

Miksi huijaukset toimivat (nyt paremmin)?

LAB-ammattikorkeakoulu

Tradenomi (AMK) tietojenkäsittely

2025

Miro Jokinen

Selvitys tekoälyn käytöstä

Tämän opinnäytetyön kirjoittaja, Miro Jokinen, vastaa koko opinnäytetyön sisällön oikeellisuudesta. Tekoälyä on käytetty kielenhuollossa satunnaisesti läpi työn, englanninkielisen yhteenvedon elävöittämisessä, sekä oppaan kuvituksessa.

Tekstiin käytetty tekoäly on OpenAI:n ChatGPT (GPT-4/ GPT-5). Kuvat on generoitu Microsoftin Bing Image Creatorilla (DALL-E 3).

Tämän opinnäytetyön aitous on tarkastettu Turnitin samankaltaisuuden tarkastusohjelmalla.

Tiivistelmä

Tekijä(t)	Julkaisun laji	Valmistumisaika
Miro Jokinen	Opinnäytetyö, AMK	2025
	Sivumäärä	
	55	
Työn nimi		
Miksi huijaukset toimivat (nyt paremmin)?		
Tutkinto ja koulutusala		
Tradenomi (AMK), tietojenkäsittely		
Toimeksiantajaorganisaatio (jos opinnäytetyöllä on toimeksiantaja)		
-		
Tiivistelmä		
<p>Työssä tutkittiin manipuloinnin, päätöksenteon ja teknologian yhteyttä. Työn tavoite oli tutustua kunkin aihepiirin kirjallisuuteen ja tunnistaa niistä keskeisimmät oivallukset, sekä esittää ne lukijalle kiinnostavalla ja helposti lähestyttävällä tavalla.</p> <p>Oletuksena oli, että keskiverto internetin käyttäjä ei ole turvallisuuden aiheista kovin kiinnostunut, vaikka riskien monimuotoisuus ja teknologian jatkuva kehitys sitä edellyttäisi. Pareton periaatteen mukaisesti työssä pyrittiin tunnistamaan maltillinen määrä kriittisimpiä havaintoja, joilla olisi suurin vaikutus lukijoiden varautuneisuuteen.</p> <p>Työ auttaa lukijaa ymmärtämään, kuinka petokset toimivat ja minkä vuoksi ne tulevat toimimaan myös tulevaisuudessa. Tavoitteena on, että työssä esiteltyihin esimerkkeihin tutustuminen auttaa tunnistamaan vastaavia tilanteita tosielämässä.</p> <p>Havainnot ja johtopäätökset herättävät myös kysymyksiä sääntelyn tarpeesta, sen nykyisestä riittävydestä sekä ajankohtaisista jatkotutkimuksista.</p>		
Asiasanat		
manipulointi, vaikuttaminen, AI, tekoäly, petos, päätöksenteko		

Abstract

Author(s)	Type of Publication	Published
Miro Jokinen	Thesis, UAS	2025
	Number of Pages	
	55	
Title of Publication		
Miksi huijaukset toimivat (nyt paremmin)?		
Degree, Field of Study		
Bachelor of Business Administration (BBA), Business Information Technology		
Organisation of the client (if the thesis work is commissioned by another party)		
-		
Abstract		
<p>This thesis looks into how manipulation, decision-making and technology are connected. The goal was to review literature on these topics and find the most important ideas, and to present them in a way that is interesting and easy to understand.</p> <p>The starting point was the assumption that the average internet user is not very interested in topics related to security, even though the risks are becoming more complex, and technology keeps developing fast. Based on the Pareto principle, the work focuses on a small number of key points that could make the biggest difference in helping readers be more prepared.</p> <p>The thesis explains how scams work and why they will likely keep working in the future. The examples and ideas shown in the text can help the reader notice similar situations in real life.</p> <p>The results also bring up questions about whether current regulations are enough and what kind of research could be useful in the future.</p>		
Keywords		
social engineering, persuasion, AI, artificial intelligence, fraud, decision making		

