rakentaa suojautumis- ja puolustautumismekanismit nimenomaan deepfakeja vastaan. Kohdassa d) käsitellään toimitusketjun turvallisuutta, jotka sisältävät turvallisuuteen liittyvät näkökohdat eri toimijoiden välillä. Mikäli yrityksen tai toimijan toimintakokonaisuus on vaarassa deepfake-huijausten vuoksi, on tarpeellista arvioida osana turvallisuusanalyysiä ja -auditointia tätä kokonaisuutta ja rakentaa siihen turvarakenteet. Kohdassa g) viitataan yleiseen kyberturvallisuushygieniaan sekä koulutukseen kyberturvallisuuden alueella. (Euroopan Parlamentti, 2024)

Monet yritykset kouluttavat jo nyt henkilöstöään IT-osaston satunnaisesti generoimilla huijausviesteillä. Vastaanottajan tulisi tunnistaa väärennetty viesti aidosta ja tunnistaessaan, ilmoittaa kalasteluyrityksestä sähköpostijärjestelmän kautta. Saman järjestelmän kautta voidaan ilmoittaa tietenkin myös aito huijausviesti. Kuva 6 on Outlook sähköpostijärjestelmään rakennetun tietojenkalastelusta ilmoittava kuvake. Tätä kuvaketta painettaessa lähtee ilmoitusviesti mahdollisesta huijausviestistä yrityksen IT-osastolle tutkittavaksi.

Kuva 6. Outlookiin rakennettu kuvake tietojenkalastelusta ilmoittamiseen.



Vuonna 2024 suurin osa koulutuksista tähtää kalastelusähköpostien tunnistamiseen, jotka muodostavat suuren osan yrityksiin kohdistuvista ja suojauksista läpi menevistä huijauksista. Huijauslinkin takaa voi löytyä sitten muuta materiaalia, jolla yritetään huijausta – vaikka deepfake-video.

Iso-Britannian hallitus julkaisee vuosittain kyberturvallisuusrikkomuksista raportin. Vuoden 2024 raportista käy ilmi, että vain keskimäärin 18 % yrityksistä on viimeisen vuoden aikana järjestänyt koulutusta tai tapahtumia, joissa kasvatetaan tietoa kybertuvallisuudesta (UK Government, 2024). Vuonna 2023 toteutetussa tutkimuksessa määrä oli sama. Vuoden 2023 raportissa todettiin niiden yritysten kasvattaneen koulutusmääriään edelliseen vuoteen verrattuna, joilla oli joko itsellä tai yrityksellä samalla alueella tapahtunut kyberturvallisuusrikos tai sellaista oli yritetty (UK Government, 2023). On huomattava, että kyberrikollisuus on kuitenkin vuoden aikana kasvanut merkittävästi kuten aikaisemmin on todettu. Näin ollen voidaan todeta, että NIS 2 -direktiivin voimaantulo tarkoittaa väistämättä yritysten tarvetta lisätä kyberturvallisuuteen liittyvän koulutuksen lisäämistä. Koulutuskokonaisuuksien sisällön rakentamiseen ei direktiivi ota kantaa, mutta deepfake-huijausten määrän ja vakavuuden kasvaessa, sen pitäisi kuulua koulutuskokonaisuuteen.

NIS 2 -direktiivin artiklassa kymmenen kohdassa h) viitataan yrityksen politiikkoihin ja toimintatapoihin liittyen kryptografiaan sekä tarvittaessa myös salaukseen. Blockchain- eli lohkoketju-teknologiaa hyödyntäen, jossa digitaalisen sisällön varmistamiseen käytetään kryptografiaa ja julkisia avaimia todentamiseen, pidetään vuonna 2024 yhtenä lupaavimpana teknologiana deepfaken tunnistamiseen (Struck, 2024). Tämä tarkoittaisi kuitenkin yrityksille valtavaa harppausta kryptografian käyttöönoton alueella. Direktiivi luo siihen kuitenkin mahdollisuuden sekä paineen.

3.2 Cybercrime Exit

Koska verkkorikoksia tekevien keski-ikä on laskussa, poliisi aloitti vuonna 2020 Keskusrikospoliisin Kyberrikostorjuntakeskuksessa Cybercrime Exit -hankkeen. Hankkeen tarkoituksena on ennaltaehkäistä nuorten tietoverkkorikollisuutta. Kohderyhmänä on 12–25-vuotiaat nuoret, jotka ovat syyllistyneet tietoverkkorikoksiin tai ovat vaarassa syyllistyä. Nuorten verkkorikoskierre alkaa yleensä lievillä rikoksilla, mutta muuttuvat ajan myötä vakavammiksi rikoksiksi. Nuorten tekemä kyberrikollisuus on kasvanut ja usein nuoret eivät itse tiedosta laillisen ja laittoman toiminnan rajoja, vaan syyllistyvät tietämättään rikoksiin. (Limnell, 2019)

Hankkeen pääpaino on rikoksissa, jotka kohdistuvat tietoverkkoon tai tietojärjestelmiin. Hankkeen toiminta perustuukin kyberrikosten ennaltaehkäisyyn, rikoskierteen katkaisuun sekä rikoksista irtaantumiseen. Hanke kannustaa nuoria ohjaamaan tietojärjestelmä- ja tietoverkko-osaamisensa erilaisiin opiskelu- ja työmahdollisuuksiin. Nuoret voivat hakeutua ohjelmaan itse tai ammattilaisen ohjaamana. (Limnell, 2019)

Cybercrime Exit -hanke on erinomainen esimerkki toimintatavasta, jossa rikoksia pyritään selvittämisen sijaan ehkäisemään ennalta. Tämä vähentää luonnollisesti tapahtuvien rikosten määrää ja lisää samalla kansalaisten ymmärrystä kyberríkollisuudesta. (Limnell, 2019)

3.3 The Joint Cybercrime Action Taskforce (J-CAT)

Suomi liittyi vuonna 2023 Europolin kansainväliseen yhteistyöryhmään The Joint Cybercrime Action Taskforceen eli J-CAT:iin, jonka tarkoitus tutkia ja estää kyberrikollisuutta. Vuonna 2014 Europoliin perustetussa ryhmässä on mukana 19 maata EU:sta ja sen ulkopuolelta. Koska tietoverkkorikollisuus on lähes poikkeuksetta valtioiden rajat ylittävää kansainvälistä rikollisuutta, sitä pystytään tehokkaammin torjumaan reaaliaikaisella kansainvälisellä viranomaisyhteistyöllä. Tämä edellyttää nopeaa lainvalvontaviranomaisten välistä yhteistyötä ja tiedonvaihtoa. J-CAT tarjoaa mahdollisuuden tällaiseen toimintaan. Yhteistyöryhmä tekee kyberrikosuhkia ja -kohteita vastaan tiedusteluun pohjautuvia toimenpiteitä. J-CAT torjuu rajat ylittäviä tietoverkkorikoksia. Tällaisia ovat esimerkiksi lasten seksuaalinen hyväksikäyttö, maksuvälinepetokset sekä muut tietoverkkoavusteiset rikokset. Ryhmä myös puuttuu pimeän verkon rikosten mahdollistamiseen. Torjuntakohteet valitaan jäsenmaiden ehdottamien tapausten pohjalta ja ryhmä myös tukee jäsenmaita tapausten tutkinnassa. Ryhmä toteutti 25 kyberoperaatiota vuonna 2021. (Poliisi, 2022)

3.4 Suojautuminen kyberrikollisuutta vastaan

Kyberrikollisuuden estämiseksi paras tapa on ehdottomasti ennaltaehkäisy. Poliisin Cybercrime Exit ja The Joint Cybercrime Action Taskforce -hankkeet ovat onnistuessaan tehokkaita ehkäisemään uusien kyberrikollisten syntyä. Hyvistä hankkeista huolimatta kyberrikollisuutta on ja tulee valitettavasti tulevaisuudessakin olemaan kasvavassa määrin. Kyberrikosten levittyä valtioiden välisiksi rikoksiksi, kansallista verkostoitumista ja yhteisiä toimia on ollut tarve rakentaa myös suojautumis- ja puolustautumisrintamalla.

Fortune Business Insightin mukaan yritykset tulevat keskittymään seuraavina vuosina erityisesti koneoppimiseen, esineiden internettiin eli Internet of Things:iin, pilviteknologiaan sekä Big Dataan rakentaessaan verkkojensa turvajärjestelmiä uusiksi. Näiden teknologioiden avulla järjestelmä kykenee automaattisesti huomaamaan erikoisia toimintoja sekä mahdollisesti tunnistamaan uhat. Toinen kasvava trendi osana kyberturvallisuuden kasvua on pilvipalveluiden tarpeen kasvuun reagointi. Suuret, globaalit palveluntarjoajat kuten IBM ja Cisco Systems kehittävät kyberturvallisuusratkaisuja pilvipalveluihin. AaaS eli Analytics as a Service tyyppisillä ratkaisuilla yritykset pyrkivät varmistamaan riittävän analytiikkakapasiteetin sekä tarjoamaan teknologian pilven kautta. (Fortune Business Insights, 2024) Kyberturvallisuuskeskuksen asiantuntijoiden mukaan samoilla alustoilla ja algoritmeilla, joilla deepfakella tuotettuja medioita jaetaan, voitaisiin valjastaa myös manipuloiden tuotetun sisällön tunnistamiseen. Tekniikka on tältä osin kuitenkin vielä kehitysasteella. (Autio ym, haastattelu, 30.8.2024)

Yritysmaailman lisäksi myös valtiolliset tahot ovat tutkineet erilaisia teknologioita turvaamaan maan kansallisia salassa pidettäviä tietoja. Intia, Saksa, Ranska, Israel ja Brasilia ovat valtiollisella tasolla investoineet internetin turvallisuusjärjestelmiin turvatakseen tietojaan. Euroopan kyberturvallisuuskeskuksen mukaan esimerkiksi Iso-Britannia käytti vuonna 2020 noin 2,3 miljardia Yhdysvaltojen dollaria verkkoratkaisuidensa ja internet turvallisuusprojektiensa rakentamiseen. (Fortune Business Insights, 2024)

Suomalaisten yritysten ja yksityisten tietoturva on heikoissa kantimissa. Yle Areenasta löytyy sarja Team Whack – kaikki on hakkeroitavissa. Sarjassa kolme ammattihakkeria näyttää miten erilaisiin järjestelmiin murtaudutaan tai miten jonkin laite otetaan hallintaan. Ryhmä näyttää, miten murtaudutaan ja otetaan haltuun taloyhtiön järjestelmät, henkilöauto, miten tehdä identiteettivarkaus tai murtaudutaan vakuutusyhtiön järjestelmiin. Monessa näistä vaaditaan laajempaa tietoteknistä osaamista ja usein heikoimpana lenkkinä on ihminen. Sarja kuitenkin näyttää miten helposti esimerkiksi suojaamattomiin järjestelmiin murtaudutaan, kun tiedetään mitä ja miten tehdään ja toimitaan (Eksymä & Sandell, 2020).

Identiteettivarkauden toteutus on toki haastavampaa eri alustojen pakottaessa käyttäjän ottamaan käyttöönsä monivaiheisen tunnistautumisen eli MFA:n.

Tarkastellaan seuraavaksi puolustautumis- ja suojautumiskeinoja erilaisia kyberrikollisuuden muotoja vastaan. Vastaavasti kuin luvussa 2.2 kerrottiin erilaisten kyberrikosten ymmärryksen olevan perustana deepfake-huijauksille, on tärkeää ymmärtää myös laajasti eri kyberrikoksien puolustautumis- ja suojautumismenetelmiä. Deepfake-rikoksen kohteeksi on suurempi todennäköisyys joutua, jos hakkeri onnistuu esimerkiksi tunkeutumaan tietoturva-aukon tai tietomurron kautta yrityksen tai yksityisen henkilön verkkoon ja saamaan pääsyn arkaluotoiseen materiaaliin.

3.4.1 Tietomurtoja vastaan suojautuminen

Tehokas tapa suojautua tietomurtoja vastaan on käyttää kaksi- tai useampivaiheista tunnistusta (MFA) järjestelmissä, joissa se on mahdollista. Monivaiheisessa tunnistuksessa salasana vahvistetaan lisäksi jollakin toisella menetelmällä. Tämä voidaan tehdä joko käyttämällä erillistä todennussovellusta tai tekstiviestitse lähettämällä kertakäyttöinen varmistuskoodi. Näin voidaan estää järjestelmään kirjautuminen myös silloin, kun kirjautumistiedot ovat joutuneet vääriin käsiin. (Poliisi, n.d.-a)

Toinen tehokas tapa suojautua tietomurtoja vastaan on pitää yllä laitteiden turvallisuutta. Tämä koskee sekä yrityksiä että yksittäisiä kansalaisia. Suojautuminen tapahtuu yleensä huolehtimalla laitteiden ja järjestelmien päivityksistä, koska merkittävä osa tietomurroista tehdään päivittämättömiin laitteisiin ja järjestelmiin tai niitä käytetään hyväksi järjestelmään tunkeutumiseen. Kyberturvallisuuskeskus ilmoitti maaliskuussa 2023 Microsoft Outlookista löytyneestä kriittisestä haavoittuvuudesta. Löydetty haavoittuvuus mahdollisti NTML Relay -hyökkäyksen, jonka avulla oli mahdollista päästä korottamaan järjestelmän käyttöoikeuksia (Traficom. 2023c).

Vaikka päivittämättömät laitteet luovat tietoturvariskin yritykselle, on päivityksiä käytetty myös rikollisiin tarkoituksiin. Vuonna 2019 toteutettiin SolarWinds hyökkäys, jossa toimitusketjuhyökkäyksen kautta yritys kuvitteli tekevänsä normaalia päivitystä. Tosiasiassa kuitenkin kyseessä oli haittaohjelma, joka avasi mahdollisuuden hakkereille päästä järjestelmään sisälle. (Khan, n.d.)

Vuonna 2017 toteutettiin hyökkäys WannaCry, jossa kohteeksi joutuivat yli 150 yritystä maalimanlaajuisesti. Pohjois-Korealainen hakkeriryhmä Lazarus-Group käytti hyväkseen Windowsista löytynyttä tietoturva-aukkoa valikoituihin yrityksiin, jotka eivät päivitystä olleet

vielä tehneet. Tappiot ovat mahdollisesti mittavat tämän kaltaisissa hyökkäyksissä. (Mitre, n.d.)

Kannattaa myös pohtia saadaanko tietoverkkoja suojauksia lisäämällä hyötyä tietojärjestelmän turvallisuudelle. Tietomurtojen vahinkoja voidaan myös minimoida käyttämällä poikkeamien hallintaa, jonka avulla ylläpitäjä saa hälytyksen, jos kirjautuminen tulee poikkeavasta IP-osoitteesta tai poikkeavaan aikaan. Poikkeamien hallinnalla voidaan myös havaita, jos hakkeri käyttää kiristyshaittaohjelmaa, jolla hän pyrkii salaamaan järjestelmän tietoja peittääkseen tunkeutumisen jäljet siten, ettei omistaja pääse niihin enää käsiksi. Lisäksi järjestelmän tapahtumista kannattaa kerätä logitietoja eli lokittaa järjestelmä tai sen osia. Näin voidaan havaita nopeammin luvattomat käyttäjät ja heidän liikkeensä järjestelmässä. (Poliisi, n.d.-a)

Tietomurtojen takana on useimmiten yksi tai useampi hakkeri. Yksi yksinkertaisimmista tavoista suojautua hakkereita vastaan on tietojen salaaminen. Salaus tapahtuu ohjelmallisesti siten, että ohjelmassa on määritetty säännöt datan koodaukselle. Lisäksi ohjelmassa on tiedot salauksen avaimesta eli mikä avaa salauksen, miten se luodaan sekä miten se todennetaan. Tyypillisesti internetin alkuaikoina käytettiin HTTP-verkkosivuja. Nykyään sivut ovat muuttuneet HTTPS-sivuiksi eli niihin on tullut mukaan salaus. Salauksen ansiosta sivustolla tapahtuvat toimet eivät ole enää julkisesti nähtävillä. Salaus on hyvä tapa suojautua hakkereita vastaan, mutta kuitenkaan varma, jos käyttäjä päätyy samaan verkkoon hakkerin kanssa. Tällöin hakkeri voi kierrättää dataliikenteen oman tietokoneensa kautta ja muuttaa liikenteen takaisin HTTP-liikenteeksi eli poistaa salaus. Sekä Google että Microsoft ovat kuitenkin kehittäneet eston omissa kirjautumisissaan tällaiselle toiminnalle. (Šimonėlytė, 2023)

Yrityksen joutuessa tietomurron kohteeksi, ovat vaikutukset kauaskantoisia ja voivat vaikuttaa syvästi yrityksen toimintaan ja maineeseen. Aiempien puhtaiden tietoturvaongelmien sijaan yritykselle saattaa tulla taloudellisia tappioita, oikeudellisia ongelmia saatujen tietojen osalta, ja kuluttajien sekä asiakkaiden luottamuksen menetyksenä. (Šimonėlytė, 2023)

3.4.2 Haittaohjelmia vastaan suojautuminen

Haittaohjelmat saattavat päätyä omalle koneelle tai oman koneen kautta yrityksen verkkoon vahingossa tai hakkerin hyökkäyksen vuoksi. Hyvä tapa suojautumiseen haittaohjelmia vastaan ovat erilaiset tietoturvallisuusohjelmistot. Esimerkki tällaisista ohjelmista ovat erilaiset virustorjuntaohjelmat eli haittaohjelmien torjuntaohjelmat. Virustorjuntaohjelmat voivat estää pääsyn ei-toivotuille sivustoilla, joissa mahdollisesti on haittaohjelmia. Lisäksi ne pystyvät skannaamaan käyttäjän laitteen löytääkseen jo mahdollisesti ladattuja tai latautuneita haittaohjelmia. Virustorjuntaohjelmien toiminta perustuu yleensä estolistaan, pilvipalvelussa säilytettävään listaan. Lista sisältää tietoja tunnetuista uhista esimerkiksi tietoja haitallisista verkkosivuista tai tiedostotyypeistä. Ohjelma käynnistää suojaustoimet havaitessaan estolistalla olevaa tietoa ja ilmoittaa tästä käyttäjälle.

Virustorjuntaohjelmat ovat hyvä ja suositeltava tapa laitteen perussuojaamiseen. Ohjelmia löytyy internetistä ilmaiseksi, mutta yleensä maksulliset ohjelmat tai versiot tarjoavat paremman ja kattavamman suojauksen (Šimonėlytė, 2023). Luonnollisesti yksinään virustorjuntaohjelma ei riitä kattavaan suojaan vaan tässäkin merkittävässä osassa on suojausketjun heikoin lenkki: ihminen.

3.4.3 Tietojenkalasteluviestejä vastaan suojautuminen

Lähes jokainen tietokoneen käyttäjä on joskus elämänsä aikana saanut tietojenkalasteluviestin. Miten näitä vastaan sitten voi suojautua? Tietojenkalasteluviestit ovat nykypäivänä yleisiä ja aina kun saa epäilyttävän viestin kannattaa kiinnittää huomiota viestiin, sen sisältöön sekä erityisesti lähettäjään. Viestit sisältävät usein kirjoitusvirheitä tai ovat konekäännöksiä ja kieli huonoa, joten viesti kannattaa lukea huolellisesti. Tyypillinen tunnistuskeino on lähettäjän sähköpostiosoite, joka kannattaa katsoa tarkasti. Sähköpostiosoite saattaa muistuttaa hämääväsi oikeaa tahoa, mutta usein näissä on vain pieniä, yhden kirjaimen eroavaisuuksia. Kannattaa muistaa, etteivät pankki, poliisi tai kukaan viranomainen koskaan kysele kirjautumistietoja sähköpostitse tai puhelimitse. Myös viestin otsikkoon kannattaa kiinnittää huomiota. Se voi olla poikkeuksellinen kirjotusvirheiden tai sisällön osalta (F-Secure, n.d.-a). Kuva 2 viesti oli helposti tunnistettavissa huijaukseksi sekä sähköpostiosoitetta että kieltä tarkasteltaessa.