Sisällys

1	Johdanto.....	2
1.1	Tausta ja merkitys	2
1.2	Aiheen rajaus	4
1.3	Käytetyt menetelmät.....	4
1.4	Työn rakenne	5
2	Tutkimuskysymykset ja työn tavoite	6
3	Keskeiset käsitteet ja termit	7
3.1	Päätöksenteon anatomia ja vaikuttamisen periaatteet.....	7
3.2	Manipuloinnin hyödyntäminen hyökkäyksissä.....	19
3.3	Tekoäly.....	27
4	Tekoäly kätyrinä.....	37
5	Turvattomuuden tunne, voimattomuus ja apatia.....	43
6	Opas.....	45
7	Yhteenveto	46
7.1	Johtopäätökset ja ehdotukset.....	46
7.2	Luotettavuus.....	47
	Lähteet	48

Liite 1. Miksi huijaukset toimivat opas .pdf -muodossa

1 Johdanto

1.1 Tausta ja merkitys

Petokset, kavallukset ja ihmisten manipulointi oman edun nimissä eivät ole uusi tutkimuksen aihe. Teknologian ja koulutuksen kehityksestä huolimatta ihmiset pysyvät sukupolvesta toiseen hyvin samankaltaisina tarpeiltaan, haluiltaan ja uskomuksiltaan. Tarvetilat, kuten nälkä, jano, seksuaalinen halu, mielialat sekä tunteet ovat yhtä läsnä kuin tuhansia vuosia aiemminkin (Stajano & Wilson, 2011).

Tämän vuoksi monet huijaukset ja petokset kestävät hyvin aikaa. Järjestelmiämme suojaavan teknologian yleistyessä ja kehittyessä sitä käyttävä ihminen voidaan usein kokea puolustusten heikoimmaksi kohdaksi. Esimerkiksi vuoden 2023 tietomurtoja käsittelevä DBIR-raportti kertoo, että 74 % tapauksista syynä on inhimillinen virhe, valtuuksien väärinkäyttö, varastetut tunnukset tai sosiaalinen manipulointi. 83 % tapauksissa tekijät olivat organisaation ulkopuolelta ja 95 % tapauksista motiivi oli taloudellinen hyöty. (Verizon, 2023). Kohde voi olla kuka tahansa, yksityishenkilöistä suuryrityksiin. Seuraamusten kirjo on laaja ja ankara; mitä tahansa mainehaitasta ja taloudellisista menetyksistä kansallisen turvallisuuden heikkenemiseen (Zaoui ym., 2024).

Tuore ongelma on, että erilaiset tekoälytyökalut täydentävät tehokkaasti olemassa olevien petosten heikkouksia ja helpottavat niiden yleistettävyyttä. Esimerkiksi tekstin muodossa edistetyn romanssihuijauksen tekijän on vielä lähivuosina tarvinnut joko keksiä tekosytä, miksi hän ei voisi näyttäytyä kuvissa tai käyttää sosiaalisesta mediasta varastettuja kuvia ominaan. Generatiivisten AI -työkalujen ja -palveluiden avulla voidaan kuitenkin helposti luoda uskottavan näköisiä, mutta täysin kuvitteellisia ihmisiä pelkän komentokehoteen perustella tai manipuloida olemassa olevia kuvia ja niissä esiintyviä ihmisiä. Deepfake -teknologian eli syvävääreännösten avulla voidaan vaihtaa videotallenteilla esiintyviä ihmisiä toisiin ja tekoälyä käyttävät filterit, suodattimet mahdollistavat jopa reaaliaikaisen kasvojen korvaamisen toisen ihmisen kasvoilla videopuhelussa. Myös tallennetun äänen kaappaaminen ja mielivaltaisten viestien tuottaminen toisen ihmisen äänellä on kaikkien saavutettavissa, jos kopioitavaa puheääntä kyetään nauhoittamaan. Äärimmäisenä esimerkkinä alkuvuodesta 2024 insinööritoimisto Arup:n on uutisoitu menettäneen 25 miljoonaa dollaria rikollisen tekeydyttyä yrityksen talousjohtajaksi videopuhelussa (World Economic Forum, 2024).

Aiheesta kiinnostumattoman ihmisen on entistä vaikeampi tuntea teknologian tuoreimpia ominaisuuksia ja mahdollisuuksia, mikä tekee petollisen toiminnan tunnistamisen yhä vaikeammaksi. Yhteiskunnan digitalisoituminen on myös lisännyt hyökkäysten tarttumapintaa yksilöiden tasolla. Raha-asioita ei tarvitse hoitaa pankeissa tai edes perinteisillä valuutoilla,

vaan rahaa voidaan lähettää esimerkiksi älypuhelimesta käsin tai jopa vaikeasti henkilöitävinä bitcoineina ja muina kryptovaluuttoina, tai vaikka kansainvälisesti toimivina lahjakortteina ja tilauksina.

Sähköpostien liitteinä kulkee esimerkiksi terveystietoja ja sopimuksia ja viestisovellukset ja sosiaalisen median yksityiset keskustelut sisältävät henkilökohtaisia keskusteluja – mahdollisesti myös arkaluontoista tietoa, kuvia tai videoita. Lisäksi potentiaalisten uhrien ja heidän perheidensä, sekä kavereidensa sosiaalisen median käyttö tuottaa paljon tietoa henkilökohtaisista tavoista, mieltymyksistä, ihmissuhteista ja sijainneista, joka puolestaan helpottaa petosten tekijöitä personoimaan profiilinsa ja tarinansa juuri uhriin vetoavaksi. Sosiaalisen median alustat myös mielellään säilyttävät käyttäjiensä tietoja kehittääkseen mainontaansa, sekä mahdollistavat esimerkiksi samoilla tunnuksilla useaan palveluun kirjautumisen, joka lisää käytettävyyttä, samalla kuitenkin tehden alustoista erityisen hyviä kohteita rikollisille (Rathod ym. 2025).

Samanaikaisesti siis tuotetaan ja säilytetään ennätysmäärä kiristykseen soveltuvaa sisältöä, jaetaan itsestämme ja läheisistämme jopa reaaliaikaisesti tietoa – kaikkea sijainnista mieltymyksiin ja kotiemme sisältöön – ja hallitaan talouttamme aina mukana olevien älylaitteiden avulla. Satunnainen *ei-kukaan* on siis parempi uhri kuin koskaan ennen. Tuotetun materiaalin lisäksi olemme myös tavoitettavissa lukuisissa erilaisissa palveluissa ja ympäristöissä, mikä puolestaan lisää pinta-alaa hyökkäyksille (Wang ym., 2021).

Osana yhteisöjä ja organisaatioita on käyttäjällä aina oltava tiettyjä oikeuksia käyttämiinsä järjestelmiin. Teknisten suojausten kehittyessä kyseinen käyttäjä voikin pahimmassa tapauksessa olla kuvainnollinen heikko lenkki olemalla yksinkertaisin ja kustannustehokkain tapa päästä luvatta käsiksi tavoiteltuun tietoon ja valtuuksiin. Tunkeutuja kun voi yksinkertaisimmillaan saavuttaa tavoitteensa vain pyytämällä oikealta ihmiseltä oikealla tavalla. Esimerkiksi tunnetun NSA tietovuotaja Edward Snowdenin kerrotaan käyttäneen 20–25 kollegan tunnuksia ja salasanoja, jotka hän oli saanut pyytämällä ja perustelemalla niiden tarvetta työtehtävillään järjestelmän ylläpitäjänä (Wang ym., 2020).

Kyberturvallisuuskeskuksen tutkijat varoittavat tekoälyjen mahdollistamia kyberhyökkäyksiä käsittelevässä julkaisussaan, että tekoälyn käyttö tulee johtamaan parempiin, nopeampiin, piiloutuvampiin ja vaikeammin ennustettaviin hyökkäyksiin (Aksela ym., 2024). Teknologian nopeat harppaukset muodostavat uudet vaatimukset verkon käyttäjien taidoille ja ymmärrykselle. Oppilaitosten ja työpaikkojen ulkopuolella ihmisiä on myös vaikeaa tavoittaa tärkeänkään aiheen nimissä, eikä oppimista voida vasten tahtoa kenellekään pakottaa. Varautumisen ja turvallisuuden aiheet voidaan kokea epäkiinnostavina tai ahdistavina ja

motivaatio niihin perehtymiseen löytyy vasta, kun uhka konkretisoituu lähipiirissä tai omalla kohdalla.

Kansan tietoisuutta pyrkii lisäämään esimerkiksi Liikenne- ja Viestintävirasto Traficom, mutta tietoisuuden lisääminen moderniin tapaan esimerkiksi sosiaalisessa mediassa ja pieninä tietoisukuina tai lyhyinä videoina on ainakin toistaiseksi lähinnä muutamien sisällöntuottajien harteilla ja tuotanto pyörii pääasiassa harrastuksen tasolla. Työtä tiedottamisen eteen on siis tehtäväksi. Tämän työn motivaatio on ennemminkin ajaa muutosta auttamalla tunnistamaan syy-seuraus -suhteita, kuin opettamalla ulkoa mitä saa tai ei saa tehdä.