Tietojenkalastelu on tyypillinen keino identiteettivarkauden yhteydessä. Tietojenkalastelun yhteydessä huijari yrittää saada kohteen luovuttamaan henkilökohtaisia tietoja, joita voi sitten käyttää hyväkseen (Microsoft, n.d.). Yle Areenasta löytyy sarja Team Whack – kaikki on hakkeroitavissa. Sarjassa näytetään muun muassa, kuinka yksinkertaista on tehdä

identiteettivarkaus. Tässäkin tapauksessa heikon lenkki, oli itse identiteettivarkauden kohteeksi joutunut henkilö. (Team Whack, 2019)

Hyviä suojautumiskeinoja tietojenkalastelua ja identiteettivarkauksia vastaan ovat esimerkiksi vahvat salasanat. Kannattaa käyttää vahvoja eli mahdollisimman pitkiä ja vaikeasti arvattavia salasanoja ja pyrkiä siihen, että käyttää kaikissa eri palveluissa eri salasanaa. Toinen suositeltava keino on tilitapahtumien tarkkailu. Omaa tiliä ja tilitapahtumia kannatta tarkkailla säännöllisesti ja jos havaitsee tilisiirtoja, joista ei itse ole tietoinen, sulkee tilin. Salasanat ja käyttäjätunnukset vaarantuvat erityisesti vieraissa verkoissa, jolloin kannattaa välttää julkisten verkkojen käyttöä, kun kirjautuu pankkitilille. Esimerkiksi yleisten wifi-verkkojen salasanat voivat olla liian helppoja tai ne ovat helposti saatavilla ja hyökkääjä pääsee sitä kautta uhrin tietoihin käsiksi. VPN-sovelluksella voi muodostaa salatun yhteyden laitteen ja VPN palvelimen välille julkisessa verkossa, mutta tämäkään ei suojaa täysin varmasti kaikilta hyökkäyksiltä. (Kyberturvallisuuskeskus, n.d.-a)

Maaliskuussa 2024 varoitettiin Applen käyttäjiin kohdistetuista viesteistä, joilla yritettiin saada käyttäjä luovuttamaan Apple ID -tunnuksensa. Tunnuksen paljastumisen jälkeen, on Apple-laite mahdollista ottaa toisaalla käyttöön. Käyttäjiä pommitettiin puhelimen käyttöjärjestelmätason viesteillä ja pyydettiin lopuksi käyttäjää resetoimaan salasananansa. Huijauksen tehostamiseksi käyttäjälle vielä soitettiin ja kerrottiin soiton tulevan Applen tuesta. Soitto Applen tukeen vahvisti, ettei heiltä oteta koskaan asiakkaaseen yhteyttä puhelimitse, ellei asiakas sitä itse pyydä. Puhelimeen kohdistuvissa tietojenkalasteluyrityksissä pätee samat turvallisuusohjeet kuin muissakin viesteissä: salasanaa tai käyttäjätunnuksia ei anneta kenellekään. Samoin epäilyttävissä viesteissä tai soitoissa pitää varmistaa kuka aidosti on soiton tai yhteydenoton takana. Tämän hyökkäyksen takana oli tietoturva-aukko, joka korjaantui iOS-käyttöjärjestelmän päivityksellä. (Krebs, 2024)

Sähköpostien mukana tulevat liitteet voivat sisältää haittaohjelman tai viruksen. Sähköpostien liitetiedostot kannatta avata vain, jos on täysin varma lähettäjästä. Pienikin epäilys sähköpostin lähteestä tai lähettäjästä tarkoittaa, että liitetiedosto voi olla haitallinen ja se kannattaa jättää avaamatta. Yksityiset dokumentit kannattaa suojata, joko siirtämällä ne ulkoiselle kiintolevylle tai USB-muistille, jota käytetään vain tarvittaessa eikä sitä pidetä laitteessa kiinni jatkuvasti. Tällöin mahdollisen murron tapahduttua, tärkeät tiedostot ovat muualla tallessa. Tavalliseen postiin kannattaa kiinnittää huomiota hankkimalla lukollinen postilaatikko, jotta kuka tahansa ei pääse postiin käsiksi. Lisäksi postin mukana tulleet arkaluontoista tietoa sisältävät paperit ja laskut kannattaa tuhota asianmukaisesti. (Kyberturvallisuuskeskus, n.d.-a)

Kaksi- tai useampivaiheista tunnistautumista kannattaa käyttää aina kun se on mahdollista. Kirjauduttaessa palveluun syötetään käyttäjätunnus ja salasana, mutta tähän liittyy vielä lisävarmistus, joka yleensä tehdään puhelimen kautta ylimääräisellä kirjautumiskoodilla (Kyberturvallisuuskeskus, n.d.-a). Luonnollisesti pitää aina varmistaa kenen kanssa on verkossa tai puhelimessa oikeasti tekemisissä. Sarjassa Team Whack – kaikki on hakkeroitavissa nähtiin, kuinka helppoa on tehdä identiteettivarkaus uskottelemalla puhelimessa olevansa joku toinen. (Team Whack, 2019)

3.4.4 Spoofing-hyökkäyksiä vastaan suojautuminen

Spoofing-hyökkäyksellä pyritään saamaan varomaton käyttäjä siirtymään hyökkääjän sivustolle. Sivuston nimi on rakennettu näyttämään joltain muulta, kuin mikä se oikeasti on.

Tilisanomat viittaa Finanssiala Oy:n julkaisuun, jossa vuonna 2022 suomalaisilta vietiin erilaisilla digitaalisilla huijauksilla yli 32 miljoonaa euroa. Suurin osa näistä perustui tietojenkalasteluun perustuneeseen spoofing-huijauksiin kohdistuen verkkopankkitunnuksiin. Rikosten kulku oli hyvin tyypillinen ja toisti itseään. Ensimmäiseksi rakennetaan tietoja kalastelevat sivut, esimerkiksi verkkopankin tunnistautumissivu. Verkkopankkitunnistautumista käytetään verkkopankkiin siirtymisen lisäksi monessa muussakin yhteydessä. OmaVero-sivusto, suomi.fi ja monet verkkokaupat käyttävät käyttäjän tunnistamiseen pankin verkkopankki -tunnistautumista. (Alanko, 2023)

Huijauksen voi tunnistaa viemällä kursori linkin päälle ennen kuin sitä klikkaa ja havaita näin mahdollisen huijauksen erikoisesta URL-osoitteesta. Lisäksi URL-osoite kannattaa syöttää käsin eikä klikata valmista linkkiä. Sivuston URL-osoite kannattaa myös tarkistaa ennen klikkausta mahdollisten kirjoitusvirheiden varalta. Lisäksi kannattaa varmistaa, että kyseessä on sivusto, jossa on salaus eli osoite on HTTPS-alkuinen. (F-Secure, n.d.-d)

Sivuston kirjoitusvirheet kannattaa tarkastaa samoin kuin sähköpostin. Jos kirjoitusvirheitä löytyy, kyseessä on luultavasti huijaus. Arkaluontoisia tietoja ei kannata paljastaa. Luotettavat tahot eivät koskaan kysy käyttäjätunnusta tai salasanaa tai muitakaan henkilökohtaisia tietoja. Kaikilla sivustoilla, joihin vaaditaan kirjautuminen, kannattaa käyttää vahvoja salasanoja ja jos mahdollista kaksi- tai useampi vaiheista tunnistautumista. Myös samaa salasanaa ei kannata käyttää useammassa palvelussa. (F-Secure, n.d.-d)

VPN- ja virustorjuntaohjelmia suositellaan käytettäväksi aina, kun se on mahdollista. VPN-ohjelmat piilottavat sijainnin ja IP-osoitteen, jolloin niiden anastaminen on vaikeampaa.

Virustorjuntaohjelmat eivät päästä haittaohjelmia laitteeseen ja estävät pääsyn haitallisille sivuille tai sivustoille. (F-Secure, n.d.-d)

Seuraavaksi huijarin pitää saada käyttäjä päätymään luodulle valesivustolle. Internetin hakukoneita on mahdollista manipuloida siten, että käyttäjän hakiessa hakutoiminnolla oman pankin sivustoa hänet ohjataan huijaussivustolle. Käyttäjä kirjaa sivustolle käyttäjätunnuksensa sekä vaadittavat salasanat antaen ne näin vääriin käsiin. Puolustautumiskeinoksi pankit ohjeistavat aina kirjautumaan sivuilleen kirjoittamalla suoraan pankin sivun osoite eikä siirtymällä minkään linkin kautta. Nopeampi ja huomattavasti tehokkaampi tapa on lähettää massaviestinä käyttäen joko sähköpostia tai tekstiviestiä ja esiintyä viranomaisena tai pankin edustajana. Viestissä kehotetaan noudattamaan annettuja ohjeita suuren vahingon välttämiseksi. Käyttäjää viestin mukaan voi uhata tilin sulkeminen, asiakkuuden menetys tai omaisuuden menetys. (Alanko, 2023)

Viestin lähetys kohdentuu suureen massaan, joista suurin osa tunnistaa huijausviestin. Pienen osan lankeaminen huijaukseen riittää kuitenkin huijareille. Tekstiviestiin reagoidaan nopeammin kuin sähköpostiin ja kun ajankohtana on ilta tai viikonloppu, ei oikea viranomainen ole helposti tavoitettavissa oikeellisuuden tarkistamiseen. (Alanko, 2023)

Kolmantena askeleena käyttäjän kirjautuessa sisään huijaussivustolla, huijari saa tunnukset, käy katsomassa mitä tunnuslukua oikea sivusto pyytää, pyytää tämän käyttäjältä ja kirjautuu sisään hänen tunnuksillaan. Näin huijari on tunnistautunut käyttäjän oikeilla tunnuksilla. Viimeisessä vaiheessa huijari siirtää rahaa omalle tililleen. Jotta jäljet haihtuvat, tilisiirtoja tapahtuu nopeaan tahtiin tililtä toiselle ja päätyen lopuksi mahdollisesti maarajojen yli. Tällöin tutkinta ja rahojen mahdollinen takaisin saanti vaikeutuu tai on jopa mahdotonta. (Alanko, 2023)

Jos huomaa joutuneensa iskun kohteeksi, on tärkeää ilmoittaa asiasta pankille välittömästi. Näin mahdollisesti saadaan jotain rahavirtoja vielä pysäytettyä. Tilisanomien artikkelissa kerrotaan pankkien monitorijärjestelmien sekä erilaisten peruutuspyyntöjen ansiosta pystyttiin pysäyttämään yli 14 miljoonaa euroa vuonna 2022. Pankin jälkeen seuraava kohde on poliisi. Verkkopankkihuijaukset tutkitaan maksuvälinepetoksina. (Alanko, 2023)

Kohteena spoofing-hyökkäyksille on useimmiten henkilöt, joiden huijarit uskovan lankeavan. Puolustautumiskeinona on olla itse valveutunut sekä huolehtia omassa lähipiirissä olevien vanhusten ja muiden ei mahdollisesti valveutuneiden ymmärryksen lisäämisestä. Kaikkia rahoja ei kannata pitää yhdellä tilillä, jolloin vahingon sattuessa, huijari ei ehdi tyhjentää kaikkia tilejä ja pankki ehtii toimia. (Alanko, 2023)

3.4.5 Informaatiovaikuttamista vastaan suojautuminen ja sen tunnistaminen

Aiemmin todettiin, että informaatiovaikuttamisen keinoja ovat esimerkiksi harhaanjohtavien tai jopa väärien tietojen levittäminen. Levittäminen voi sisältää myös painostamista sekä faktojen irrottamista asiayhteydestä. Toiminnan tavoitteena on saada kohde tekemään itselleen haitallisia päätöksiä, toimimaan omaa etuaan vastaan tiedostamattaan. Erilaiset mielipiteet ja näkemykset kuuluvat luonnollisena osana demokraattiseen yhteiskuntaan - ne eivät ole informaatiovaikuttamista.

Informaatiovaikuttamista voi tulla vastaan missä tahansa, joten kriittisyys julkaisuja kohtaan on ensimmäinen askel puolustautumisessa. Jos epäilee informaatiovaikuttamista jossain julkaisussa, kannattaa lukea koko julkaisu ja tarkastella kokonaisuutta sekä miettiä, mihin ja miten sillä pyritään vaikuttamaan ja minkälainen viesti on. Julkaisijan tarkastelu on syytä tehdä tarkasti: jos julkaisija on uusi ja julkaisee jatkuvasti jotain, on syytä epäillä automatisoitua tiliä. Julkaisijan profiilia kannattaa myös katsoa tarkasti. Onko profiilikuva ja julkaisussa käytetyt kuvat aitoja vai löytyvätkö ne helposti hakukoneiden kuvahauilla? Myös kieleen ja kielioppiin kannattaa kiinnittää huomiota. Jos tekstissä on paljon virheitä tai epäjohdonmukaisia lauseita, voi kyseessä olla automatisoitu käännös. Myös julkaisijasta kannatta yrittää etsiä tietoja ja yrittää siten varmistua, että kyseessä on luotettava julkaisija. Jos kyseessä on verkkosivu, kannattaa sivun URL-osoitteeseen kiinnittää huomiota. Jos osoitetta on muutettu kyseessä saattaa olla informaatiovaikuttamisen keino. Myös julkaisun motiivia tai tavoitetta kannattaa miettiä, sekä sitä miksi julkaisu on tehty juuri nyt. Viestin lähteet kannattaa myös tarkastaa ja miettiä onko niitä käytetty oikein. (Traficom, 2022)

Osa informaatiovaikuttamisesta on disinformaatioita, väärää tietoa, mutta suojautuminen informaatiovaikuttamista vastaan on myös muuta kuin vain oikean ja väärän tiedon erottamista. Disinformaatioksi voi kutsua vaikka valkoista valhetta, jonka henkilö päästää suustaan, mutta välttämättä kyse ei ole informaatiovaikuttamisesta. Koska informaatiovaikuttaminen on aina kohteensa etujen vastaista, se pyrkii aiheuttamaan vahinkoa joko yksilölle, ihmisryhmälle, yritykselle tai jopa valtioille. Ajatusten, asenteiden ja sitä kautta tekojen muuttaminen väärän tiedon levittämisen takia, on informaatiovaikuttamista. Kyse on siis pitkälti psykologisesta vaikuttamisesta kohteeseen. Koska kohde on yhteiskunnan jäsen, vaikuttavat hänen tekonsa pahimmillaan koko yhteiskunnan toimintaan. Tämän vuoksi puolustautumisessa on viranomaisilla tärkeä rooli. Viranomainen voi viestiä julkisesti, mitä disinformaation levittäjällä on tavoitteena sekä kertoa, miten ihmisten tulisi toimia.

Äärimmäisellä valtioiden välisellä informaatiovaikuttamisen tasolla, voi toinen valtio pyrkiä vaikuttamaan toisen valtion poliittiseen ilmapiiriin, vaikka vaalien kautta tai poliittisen status quon kautta. Tammikuussa 2024 levisi mediassa puhelu, jossa Yhdysvaltain presidentti Joe Biden kehotti ihmisiä olemaan äänestämättä New Hampshiren osavaltion esivaaleissa. Niin sanottu Robocall ei kuitenkaan ollut presidentti Bidenin tekemä ja se paljastettiin startup-yritys ElevenLabsin ohjelmalla tehdyksi deepfakeksi (Bergen, Mets, Murphy, 2024). Vaalien lähestyessä Suomen presidentin vaalien aikaan vuonna 2024, ennen varsinaisia vaaleja levisi sosiaalisen median alustoilla deepfake-video, jossa presidenttiehdokas Alexander Stubb mainosti kryptovaluuttaa (Korhonen, 2024). Informaatiovaikuttamisen ohella deepfake-videoista on vaarallisempiakin esimerkkejä. Vuonna 2019 Gabonin presidentti Ali Bongon oltua mediasta poissa pitkähkön ajan sairauden ja hoitojen vuoksi, hänen videoesiintymisensä sai aikaan epäilyt deepfake-videosta, jolla pyritään salaamaan presidentin kuolema. Epäilyt olivat niin vahvat, että armeija teki vallankaappausyrityksen (Breland, 2021). Suojautuminen ja puolustautuminen kuuluvat tällöin puhtaasti viranomaisille maan lakien antamien valtuuksien mukaisesti.

Yksilöinä yksinkertaisin suojautumisen muoto, on välttää lähteitä ja informaatiota, jotka vaikuttavat itseen mahdollisesti haitallisesti. Käytännössä tämä on helpommin sanottu kuin tehty: vuonna 2024 uutisvirta on täynnä Ukrainan sotaa sekä Gazan konfliktia. Moni jakaa uutisvirrasta näkemiään tekstejä ja videoita eteenpäin ja osallistuu näin välillisesti disinformaation levittämiseen. Videot herättävät tunteita ja vaikuttavat meihin psykologisesti. Kumman tahansa osapuolen kuvaamat videot näyttävät asian tietyssä valossa ja pyrkivät näin saamaan katsojan valitsemaan puolen. Jokaiseen meistä voidaan vaikuttaa ja sitä tehokkaampaa vaikuttaminen on, kun se löytää meistä jonkun haavoittuvaisuuden. Onko se sodassa kuolevia lapsia, kärsiviä eläimiä, voimakkaampi sortaa heikompaa tai epäoikeudenmukaisuutta; kaikkiin vaikuttaa jokin asia voimakkaasti. Kuka on lopulta tarinan paholainen ja kuka viaton uhri? Kasvattamalla resilienssiään oman haavoittuvuuden osalta, on jo huomattavasti paremmassa suojautumisen tilassa pelkän tiedon välttämisen sijaan.

3.4.6 Kiristyshaittaohjelmia vastaan suojautuminen

Kiristyshaittaohjelma tulee koneelle mahdollisesti hakkerin tunkeutumisen seurauksena. Kun koneeseen on isketty, on haittaohjelman poistaminen vaikeaa ja jopa mahdotonta. Kiristyshaittaohjelmia kuten muitakin haittaohjelmia on tehty myös puhelimille ja muille älylaitteille. On mahdollista käyttää salauksenpurkutyökalua, jolla voi poistaa kyseisen salauksen, mutta usein tämäkään ei välttämättä onnistu. Näin ollen puolustautuminen alkaa suojautumisella niitä vastaan. (F-Secure, n.d.-c)

Laitteessa kannattaa pitää luotettava ja ajantasainen suojaus esimerkiksi virustorjuntaohjelma, joka mahdollisesti estää haittaohjelman pääsyn koneelle. Toinen hyvä suojautumiskeino on varmuuskopioida säännöllisesti ja pitää varmuuskopiot erillään verkosta esimerkiksi USB-muistilla tai ulkoisella kovalevyllä. Näin toimiessa, ei tietoja menetetä hyökkäyksen lukitessa varsinaisen koneen. Laitteen ohjelmistot kannatta pitää ajan tasalla eli päivittää ohjelmat säännöllisesti tai käyttää automaattisia päivityksiä. (F-Secure, n.d.-c)

Kun yrityksessä on laajalti ymmärrys mahdollisista Ransonware-hyökkäyksistä ja niiden seurauksista, kaikki ymmärtävät, että sähköpostien liitetiedostoihin kannattaa suhtautua erityisellä varovaisuudella ja avata vain luotettavasta kohteesta tulleet liitetiedostot. Erityisesti jos liitetiedostot haluavat käyttää makroja tai muuta laitteen sisältöä. Teknisesti selainliitännäisten käyttöä voidaan myös rajoittaa, jos mahdollista, ja estää haavoittuvimmat liitännäiset, kun niitä ei käytetä.

Monet organisaatiot käyttävät henkilöstön kouluttamiseen pelillisiä keinoja: tietoturvaohjelmisto lähettää vaarattomia kalastelusähköpostiviestejä henkilöstölle ja kerää samalla dataa, miten paljon viestejä avataan ja linkkejä avataan. Näin saatua analysoitua dataa voidaan sitten käyttää jatkossa kouluttamiseen ja suojauksen tason nostamiseen entisestään.

3.4.7 Social engineering hyökkäyksiä vastaan suojautuminen ja puolustautuminen

Vaikka yritykset ovat käyttäneet suojautumiseen potentiaalisia kyberhyökkäyksiä vastaan runsaasti resursseja, yksi elementti on suhteellisen helposti hakkeroitavissa: ihminen. Tunnistettaessa yksittäisen ihmisen heikko kohta, voidaan hänen kauttaan päästä sisälle yritykseen. Tärkein keino yrityksen suojautua social engineeringiä vastaan on pyrkiä suojaamaan yksittäinen työntekijä manipulointiyrityksiltä.

Yle Areenan ohjelmistossa Team Whack – kaikki on hakkeroitavissa, jaksossa neljä Team Whack murtautuu Lähi-Tapiola vakuutusyhtiöön. Siinä ensimmäinen askel on social engineering, jolloin sosiaalisen median alusta LinkedIniä käytettiin luomaan uskottavuutta luoduilla henkilöhahmoilla. Kun ihmisillä on jo valmiiksi joku ”tutun tuttu”, tähän luotetaan ja suojaukset tippuvat. Tyypillistä social engineering tapauksille onkin, että ensin saavutetaan luottamus ja sen jälkeen uhri avaa varomattomasti ”tutulta” tulevan linkin, laittaa koneeseensa vieraan USB-muistin tai luovuttaa tietojaan varomattomasti. (Team Whack, 2020)

Mitä sitten verkossa kannattaa kertoa itsestään? Vastaus: mahdollisimman vähän, jos yhtään mitään. Paras suojautumiskeino on antaa hyökkääjille mahdollisimman vähän tietoa, jota hän voisi käyttää. Sosiaalisessa mediassa jaettavien tietojen kanssa kannattaa olla varovainen, koska hyökkääjä voi myös päästä niihin käsiksi ja hyödyntää niistä saamiaan tietoja hyökkäyksiin. (F-Secure, n.d.-d)

Ehdottomasti paras tapa Social engineering hyökkäyksiä vastaan on suojautua käyttämällä riittävän vahvoja salasanoja sekä kaksi- tai useampivaiheista tunnistautumista (MFA) kaikkiin palveluihin, joissa se on mahdollista. Luodun huijaussivuston kautta pystytään toki ohittamaan myös monivaiheiset tunnistukset. Mahdollisten hyökkäysten onnistuessa, on tärkeää seurata jatkuvasti kriittisiä ohjelmia ja niissä tapahtuvaa poikkeuksellista toimintaa. (F-Secure, n.d.-d)

Hyväksyttäessä uusia kontakteja tai luovutettaessa henkilökohtaisia tietoja pitää kohteen, jolle tietoja luovutetaan, henkilöllisyys varmistaa ennen luovutusta. Kirjautumistietoja, salasanoja ja mahdollisia vahvistuskoodeja ei pidä koskaan luovuttaa kenellekään. Jos ulkoinen laite (yleensä USB-muisti tai ulkoinen kovalevy) yhdistetään laitteeseen, pitää olla täysin varma lähettäjästä sekä siitä, että laite on turvallinen ja sisältää turvallisia tietoja. (F-Secure, n.d.-d)

Suojautumiskeinona yksityiselämässä erityisesti lapsille ja vanhuksille kannatta internetin turvallinen käyttö opettaa hyvissä ajoin. Samoin on tärkeää pitää niitäkin henkilöitä tietoisina mahdollisista vaikuttamisyrityksistä, jotka eivät seuraa uutisia säännöllisesti. (F-Secure, n.d.-d)

3.4.8 WhatsApp-huijauksia vastaan suojautuminen

WhatsApp-huijauksia vastaan voidaan suojautua samoin kuin muitakin huijauksia vastaan. WhatsAppissa on tunnuksen vahvistustoiminto, jonka avulla pääsee käyttäjän WhatsApp-tiliin käsiksi. (Augustėnė, 2023)

Rahanpyyntöviesteihin ei tule reagoida. Kannattaa myös pohtia onko rahanpyyntöviesti henkilöltä, joka oikeasti pyytäisi rahaa ja kuinka aidolta viesti näyttää. Kannattaa käyttää kaksi- tai useampivaiheista tunnistautumista (MFA) aina kun se on mahdollista. Viestin lähettäjältä kannattaa kysyä kysymyksiä varmistuakseen viesti aitoudesta. Parhaita kysymyksiä ovat ne, joihin vain viestin kohdehenkilö osaa vastata. Huijarikin voi tehdä taustatutkimusta uhrista ja näin osata vastata helpompiin kysymyksiin. Aina voi myös soittaa

viestin lähettäjälle numeroon, jonka tietää ja varmistua näin viestin aitoudesta. (Augustėnė, 2023)