1.2 Aiheen rajaus

Työ keskittyy inhimillisiin mekanismeihin ja tekoälytyökalujen epärehelliseen käyttöön. Tekoälyjen perusteet esitellään lyhyesti, mutta tarkemmat toimintaperiaatteet, matematiikka tai teknologia työkalujen toiminnan takana eivät kuulu työn sisältöön. Kyberturvallisuus liittyy aiheeseen tiiviisti, mutta sitä sivutaan varsin pinnallisesti, painopisteen ollessa inhimillisyydessä. Työssä ei myöskään oteta kantaa tekoälytyökalujen odotettuihin, turvallisuutta parantaviin mahdollisuuksiin, joita myös kehitetään. Rikolliseen toimintaan viitataan pääasiassa hypoteettisin esimerkein. Rikollisuutta liiketoimintana ei esitellä tarkemmin, mutta huijauksen yleistettävyyden ja automatisointi kuuluu vielä työn rajauksiin. Rikollinen toiminta nähdään tuomittavana ja sitä halutaan ymmärtää nimenomaan suojautumisen ja ehkäisemisen vuoksi. Aihetta koskevaa lainsäädäntöä ei ole tarkoitus esitellä yksityiskohtaisesti, sen ollessa monimutkainen ja jopa keskeneräinen kokonaisuus. AI-huijauksilla on vahva yhteys myös informaatiovaikuttamisen ja disinformaation kentässä, mutta työn aiheisiin niitä ei sisällytetä.

1.3 Käytetyt menetelmät

Työssä käytetään laadullista tutkimusta ja kirjallisuuskatsausta, selvittämään psykologian ja teknologian synergiaa. Petokset ja manipulointi ovat aiheena vähemmän tunnettuja, sillä niihin lankeaminen aiheuttaa esimerkiksi taloudellisten seuraamusten lisäksi uhrissaan häpeää, jos kyseessä on yksityishenkilö tai mainehaittaa, jos uhrina on yritys. Petosten uhreilta jääkin usein ilmoittamatta niin viranomaisille kuin mahdollisille asiakkailleen, mitä on tapahtunut. Tämä on keskeisiä syitä, minkä vuoksi tutkimusta tehdään olemassa olevan kirjallisuuden avulla, eikä esimerkiksi kyselyinä ja haastatteluina. Erilaiset petokset eivät myöskään ole hyvin verrattavissa toisiinsa, kun seuraukset voivat olla mitä tahansa mainehaitasta konkurssiin ja sydänsuruista työkyvyttömyyteen.

Huomionarvoista työhön liittyen on, että valtaosalle terminologiasta ei ole virallista tai laajalle levinnyttä suomennosta. Parhaaksi käytännöksi on katsottu, että paremmin tunnettu englanninkielinen käsite kulkee aina mukana mahdollistaen näin aiheeseen syventymisen esimerkiksi hakukoneiden avulla. Suomenkieliset käännökset sisällytetään tekstiin, mutta ne vaihtelevat vakiintuneiden termien ja suorien käännösten välillä. Yhtenäisten standardien puutteen ja työn perehdyttävän tavoitteen vuoksi lyhyet kuvaukset suuresta määrästä havaintoja on koettu tarkoituksenmukaiseksi.

1.4 Työn rakenne

Kappaleessa kaksi kerrotaan, mitä tehdään ja miksi. Työn tavoitteiden ja tutkimuskysymysten sanoittaminen määrittelee vaatimukset kolmannen kappaleen sisällölle. Se myös asettaa raamit, joiden avulla työn merkitystä ja onnistumista voidaan tarkastella.

Kappale kolme esittelee keskeisiä käsitteitä ja tavoitteiden kannalta merkittäviä tutkimuksia. Työn poikkitieteellisyyden vuoksi aiheiden kirjo on kattava. Käsitteet eivät ilmene tärkeysjärjestyksessä, eikä niitä ole tarkoitus arvottaa toisiaan korkeammalle. Kunkin osan syvyys ja laajuus on pyritty rajaamaan siten, että se tuottaa alustavan ymmärryksen aiheista vaatimatta tiettyä ennakkotiedon tasoa.

Kappaleet neljä ja viisi käsittelevät aiemman tietopohjan ja teorian avulla aiheen haasteita ja riskejä nykyhetkessä ja tulevaisuudessa. Huomionarvoista on, että näiden kappaleiden sävy on tietoperustaa hypoteettisempi. Kappaleet sisältävät omaa pohdintaa ja yksilön arvioita tulevasta, jota ei ole ennustettu tieteellisesti.

Kappale kuusi avaa ajatusprosessia työn osana toteutetun oppaan taustalla ja esittelee sen rakenteen.

Seitsemäs ja viimeinen kappale vetää yhteen mitä työstä on opittu, arvioi havaintojen luotettavuutta ja esittää aiheeseen liittyviä jatkotutkimuksia, sekä toimenpiteitä.

2 Tutkimuskysymykset ja työn tavoite

Työn tavoite on tunnistaa, miksi petokset toimivat ja kuinka petollista toimintaa voidaan tehostaa teknologian avulla. Opinnäytetyön osana toteutetaan helppolukuinen ja kansantaajuinen opas manipuloinnin tunnistamisesta. Oppaan tavoite on tuoda niin sanotusti tavallisen ihmisen, internetin ja palveluiden käyttäjän tietoon manipuloinnin perusteet ja alleviivata epäilyksen sekä varmistamisen roolia omassa toiminnassa, asemasta riippumattomasti. Oppaan sisältö esitetään mahdollisimman yksinkertaisin termein, jotta niistä voisi hyötyä kuka tahansa, taustastaan riippumatta. Työn ei ole tarkoitus pelotella lukijaansa tai yksityiskohtaisesti kieltää tekemästä erilaisia asioita. Päätöksentekoon vaikuttavien tekijöiden näkyväksi tekeminen ja esittely jättää lukijaansa muistijäljen, joka toivottavasti aktivoituu, kun he kohtaavat jotain potentiaalisesti epäilyttävää.

Tutkimuskysymykset tavoitteiden saavuttamiseksi ovat seuraavat;

Mitkä inhimilliset tekijät tekevät ihmisistä alttiita huijauksille?

Miten tekoäly voi tehostaa sosiaalista manipulointia?

3 Keskeiset käsitteet ja termit

3.1 Päätöksenteon anatomia ja vaikuttamisen periaatteet

Ihmisten päätöksenteko on monitahoinen aihealue. Erilaisten valintojen ja päätösten muuttajat eivät välttämättä ole kirkkaita edes päättäjälle itselleen, sillä niiden kirjo on valtava ja vaikeasti todennettava. Pelkistettynä päätöksenteon voidaan kuvailla koostuvan kahdesta puolesta; ensimmäinen on tunnepohjainen, automaattinen ja nopea ja toinen puolestaan tiedollinen, hidas ja harkittu (Wang ym., 2021). Yksilöllisten kokemusten, arvojen ja opittujen tapojen rinnalla meihin kohdistuu esimerkiksi yhteisöllisiä odotuksia siitä, mikä on normaalia ja kuinka kuuluu toimia. Myös levon määrä ja laatu sekä ruokailut vaikuttavat sekä tarkkaavaisuuteen ja mielialaan että maailmankuvaan ja sitä kautta päätöksentekoon. Tähän yhtälöön lisätään vielä ainakin tunteet, jotka voimakkaimmillaan voivat olla kuin vastakohta rationaaliselle päätöksenteolle ja ovat vain osittain omassa hallinnassamme.

Päätöksenteko voisi olla haastavaa tyhjiössäkin, ilman inhimillisiä tuntemuksia. Joissain ongelmissa voimme käyttää hyväksemme oppimiamme asioita, kuten matematiikkaa ja sillä tavoin saavuttaa hyvin johdonmukaisia ja perusteltuja tuloksia. Voimme myös kohdata ongelmia, joissa muuttujien määrä tai monimutkaisuus on niin suuri, että kaiken huomioon ottaminen tarjoaa lopulta vain näennäisen varmuuden tunteen eli pelkkä rationaalinen kaavan seuraaminen ei riitä (Shabnam & Gerd, 2014). Suuri osa elämän kokemuksista on kuitenkin niin vieraita, odottamattomia tai muuten ainutlaatuisia, ettei sopivaa verrokkia tai sovellettavaa kaavaa ei löydy. Suuri osa päätöksistä tehdäänkin intuitiolla.