VPN-verkon käyttäminen salaa verkkoliikenteen ja suojaa IP-osoitteen. Tämä tekee myös keskustelujen seuraamisesta vaikeampaa. Epäilyttäviä linkkejä ei kannata klikata WhatsAppissa vaan kirjoittaa haluttu osoite suoraan osoiteriville. Viestin kieliasuun kannattaa kiinnittää erityistä huomiota. Jos viestissä on virheitä tai sanavalinnat ovat outoja viesti saattaa olla käännetty automaattisella kääntäjällä ja kyseessä on huijaus. (Augustėnė, 2023)

WhatsAppin vahvistuspyyntöjen kanssa kannattaa olla tarkkana ja lukea ne huolellisesti, ettei hyväksy vahingossa esimerkiksi huijareiden tilisiirtoja tai vahvistuskoodeja. Ajantasaisen virustorjuntaohjelman käyttö on hyvä tapa suojata laitetta. (Augustėnė, 2023)

Kun huomaa joutuneensa WhatsApp-huijauksen kohteeksi, kannattaa heti vaihtaa salasana. Myös kaksi- tai useampivaiheinen tunnistautuminen kannattaa ottaa heti käyttöön. Jos salasanoja on vaikea muistaa, kannattaa käyttää salasanojen hallintaohjelmaa. Tällaisen ohjelman voi ladata yleensä ilmaiseksi sovelluskaupasta. Jos huijaukseen liittyy rikos kannattaa tehdä poliisille rikosilmoitus mahdollisimman nopeasti. Tällaisessa tapauksessa myös WhatsAppille kannattaa tehdä ilmoitus. Jos epäilee tulleensa huijatuksi ja menettäneensä tilinsä WhatsAppissa huijareille, kannattaa aina tarkistaa profiilin tiedot. Huijarit voivat muuttaa profiilin tietoja. Kannattaa myös tarkistaa onko tilillä aloitettu uusia outoja keskusteluja tai puheluita. (Augustėnė, 2023)

3.4.9 Deepfake-väärennösten tunnistaminen ja niitä vastaan suojautuminen

Kyberrikosten määrissä deepfake-huijausten määrä on vielä ollut suhteellisen pieni verrattuna muihin kyberhuijauksiin. Niiden määrä on kuitenkin huimassa kasvussa osana kyberrikollisuuden kasvua. Asiantuntijat ovat arvioineet internetin eri alustoilla olleiden deepfake-videoiden määrän olleen vielä vuonna 2021 noin 14 000 kun vuonna 2023 niiden määrä oli jo noin 500 000. Vuonna 2019 arvioitiin yli 95 % deepfake-videoista olleen pornografisia ja niiden kohteena oli joku julkisuuden henkilö. Lisäksi kohteena oli muita julkisuuden henkilöitä kuten poliitikkoja. Kun tähän lisätään arvio, että vain noin 30 % ihmisistä edes tietää, mikä deepfake on ja näistä arviolta hiukan yli puolet (57 %) uskoo tunnistavansa sellaisen, voisi deepfake-huijausten sanoa uppoavan kuin ”kuuma veitsi voihin” useimmissa tapauksissa. Lähteenä käytetyssä tutkimuksessa etenkin Yhdysvalloissa on arvioitu Youtube-kanavista yli 70 % sisältävän deepfake-materiaalia. (McGill, 2024)

2019 tapahtui arvatenkin ensimmäinen raportoitu huijaus, jossa deepfake äänen avulla onnistuttiin saamaan yritykseltä noin 220 000 euroa. Kyberrikollinen käytti AI ääniteknologian avulla muokattua ääntä. Tapauksen kulku oli oppikirjamainen: toimitusjohtajalta kiireellinen soitto, jossa pyydettiin siirtämään rahaa ulkomaiselle tilille. Lähes välittömästi rahat siirtyivät eteenpäin ja myöhemmin hajautettiin useampiin lähteisiin. Tässä tapauksessa huijaus onnistui vain osittain tai huijarilla kasvoi ahneus liikaa onnistuessaan. Hän soitti uudestaan uhrille ilmoittaen, ettei maksu tullut perille ja vielä kolmannenkin kerran pyytäen uutta suoritusta. Huijattu näki ensimmäisen erän siirtyneen ja huomasi soiton tulevan itävaltalaisesta numerosta, kun puhelun piti tulla Saksasta. Onnistunut deepfake-äänihuijaus maksoi kuitenkin yritykselle yli 200 000 euroa. (Damiani, 2019)

Arupin kaltaisessa tapauksessa, jossa deepfake-huijausta oli yritetty onnistumatta, olisi yrityksen kannattanut viestiä laajasti huijausyrityksestä yrityksen sisällä (Tekniikan Maailma, 2024). Tällä tavoin henkilöstö olisi ollut varuillaan ja mahdollisesti onnistunut huijaus olisi onnistuttu välttämään. Samoin jakamalla tietoa muille yrityksille, olisi ennakointia parannettu. Heikoin lenkki kaikissa onnistuneissa tapauksissa oli ihminen.

Deepfake-väärennöksen tunnistaminen on entistä vaikeampaa. Tekoälyn, neuroverkkojen niiden kautta syväoppimisen kehittyessä deepfake-huijauksia on yhä vaikeampi tunnistaa. On kuitenkin olemassa useita kohtia, joista deepfake on tunnistettavissa. Deepfake-videossa väärennetyt kasvot voivat näyttää oudon tasaisilta, kun ne asettuvat alla olevien kasvojen päälle. Kasvojen sumeita ääriviivoja kannattaa myös seurata, nämä yleensä paljastuvat äkillisten liikkeiden tai muuttuvan valaistuksen seurauksena. Jos kyseessä on julkisuuden henkilöä tai poliitikkoa esittävä video, kannattaa etsiä henkilöstä aitoja videoita ja tarkastella muita yksityiskohtia kuin kasvoja. Esimerkiksi kädet, hiukset tai vartalon muodot voivat paljastaa väärennöksen. (Zieniūtė, 2022)

Videosta kannattaa katsoa tarkasti kohteen liikettä. Jos kohde liikkuu luonnottomasti tai kohteen pää tai keho näyttävät oudon jähmeiltä, voi kyseessä olla kyseessä väärennös. Toistaiseksi deepfake-huijauksissa ääni on niin sanottu ”heikoin lenkki”. Sen väärentäminen on selvästi vaikeampaa kuin videon tai kuvan. Tämän vuoksi ääneen kannattaa kiinnittää erityistä huomiota. Väärentäjät voivat käyttää joko näyttelijän ääntä, joka matkii kohdetta tai tekoälyn avulla luotua ääntä. Kuuntelemalla kohteen aitoa ääntä, voi havaita eroja. (Zieniūtė, 2022)

Deepfake-huijaukset eivät toistaiseksi ole tavallisille internetin käyttäjille suurin tietoturvariski. Deepfake-huijausten välineenä käytetään usein julkisuuden henkilöä ja henkilöitä, jotka ovat uskottavia muiden silmissä. Deepfaken avulla haetaan taloudellista hyötyä, pyritään

informaatiovaikuttamaan tai saamaan deepfakessa käytetty henkilö mustamaalattua. (Zieniūtė, 2022). Kyberturvallisuuskeskuksen mukaan tapauksia, joissa olisi käytetty deepfakeja Suomalaisissa yrityksissä, on vähän. Virallisen arvion mukaan uhan taso ja vaikutukset ovat suhteellisen matalat. Uhan suuruus on kasvaa, kun kohteena on yksityisen henkilön sijaan suurempi kansalaisjoukko. (Autio ym, haastattelu, 30.8.2024)

Oman yksityisyyden suojaaminen on tärkein puolustautumiskeino, jotta vältytään joutumasta deepfaken välineeksi. Omat sosiaalisen median tilit kannattaa suojata tarkoin ja pitää ne yksityisinä. Kannattaa myös välttää omien kuvien ja videoiden jakamista, erityisesti kasvokuvien jakamisessa kannattaa olla tarkkana. Aidolta vaikuttavien kuvien tai videoiden kanssa kannattaa olla tarkkana, koska ne saattavat olla yritys kerätä rahaa esimerkiksi hyväntekeväisyyden nimissä. Julkisuuden henkilöille tämä on vaikeaa, koska osa heidän ammattiaan on olla julkisuudessa. Siksi he ovatkin otollisia kohteita deepfakejen lähteiksi. (Zieniūtė, 2022)

Samalla nopeudella, kun tunnistuskeinot deepfake-huijauksiin kehittyvät, myös teknologiat niiden luomiseen kehittyvät. Kissa ja hiiri -leikistä hyvänä esimerkkinä oli vuonna 2020 tutkijoiden huomaama heikkous deepfake-algoritmeissä, jossa ihmisen silmien räpäyttämistä ei ollut ohjelmoitu kunnolla. Tämä tarkoitti deepfake-videoissa usein joko epänormaalia silmien räpyttämistä tai kokonaan silmien aukipitämistä. Tämä anomalia oli helppo huomata yksinkertaisella tunnistus algoritmilla. Muutaman viikko julkistuksen jälkeen ilmestyi seuraavan sukupolven deepfake-algoritmit, jotka osasivat emuloida silmien räpäyttämistä huomattavasti luonnollisemmin. Sykli muistuttaa alkuperäistä GAN prosessia syväoppimisessa. (Struck, 2024)

3.4.10 Työkaluja Deepfake-videoiden tunnistamiseen

Deepfake-videoita vastaan on myös kehitetty tekniikoita ja työkaluja. Näistä lupaavimmat käytössä olevat ovat Googlen, Sentinelin, Microsoftin, Intelin ja muun muutaman organisaation luomat tunnistusteknologiat, joita käydään seuraavassa läpi.

Google kehittämä **SynthID** perustuu kuvaan tai ääneen lisättävään vesileimaan tekoälyllä tuotetuissa kuvissa. Digitaalinen vesileima upotetaan kuvaan suoraan kuvan pikseleihin ja se on ihmissilmälle huomaamaton, mutta algoritmi tunnistaa sen. Työkalu tukee vain Googlen Imagenia, eli tekstistä kuvaksi työkalua. Äänessä vesileima laitetaan suoraan ääniaaltoon. Vesileima pysyy paikallaan kuvan tai äänen muokkaamisesta tai pakkaamisesta huolimatta, eikä se vaikuta kuvan tai äänen laatuun. Työkalu perustuu kahteen tekoälymalliin, joista toinen tekee vesileiman ja toinen tunnistaa sen. (Gowal&Kohli, 2023)

Sentinel on kehittänyt tekoälyyn perustuvan deepfake tunnistuksen, joka auttaa muun muassa hallituksia taistelussa deepfakeja vastaan. Sentinelin päämaja on Tallinnassa ja sitä käyttävät useat Euroopan johtavat organisaatiot. Järjestelmä toimii siten, että se sallii käyttäjien ladata digitaalista sisältöä verkkosivustonsa tai API:n kautta. Ladattu sisältö tarkastetaan automaattisesti ja järjestelmä määrittää, onko sisältö deepfake vai ei. Tunnistustekniikka on suunniteltu digitaalisen median eheyden suojaamiseen. Kehittyneet tekoälyalgoritmit analysoivat ladatun sisällön ja määrittävät onko sisältöä manipuloitu. Järjestelmä antaa tarkan raportin ja visualisoinnin manipuloinnista. Tästä käyttäjä näkee tarkasti, miten ja missä mediaa on manipuloitu. (McFarland, 2024)

Intelin kehittämä reaaliaikainen deepfake-tunnistus on nimeltään **FakeCatcher**, joka tunnistaa väärennetyn videon Intelin mukaan 96 % tarkkuudella. New Yorkin osavaltion yliopistossa Binghamtonissa yhdessä Umur Ciftcin kanssa suunniteltu järjestelmä toimii palvelimella, joka käyttää Intelin laitteistoa ja ohjelmistoa ja se liitetään verkkopohjaisen alustan kautta. FakeCatcherin tekniikka perustuu siihen, että se etsii inhimillisiä aitoja vihjeitä oikeista videoista. Se tutkii hienovaraista ”verenvirtausta” videon pikseleistä. Kun sydän pumppaa verta, suonet vaihtavat väriä. Verenvirtaussignaalit kerätään kaikkialta kasvoista ja algoritmit muuttavat ne spatiotemporaalisiksi kartoiksi. Tämän jälkeen syväoppimisen avulla video voidaan tunnistaa aidoksi tai väärennökseksi. (McFarland, 2024)

Älykkäitä in-the-loop-sisällönvarmistus- ja disinformaatioanalyysimenetelmiä ja -työkaluja kehittävä projekti on nimeltään **WeVerify** -projekti. Kontekstualisointiin verkkoympäristössä ja sosiaalisen median ja verkkosisällön analysointiin keskittyvä projekti valmistetun sisällön paljastamiseen. Sosiaalisen verkoston analyysi, monimuotoinen sisällön todentaminen, mikrokohdistettu paljastus ja lohkoketjupohjainen julkinen tietokanta auttavat saavuttamaan tämän tunnetuista väärennöksistä. (McFarland, 2024)

Microsoftin kehittämä työkalu still-kuvan tai videon analysointiin on nimeltään **Video Authenticator**. Työkalu antaa luottamuspisteet, joka kertoo, onko media väärennetty. Se havaitsee hienovaraisten harmaasävyelementtien sekoittumisrajan. Ohjelman tarjoama reaaliaikainen luottamuspisteen ansiosta väärennös voidaan havaita heti. Video Authenticator käyttää median analysointiin kehittyneitä tekoälyalgoritmeja, jotka etsivät harmaasävyelementeistä hienovaraisia muutoksia, jotka ovat usein syväväärennöksen merkki. Luottamuspisteen avulla käyttäjä voi määrittää onko media aito vai ei. (McFarland, 2024)

Stanfordin ja Kalifornian yliopiston tutkijat ovat kehittäneet innovatiivisen tekniikan nimeltään **foneemi-viseemi-epäsopivuus**. Viseemit kertovat suun muodon dynamiikasta ja foneemit

puhutusta. Nämä ovat joskus epäjohdonmukaisia tai erilaisia. Tämä epäjohdonmukaisuus on syvävääräennöksien yleisin ongelma, tekoälyn pyrkiessä sovittamaan täydellisesti yhteen puhutut sanat ja suun liikkeet. Foneemi-viseemi-tekniikka analysoi videoita edistyneiden tekoälyalgoritmien avulla pyrkien löytämään epäjohdonmukaisuuksia. Tekniikka vertaa suun liikkeitä (viseemit) ja puhuttuja sanoja (foneemit) pyrkien etsimään ristiriitoja. Jos havaitaan epäsuhta, se on yleensä osoitus siitä, että kyseessä on väärennös. (McFarland, 2024)

3.5 Puolustautumismenetelmien toimivuus ja luotettavuus

Tässä luvussa tarkastellaan edellä mainittujen puolustautumismenetelmien toimivuutta ja luotettavuutta. Kaksi tai useampi vaiheinen tunnistus on todettu hyväksi suojautumistavaksi useimpia kyberrikoksia vastaan ja se onkin käytössä useissa eri palveluita tarjoavissa yrityksissä. Myös eri järjestelmien päivitykset ovat suositeltava ja tehokas tapa suojautua. Ongelma päivityksissä on se, että ne tulevat viiveellä. Mahdollista tietoturva-aukkoa voidaan käyttää hyökkäykseen ennen päivityksen tuloa tai mikäli päivitystä ei ole tehty. Tietoverkkojen suojauksia lisäämällä esimerkiksi työpaikoilla saatetaan tulla siihen tilanteeseen, että suojaus menee liian tehokkaaksi ja ihmiset eivät pysty enää tekemään työtään. On myös mahdollista, että liika suojautuminen aiheuttaa sen, että jokin tai jotkut ohjelmat eivät enää toimi. Tämän vuoksi onkin jatkuvaa tasapainoilua yrityksillä saada suojaukset oikeanlaisiksi. Tietoturvaohjelmistoja ja virustorjuntaohjelmia käytettäessä on myös mahdollista, että ne antavat niin kutsuttuja vääriä hälytyksiä. Tämä tarkoittaa, että ohjelma on tunnistavinaan viruksen, vaikka sellaista ei ole. Väärä hälytys saattaa aiheutua myös laillisen ohjelman yrittäessä päästä internettiin vain tarkistaakseen esimerkiksi päivityksiä. On myös mahdollista, että esimerkiksi virustorjuntaohjelmisto ei päästä käyttäjää jollekin sivustolle, joka on turvallinen, koska se olettaa kyseessä olevan haitallinen sivusto.

Myöskään vaikeinkaan salasana ei anna täydellistä suojausta ja turvaa. Hakkereilla on käytössään ohjelmia, jotka testaavat kaikki mahdolliset merkit jokaiseen salasanan merkin kohdalle ja löytää lopulta oikean merkin ja siten salasanan. Salasanan ollessa pitkä ja sisältäen erikoismerkkejä, prosessi vie aikaa. Jos käyttäjä sattuu vaihtamaan salasanaa oikeaan aikaan koko hakkerin työ saattaa mennä hukkaan ja prosessi on aloitettava alusta.

Deepfakeja tunnistavien työkalujen toiminta ei suinkaan ole aukotonta. Eri ohjelmistot tunnistavat eri asioita kuvista ja videoista ja näin ollen onkin mahdollista, että deepfake on muokattu siten, ettei juuri kyseinen ohjelma tunnista sitä väärennökseksi. Jokin toinen ohjelma voisi tunnistaa, mutta yleensä yksi yritys tai ihminen käyttää vai yhtä ohjelmaa. Lisäksi ongelma tunnistusohjelmissa on se, että tekoälyteknologian kehittyminen on nopeaa ja jos päivityksiä kyseiseen ohjelmaan ei tule jatkuvasti saattaa sen tunnistus heikentyä.

Deloitten mukaan vuonna 2024 on kaksi muita parempaa teknologiaa deepfakejen tunnistamiseen. Ensimmäinen on vesileimateknologia, jota Google käyttää omassa SynthID deepfake-tunnistusohjelmassaan ja toinen on photo plethysmography-teknologia, joka on puolestaan käytössä Intelin FakeCatcher tunnistusohjelmassa. (Deloitte, 2023)

Vuonna 2021 Grohin ryhmä teki tutkimuksen, jonka mukaan ihmiset tunnistivat väärennökset yhtä hyvin kuin konenäköohjelmisto, vaikka ihmiset tekivät erilaisia virheitä konenäköön verrattuna. Tutkimuksessa havaittiin myös, että ihmiset havaitsivat koneita huonommin väärennöksen, jos kuva oli käännetty ylösalaisin ja silmät peitetty. (Growth& Epstein& Firestone&Pickard, 2021)

Mihin kyberturvallisuuden osa-alueeseen sitten yritysten sekä kansallisen toimijan tulisi keskittyä koulutuksessa ja ihmisten valmistamisessa deepfake-huijauksiin? Tähän ei ole yksiselitteistä vastausta. Minkä tahansa puolustautumisteknologian tai -menetelmän toimivuus perustuu siihen, että se on oppiva järjestelmä sekä sen kyvykkyyttä arvioidaan jatkuvasti. Fyysikko Stephen Hawkings kertoi jo vuonna 2017 Wired-lehden julkaisussa, että AI tullee korvaamaan ihmiset. Hän totesi haastattelussa, että koska ihminen pystyy ohjelmoimaan tietokoneviruksen, joku kykenee luomaan tekoälyn, joka replikoi itsensä (Medeiros, 2017). Uhka on todellinen, koska neuroverkkojen kehittyessä äärimmilleen ne eivät eroa ihmisaivoista kuin ylivoimaiselta prosessointikyvyillään.

Yritysten kannalta suojautumisessa deepfakeja vastaan olennaista on ymmärtää kyberrikollisuuden moninainen spektri. Koko henkilöstön eikä vain IT-osaston tulee käsittää, että kyberturvallisuushygienian pohjana on yrityksen tietoturvallisuusympäristö ja -arkkitehtuuri. Viaton Whatsapp-huijaus, informaatiovaikuttaminen, sosiaalinen manipulointi voivat kaikki olla pohjana hyvin valmistelulle Teams-puhelulle, jossa toinen osapuoli onkin deepfake-huijauksen kohteena. Kouluttaminen, tiedon jakaminen ja valistaminen kyberuhkista sekä yrityksen työntekijöille että yksittäiselle kansalaiselle on tärkein suojautumiskeino. Ihminen on koko ketjussa heikoin lenkki ja näin tulee olemaan jatkossakin deepfake-teknologian kehittyessä.

Kyberturvallisuuskeskuksen asiantuntijoiden mukaan organisaatioissa on pääsääntöisesti varauduttu kyberuhkiin hyvin. Myös asiantuntijatiimi korostaa riskien tiedostamisen, niihin varautumisen, henkilöstön kouluttamisen sekä selkeiden prosessien olevan avainasemassa deepfakejen uhkaa torjuttaessa. Valveutuneiden työntekijöiden sekä laskumaksuprosessin ollessa selkeä, on Suomessakin yritettyjen ääniväärennösten läpimeno haastavaa. Yhteiskunnallisesti tärkeää olisi yleisen medialukutaidon kasvattaminen sekä tietoisuuden lisääminen nykyisestään mediaväärennösten suhteen. (Autio ym, haastattelu, 30.8.2024)

4 Deepfake-huijauksen tekeminen ja tunnistaminen

Tässä kappaleessa kerrotaan työn empiirisestä osuudesta, jossa luotiin koetilanne 25 koehenkilölle. Testissä oli tarkoitus tunnistaan tekoälyllä muokatut kuvat, deepfake ääninäytteet ja deepfake-videot. Kerätystä datasta tehtiin tilastollinen analyysi, kuka tai ketkä tunnistavat aidon AI:n tekemästä tallenteesta.