Heuristiikat ja kognitiiviset vinoumat

Rationaalisen päättelyn teorian mukaan, ihmiset etsivät parhaan ratkaisun käytettävissä olevan tiedon perusteella. On kuitenkin esitetty, että usein ihmiset päätyvät *riittävän hyvään* ratkaisuun silläkin riskillä, että lopputulos ei välttämättä ole paras mahdollinen (Stajano & Wilson, 2011). Heuristiikat, *heuristics*, ovat aivojemme tapa keventää päätöksenteon kuormaa ja nopeuttaa prosessia. Jokaisen ainutlaatuisen muuttujan huomioimisen sijaan voimme perustaa arviomme samankaltaiseen tilanteeseen, jonka olemme kokeneet aiemmin. Ne ovat on päätöksenteon strategia, jota käytetään joko tietoisesti tai tiedostamatta. Tiedostamaton käyttö tunnetaan intuitionä eli tilanteena, jossa tietää mitä tehdä, vaikkei osaa selittää miksi (Shabnam & Gerd, 2014). Heuristiikat toimivat parhaiten, kun käyttäjällä on asiantuntemusta tai johdonmukaista kokemusta ratkaistavan ongelman aihealueesta (Korteling & Toet, 2020).

Heuristiikat eivät itsessään ole kielteinen asia, mutta ne voivat altistaa kognitiivisille vinoumille, *cognitive bias*. Olemme esimerkiksi alttiita etsimään ja hyväksymään sellaista

tietoa, joka vastaa aiempia uskomuksiamme, sekä yliarvioimaan kuinka ennustettavissa jokin tapahtuma oli, kun se on jo tapahtunut. (Berthet, 2022) Vinouman tai vääristymän, *bias*, yleinen määritelmä on vahva tunne jonkin ihmisryhmän tai väittelyn osapuolen puolesta tai sitä vastaan, usein ilman reilua ja puolueetonta harkintaa. Vinoumia ilmenee eri muodoissa, esimerkiksi ennakkoluuloina, eturistiriitoina, tilastollisia vinoumina sekä kognitiivisina vinoumia (Soprano ym., 2024).

Aiheen perustasoinen ymmärrys auttaa paitsi tunnistamaan manipuloinnin yrityksiä, myös ymmärtämään ihmisten toimintaa yleisesti. Vinoumat voivat vääristää päätöksentekoamme ja niitä esiintyy meissä kaikissa. Vaikka olemme yksilöitä, ovat sisäiset prosessimme hyvin samankaltaisia. Tästä syystä vinoumat ovat hyvin ennustettavissa ja siten soveltuvat manipuloinnin välineiksi. Esimerkkeinä yleisistä ja arkisista vinoumista voidaan mainita vaikka oletus siitä, että kalliimpi tarkoittaa parempaa ja edullisempi huonompaa (Cialdini, 1993).

Aivomme pyrkivät optimoimaan ja nopeuttamaan tiedon prosessointia ja päätöstentekoa erilaisilla oikopoluilla, kuten heuristiikoilla. Tämä tekee tiedon käsittelystä nopeampaa, mutta voi heikentää tarkkuutta. Toinen merkittävä tekijä on, että emme ole puhtaan rationaalisia olentoja, vaan tunteilla on merkittävä vaikutus päätöksentekoomme. Mahdollisesti jopa enemmän kuin rationaalisella ajattelulla. Kolmantena, olemme niin sanotusti ympäristömme tuote. Aiemmat kokemukset, kulttuuri, yhteisöt, joihin kuulumme ja muiden odotukset ovat kaikki esimerkkejä muuttujista, joilla on vaihteleva painoarvo sekä päätöksissämme että tavassamme katsoa maailmaa. Vinoumien olemassaolon tiedostaminen on olennaisempaa kuin niiden ulkoa opettelu ja toisistaan erottaminen.