Finanssiala ry:n tekemässä tutkimuksessa todettiin, että valtaosa suomalaisista, lähes 70 % on joutunut huijausyrityksen kohteeksi. Analyysin mukaan yli 160 000 suomalaista on myös menettänyt itselleen merkittävän summan rahaa. Tutkimuksen mukaan huijauksen kohteeksi joutuu kaiken ikäisiä, sukupuolisia ja kaikkien ammattiryhmien edustajia. Tutkimuksessa selvisi myös, että nuoret ikäpolvet joutuvat huijausten kohteeksi hieman varttuneempia useammin (Finanssiala ry, 2024). Myös Keskusrikospoliisin asiantuntijat vahvistavat, että huijausten ja niiden yritysten kohteeksi on joutunut kaikista ikäryhmistä suomalaisia. Heidän mukaansa myös ns. diginatiivisukupolven edustaja voi haksahtaa huijaukseen yhtä helposti kuin varttuneemmatkin kansalaiset. (Keskusrikospoliisin kyberrikostorjuntayksikkö, 2024)

4.1 Deepfake-videon ja äänitallenteen luominen

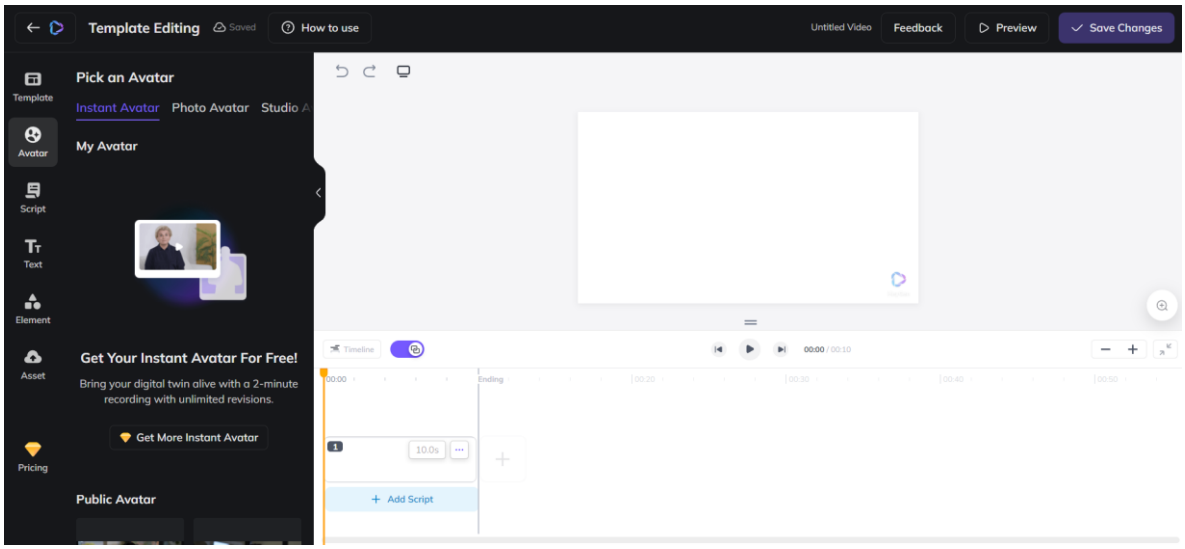
Kuinka helppoa sitten on deepfake-videoiden ja äänitalleteiden tekeminen? Kuinka hyvin ihmiset tunnistavat nämä? Työn tässä osuudessa on testattu kahta julkisesti saatavilla olevaa videoiden tekemiseen tarkoitettua sivustoa.

Testiä varten luotiin samalla sivustolla äänitallenteita ja videoita. Internetissä on saatavilla useita ohjelmia ja tässä työssä testattiin kahta. Ensimmäinen eli Heygen valittiin Ismo Turpeen, HAMKn ohjaajan suosituksesta ja toinen Deepbrain AI Aleksi Blomqvistin Kyberturvallisuuskeskuksen ohjaajan suosituksesta. Kuvat olivat satunnaisesti valittuja internetistä.

4.1.1 Heygen - www.heygen.com

Ensimmäiseksi testattu ohjelma deepfake-videoiden tekemiseen oli Heygen. Ohjelma vaati kirjautumisen Google- tai Facebook-tunnuksilla ja tunnuksen voi luoda nopeasti. Tunnuksen voi luoda myös suoraan sähköpostiosoitteella ja salasanalla. Rekisteröintivaiheessa Heygenissä oli runsaasti kysymyksiä tiliin liittyen ja tämä hidasti tunnuksen luomista. Tilin tekemisen jälkeen päästiin pääsivulle ja videon luominen voi alkaa. Videon tekeminen Heygenissä tapahtuu AI Studiolla, jonka käynnistäminen on suhteellisen vaivatonta ja helppoa. Kuva 7 on AI Studion aloitusnäkymä videoeditorin käynnistyessä.

Kuva 7. Heygenin AI Studio -videoeditori.



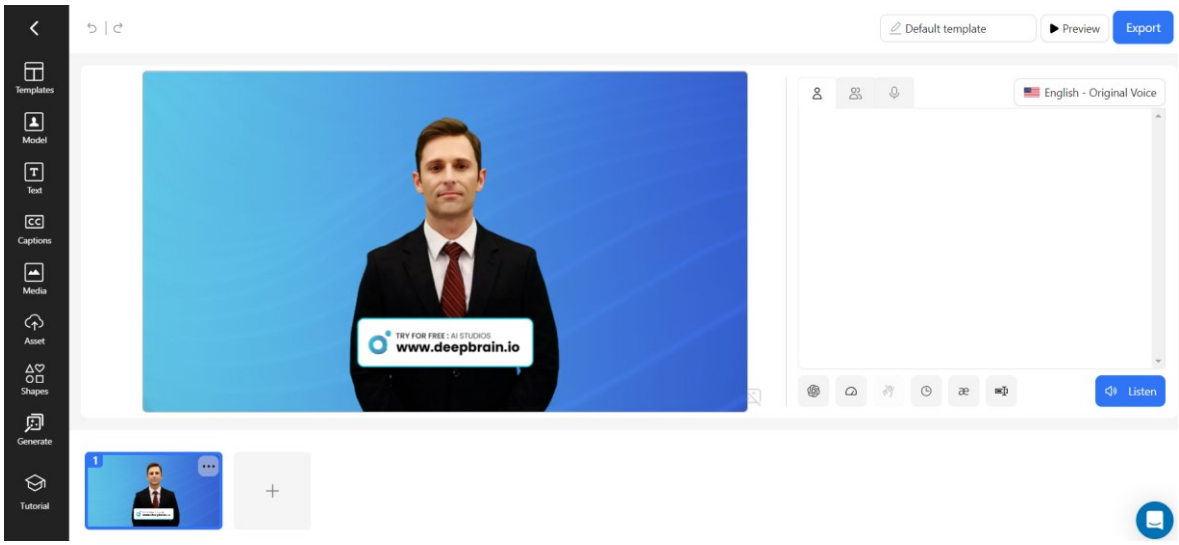
AI Studion käyttöliittymä on yksinkertainen ja helppo käyttää. Videon Heygenissä voi luoda joko valmiiseen tai tyhjään videopohjaan eli templateen. Ilmaisversiossa on muutamia valmiita avattaria eli kasvoja, joita voi käyttää. Myös ääniä on useilla kielillä, ne ovat monipuolisia ja mukana oli myös kaksi suomenkielistä ääninäyttelijää. Skriptin eli hahmon puheen luominen ja kirjoittaminen on helppoa ja kirjoitetun puheen pystyy kuuntelemaan etukäteen esikuuntelusta.

Heygenissä on käytössä krediitit, joilla maksetaan videon tekemisestä. Käyttäjätunnuksen luonnin yhteydessä saa yhden krediitin ja jokaisen cyclen eli tietyn ajanjakson lopussa saa yhden krediitin lisää. Yhdellä krediitillä voi tehdä minuutin pituisen videon. Ilmaiskrediiteillä Heygen ei kuitenkaan anna tehdä videota, vaikka krediitit siihen riittäisivätkin. Videon luomiseen vaaditaan Creator-tason tunnus, joka maksaa 29 dollaria kuukaudessa. Tässä vaiheessa päättyi videon tekeminen ilmaiseksi ja siirryttiin kokeilemaan seuraavaa ohjelmavaihtoehtoa. Haluttiin nähdä videon laatu ja sivuston toimivuus ennen ohjelmasta maksamista. Ohjelma ei kuitenkaan antanut tehdä ilmaiseksi videota ja näin ollen videon laatua ei voinut ennen maksamista varmistaa.

4.1.2 Deepbrain AI – www.deepbrain.io

Seuraavaksi vaihtoehdoksi valikoitui Deepbrain AI. Myös Deepbrain AI vaatii kirjautumisen ja kuten Heygenissäkin tunnuksen voi tehdä Googlen tai Microsoftin tunnuksilla. Tunnus voidaan luoda myös pelkällä sähköpostiosoitteella. Kun tunnus on luotu, voidaan videon tekeminen aloittaa helposti. Kuva 8 on juuri aloitettu uusi projekti ja videoeditorin aloitusnäkymä.

Kuva 8. Deepbrain AI Studios -videoeditori.



Video voidaan luoda valmiista tai tyhjästä templatesta ja kuten Heygenissäkin editorin käyttö on helppoa ja vaivatonta aloittelijallekin. Deepbrain tarjoaa muutamia malleja eli hahmovaihtoehtoja ilmaiseksi. Myös ääniä eri kielillä on tarjolla runsaasti ja käytössä on Heygenin tapaan kaksi suomenkielistä ääninäyttelijää. Skriptin eli hahmon puheen kirjoittaminen on helppoa, nopeaa ja äänen pystyy kuuntelemaan esikatselusta ennen varsinaisen videon tekemistä. Kun videon luominen aloitetaan, voidaan luoda koko video tai pelkkä ääni. Myös Deepbrainissa on käytössä krediitit videoiden luomiseen. Ilmaiseksi saa tehdä yhden minuutin edestä videoita ja Deepbrain tekee videon ilmaiseksi vähentäen videon pituuden jäljellä olevista krediiteistä. Noin 5 sekunnin pituisen videon luomiseen kului aikaa noin 15 minuuttia ja Deepbrain lähetti sähköpostin, kun käännös oli valmis. Koska käytettiin ilmaisversiota Deepbrainista, Deepbrain lisää kaikkiin videoihin oman vesileimansa. Vesileiman saa pois, kun ostaa personal tunnuksen. Tunnus maksaa halvimmillaan kuukausilaskutuksella 29 dollaria (noin 27 euroa) ja sillä saa luoda 15 minuutin edestä videoita kuukaudessa. Vuosilaskutuksella hinta on 24 dollaria kuukaudessa. Ostaessa tunnusta voi valita myös enemmän aikaa luotaviin videoihin kuukaudessa. Maksimi tälle on 60 minuuttia ja hinta toki nelinkertainen tuohon halvimpaan eli 15 minuuttiin nähden (119 dollaria).

Deepfake-videon luominen ilmaiseksi kuitenkin onnistui, joten suhteellisen pitkälle pääsi ilman maksamista. Video on melko hyvänlaatuinen ja animaatiot pääosin uskottavia. Hahmon suun liikkeet ääneen nähden eivät toimi täydellisesti ja ne mahdollistavat väärennöksen tunnistamisen helpommin. Koska Deepbrain tarjoaa tuolla halvimmalla personal tunnuksella mahdollisuuden lisätä oman avattaren ja oman äänen, valittiin se pohjaksi empiiriselle kokeelle. Kuukausimaksun maksamisen jälkeen Deepbrain aktivoi

automaattisesti personal tunnuksen. Maksutapahtuma oli helppo ja nopea. Maksun jälkeen Deepbrain päivitti heti krediitit sekä jäljellä olevan ajan tilille.

Maksun maksamisen jälkeen tarkoitus oli tehdä oma avatar eli oma hahmo. Deepbrainin ohjeissa luki, että itsestä pitää tehdä kahden minuutin pituinen video, jossa lauseiden välissä pitää suu laittaa kiinni. Videon voi tehdä joko Deepbrainin nettisivuilla tai lähettämällä Deepbrainiin valmiiksi tehdyn videon. Etukäteen tehdyn videon lähettämiseen Deepbrainiin kului jonkin aikaa, mutta se onnistui lopulta hyvin.

Seuraavaksi oli tehtävä puheen eli skriptin kirjoitus avattarelle. Skriptiä testatessa huomattiin, ettei avatar osaa toistaa numeroita, jos numeron antaa numerona, mutta kun kirjoitti numeron sanana, niin avatar toisti sen oikein. Deepbrainissa on tekstinluontityökalussa myös joitain kiellettyjä sanoja. Tällaisen sanan ollessa tekstissä, ohjelma antaa virheilmoituksen ja sanaa ei toisteta. Yllätykseksi tällaisia sanoja eivät ole esimerkiksi suomenkieliset kirosanat, mutta yksi tällainen kielletty sana on ”natsi” ja kaikki siihen liittyvät muodot. Puheen valmistuttua laitettiin video valmistumaan. Videon valmistuminen kesti vajaan puoli tuntia ja sillä oli pituutta 24 sekuntia. Voidaan siis karkeasti sanoa, että yhden sekunnin videon valmistuminen kestää minuutin ja tästä voi laskea käännökseen kuluvan ajan eli esimerkiksi yhden minuutin pituisen videon valmistuminen kestää yhden tunnin.

Videon laatu oli kaiken kaikkiaan hyvä ja avattaren liikkeet luonnollisia suun ja äänen synkronointia lukuun ottamatta. Väärennöksen voi siis tarkasti katsomalla erottaa suun liikkeistä puheen suhteen. Niissä on epätasaisuutta puheen kanssa. Myös pelkän äänen tekeminen sujui helposti ja tämän tunnistaminen väärennökseksi saattaa ollakin hankalampaa, koska kuvaa ja erityisesti suun liikettä ei testattava näe. Toisen näyttelijän luvalla tehtiin myös avatar, jota käytettiin myös ääniin ja videoihin laajemman ja erilaisemman testimateriaalin saamiseksi empiiriseen osuuteen.

4.2 Deepfake-videoiden ja äänien tunnistamisen testaaminen

Koetilanteeseen luotiin deepfake-videoita, -äänitallenteita sekä internetistä löytyviä kuvapareja. Koehenkilöjoukkoon valittiin lähipiiristä henkilöitä ja heille näytettiin valvotuissa olosuhteissa aineisto tunnistamista varten. Kaikki koehenkilöt tunsivat ainakin toisen ääni- ja videonäytteen tekijän. Kuvaparit olivat kaikille entuudestaan vieraita.

Empiiriseen kokeeseen luotiin koeaineisto ja tehtyjen kokeiden perusteella on arvioitu deepfake-videon ja -kuvan erottamista aidoista. Koehenkilöjoukko oli anonymisoitu.

4.2.1 Testiaineiston luominen empiiriseen kokeeseen

Testiaineisto koostui kolmesta kokeesta: viidestä kuvaparista, kahdesta äänitallenneparista eli neljästä äänitallenteesta ja neljästä videotallenneparista eli kahdeksasta videosta. Kuvat olivat valmiita ja ääni- sekä videotallenteet oli luotu itse.

Ensimmäistä koetta varten etsittiin internetistä viisi kuvaparia, joissa toinen kuva oli aito ja toinen muokattu. Kuvaparit oli haettu Youtube-videoista ja ne löytyvät liitteestä 2. Kuviin pyrittiin valitsemaan muutama tunnettu henkilö, jotta koetta tehtäessä testattava tunnistaisi henkilön. Yksi kuvapareista oli ei-tunnettu henkilö. Kuvapariin 1A ja 1B valittu tarkoituksella kuvat siten, että muokattu kuva oli alkuperäistä huonomman näköinen. Oletusarvo on testitilanteessa, että manipuloitu kuva on alkuperäistä parempi ja henkilö paremman näköinen.

Deepfake-videot ja deepfake-äänitallenteet tehtiin Deepbrain AI sivustolla, aidot videot Microsoftin Kamera -ohjelmalla ja aidot äänet Microsoftin Sound Recorder -ohjelmalla. Deepfake- ja aitoja videoita luodessa pyrittiin siihen, että kaikki ulkoinen näkymä videoissa olisi mahdollisimman samanlaista. Deepfake-videota varten tehty avatar sekä aito video kuvattiin samassa kohdassa, jotta tausta olisi sama. Samoin kummassakin tilanteessa näyttelijöillä oli samat vaatteet päällään, hiukset samalla tavalla, jotta mikään muukaan ulkoinen seikka erottaisi tilanteita. Ääninäyttelijät lukivat tekstit sekä aitoihin videoihin että Deepbrain AI:n avattaren luontivideoon mahdollisimman samanlaisella äänellä sekä kasvojen ilmeillä.

Sekä video- että äänitallenteita varten kirjoitettiin tarinat aiheista, joista ihmisillä on yleensä selkeä mielipide. Jokaista tarinaa varten kirjoitettiin yhdestä aiheesta vastakkaiset mielipiteet tekstiin. Tällä pyrittiin siihen, että kukaan testattava ei arvioisi tarinan perusteella onko tallenne aito vai epäaito; onko puhuja aidosti tätä mieltä vai ei. Liitteessä 3 on äänitallenteiden ja liitteessä 4 videotallenteiden tekstit.

4.2.2 Koetilanne ja testiaineiston kerääminen

Koehenkilöt eivät olleet satunnaisesti valittuja vaan kokeeseen valittiin tuttuja henkilöitä. Osa koehenkilöistä osallistui Teams-kokouksen kautta testiin. Jokainen koehenkilö tunsi ainakin toisen ääninäyttelijän jollain tasolla. Tämä saattoi myös vaikuttaa testin tulokseen. Tätä on analysoitu tarkemmin luvussa 4.3.

Testiaineiston keruuta varten luotiin Excel-taulukko, joka on liitteessä 5. Kokeen aluksi kerrottiin henkilölle kokeen tulevan osaksi Sami Rintalan opinnäytetyötä, tietojen olevan anonymisoituja ja tärkeää ei ole, menevätkö arvioit oikein vai väärin. Aluksi kerättiin henkilön tiedot, jolla data saatiin ryhmiteltyä. Koeaineiston perusteella haluttiin tietää, vaikuttaako koehenkilön ikä, sukupuoli, koulutustaso tai aiempi tietämys deepfakesta tunnistamiskyvykkyyteen. Testi aloitettiin kysymällä tietäänkö henkilö, mitä tarkoittaa deepfake. Mikäli henkilö ei tiennyt vastausta, hänelle kerrottiin mitä termillä tarkoitetaan. Ennen ensimmäisen testin aloittamista, henkilölle kerrottiin koejärjestelystä.

Ensimmäinen testi aloitettiin kuvilla ja koehenkilölle kerrottiin, että toinen kuvaparin kuvista on aito ja toinen on muokattu. Hänen tehtävänsä oli kertoa, kumman arveli olevan kumpi. Kuvat näytettiin yksi kuvapari kerrallaan, siten että henkilö näki vain yhden kuvaparin kerrallaan. Tulokset tallennettiin kunkin kuvaparin jälkeen. Mikäli henkilö halusi muuttaa vastaustaan, se oli mahdollista.

Toinen testi oli tunnistaa ääninäytteet. Koehenkilölle kerrottiin ennen ensimmäistä ääntä, että kuulet seuraavaksi neljä ääninäytettä, joissa kahdessa ensimmäisessä on ääninäyttelijä numero yksi ja kahdessa seuraavassa ääninäyttelijä numero kaksi. Koehenkilölle kerrottiin, että äänistä saattavat kaikki olla aitoja, kaikki AI:lla luotuja tai osa aitoja ja osa AI:lla luotuja. Henkilölle kerrottiin myös, että ääninäytteiden tarinat ovat kaksi vastakkaista mielipidettä ja hänen tunnistuksensa ei pitäisi riippua, mitä mieltä hän itse on asiasta. Äänet soitettiin järjestyksessä ja henkilön arviot tallennettiin jokaisen ääninäytteen jälkeen. Henkilöllä oli mahdollista muuttaa aiemman ääninäytteen arvioitaan myöhemmin kuullun ääninäytteen jälkeen. Hänellä oli myös mahdollista kuunnella ääninäyte useampaan kertaan.

Kolmas testi oli tunnistaa deepfake-videot aidoista videoista. Samoin kuin äänitallenteissa henkilölle kerrottiin ennen ensimmäistä videota, että kaikki videot saattavat olla aitoja, kaikki AI:lla luotuja tai osa aitoja ja osa AI:lla luotuja. Henkilöllä oli mahdollisuus muuttaa mielipidettään myöhemmin ja heti henkilön kertoessa arvionsa, video pysäytettiin ja siirryttiin seuraavaan.

Yksittäinen testi kesti keskimäärin noin kymmenen minuuttia. Koetilanteessa ääni- ja videonäytteiden jälkeen koehenkilöiltä ei kysytty ääninäytteen tai videon sisällöstä. Tällä kysymyksellä olisi voitu varmistaa henkilön keskittyneen sekä asiasisältöön eikä pelkästään mahdolliseen deepfaken tunnistamiseen. Suurin osa koehenkilöistä kuitenkin kommentoi sisältöä, joten tästä voidaan päätellä henkilön myös kuunnelleen itse sisältöä.

Otosmäärä tässä tutkimuksessa jäi suhteellisen pieneksi, n=25. Tämän vuoksi tulosten pohjalta ei ole mahdollista tehdä johtopäätöksiä, joita voitaisiin käyttää pohjana esimerkiksi kansallisella tasolla.

4.3 Testitulosten analysointi

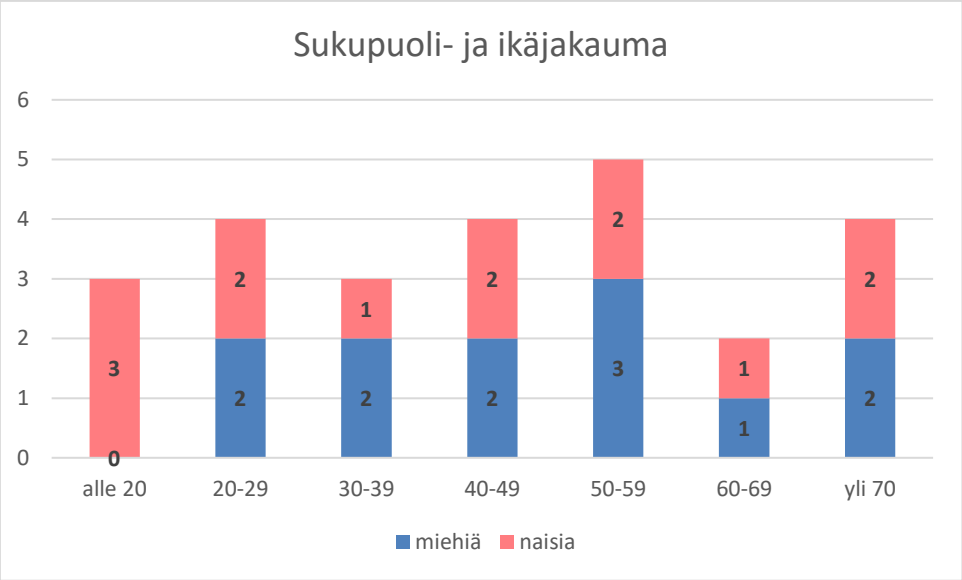
Testit tehtiin aikavälillä 1.7. – 17.7.2024. Kaksi testeistä tehtiin Teams-yhteyden välityksellä ja kaikissa muissa testattava oli samassa tilassa testaajan kanssa. Ensimmäiseksi testattavilta kysyttiin, onko termi deepfake entuudestaan tuttu ja mitä he sillä ymmärtävät: Mikäli termi ei ollut tuttu, se selitettiin.

Toisena testin tehtävänä pyrittiin tunnistamaan kuvaparit. Tämän jälkeen ääninäytteet ja lopuksi vielä videotallenteet. Testiin osallistuneiden kokonaismäärä oli 25. Jokaisessa koulutustasossa ja ikäluokassa oli useampi vastaaja. Testiin osallistuneiden sukupuolijakauma oli tasainen: vastaajista 12 oli miestä ja 13 naista. Kukaan vastaajista ei identifioinut itseään muunsukupuoliseksi.

4.3.1 Deepfake-käsitteen ja kuvaparien tunnistaminen

Etukäteisymmärrys deepfakesta jakautui tasaisesti sukupuolten kesken. Noin puolet vastaajista ei tunnistanut deepfake-termiä. Tunnistaneiden osuudesta noin puolet oli naisia ja puolet miehiä. Tarkasteltaessa koulutustasoa selittävänä tekijänä, ei myöskään sillä nähdä olevan vaikutusta. Kummassakin kategoriassa – tietäänkö mikä deepfake on – oli testattavia kaikista koulutusryhmistä. Niin ikään käytettäessä ikää selittävänä tekijänä, myös tällöin termin tunteminen jakautui kaikkiin ikäluokkiin. Alle 20 vuotiaista ei kukaan tunnistanut termiä etukäteen, mutta koska heistäkin jokainen selityksen jälkeen tunnisti asian, ei ikäkään ole selittävä tekijä. Voidaan siis todeta, että sukupuoli, ikä tai koulutustaso eivät olleet selittäviä tekijöitä termin deepfake-tunnistamisessa otoksessa. Kuva 9 nähdään testiin osallistuneiden sukupuoli- ja ikäjakauma.

Kuva 9. Testattujen sukupuoli- ja ikäjakauma.



4.3.2 Kuvaparien tunnistaminen

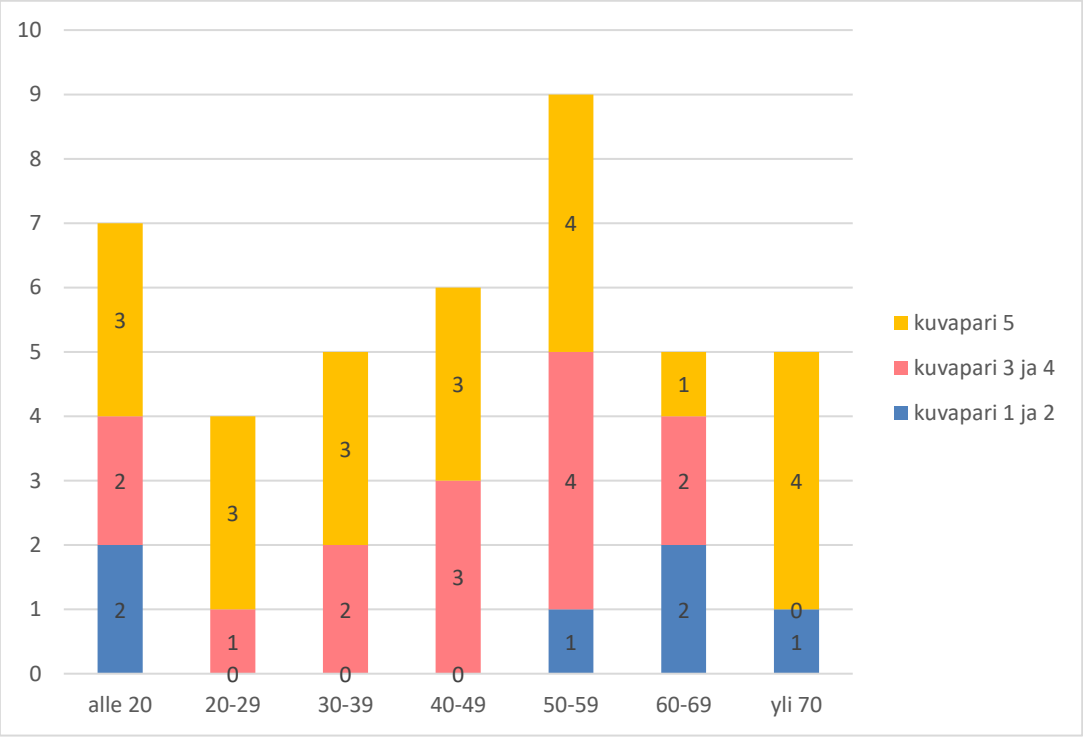
Kuvaparit oli valittu siten, että kuvapareista parit yksi ja kaksi olivat poliitikkoja useamman vuoden takaa. Oletuksena oli, että nuoremmat vastaajat ikäluokasta alle 20 ikäluokkaan 30–39 eivät ehkä tunnista kuvan henkilöitä. Näin myös kommenteista kävi ilmi. Tätä tietoa ei kuitenkaan tallennettu. Vastaavasti kuvapareissa pari kolme ja neljä olivat nyt mediassa paljon esillä olevia poliittisia henkilöitä, jotka kaikkien vastaajien olisi ajateltu henkilöinä tunnistavan. Kuvapari viisi oli taas nimetön nuori nainen, jonka kuva on muokattu nukeksi tai mallikuvaksi. Kuvapareissa numero kaksi osoittautui kaikista vaikeimmaksi ja kuvapari viisi kaikista helpoimmaksi. Alle puolet eli 10 vastaajaa tunnisti Iso-Britannian entisen pääministeri Theresa Mayn oikean kuvan muokatusta. Vain neljä vastaajaa ei tunnistanut kuvaparissa viisi ollutta muokattua nuoren naisen kuvaa aidosta. Kolme vastaajista tunnisti kaikki kuvaparit oikein ja kaikilla vastaajilla oli vähintään kaksi kuvapareista tunnistettu oikein.

Yksi vastaaja kommentoi kuvaparien tunnistamisen jälkeen, että hän etsi kuvista deepfakella tehtyjä muutoksia. Hän ei ollut ymmärtänyt, että kuvat voivat olla manipuloituja myös jollain kuvanmuokkausohjelmalla. Tätä selitystä tarkennettiin seuraaviin testeihin. Hän arveli tämän vuoksi itse tunnistaneensa kolme kuvaparia viidestä väärin, koska etsi deepfakella tehtyjä muokkauksia kaulan ja sormien alueelle; nämä ovat haastavimpia tietokoneelle muokata.

Hypoteesi, että vanhempien vastaajien kategoriat tunnistaisivat helpommin kuvaparit yksi ja kaksi, nuoremmat kuvaparin viisi sekä kummankin tunnistaessa yhtä hyvin kuvaparit kolme ja neljä, toteutui osittain. Ikäluokista 20–29, 30–39 ja 40–49 ei kukaan tunnistanut aitoa

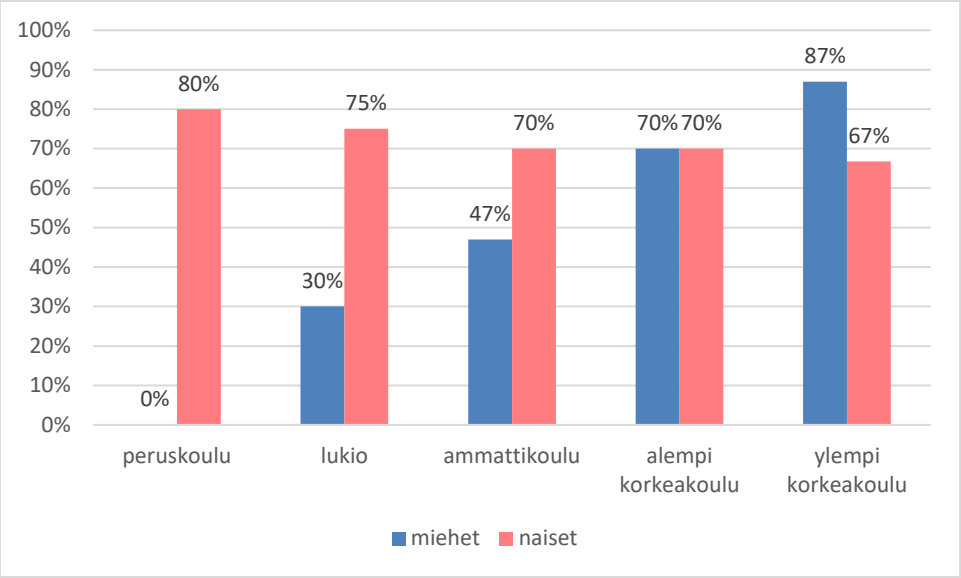
kuvaa kuvaparista yksi ja kaksi, mutta yli 70-vuotiaat eivät myöskään tunnistaneet aitoja väärennöksistä. Sen sijaan kuvaparia viisi tunnistettiin kaikissa ikäluokissa. Kuva 10 nähdään, miten hyvin eri ikäluokat tunnistivat testikuvat.

Kuva 10. Testikuvien tunnistaminen eri ikäluokissa.



Testikuvien tunnistamisen osalta tarkasteltiin lopuksi, onko sukupuolella ja koulutuksella merkitystä kuvien tunnistamisen onnistumisessa. Kuva 11 nähdään, että naisilla koulutustasolla ei juurikaan ole vaikutusta testikuvien tunnistamisen osalta, mutta miehillä selkeästi koulutustaso korreloi muokattujen kuvien tunnistamisessa. Mitä korkeampi koulutus, sitä paremmin testikuvien tunnistaminen onnistui.

Kuva 11. Koulutustaso sukupuolittain testikuvien tunnistamisessa.



Kuvaparien tunnistamisen osalta ainoana korreloivana tekijänä voidaan tämän otoksen perusteella todeta miehillä olevan koulutustason.

4.3.3 Deepfake ääninäytteet ja niiden tunnistaminen

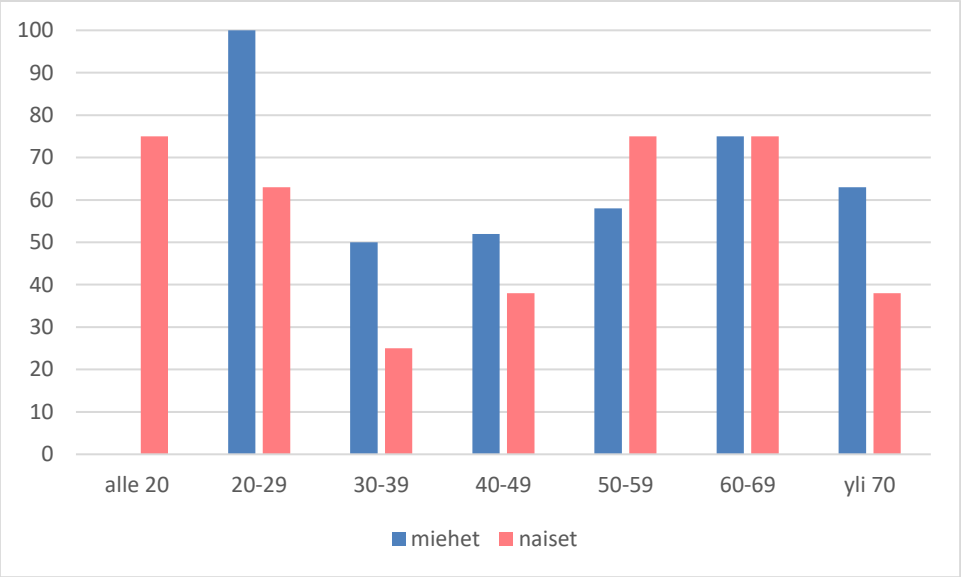
Ääninäytteiden osalta testattavilta tulleita kommentteja oli muun muassa, että ne olivat koko sarjasta vaikeimmat. Deepbrain AI on suhteellisen hyvin onnistunut äänen väärentämisessä. Testituloksiin ei tallennettu, kuinka hyvin henkilö tunsi ääninäytteessä puhuneen henkilön. Tämä saattoi jollain tasolla vaikuttaa tuloksiin: mitä tutumpi henkilö, sitä paremmin erotti aidon äänen koneen luomasta. Ääninäytteitä kuunnellessa Deepbrain AI:n tekemät äänet olivat hiukan robottimaisia, mutta testejä tehdessä sekoittui välillä se, oliko kyseessä AI:n luoma vai aito ääni.

Kaikista koehenkilöistä vain viisi pystyi erottamaan kaikki aidot äänet vääristä ja kuusi koehenkilöä sai vain yhden oikein. Kaikista ääninäytteistä äänet numero yksi ja kolme olivat vaikeimmin tunnistettavia ja ne erottivat koneen luomaksi ääneksi 14 testattavaa eli 56 %. Helpoimmaksi osoittautui ääni numero kaksi, jonka tunnisti aidoksi 19 testattavaa eli 76 % kaikista testattavista. Äänet yksi ja kaksi oli puhunut sama henkilö. Samoin äänet kolme ja neljä puhui sama henkilö. Ääninäytteet yksi, kolme ja neljä olivat tietokoneen luomia ja vain ääninäyte kaksi oli aito. Kuusi henkilöä tulkitsi aidon ääninäytteen koneella tehdyksi mukaan lukien lukijan oma äiti.

Vastaavasti kuin kuvaparien tunnistamisessa lähdettiin ääninäytteissä liikkeelle tutkimalla, korreloiko sukupuoli, ikä tai koulutustaso siihen, miten hyvin koehenkilö tunnisti aidot äänet.

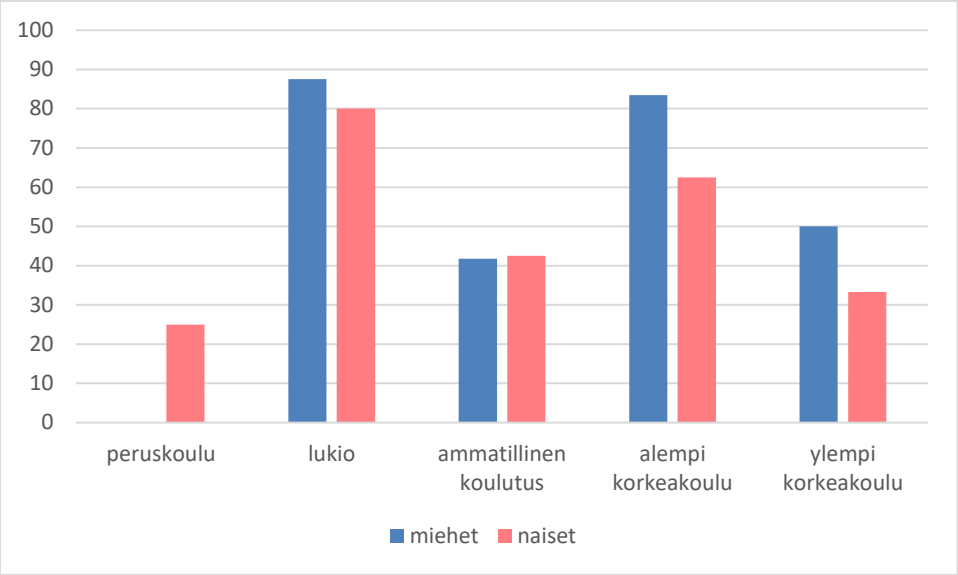
Kuva 12 on kuvattu miesten ja naisen kyvykkyyttä tunnistaa ääninäytteet oikeiksi ja tietokoneen tekemiksi. Y-akseli kuvaa prosentuaalista osuutta ikäluokasta, jotka tunnistivat äänet oikein. Ikäluokissa 20–29, 30–39, 40–49 sekä yli 70-vuotiaat, miehet ovat tunnistaneet hiukan naisia paremmin aidot äänet koneen tekemistä. Ainoastaan ikäluokassa 50–59 naiset ovat erottaneet äänet miehiä paremmin.

Kuva 12. Sukupuolten välinen ero ikäluokittain ääninäytteiden tunnistamisessa.



Tutkittaessa koulutuksen vaikutusta oikeiden äänien erottamisessa, miehet koulutustasosta riippumatta onnistuivat hieman naisia paremmin erottamaan koneen luomat äänet aidoista. Kuva 13 nähdään, että ainoastaan ammatillisen koulutuksen omaavilla naiset ovat tunnistaneet näytteet hieman miehiä paremmin. Pystysuoralla akselilla on prosenttiosuus koulutusluokkiin kuuluvista, jotka tunnistivat ääninäytteet oikein.

Kuva 13. Koulutustaustan vaikutus ääninäytteiden tunnistamiseen.

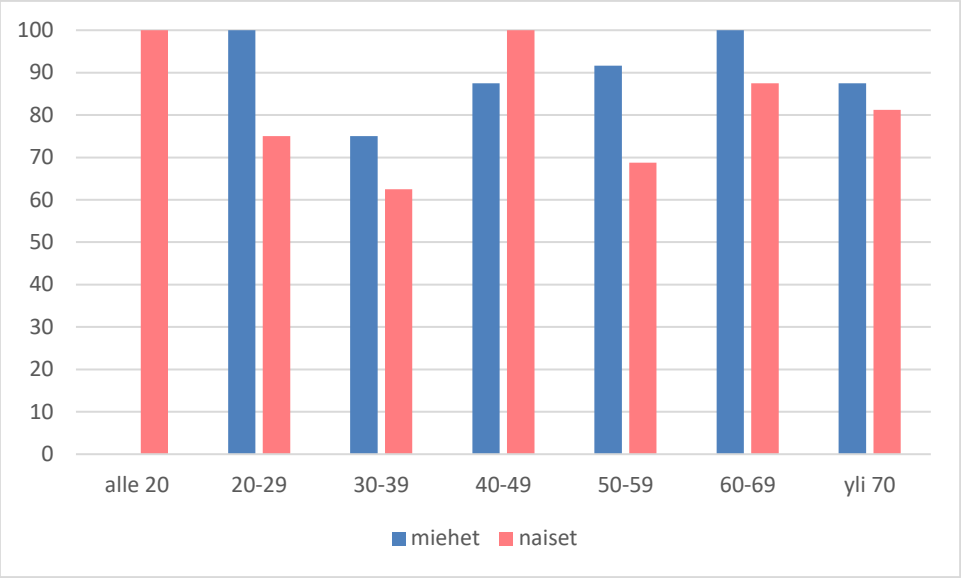


4.3.4 Deepfake-videot ja niiden tunnistaminen

Viimeisenä kokeena koesarjassa oli kahdeksan videon osalta tunnistaa videoista väärät. Videot olivat koesarjassa vastaajien mielestä kaikista helpoimmin tunnistettavia. Kahdeksasta videosta peräti 14 tunnisti kaikki aidot ja väärät. Puolet kaikkien videoiden tunnistaneista oli miehiä ja puolet naisia. Eniten virheitä tehnyt tunnisti hänkin kolme videota kahdeksasta. Keskimäärin tunnistettiin seitsemän videota (6,9). Videoista kolme oli aitoa ja viisi AI:n tekemää. Etenkin videota numero yksi, joka oli aito, arveli viisi vastaajaa AI:n tekemäksi.

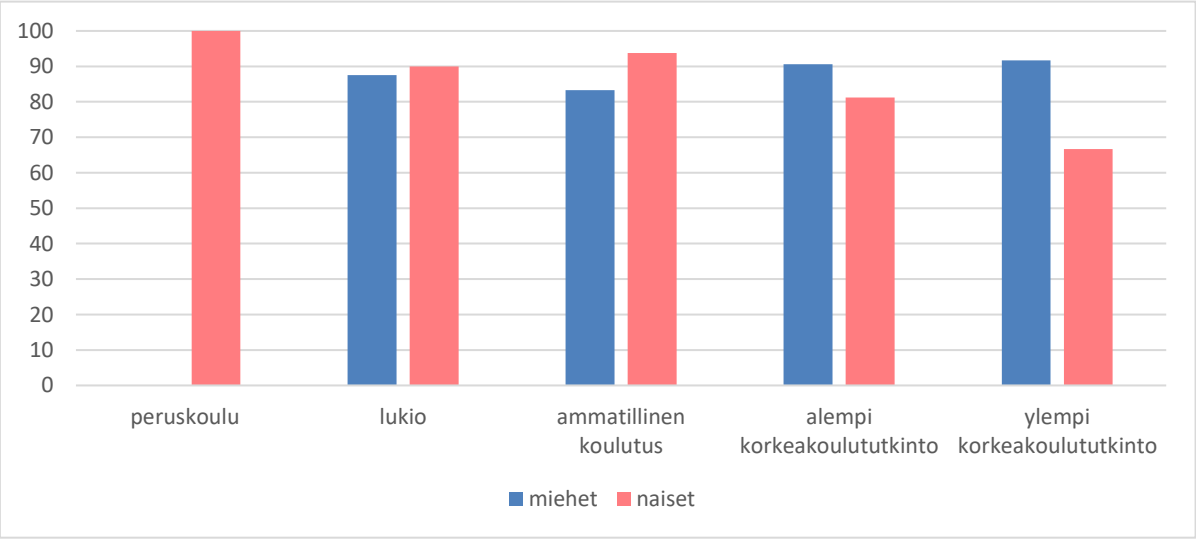
Tarkasteltaessa sukupuolten välisiä eroja eri ikäryhmissä deepfake-videoita tunnistettaessa, voidaan Kuva 14 perusteella arvioida miesten ehkä hieman naisia paremmin tässä testissä erottaneen aidot ja AI:n tekemät. Ainoastaan ikäluokassa 40–49 naisten osuus tunnistaneista oli hieman miehiä suurempi. Kaikissa muista ikäluokissa miehet erottivat videonäytteet naisia paremmin.

Kuva 14. Sukupuolten välinen ero ikäluokittain videonäytteiden tunnistamisessa.