Tunnetuille vinoumille ei ole virallista määrää, vaan se vaihtelee lähdemateriaalin mukaan. Lähteestä riippumatta, puhutaan kuitenkin runsaasta kokoelmasta. Esimerkiksi *Cognitive Bias Codex*, epävirallinen kokoelma tunnettuja vinoumia, listaa niitä 160 erilaista. Soprano ym. (2024) tutkimuksessaan vinoumien roolista faktan tarkistukseen nimeävät 221 erilaista vinoumaa, joista ainakin 39 todetaan vaikuttavaan kyseiseen tehtävään. Ammattimaista faktan tarkastusta ei voida suoraan verrata yksilön päätöksentekoon, mutta aiheissa voidaan katsoa olevan päällekkäisyyttä. Myös Vincent Berthetin (2022) tutkimus kognitiivisten vinoumien vaikutuksista asiantuntijatyöhön ja sen kymmenen keskeisintä tunnistettua vinoumaa, joista kuusi nousee esille myös ensin mainitussa tutkimuksessa, jakavat yhtäläisyyksiä yksilön päätöksenteon, sekä tietoisien manipuloinnin kanssa. Molempien tutkimusten löydösten koetaan tarjoavan käyttökelpoista näkökulmaa myös tähän työhön. Vinoumat, jotka ilmenevät molemmissa töissä on korostettu asteriskilla.

<p>Kognitiiviset vinoumat ammattimaisessa faktan tarkastuksessa (Soprano ym., 2024)</p>
<p>Affect Heuristic – vaikutelmaheuristiikka. Tunteet vaikuttavat usein päätöksentekoon. Esimerkiksi jos kuulija pitää puhujasta voi objektiivisuus kärsiä.</p>
<p>Anchoring bias – ankkurointivaikutus*. Viittaa taipumukseen tehdä arvioita – varsinkin numeerisia – nojautuen ensimmäiseen saatavilla olevaan tietoon, jonka pohjalta myöhempiä arvioita mukautetaan.</p>
<p>Attentional Bias – tarkkaavaisuusharha. Viittaa taipumukseen havaita ja painottaa toistuvasti mieleen palautuvia tai emotionaalisesti merkittäviä asioita muiden kustannuksella. Tämä voi johtaa virheellisiin arvioihin esimerkiksi silloin, kun mediassa korostetaan jatkuvasti tiettyjä aiheita.</p>
<p>Authority Bias – auktoriteettiharha. Viittaa taipumukseen antaa suurempi painoarvo auktoriteettihenkilön mielipiteelle tai näkemykselle – riippumatta sen sisällöllisestä oikeellisuudesta. Arviointiin voi vaikuttaa esimerkiksi se, kuka puhuu tai mikä taho on tiedon esittänyt. Tämä vinouma voi ilmetä myös siten, että taho, joka on aiemmin esittänyt uskottavia väitteitä, saa enemmän uskottavuutta uusille väitteilleen ilman erillistä tarkastelua.</p>
<p>Automation Bias – automaatioharha. Tarkoittaa taipumusta luottaa liiallisesti automaattisiin järjestelmiin, jopa silloin kun ne tekevät virheitä tai ristiriitaisia päätelmiä ihmisen oman arvioinnin kanssa.</p>
<p>Availability Cascade – saatavuuden kierre. Uskomus saattaa alkaa tuntua todennäköisemmältä tai oikealta pelkästään siksi, että se on laajasti esillä ja yleisesti tunnettu. Toisto ja näkyvyys lisäävät uskomuksen uskottavuutta, vaikka sille ei olisi vahvaa näyttöä.</p>
<p>Availability bias – saatavuusharha*. Kuvaa ihmisen taipumusta arvioida tapahtumien todennäköisyyttä sen perusteella, kuinka helposti vastaavia esimerkkejä tulee mieleen.</p> <p>Esimerkiksi lento-onnettomuuksien laaja uutisointi voi saada lentomatrustamisen tuntuun paljon turvattomammalta kuin mitä se todellisuudessa on.</p>
<p>Backfire Effect – reaktiivinen torjunta. Kun omaa uskomusta vastaan esitetään ristiriitaista tietoa, usko alkuperäiseen näkemykseen saattaa jopa vahvistua sen horjumisen sijaan. Ilmiö on yleinen erityisesti vahvasti ideologisissa tai tunteita herättävissä kysymyksissä.</p>

Bandwagon Effect – laumavaikutus. Ihminen alkaa uskoa tai tehdä jotakin, koska muutkin tekevät tai uskovat niin. Mitä useampi ihminen kannattaa väitettä tai ideaa, sitä houkuttelevammalta se alkaa vaikuttaa – riippumatta sen todellisesta paikkansapitävyydestä.