Viimeiseksi analysoitiin kerätyn datan perusteella, voidaanko koulutusta pitää selittävänä tekijänä videoiden tunnistamisessa. Kuva 15 näkyy keskiarvot kummankin sukupuolen osalta koulutustason mukaan. Tässä otoksessa näyttäisi, että miehillä koulutustaso ei juurikaan korreloi tunnistamiseen. Naisten kohdalla silmämääräisesti piirretty sovitesuora olisi laskeva ja näin ollen alemman koulutuksen omaavat, tunnistivat videot korkeammin koulutettuja paremmin. Tässä saattaa olla yhtenä selittävänä tekijänä myös se, että nuoremmat testiin osallistuneet olivat pääsääntöisesti koulutuksensa alkutaipaleella ja näin ollen kuitenkin valveutuneempia kuin opintonsa jo päättäneet ja vanhemman sukupolven edustajat.

Kuva 15. Koulutustason vaikutus sukupuolittain videoiden tunnistamisessa.



4.3.5 Yhteenveto tuloksista

Otos määrän oltua hyvin maltillinen, ei tämän testin tuloksista voida tilastollisesti tehdä päätelmiä. Kuten aiemmissa luvuissa on jo mainittu, kuvien tunnistamisen osalta miehillä koulutustaso vaikutti positiivisesti, kun vastaavasti naiset tunnistivat kuvat koulutustasosta riippumassa yhtä hyvin. Kaikkiaan ikäluokka 50–59 pärjäsi testeissä kaikista parhaiten. Ääninäytteissä miehet iästä ja koulutustasosta riippumatta tunnistivat aidot AI:n tekemistä äänistä naisia paremmin. Videoiden osalta miehillä vaikutti ikä hieman positiivisesti tunnistamiseen, kun taas naisilla hienoinen positiivinen korrelaatio on nähtävissä koulutuksen osalta.

Tilastollisiin tuloksiin vaikutti luonnollisesti se, että kaikki koehenkilöt tunsivat toisen ääni- ja videonäytteiden tekijöistä. Tätä ei kuitenkaan tilastoitu, joten sen merkitystä ei tässä otoksessa tarkasteltu. Suurin vaikutus lienee kuitenkin ollut itse testitilanteella. Koehenkilöille kerrottiin mahdollisista AI:lla muokatuista kuvista, AI:n luomista äänistä ja videoista, joten he osasivat odottaa muokattua materiaalia. Jokaisen testin jälkeen koehenkilöt spontaanisti kertoivat testin vaatineen keskittymistä ja mikäli kyseinen tuotos olisi tullut normaalissa tilanteessa mediavirrassa vastaan, ei sitä välttämättä olisi tunnistanut. Yksi koehenkilö sanoi jopa, että tämän jälkeen hän ei uskalla luottaa mihinkään.

Oma arvio AI:n tekemistä videoista oli, että videot viisi, seitsemän ja kahdeksan olivat laadullisesti parempia kuin videot kaksi ja neljä. Videoissa kaksi ja neljä suun ja puheen synkronoinnin epätasaisuus oli selkeästi nähtävissä. Kolmessa loppupään videossa synkronointi toimi paremmin. Silmien räpyttämiseen ei kukaan kiinnittänyt huomiota, joka sekin oli puutteellista Deepbrain AI:n tekemissä videoissa. Tästä huolimatta jokainen testin tehnyt kertoi suhtautuvansa tulevaisuudessa varovaisemmin kuviin, videoihin ja ääniin, jotka tulevat mediassa vastaan. Kaikki myös ymmärsivät, että kalliimmalla deepfake-ohjelmalla luotujen videoiden ja äänien laatu paranee ja on entistä vaikeammin tunnistettavissa.

5 Tulokset

Fyysikko Stephen Hawkingsin ennusteeseen, jossa AI tulee korvaamaan ihmiset, viitattiin kappaleessa 3. Tulevaisuudenskenaarioon, jossa tietokoneet ottaisivat vallan yli ihmiskunnasta, on kuitenkin vielä matkaa. Tätä päivää on kuitenkin kyberrikollisuus, joka kohdistuu yksittäisiin ihmisiin, ihmisryhmiin, yrityksiin ja valtioihin. Näiden rikosten takana on aina toinen ihminen, ihmisryhmä tai valtio. On arvioitu olevan tapauksia, joissa valtio ylläpitää yrityksen kaltaista organisaatiota keskittyen kyberrikollisuuteen. Tällöinkin takana on aina ihminen tai ihmisryhmä.

Tässä työssä tarkasteltiin yksityiskohtaisesti erilaisia kyberrikollisuuden muotoja WhatsApp-huijauksista identiteettivarkauksiin ja kiristyshaittaohjelmista spoofing-hyökkäyksiin. Myös toimeksiantajan mielestä, työssä eriteltiin ajakohtaiset ja keskeiset kyberrikollisuuden trendit ja muodot kattavasti. Toimeksiantaja myös totesi, että tämän avulla vastattiin tutkimuskysymykseen kyberrikollisuuden yleiskuvasta sekä annettiin kattava kuvaus nykyisistä yleisimmistä kyberrikollisuuden trendeistä.

On tärkeää ymmärtää, että kyberrikollisuuden eri muodot ovat eräänlaista jatkumoa toisilleen. Yksinkertainen, yhteen ihmiseen kohdistettu huijaus on muuttanut muotoaan kohdistuen massoille ja tavoitteena on saada rikollisille mahdollisimman suuri näkyvyys, vaikuttavuus tai rahallinen tuotto. Tästä esimerkkinä voidaan mainita, vaikka yhdelle ihmiselle kohdistettu rahahuijausyritys tekstiviestin kautta, joka nyt voidaan masinoida systemaattisesti sadoilletuhansille ihmisille samanaikaisesti. Deepfake-huijaukset ovat yksi kyberrikollisuuden uusimpia ilmenemismuotoja. Deepfakesta riippuen siihen voi liittyä kohdistettu avoimesta lähteestä etsitty tieto, jota seuraa kohdennettu tietojenkalastelu, identiteettivarkauden yritys deepfake-kokouksen kautta ja lopputulemana vaikka Arupin kaltainen rikos. Tässä työssä käytettiin Arupia esimerkkinä mihin uusimmalla deepfake-teknologialla on mahdollista päästä, mutta vastaavia tapauksia on muitakin.

Deepfake-videoita tai ääniä jaetaan sosiaalisen median erilaisilla alustoilla, mutta niitä jaetaan myös WhatsApissa, Telegramissa tai missä tahansa viestintäsovelluksessa, jossa lähettäminen onnistuu massoille samaan aikaan ja kustannukset ovat lähes olemattomat. Vaikka vain murto-osa kohteista tarttuisi syöttiin, onnistuneet rikokset tuovat riittävän tuoton. Yhteistä näille kaikille rikoksille on myös se, että mahdollinen tuotto hajaantuu nopeasti useammalle taholle ja maan rajojen ulkopuolelle, joten poliisin tutkinta vaikeutuu ja muuttuu joissain tapauksissa jopa mahdottomaksi. Lisäksi massarikollisuudelle on tyypillistä, että rikosten kohteiden määrän ollessa sadoissatuhansissa, poliisin rajalliset resurssit eivät kykene selvittämään kaikkia rikoksia. Näihin myös rikolliset luottavat ja toiminta voi jatkua.

Tehtäessä kohdistettua tietojenkalastelua, luottamus saavutetaan huomattavasti helpommin, kun pelkän viestittelyn sijaan toisella osapuolella on myös kasvot ja ääni. Nämä voidaan taas luoda deepfakella, jolloin varsinainen rikoksentekijä jää pimentoon ja kohde luottaa asioivansa oikean henkilön kanssa. Mielenkiintoista on kuitenkin myös se puoli ihmisen mielessä, että vaikka tiedetään toisen osapuolen olevan tietokoneen luoma hahmo, tämän kanssa jatketaan yhteydenpitoa. Tästä oli esimerkkeinä 24-vuotias tekoälyn luoma Milla-Sofia kappaleessa 2, jolla oli vuonna 2023 kymmeniä tuhansia seuraajia sosiaalisen median tileillään osan näistä uskoessa hahmon olevan aito. Muokattu kuva, joka perustuu puhtaasti

tekoälyn tekemään hahmoon uskotellen tämän olevan aito tai identiteettivarkaus sosiaalisen median tilin avulla tunnetusta henkilöstä, on deepfake.

Deepfaket eivät aina ole negatiivinen asia, mikäli ihminen niiden takana ei käytä sitä rikokseen. Tekoälyn käyttäminen elokuvateollisuudessa luomaan tehosteita valkokankaalle tai herättämään henkiin jo kuolleita elokuvamaaliman tähtiä tai Milla-Sofian kaltainen luotu hahmo tuomaan yksinäiselle ihmiselle seuraa ja yhteenkuuluvuuden tunnetta, ovat positiivisia asioita. Näiden määrä artikkeleissa ja tutkimuksissa on kuitenkin toistaiseksi huomattavasti pienempi kuin kyberrikollisuuteen liitettyjen deepfakejen sekä niiden negatiivisten vaikutusten.

Kyberrikollisuus on tullut jäädäkseen ja tulee myös tulevina vuosina kehittymään ja kasvamaan sekä deepfakejen että muiden muotojen osalta. Deepfakejen kehittyessä käydään samanlaista kilpajuoksua rikollisten ja rikoksen tunnistajien välillä kuin GAN teknologiassa deepfake-teknologian alussa. Yhtenä lupaavana skenaariona asiantuntijat olivat tarkastelleet lohkoketju-teknologiaa sekä yleisten avainten käyttöä yritysten välisissä kanssakäymisissä. Näillä voitaisiin varmemmin tunnistaa toinen osapuoli ja vältettäisiin Arupin kaltaisten deepfakejen onnistuminen - vai vältettäisiinkö.

Ihmisen ollessa heikon lenkki suojautumisessa ja puolustautumisessa ja teknologian ollessa tärkeässä roolissa, on aina mahdollista keksiä ohituskaista tai löytää porsaanreikä. Psykologinen vaikuttaminen on tehokasta. Ihmisen mieleen liittyvä alue oli kuitenkin rajattu tästä työstä pois, mutta siihen on ollut viitteitä useassa työssä käytetyssä lähteessä. Tämän vuoksi deepfakeja vastaan puolustautumisen ja suojautumisen kannalta on ensiarvioisen tärkeää kouluttaminen ja valistaminen teknologisten ratkaisuiden ohella. ISO 27001 ja NIS 2 kaltaiset standardit ja direktiivit, luovat kehyksen tekemiselle, mutta onnistuvat tavoitteessaan vain, mikäli yritykset ottavat ne aidosti pohjaksi omalle kyberturvallisuusarkkitehtuurilleen.

Nopea vilkaisu työmarkkinoilla kyberturvallisuusosaajille paljastaa karun totuuden: paikkoja on vielä tarjolla suhteellisen niukasti ja niissäkin haetaan ansioituneita osaajia useamman vuoden kokemuksella ja kasalla alan sertifikaatteja varustettuna. Onko niin, että yritykset eivät ole valmiita kouluttamaan riittävästi osaajia edes omiin tarpeisiinsa vaan luotetaan osaamistason kehittyvän toisaalla? Olisi äärimmäisen tärkeää sekä koulumaailman, yhteiskunnan avainpelaajien sekä yritysmaailman pitää yhdessä huolta riittävän osaamisen kasvattamisesta, ylläpitämisestä sekä kehittämisestä. Luvussa 3 viitattiin Iso-Britannian hallituksen ylläpitämään aineistoon, jossa yhtenä tarkastelukohteena oli yritysten ilmoittama kybertuvallisuuskoulutusten määrä edellisenä vuonna. Deepfakejen ja muiden kyberteknologioiden kehitysvauhdilla sekä laajenevan kyberrikollisuuden kasvutahdilla, alle

20 % vuosittainen koulutusmäärä tuntuu varsin vaatimattomalta. Hankkeet kuten Cybercrime Exit ja The Joint Cybercrime Taskforce mainittiin kappaleessa 3 vievät kehitystä ehdottomasti oikeaan suuntaan, mutta pelkällä ennaltaehkäisyllä ei päästä maaliin asti.

Empiirinen koe osoitti, että deepfake-videoiden ja äänien tekeminen on helppoa ja halpaa. Vaikka koejoukolle videoissa esiintyneet ihmiset olivat tuttuja, silti tunnistusprosentti ei ollut 100. Jos videot ja äänet olisi tehty kalliimmalla ohjelmalla, käytetty suuren yleisön tuntemaa henkilöä, jaettu sosiaalisen median alustalle ja aiheena olisi ollut yhteiskunnan toiminnan kannalta kriittinen kohde tai pyydetty rahaa, eivätkä ihmiset olisi tienneet mahdollisesti kyseessä olevan tekoälyn tuotoksen, olisi kyse ollut deepfakesta ja massarikoksesta. Tai jos kohteena olisi ollut yritys ja siellä henkilö, jolla on mahdollisuus tehdä yrityksen puolesta maksatuksia tai päästä kiinni sensitiiviseen materiaaliin ja hänestä olisi kiristysmielessä tehty arveluttava deepfake-video, olisivat seuraukset voineet olla merkittävät sekä yritykselle että henkilölle itselleen. Deepfake-huijauksissa on live-videoiden myötä jo nyt vain taivas rajana ilman kunnollisia teknologisia suojauksia sekä osaavia ihmisiä. Osaamisen kasvattaminen deepfakeja vastaan alkaa ymmärryksestä esimerkiksi WhatsApp-huijauksista, miten kuka tahansa voi joutua sosiaalisen manipuloinnin tai tietojen kalastelun kohteeksi ja miten suojautua tietomurtoja tai haittaohjelmia vastaan.

Toimeksiantajan näkökulmasta työ onnistui nostamaan esiin useita konkreettisia ja tärkeitä havaintoja, kuten uusien teknologioiden luomien muutosten jatkuvan huomioimisen turvallisuuskoulutusten kehittämisessä. Kokonaisuutena tarkastellen toimeksiantajan mielestä työ onnistui vastaamaan jokaiseen asetettuun tutkimuskysymykseen hyvin.

Stephen Hawkingsin ennusteeseen tekoälyn ylivoimaisesta kehityksestä on vielä matkaa. Skenaario ei kuitenkaan ole pelkkä kaukaisen tulevaisuuden science fiction ajatus, vaan mahdollisesti totta jo muutaman seuraavan vuosikymmenen aikana. Tämän vuoksi on maailmanlaajuisesti yhteiskuntien pidettävä huoli, että deepfakejen ja muiden kyberrikosten tunnistaminen ja torjunta, on aina useamman askeleen edellä rikollisia.

6 Johtopäätökset ja pohdinta

Kokonaisuutena opinnäytetyöstä tuli onnistunut ja tyytyväisyyttä lisäsi aikataulussa pysyminen. Vaikka työn tekeminen tapahtui kesän aikana, opinnäytetyön ohjaaja sekä toimeksiantaja olivat kohtuullisen hyvin tavoitettavissa. Aihe oli erittäin mielenkiintoinen ja HAMKn kurssi ”Kyberturvallisuus” antoi hyvän pohjan opinnäytetyöhön. Kyberturvallisuuteen ja erityisesti deepfake-huijauksiin kiinnostus heräsi Arupin tapauksesta, jota on tässä työssä käytetty yhtenä lähteenä.

Työhön liittyviä deepfakeja tehdessä piti pohtia eettistä puolta, vaikka kyseessä oli itsetehdyt video- ja ääninäytteet. Tämän vuoksi tässä työssä on liitteenä vain käytetyt kuvaparit, jotka ovat vapaasti internetistä löydettävissä. Ääni- ja videotallenteet säilytetään HAMKn verkkolevyllä vuoden ajan julkaisusta, jonka jälkeen ne tuhotaan. Näin toimien ne eivät päädy vahingossa julkiseen levitykseen. Itse deepfake-videoiden ja deepfake-äänien tekeminen yllätti: lopputulos oli halvallakin ohjelmalla sangen aidon tuntuinen. Panostamalla deepfake-ohjelmaan, olisi mahdollista saada luotua vielä huomattavasti parempia ja jopa reaaliaikaisia videoita.

Itse opinnäytetyön osalta tai sen käytölle ei ole toistaiseksi suunnitelmia. Toimeksiantajalla on luonnollisesti oikeus käyttää työtä tai sen osia parhaaksi näkemällään tavalla. Aihealueena deepfake-teknologia ja sen kehittyminen on äärettömän mielenkiintoista. Photoplethysmography-teknologia, jossa tutkitaan ihon verisuonien muutoksia, lohkoketju-teknologia ja yleiset avaimet tai videotallenteisiin vesileimojen laittaminen, ansaitsisivat jokainen oman tutkimuksen. Mielenkiintoista olisi myös lähestyä aihetta psykologiselta tai sosiologiselta kannalta. Tässä työssä tehty empiirinen tutkimus antoi hieman viitteitä iän, sukupuolen ja koulutustason osalta, miten hyvin henkilö tunnistaa deepfaket. Tämänkaltaista tutkimusta kannattaisi laajentaa ja syventää, ja tuloksia käyttää sitten kouluttamiseen ja huijausyritysten torjuntaan. Aihealueita deepfaken alueella riittää ja teknologioiden kehittyessä, niitä tulee lisää.

Tutkimuskysymykset olivat loppuun asti keskeisinä avattaessa aluksi teoreettista pohjaa, tutkittaessa erilaisia hyökkäys- ja puolustusmenetelmiä sekä empiiristä koetta tehtäessä. Niihin vastaaminen onnistui hyvin ja voidaan todeta, että kysymykset oli valittu keskeisistä teemoista. Laajemmassa työssä olisi tutkimuskysymyksiä voitu vielä laajentaa sekä lisätä.

7 Yhteenveto

Tutkimuskysymyksiin vastaaminen onnistui mielestäni erinomaisesti. Saatavilla olleen materiaalin suuren määrän vuoksi kaikkien tutkimuskysymysten osalta piti tehtyjä rajauksia miettiä tarkasti. Materiaalia löytyy runsaasti akateemisista tutkimuksista ympäri maailmaa pohtien deepfakeja ja kyberrikollisuutta psykologiselta, sosiaaliselta, teknologiselta ja jopa teologiselta kannalta. Aihetta on pohdittu yksityisten tutkijoiden, tutkijaryhmien, teknologiayritysten ja valtioiden kannalta. Jokainen näistä lähestymistavoista olisi tuottanut hieman erilaisen lopputuloksen. Koska tulevaisuuden skenaarioita deepfake-huijausten osalta on useita, yhdistin työssä näistä oman arvioni mukaan todennäköisimmän polun sekä pyrin pitämään riittävän geneerisen katsantokannan. Halusin opinnäytetyön olevan kuitenkin sellainen, ettei asian ymmärtämiseen tarvita syvällistä osaamista erityisesti deepfake-teknologiasta, psykologiasta tai muustakaan vastaavasta. Tulevaisuuden kehityspolkuja voi vastaavasti olla useita riippuen suojautumis- ja puolustautumiskeinojen kehittymispoluista. Tässäkin tapauksessa pyrin pitämään lähestymisen riittävän yleisellä mutta kuitenkin laajalla tasolla.

Oma osaamiseni ylipäätään kyberrikollisuudesta, sen eri muodoista, tulevaisuuden kehitysskenaarioista sekä miten deepfake-huijausten kehitys nykypisteeseen on vaatinut muiden kyberrikosten kehitystä, oli mielenkiintoinen polku. Erityisen mielenkiintoista oli tutustua teknologisen ymmärryksen kasvattamisen jälkeen deepfaken tekemiseen ja niiden testaamiseen koejoukolla. Jos nyt tekisin vastaavan testin uudelleen, käyttäisin enemmän rahaa itse sovellukseen ja hioisin videoita vielä enemmän saavuttaakseni vielä todenmukaisemman lopputuloksen. Mielenkiintoinen tutkimushaara tulevaisuudessa itselleni on psykologinen puoli nimenomaan deepfakejen jakamisessa eteenpäin. Akateemisissa tutkimuksissa, joita en tässä työssä käyttänyt aineistona, oli useita viiteitä, joissa ihmiset jakavat deepfake-materiaalia, vaikka tietävät niiden olevan tietokoneen tekemiä. Tämä toteutui erityisesti, mikäli deepfaken sisältö tuki heidän mielipiteitään asiasta. Tästä syystä tein omat deepfake-videot ja -äänet kahdesta vastakkaisesta mielipiteestä asiaan. Tulevaisuudessa voisi olla kiinnostavaa tutkia yhteyttä koehenkilön oman mielipiteen ja tunnistuksen välillä.

Kaikkien koehenkilöiden osaamista kyberhuijausten ja erityisesti deepfakejen osalta kasvatettiin testin avulla. Olisiko mahdollista käyttää vastaavaa lähestymistapaa koulutettaessa nykyistä ja seuraavaa sukupolvea tunnistamaan verkossa vaanivat vaarat? Se jää nähtäväksi ja riippuu minkä kehityssuunnan deepfaket tulevaisuudessa saavat.

Lähteet

Alanko, E. (27.11.2023). *Kuinka nyky-aikainen verkko-pankki-huijaus toimii ja miten siltä voi suojautua?* https://tilisanomat.fi/teknologia/kuinka-nykyaikainen-verkkopankkihuijaus-toimii-ja-miten-silta-voi-suojautua

Augustėnė, A. (7.5.2023). *Yleisimmät Whatsapp-huijaukset – näin vältät ne*. https://nordvpn.com/fi/blog/whatsapp-huijaus/

Autio, S. tietoturva-asiantuntija & Mesiä, M. tietoturva-asiantuntija & Tretjakov, J. erityisasiantuntija. Kyberturvallisuuskeskus, Helsinki. Sähköpostihaastattelu 30.8.2024, haastattelijana Sami Rintala. Tallenne liitteessä 6.

Bergen, M.& Metz, R.& Murphy, M. (27.1.2024). *AI Startup ElevenLabs Bans Account Blamed for Biden Audio Deepfake*. https://www.bloomberg.com/news/articles/2024-01-26/ai-startup-elevenlabs-bans-account-blamed-for-biden-audio-deepfake?embedded-checkout=true%5d

Breland, A. (15.3.2019). *The Bizarre and Terrifying Case of the "Deepfake" Video that Helped Bring an African Nation to the Brink*. https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B.& Bharath, A. (19.10.2017). *Generative Adversarial Networks: An Overview*. https://arxiv.org/pdf/1710.07035

Damiani, J. (3.9.2019). *A Voice Deepfake Was Used To Scam A CEO Out Of $243,000*. https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

Deloitte. (3.2023). *How to safeguard against the menace of deepfake technology The battle against digital manipulation*. https://www2.deloitte.com/content/dam/Deloitte/in/Documents/risk/in-ra-safeguarding-against-deepfake-technology-noexp.pdf

Eduskunta. (1.2.2019). *Vastaus kirjalliseen kysymykseen massarikosten tutkinnasta*. https://www.eduskunta.fi/FI/vaski/Kysymys/Documents/KKV_560+2018.pdf

Eksymä, A-R & Sandell, E. (4.5.2020) *Miksi alkaa hakkeriksi? – Team Whackin valkohattuhakkerit tuovat kyberuhat arjen tasolle*. Haettu 1.6. osoitteesta https://yle.fi/aihe/artikkeli/2019/03/04/miksi-alkaa-hakkeriksi-team-whackin-valkohattuhakkerit-tuovat-kyberuhat-arjen

Enisa. (n.d.). *NIS Directive*. https://www.enisa.europa.eu/topics/cybersecurity-policy/nis-directive-new

European Parliament. (2023). *The NIS2 Directive*. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/689333/EPRS_BRI(2021)689333_EN.pdf

Finanssiala Ry (17.5.2024). Valtaosa suomalaisista joutunut huijausyrityksen kohteeksi – Finanssiala aloittaa torjuntakampanjan ”huijarille luu kurkkuun”. https://www.finanssiala.fi/uutiset/valtaosa-suomalaisista-joutunut-huijausyrityksen-kohteeksi-finanssiala-aloittaa-torjuntakampanjan-luu-kurkkuun-huijareille/

Fortune Business Insight. (17.6.2024). *Technology/Cyber Security Market*. Haettu 20.6. osoitteesta https://www.fortunebusinessinsights.com/industry-reports/cyber-security-market-101165

F-Secure. (n.d.-a). Mitä on tietojenkalastelu? https://www.f-secure.com/fi/articles/what-is-phishing

F-Secure. (n.d.-b). *Mikä on identiteettivarkaus? Kaikki mitä pitää tietää*. https://www.f-secure.com/fi/articles/what-is-identity-theft

F-Secure. (n.d.-c). *Mikä on ransomware?* https://www.f-secure.com/fi/articles/what-is-a-ransomware-attack

F-Secure. (n.d.-d). *Mitä on spoofing?* https://www.f-secure.com/fi/articles/spoofing

F-Secure. (n.d.-e). *Mitä on käyttäjän manipulointi?* https://www.f-secure.com/fi/articles/what-is-social-engineering

Groth, M., Epstein, Z., Firestone, C, Pickard, R. (28.12.2021). *Deepfake detection by human crowds, machines, and machine-informed crowds*. https://www.pnas.org/doi/10.1073/pnas.2110013119

Gowal, S., Kohli, P. (29.8.2023). *Identifying AI-generated images with SynthID*. https://deepmind.google/discover/blog/identifying-ai-generated-images-with-synthid/

Haajanen, A & Tuominen, A. (21.2.2024). *Lintukodon aika on ohi – deepfake-huijaukset ovat uhka suomalaisyrityksille*. https://www.ey.com/fi_fi/cybersecurity/deepfake-huijaukset-ovat-uhka-suomalaisyrityksille

ISO. (n.d.). ISO/IEC 27001:2022. https://www.iso.org/standard/27001

Kymäläinen, S. (2.9.2023). *Virheettömästä Milla-Sofiasta tuli Tiktok-tähti – tilin takaa paljastuu 44-vuotias mies Vantaalta*. https://yle.fi/a/74-20046003

Järveläinen, V. (26.6.2024). *Somessa leviää deepfake-video Kekkosesta*. https://www.tivi.fi/uutiset/somessa-leviaa-deepfake-video-kekkosesta/521a4b5c-ff5b-4d3c-b5bc-5b03b7e5d730

Keskusrikospoliisin kyberrikostorjuntayksikkö. Keskusrikospoliisi, Helsinki. Sähköpostihaastattelu 27.9.2024, haastattelijana Sami Rintala. Tallenne liitteessä 7.

Khan, U. (n.d.). *The SolarWinds Sunburst Supply Chain Attack.* SolarWinds Sunburst Supply Chain Attack - Security Overview (mindpointgroup.com)

Korhonen, M. (12.4.2024). *Presidentti Alexander Stubb sotkettiin kryptohuijaukseen*. https://fi.cryptonews.com/news/presidentti-alexander-stubb-kryptohuijaus.htm

Krebs, B. (26.3.2024). Recent 'MFA Bombing' Attacks Targeting Apple Users. https://krebsonsecurity.com/2024/03/recent-mfa-bombing-attacks-targeting-apple-users/

Kyberturvallisuuskeskus. (n.d.). *Tietojenkalastelu ja identiteettivarkaudet verkossa*. https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/file/Tietojenkalastelu%20ja%20identiteettivarkaudet.pdf

Lanham, M. (2012). *Generating a New Reality. From Autoencoders and Adversarial Networks to Deepfakes*. https://learning.oreilly.com/library/view/generating-a-new/9781484270929/html/502181_1_En_BookFrontmatter_OnlinePDF.xhtml

Limnell, A. (9.8.2023). *Ajankohtainen katsaus kyberrikollisuuteen*. https://poliisi.fi/blogi/-/blogs/ajankohtainen-katsaus-kyberrikollisuuteen-

McFarland, A. (2.7.2024). *5 parasta Deepfake Detector -työkalua ja -tekniikkaa*. https://www.unite.ai/fi/best-deepfake-detector-tools-and-techniques/

McGill, J. (23.5.2024). *24 Deepfake Statistics – Current Trends, Growth, and Popularity (December 2023)*. https://contentdetector.ai/articles/deepfake-statistics/

Medeiros, J. (28.11.2017). *Stephen Hawking: 'I fear AI may replace humans altogether'*. https://www.wired.com/story/stephen-hawking-interview-alien-life-climate-change-donald-trump/

Mitre. (n.d.). WannaCry. https://attack.mitre.org/software/S0366/

Microsoft. (n.d.). *Tietojen kalastelulta suojautuminen*. https://support.microsoft.com/fi-fi/windows/tietojen-kalastelulta-suojautuminen-0c7ea947-ba98-3bd9-7184-430e1f860a44

Mäntysalo, J. (27.8.2022). *Tietomurtoja tapahtuu nyt enemmän kuin koskaan, mutta rikollisia ei saada kiinni – viime vuonna 96 prosenttia tapauksista jäi selvittämättä*. https://yle.fi/a/3-12596198

NIS2 Directive. (n.d.) *The NIS2 Directive Explained*. https://nis2directive.eu/

Organization for Social Media Safety. (n.d.). *Deepfake Technology*. https://www.socialmediasafety.org/advocacy/deepfake-technology/

Palmgren, J. (6.2.2024). *Huijareilla oli aktiivinen vuosi 2023 – Pankit saivet estettyä digihuijauksia lähes 33 miljoonan euron edestä*. https://www.finanssiala.fi/uutiset/huijareilla-oli-aktiivinen-vuosi-2023-pankit-saivat-estettya-digihuijauksia-lahes-33-miljoonan-euron-edesta/

Poliisi. (n.d.-a). *Tietomurrot.* Haettu 30.5.2024 osoitteesta https://poliisi.fi/tietomurrot

Poliisi. (20.12.2022). *Poliisi tehostaa kyberrikostorjuntaa: Suomi mukaan Europolin yhteistyöryhmään*. https://poliisi.fi/-/poliisi-tehostaa-kyberrikostorjuntaa-suomi-mukaan-europolin-yhteistyoryhmaan

Poliisi. (26.32024). *Nettituttavuus voi johtaa kiristykseen*. https://poliisi.fi/-/nettituttavuus-voi-johtaa-kiristykseen

Polisen. (n.d.). Näin poliisi torjuu massarikoksia. Haettu 30.5.2024 osoitteesta https://polisen.se/fi/lait-ja-saannot/massarikollisuus/

Šimonėlytė, M. (11.1.2023). *Kyberturvallisuuden historia*. https://nordvpn.com/fi/blog/kyberturvallisuuden-historia/

Sisäministeriö. (n.d.). *Kyberrikollisuus ylittää rajat tietoverkoissa*. https://intermin.fi/poliisiasiat/kyberrikollisuus

Struck. (17.1.2024). *DEEPFAKES AND BLOCKCHAIN*. https://struckcapital.com/deepfakes-and-blockchain/

Team Whack – kaikki on hakkeroitavissa. (4.5.2020). *Kausi 2, Jakso 4: Team Whack murtautuu vakuutusyhtiöön*. https://areena.yle.fi/1-50418776

Team Whack – kaikki on hakkeroitavissa. (4.3.2019). Kausi 1, Jakso 2: *Team Whack varastaa identiteettisi*. https://areena.yle.fi/1-4664684

Tekniikan Maailma. (21.5.2024). *Isoin vedätys koskaan? Deepfake-huijarit veivät brittifirmalta 23 miljoonaa euroa ovelalla videopuhelulla*. https://tekniikanmaailma.fi/isoin-vedatys-koskaan-deepfake-huijarit-veivat-brittifirmalta-23-miljoonaa-euroa-ovelalla-videopuhelulla/

Traficom. (16.6.2023 -a). *Pornokiristyksiä runsaasti liikkeellä – älä usko huijareiden väitteitä*. https://www.kyberturvallisuuskeskus.fi/fi/ajankohtaista/pornokiristyksia-runsaasti-liikkeella-ala-usko-huijarien-vaitteita

Traficom. (2023-b). *NIS-koordinointi ja viranomaisyhteistyö*. https://www.kyberturvallisuuskeskus.fi/fi/palvelumme/tilannekuva-ja-verkostojohtaminen/nis-koordinointi-ja-viranomaisyhteistyo

Traficom. (15.3.2023 -c). *Kriittinen haavoittuvuus Microsoft Outlookissa*. https://www.kyberturvallisuuskeskus.fi/fi/kriittinen-haavoittuvuus-microsoft-outlookissa

Traficom. (14.4.2022). *Vinkkejä informaatiovaikuttamisen tunnistamiseksi - Ole tarkkana ja toimi vastuullisesti*. https://www.kyberturvallisuuskeskus.fi/fi/ajankohtaista/vinkkeja-informaatiovaikuttamisen-tunnistamiseksi-ole-tarkkana-ja-toimi

UK Government. (9.4.2024). *Cyber security breaches survey 2024*. https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2024/cyber-security-breaches-survey-2024#summary

UK Government (19.4.2023). *Cyber security breaches survey 2023.* https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2023/cyber-security-breaches-survey-2023#summary

Valtioneuvoston kanslia. (5.4.2019). *Informaatiovaikuttamiseen vastaaminen. Opas viestijöille*. https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161512/VNK_11_2019_Informaatiovaikuttamisen%20vastaaminen_web.pdf?sequence=1&isAllowed=y

Yle. (6.9.2019). *Syvä Huijaus*. https://yle.fi/a/3-10955498

Zieniūtė, U. (21.8.2022). *Mikä on deepfake?* https://nordvpn.com/fi/blog/mika-on-deepfake/

Liite 1: Aineistonhallintasuunnitelma

Tutkimuksellinen työ

Tässä opinnäytetyössä kerätty aineisto tallennetaan Hämeen Ammattikorkeakoulun palvelimelle. Työstä tehdään säännöllisin väliajoin varmuuskopio ulkoiselle kiintolevylle sekä tekijän tietokoneen C-asemalle. Tällä varmistetaan se, että jos yksi näistä tuhoutuu, ovat muut aineistolähteet edelleen saatavilla.

Opinnäytetyössä kerätty aineisto on anonymisoitu eikä se sisällä minkäänlaisia henkilötietoja. Näin ollen työn aikana sekä sen jälkeen aineistojen jakamiseen ei liity eettisiä kysymyksiä. Opinnäytetyön tekijä Sami Rintalalla on omistajuus ja tekijänoikeudet kaikkiin kerättyihin ja tuotettuihin aineistoihin. Opinnäytetyön aikana kerätty materiaali sekä siitä tehdyt analyysit ovat vapaasti jatkokäytettävissä.

Opinnäytetyöhön sisältyy 25 kappaletta haastattelujen avulla tehtyä testiä. Koska nämä tiedot on anonymisoitu, testeihin osallistuneet henkilöt ovat antaneet luvan vastauksiensa olla osana kerättyä data-aineistoa.

Opinnäytetyössä käytetyt ovat joko julkisista lähteistä otettuja vapaasti käytettävää materiaalia tai itseluotuja joihin työn tekijällä on omistusoikeus. Opinnäytetyöhön sisältyy edellämainittujen haastattelujen lisäksi Kyberturvallisuuskeskuksen asiantuntijoiden haastattelu. Haastattelu on litteroituna liitteessä 6. Haastattelun kohteelta on saatu lupa jättää aineisto julkiseksi sekä jatkokäyttää sitä halutessa.

Opinnäytetyöaineiston jatkokäyttö työn valmistumisen jälkeen

Haluan luovuttaa opinnäytetyöni aineiston jatkokäyttöön. Käytettäessä Kyberturvallisuuskeskuksen asiantuntijoiden haastattelua tulee siihen tehdä normaalit lähdeviittaukset. Toimeksiantajan kanssa on sovittu opinnäytetyön jatkokäytöstä. Opinnäytetyöhön kerättyä tutkimusaineistoa ei ole sovittu luovutettavan Hämeen Ammattikorkeakoululle jatkokäyttöön.

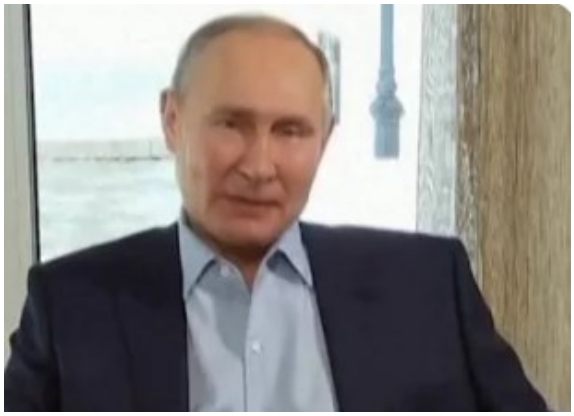
Liite 2. Testiaineiston kuvaparit

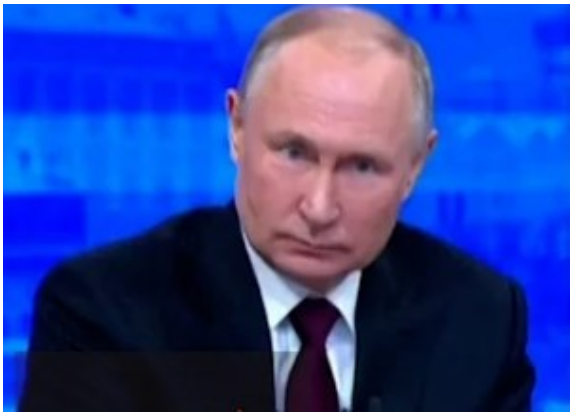
Toinen on aito ja toinen väärennös, mutta kumpi on kumpi?

1A **1B**

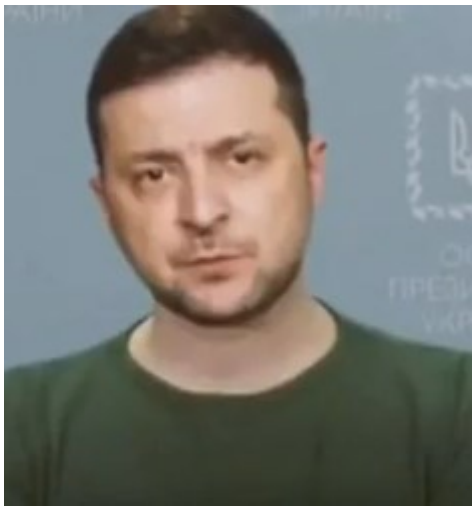
2A **2B**

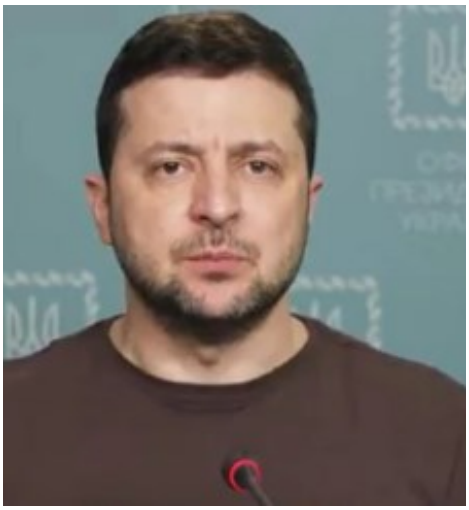
3A **3B**





4A **4B**





5A **5B**

Liite 3. Äänitallenteiden tekstit

Ääni 1 - Pakkoruotsia vastaan

Ruotsin kielen opiskelu pitäisi muuttaa vapaaehtoiseksi Suomessa. Ruotsia puhuu äidinkielenään meillä viisi prosenttia väestöstä ja heistäkin suurin osa puhuu Suomea sujuvasti. Miksi muutenkin syvässä kuilussa olevan maan opetusjärjestelmä ja heikko valtiontaloutemme käyttää älyttömiä summia marginaalisen kansanryhmän kielen pakko-opetukseen? Käytännössä suurin osa suomalaisista ei kouluaikojensa jälkeen käytä sanaakaan ruotsia ja ulkomailla käydessään puhutaan englantia. Täysin hukkaan heitettyä rahaa ja ajan voisi käyttää jonkun maailmankielen opiskeluun.

Ääni 2 - Pakkoruotsin puolesta

Suomessa on rikkautena moninainen kulttuuri ja siihen kuuluu kaksikielisyys. Helsingin yliopistossa tehdyssä tutkimuksessa yli seitsemänkymmentä prosenttia suomalaisista kokee ruotsin kielen myönteisenä Suomalaisessa yhteiskunnassa. Ruotsin kielen osaaminen tarjoaa monelle suomenkieliselle upeita mahdollisuuksia sekä työ että opiskelurintamalla. Kaikki kieltenopiskelu on plussaa ja auttaa jatkossa muiden vieraiden kielten opiskelussa.

Ääni 3 - Orpon hallituksen politiikan puolesta

Marinin hallitus jätti melkoisen korjaus- ja siivousvelan seuraavalle hallitukselle. Tehtiin velanottoon perustuva budjetti, vaikka velkaa oli jo ennestään niin että kipeää teki. Valtion velka on yli 160 miljardia euroa ja osa edellisen hallituksen tekemistä mm. sitoumuksista on sellaisia, että irti ei päästä. Samoin Suomi sidottiin Euroopan Unionin yhteisvelkaan, vaikka sen piti olla kertaluontoinen hanke. Hyvinvointialueuudistus oli susi jo syntyessään ja kallis. Marinin ja vasemmistohallituksen ohjenuora oli joku muu maksaa laskun ja rikkailta voidaan aina kiskoa enemmän veroja. Orpon hallitus joutuu tekemään kipeitä päätöksiä saadakseen meidät ylös tästä suosta. Nyt vain vyötä kireämmälle.

Ääni 4 - Orpon hallituksen politiikkaa vastaan

Köyhät kyykkyyn on nykyisen hallituksemme linja. Valtiovarainministeriö antoi useita ehdotuksia Suomen talouden tasapainottamiseen ja Orpon hallitus päätti valita ne, jotka iskevät kaikista kovimmin köyhiin ja heikossa asemassa muutenkin oleviin. Hallituksen köyhien kyykytysohjelma ei toimi, koska velkaa otetaan enemmän kuin aikaisemmin. Pitäisi ottaa rikkaat mukaan talkoisiin, eli veronalennusten peruuttaminen ja verotuksen tason nostaminen tavalla tai toisella. Se olisi mahdollista, mutta se olisi kokoomuksen vaalilupausten pettämistä, joten niin ei tule käymään. Köyhät köyhtyvät ja rikkaat rikastuvat entisestään.

Liite 4. Videotallenteiden tekstit

Video 1 - Huumeiden vapauttamisen puolesta

Kansalaisaktivisti Saija Rintala tässä hei. On aivan päivänselvää, että suuri raha hyötyy huumeiden laittomuudesta - ei vain marihuanan. Kaikki huumeet pitää vapauttaa, silloin murtuu yritysten ja valtioiden kaksinaismoralistinen pelailu. Huumesotaa ylläpidetään vetoamalla absurdeihin asioihin, aivan kuin kaikki olisivat heti tykittämässä itsensä sekopääksi. Tekisitkö sinä niin, minä en. Koko päihdepolitiikka on ihmistä aliarvioivaa ja kaikenlisäksi se syö valtavasti rahaa. Ainoa tie voittaa huumesota, on korjata valheet ja se tärkein asennemuutos eli vapauttaa kaikki huumeet.

Video 2 - Huumeiden vapauttamista vastaan

Kansalaisaktivisti Saija Rintala tässä hei. Käsitykset mietojen huumeiden todellisista vaikutuksista ovat edelleen niin ristiriitaisia, että vapauttaminen ei tule kysymykseen. Jatkuvasti käytettynä miedotkin aineet ovat haitallisia ja todistetusti aiheuttavat riippuvuutta sekä monessa tapauksessa siirtymistä vahvempiin aineisiin. Suomessa huumeiden käyttö on tutkimusten mukaan pysynyt samalla tasolla 2000-luvulla. Jos miedot huumeet vapautettaisiin, käyttö lisääntyisi räjähdysmäisesti. Samoin huumeongelmat.
Kannabis suorastaan tyhmentää: säännöllinen käyttö vaurioittaa kasvavan aivosoluja, runsas käyttö passivoittaa ja heikentää lähimuistia sekä keskittymiskykyä. Tällaista yhteiskuntaa emme halua.

Video 3 - Rokotteiden puolesta

Kauhavalaisessa koulussa levisi hallitsemattomasti tuhkarokkoepidemia. Lapsia joutui sairaalaan ja osa oli hengenvaarassa. Miksi? Koska osa vanhemmista oli päättänyt olla vapaamatkustajia muiden kustannuksella ja jättää lapsensa rokottamatta. Joukkosuoja syntyy vasta, kun riittävä määrä on saanut suojan rokotteesta ja tauti ei pääse leviämään. Ihmiset matkustavat, tuovat meillä jo kuolleita tauteja tullessaan ilman rokotuksia ja levittävät niitä. On edesvastuutonta jättää rokotteet ottamatta ja levittää tauteja.

Video 4 - Rokotteita vastaan

Oletko koskaan miettinyt, miksi yhteiskunta haluaa tunkea meihin rokotteita? Kuka niistä hyötyy? Koska olet viimeksi kohdannut ihmisen, jolla oli tuhkarokko? Niinpä. Rokotteista hyötyvät lähinnä rikkaat lääketehtaat ja samalla yhteiskunta lisää holhoustaan meihin. Monet rokotteet on todettu ihmisille jopa vaarallisiksi ja seurauksena on ollut koronarokotteesta

keuhkoveritulppia ja sikainfluenssarokotteesta narkolepsiaa. Viisas jättää rokotteet ottamatta ja rahat käytetään johonkin aidosti hyödylliseen.

Video 5 – Sähköauton puolesta

Hei, tässä Sami Rintala automyyjä. Oletko hankkimassa itsellesi uutta autoa? Järkevän ihmisen valinta tänä päivänä on sähköauto. Ainoastaan vähentämällä fossiilisten polttoaineiden kulutusta säästämme maapallon tuleville sukupolville. On edesvastuutonta hankkia polttomoottoriauto ilmaston muutoksen syventyessä hurjalla vauhdilla.

Video 6 – Sähköautoa vastaan

Hei, tässä Sami Rintala automyyjä. Oletko hankkimassa itsellesi uutta autoa? Järkevän ihmisen valinta tänä päivänä on polttomoottoriauto. Mediassa on tuutin täydeltä väärää informaatiota, miten sähköautot auttavat ilmastonmuutoksen torjunnassa. Polttomoottoriautoissa käytetään biopohjaisia polttoaineita, jotka eivät saastuta luontoa. Sähköautoissa käytetään litiumakkuja ja litiumin louhinnasta koituu valtavia ympäristöongelmia. On edesvastuutonta hankkia sähköauto tietäen siitä koituvat ympäristöhaitat.

Video 7 – Ukrainan puolesta

Hei, tässä sotareportteri Sami Rintala. Ukrainan sota on jatkunut jo yli kaksi vuotta. Hirviömäinen Venäjä pommittaa säälimättä naisia ja lapsia ja muita viattomia ukrainalaisia. Urheat ukrainalaiset sotilaat puolustavat urheasti maansa infrastruktuuria, sairaaloita, ruokavarastoja ja muuta elintärkeää raukkamaista vihollista vastaan. Liittoudutaan yhdessä venäjän hirmuvaltaa vastaan.

Video 8 – Ukrainaa vastaan

Hei, tässä sotareportteri Sami Rintala. Venäjän erikoisoperaatio on jatkunut jo yli kaksi vuotta. Euroopan valtiot kieltäytyvät jääräpäisesti näkemästä nousevaa uhkaa Ukrainassa. Puolustaessaan maailmanrauhaa venäjä on onnistunut saamaan Valko-Venäjän ja Pohjois-Korean liitolaisikseen. Liittoudutaan yhdessä uhkaa vastaan.

Liite 5. Testien tulokset

hlö#	sukup	mikä deepfake on	ikä	koulutus	kuva 1 A	kuva 1 B	Kuva 2 A	Kuva 2 B	Kuva 3A	Kuva 3B	Kuva 4A	Kuva 4B	Kuva 5A	Kuva 5B	Ääni 1	Ääni 2	Ääni 3	Ääni 4	Video 1	Video 2	video 3	video 4	video 5	video 6	video 7	video 8
1	nainen	ei	<20	lukio	aito	fake	fake	aito	fake	aito	fake	aito	fake	aito	fake	aito	fake	fake	aito	fake	aito	fake	fake	aito	fake	fake
2	mies	kyllä	20-29	alempi korkeakoulu	fake	aito	aito	fake	fake	aito	aito	fake	fake	aito	fake	aito	fake	fake	aito	fake	aito	fake	fake	aito	fake	fake
3	nainen	ei	40-49	ylempi korkeakoulu	aito	fake	fake	aito	fake	aito	fake	aito	fake	aito	fake	aito	aito	aito	aito	fake	aito	fake	fake	aito	fake	fake
4	nainen	ei	30-39	alempi korkeakoulu	aito	fake	fake	aito	fake	aito	fake	aito	fake	aito	aito	aito	aito	aito	fake	fake	fake	fake	fake	fake	fake	fake
5	mies	kyllä	20-29	lukio	aito	fake	aito	fake	fake	aito	fake	aito	aito	fake	fake	aito	fake	fake	aito	fake	aito	fake	fake	aito	fake	fake
6	nainen	kyllä	>70	lukio	fake	aito	aito	fake	fake	aito	fake	aito	fake	aito	fake	aito	aito	fake	aito	fake	aito	fake	aito	aito	aito	aito
7	mies	kyllä	50-59	ylempi korkeakoulu	fake	aito	aito	fake	fake	aito	fake	aito	fake	aito	fake	aito	aito	fake	aito	fake	aito	fake	fake	aito	fake	fake
8	mies	kyllä	40-49	ylempi korkeakoulu	fake	aito	aito	fake	fake	aito	fake	aito	fake	aito	aito	aito	aito	fake	fake	fake	aito	fake	aito	aito	fake	fake
9	mies	kyllä	30-39	ammatillinen koulutus	aito	fake	aito	fake	aito	fake	fake	aito	fake	aito	aito	aito	aito	aito	aito	fake	fake	fake	fake	aito	fake	fake
10	mies	kyllä	30-39	alempi korkeakoulu	aito	fake	aito	fake	fake	aito	fake	aito	fake	aito	fake	fake	fake	fake	aito	aito	fake	fake	fake	fake	fake	fake
11	mies	kyllä	50-59	lukio	fake	aito	aito	fake	fake	aito	fake	aito	aito	fake	aito	aito	fake	fake	aito	fake	fake	fake	aito	aito	fake	fake
12	mies	kyllä	60-69	alempi korkeakoulu	fake	aito	fake	aito	fake	aito	fake	aito	fake	aito	fake	aito	aito	fake	aito	fake	aito	fake	fake	aito	fake	fake
13	mies	kyllä	>70	alempi korkeakoulu	fake	aito	fake	aito	aito	fake	aito	fake	fake	aito	fake	fake	fake	fake	aito	fake	aito	fake	fake	aito	fake	fake
14	nainen	ei	>70	peruskoulu	fake	aito	aito	fake	fake	aito	fake	aito	fake	aito	aito	aito	aito	aito	aito	fake	aito	fake	fake	aito	fake	fake
15	nainen	kyllä	20-29	lukio	fake	aito	aito	fake	aito	fake	aito	fake	fake	aito	fake	aito	fake	fake	fake	fake	aito	fake	fake	aito	fake	fake
16	mies	ei	>70	ammatillinen koulutus	aito	fake	fake	aito	fake	aito	aito	fake	fake	aito	aito	fake	fake	fake	aito	fake	aito	fake	aito	aito	fake	aito
17	nainen	ei	60-69	ammatillinen koulutus	fake	aito	fake	aito	fake	aito	fake	aito	aito	fake	fake	aito	fake	aito	aito	fake	aito	fake	fake	aito	fake	aito
18	nainen	ei	40-49	ammatillinen koulutus	fake	aito	aito	fake	fake	aito	fake	aito	aito	fake	aito	fake	fake	aito	aito	fake	aito	fake	fake	aito	fake	fake
19	mies	ei	40-49	ammatillinen koulutus	aito	fake	aito	fake	fake	aito	aito	fake	fake	aito	aito	fake	fake	fake	aito	fake	aito	fake	fake	aito	fake	fake
20	mies	kyllä	50-59	ylempi korkeakoulu	fake	aito	fake	aito	fake	aito	fake	aito	fake	aito	aito	aito	aito	aito	aito	fake	aito	fake	fake	aito	fake	fake
21	nainen	ei	50-59	alempi korkeakoulu	aito	fake	aito	fake	fake	aito	fake	aito	fake	aito	fake	aito	fake	fake	aito	fake	aito	fake	fake	aito	fake	fake
22	nainen	kyllä	50-59	ylempi korkeakoulu	fake	aito	aito	fake	aito	fake	fake	aito	fake	aito	fake	aito	aito	aito	fake	fake	aito	fake	aito	fake	aito	aito
23	nainen	ei	<20	lukio	fake	aito	fake	aito	aito	fake	fake	aito	fake	aito	aito	fake	fake	fake	aito	fake	aito	fake	fake	aito	fake	fake
24	nainen	ei	<20	lukio	fake	aito	fake	aito	fake	aito	fake	aito	fake	aito	fake	aito	fake	aito	aito	fake	aito	fake	fake	aito	fake	fake
25	nainen	ei	20-29	ylempi korkeakoulu	fake	aito	aito	fake	fake	aito	aito	fake	fake	aito	aito	aito	aito	aito	fake	fake	aito	fake	aito	aito	aito	fake

Liite 6. Kyberturvallisuuskeskuksen asiantuntijoiden haastattelu litteroituna

Kysymys 1a:

Lähes kaikki kyberrikollisuus on tavalla tai toisella kansainvälistä - onko tunnistettavissa joitain tiettyjä keinoja, joilla voidaan vastata syväväärennösten muodostamaan uhkaan?

- Viestintä ja koulutus korostuvat -> yhteiskunnan median lukutaidon kasvattaminen ja tietoisuuden lisäämisen mediaväärennöksistä

- Mediaväärennöksien tunnistusmenetelmien kehittäminen ja jakaminen kansainvälisesti sekä yksityisten/julkisten organisaatioiden välillä. Samat algoritmit ja alustat, joilla väärennettyä sisältöä voidaan tuottaa ja julkaista, voi valjastaa myös manipuloidun sisällön tunnistamiseen automaattisesti. Tekniikka tältä osin on nuorta ja jatkuvasti kehittyvää.

- Viranomaisyhteistyö kansallisesti ja kansainvälisesti, erilaisten mediaväärennöksiä käyttävien kampanjoiden tunnistaminen ja tiedottaminen niin kotimaassa kuin kansainvälisestikin.

Juha Tretjakovin kommentaari terminologiasta:

“Deep fake” on englannin kielessä vakiintunut tarkoittamaan koneoppivalla tekoälyllä tuotettua kuva-, video- tai ääniväärennöstä. Suomen kieleen käsitteelle on rantautunut laiskasti roiskaistu suora käännös “syväväärennös”. Termi on mielestäni huono, koska se ei suomeksi selity itse, vaan vaatii että se pitää opetella tai hakea englannin kielen kautta. Kyse on väärennöksestä, mutta mikä siinä on syvää?

Englannin kielen “deep” tulee samasta taustasta kuin neuroverkkojen koneoppimisalgoritmeja tarkoittava “deep learning”. Siinäkin epäselväksi jää, mikä tässä on niin diippiä, neuroverkkojen kaninkoloko.

“Syvä” suomen kielessä voi tarkoittaa konkreettisen kaivetun kuopan syvyyden lisäksi myös kuvainnollista syvyyttä (syvä hiljaisuus, syvään hengitys, syvä rauha) tai synonyymiä syvällisyydelle (syvät mietteet, syvä viisaus), tai äänen mataluutta (syvä basso). Mikään näistä merkityksistä ei kerro, mikä koneoppimisalgoritmeilla tehdyssä väärennöksessä on syvää. Vai onko se syvältä?

Tapauksesta riippuen olen kuvannut tekstissä tarkemmin, mistä on kysymys: ääniväärennös, kuvamanipulaatio, videoväärennös, -huijaus tai -manipulaatio. Jos näille tarvitaan yhteinen kattokäsite, voisiko käyttää “tekoälyväärennöstä”? Se ei ole niin napakka, mutta kuitenkin kuvaavampi kuin syvävääärennös. Ääni, kuva ja video voitaisiin myös niputtaa kattokäsitteeksi “media”, mistä voi johtaa termin “mediavääärennös”. Se kertoo jo itsessään, mitä on tehty (väärennös) ja mihin se kohdistuu (mediaan). Tässä yhteydessä harvoin on oleellista, miten syvää tai matalaa se nyt sattuu olemaan.

Hesarin uutisessa https://www.hs.fi/maailma/art-2000010595322.html kerrotaan manipuloidusta Harrisia imitoivasta videosta, väärennetystä videosta, manipuloidusta ääniraidasta ja tekoälyn tekemästä äänestä. Huom: ei sanaakaan mistään syvästä eikä deepistä. Tätä pidän asianmukaisena ja onnistuneena uutisointina ja otan tästä oppia.

Kysymys 1b:

Minkälaista kansainvälistä yhteistyötä kyberturvallisuuskeskus tekee kyberrikosten osalta?

- Aktiivisesti ja säännöllisesti tiedonvaihtoa erityisesti muiden maiden cert-toimijoiden kanssa. Kyberrikosten osalta avataan tilannekuvaa Suomesta kertomalla esimerkiksi minkä tyyppisiä ilmiöitä ja missä mittakaavassa Suomessa niitä on esiintynyt. (lisäksi myös IoC-tietojen vaihto ja yksittäisistä tapauksista tiedottaminen)

- Varsinkin vaalivaikuttamisen tiimoilta suomalaiset viranomaiset (Kyberturvallisuuskeskus yhdessä vaaliprosessia hallitsevan Oikeusministeriön kanssa) ovat tehneet yhteistyötä ja vaihtaneet tietoa sekä kokemuksia muiden maiden viranomaisten kanssa. Erityisesti eurovaalien yhteydessä on tavattu myös kansainvälisten mediajättien edustajia ja keskusteltu valemedian käytöstä informaatiovaikuttamisessa.

Kysymys 2a:

Miten hyvin nykyiset ohjeistukset ja standardit kykenevät käsittelemään syvävääärennösten erityispiirteitä tietoturvauhkien torjunnassa?

- Usein sääntely kehittyy jälkijunassa, etenkin kybermaailman ilmiöihin nähden. Jonkin verran olemassa olevaa lainsäädäntöä on jo olemassa. Syvävääärennöksen tekijä ja etenkin levittäjä voi syyllistyä rikokseen (esim.

kunnianloukkaus, identiteettivarkaus), mikä saattaa ennalta estää syvävääärennösten väärinkäyttöä.

- EU:n tekoälyasetus (ja ehkäpä myös digipalvelusäädös) tuo hieman lisää sääntelyä ja läpinäkyvyyttä. Sääntelyn myötä palveluntarjoajien on esimerkiksi varmistettava, että tekoälyn tuottama sisältö on tunnistettavissa. Mediamanipulaatioiden kohdalla materiaali pitää merkitä selkeästi manipuloiduksi/keinotekoisesti tuotetuksi. Mielenkiintoista on tietysti nähdä, miten tämä sitten toteutuu. Ja kääntöpuolena voi olla se, että ihmiset luottavat mediavääärennösten sisältävän merkinnän ja valppaustaso laskee tämän myötä. Täältä löytyy myös lisätietoa: https://digital-strategy.ec.europa.eu/fi/policies/regulatory-framework-ai

Kysymys 2b:

Tuoko NIS2 tähän jotain muutosta?

- Toki sillä tasolla, että niiden toimijoiden osalta, joita NIS2 koskettaa, kyberuhkiin varaudutaan entistä paremmin riskienhallinnalla ja raportointivelvollisuus myös tiukentuu, jolloin näitä tapauksia tulee ehkä enemmän viranomaisten tietoon ja tilannekuva paranee.

- En osaa sanoa, otetaanko mediavääärennökseen suoraa kantaa NIS2 -osalta.

Kysymys 3a:

Miltä suomalaisten yritysten valmius kohdata kyberuhkia ja erityisesti syvävääärennöksiä näyttää resurssien kannalta? (teknologia, osaaminen, henkilöt)

- Riskin tiedostaminen ja siihen varautuminen, henkilöstön koulutus ja yrityksen selkeät prosessit ovat avainasemassa tätäkin kyberuhkaa torjuttaessa. Äänivääärennöstä on käytetty esimerkiksi laskutushuijausyrityksessä. Kun työntekijät ovat valveutuneita ja yrityksen maksuprosessi selvä, on vakuuttavaakin huijausta haastavampaa saada läpi.

- Yleisesti organisaatioissa on pääsääntöisesti hyvin varauduttu kyberuhkiin.

Kysymys 3b:

Onko Kyberturvallisuuskeskuksella arviota syväväärennösten vaikutuksesta suomalaiseen yhteiskuntaan?

- Tällä hetkellä tapauksia, joissa on tiedettävästi käytetty mediaväärennöstä, on tullut Kyberturvallisuuskeskuksen tietoon varsin vähän. Pitäisimme uhkaa ja vaikutuksia tällä hetkellä matalana. Tulevaisuudessa tekoälyllä tehdyt väärennökset lisääntynevät ja kehittyvät samalla entistä uskottavammiksi, jolloin tunnistaminen voi vaikeutua ja vaikutukset vakavoitua. Uhkaa arvioidessa on tarpeen ottaa huomioon erityisesti isommille joukoille kohdistetut väärennökset, joilla voidaan muun muassa vaikuttaa ihmisten mielipiteisiin, lisätä vastakkainasettelua ja horjuttaa luottamusta esimerkiksi instituutioita kohtaan. Mediaväärennösten käyttö esimerkiksi vaalivaikuttamisessa ja disinformaatiossa ovat jo tunnistettuja ilmiöitä, joita on tarpeen seurata myös Suomessa herkällä korvalla.

- Tekoälyn käyttö on kuitenkin edelleen kalliimpaa kuin pelkän geneerisen huijausviestin massalähetys eikä saavutettu tuotto välttämättä ole sen parempi. Esimerkiksi uhrille suunnattu juuri häntä varten räätälöity videomanipulaatio tai ääniväärennös sukulaisen tai työtoverin nimiin vaatii enemmän resursseja kuin massahuijausviesti, joten silloin tuotto-odotuksenkin pitäisi olla suurempi, jotta rikolliselle olisi kannattavaa käyttää tekoälyä. Kyberturvallisuuskeskuksella ei ole edellytyksiä arvioida rikollisen toiminnan psykologisia profiileja, mutta voidaan olettaa, että rikollinen liiketoiminta harvemmin panostaa kunnianhimoiseen tuotekehitykseen tai tekniseen edistykseen välittömän rikoshyödyn sijasta.

Kysymys 4:

Millaisia toimia Kyberturvallisuuskeskuksen tai muiden toimijoiden tulisi tehdä tehostaakseen syväväärennöksiin liittyvän massarikollisuuden kitkemiseksi?

- Riskien ja erilaisten tekotapojen tunnistaminen
- Tiivis yhteistyö
- Matala tiedonvaihtokynnys kansallisesti ja kansainvälisesti

- Viestintä

Kysymys 5:

Kuinka hyvin nykyinen koulutus ja kansalaisten/yritysten tiedotus syväväärennöksistä ja kyberrikollisuudesta tavoittaa ihmiset/yritykset?

- Tavoittavuutta voi olla hankalaa arvioida, meillä ei taida olla siihen mittaria. Mutta mediaväärennökset kiinnostavat ja tällä hetkellä mediaväärennöksiä koskeva tiedotus saa hyvin näkyvyyttä, koska tunnistettuja tapauksia on vähän. Tavoittavuuden osalta tiedotusta olisi hyvä suunnata erityisesti sinne, missä mediamanipulaatiota hyödyntävää massarikollisuutta esiintyy, esimerkiksi some-maailmaan.

Kysymys 6:

Muita kommentteja/ajatuksia kyberrikollisuudesta ja erityisesti syväväärennöksistä

- Kuten kyberrikollisuudessa yleisesti, myös mediaväärennösten osalta ongelmalliseksi tulee tunnistaa tekijät väärennöksien taustalla. Laajasti leviävien kampanjoiden osalta voi olla haastavaa tunnistaa, mistä väärennös on alunperin lähtenyt liikkeelle. Matala kiinnijäämisriski ja pienet tai olemattomat seuraukset ovat omiaan ruokkimaan rikollisuutta, mutta sama ongelma on yleinen kaikessa muussakin verkkohuijaamiseen ja tietojenkalasteluun liittyvässä rikollisuudessa.

Liite 7. Keskusrikospoliisin asiantuntijoiden haastattelu litteroituna

1. Lähes kaikki kyberrikollisuus on tavalla tai toisella kansainvälistä - onko tunnistettavissa joitain tiettyjä keinoja, joilla voidaan vastata syväväärennösten muodostamaan uhkaan?

Minkälaista kansainvälistä yhteistyötä keskusrikospoliisi tekee kyberrikosten osalta?

Vastaus: Tietoverkkoavusteisten petosten (TVA-petokset) Keskusrikospoliisi tekee aktiivista yhteistyötä muun muassa EUROPOLin kanssa. Eri maiden viranomaisten yhteistyöverkostot ovat keskiössä vaihdettaessa niin tutkinnallista kuin tiedustelullista tietoa.

2. Miten hyvin nykyiset ohjeistukset ja standardit kykenevät käsittelemään syväväärennösten erityispiirteitä tietoturvauhkien torjunnassa?

Tuoko NIS2 tähän jotain muutosta?

3. Miltä suomalaisten yritysten valmius kohdata kyberuhkia ja erityisesti deepfakeja näyttää resurssien kannalta? (teknologia, osaaminen, henkilöt)

Onko keskusrikospoliisilla arviota syväväärennösten vaikutuksesta suomalaiseen yhteiskuntaan?

4. Millaisia toimia keskusrikospoliisin tai muiden toimijoiden tulisi tehdä tehostaakseen syväväärennöksiin liittyvän massarikollisuuden kitkemiseksi?

5. Kuinka hyvin nykyinen koulutus ja kansalaisten/yritysten tiedotus syväväärennöksistä ja kyberrikollisuudesta tavoittaa ihmiset/yritykset?

6. Muita kommentteja/ajatuksia kyberrikollisuudesta ja erityisesti syväväärennöksistä

7. Minkäikäiset/minkälaiset ihmiset menevät eniten kyberhuijauksiin?

Kommentti: TVA-petoksissa uhrien profiili vaihtelee petostyypeittäin. Yleisellä tasolla voidaan kuitenkin todeta, että TVA-petoksiin uhriutuvia löytyy kaikista ryhmistä. Myös ns. diginatiivit uhriutuvat taitavasti tehtyihin petoksiin.

8. Paljonko rahaa suomalaiset yritykset menettävät vuodessa huijauksiin?

Kommentti: Finanssiala ry on 17.9 julkaissut uudet puolivuotisluvut koskien keskeisiä petosilmiöitä. Ne löytävät FA:n sivuilta.