Barnum Effect / Forer Effect – Barnum-ilmiö / Forer-vaikutus. Ihminen tunnistaa itseensä sopivia merkityksiä epämääräisistä ja yleisluontoisista väitteistä. Henkilö tulkitsee viestin osuvaksi, koska täyttää aukot omilla kokemuksillaan tai uskomuksillaan – vaikka sama väite voisi sopia lähes kenelle tahansa. Ilmenee esimerkiksi persoonallisuustestien ja horoskooppien yhteydessä.

Base Rate Fallacy – perustason harha. Tarkoittaa sitä, että kokonaisuuden tarkastelun sijaan keskitytään liiaksi johonkin yksittäiseen tietoon, mikä voi vääristää päätöksentekoa. Esimerkiksi rikosta tutkittaessa, epäillyn aiemman tuomion ei pitäisi vaikuttaa siihen, miten hänen syyllisyyttään arvioidaan nykyisessä tapauksessa.

Belief Bias – uskomusharha. Viittaa siihen, että arvioija antaa liikaa painoarvoa jonkun argumentin loogisuudelle sen vuoksi, että lopputulos tai johtopäätös tuntuu oikealta tai uskottavalta. Esimerkiksi rikkaan ihmisen taloudelliset vinkit saattavat tuntua merkityksellisiltä hänen vaurautensa perusteella, vaikka ne eivät välttämättä kestäisi tarkempaa arviointia.

Choice-Supportive Bias – valintamyönteisyys. Eli henkilön taipumus muistaa omat valintansa parempina kuin ne todellisuudessa olivat. Esimerkiksi impulsiivisen tai huonosti harkitun hankinnan jälkeen henkilö voi pyrkiä keskittymään sen positiivisiin puoliin ja keksimään itselleen selityksiä, miksi valinta oli oikea.

Compassion Fade – myötätunnon väheneminen tarkoittaa sitä, että ihmiset kokevat enemmän myötätuntoa pienempää uhriryhmää kohtaan, mutta heidän tunteensa heikenevät, kun kyseessä on suurempi ryhmä. Esimerkiksi yksittäinen tulipalossa menehtynyt perhe voi tuntua uutisena paljon koskettavammalta kuin uutiset, joissa kerrotaan suurista konflikteista, joissa uhreja on kymmenistä satoihin.

Confirmation bias – vahvistusharha*. Tarkoittaa taipumusta etsiä, tulkita, painottaa ja muistaa tietoa tavalla, joka tukee henkilön aiempia uskomuksia.

Conjunction Fallacy /Linda Problem – yhdistelmätodennäköisyyden harha /Linda-ongelma. Usein arvioidaan, että kahden samanaikaisen tapahtuman todennäköisyys on korkeampi, kuin yksittäisen tapahtuman.

Linda -ongelmalla tarkoitetaan usein aiheesta käytettyä esimerkkiä:

Linda on 31-vuotias, älykäs ja kantaottava nainen, joka opiskeli filosofiaa ja osallistui opiskeluaikana mielenosoituksiin. Kumpi on todennäköisempää?

- a) Linda on pankkivirkailija
 b) Linda on pankkivirkailija ja feministi

Monet valitsevat vaihtoehdon b, vaikka yksittäinen tapahtuma on aina todennäköisempi kuin yhdistelmä.

Conservatism Bias – konservatismiharha. Uutta tietoa ei oteta riittävästi huomioon uskomusta päivitettäessä – mielipidettä ei muuteta yhtä paljon kuin olisi perusteltua tiedon määrän ja laadun perusteella.

Esimerkiksi siten, että henkilö uskoo tietyn ruoka-aineen tai lääkkeen haitallisuuteen. Hän saa käsiinsä laajan ja korkealaatuisen tutkimusaineiston, joka osoittaa sen olevan turvallinen. Silti hän muuttaa kantaansa vain vähän tai ei lainkaan, pitäytyen lähes alkuperäisessä uskomuksessa. Kyseessä ei välttämättä ole päätöksenvaarainen itsepäisyys vaan ajattelun vinouma. Uutta tietoa painotetaan vähemmän kuin aiempia käsityksiä.

Consistency Bias – johdonmukaisuusharha. Ihminen muistaa tai tulkitsee aiempia arvioitaan ja tekojaan nykyisen käyttäytymisensä tai uskomustensa mukaisesti – ikään kuin ne olisivat aina olleet yhdenmukaisia.

Courtesy Bias – kohteliaisuusharha. Todellisen vastauksen tai mielipiteen sijaan annetaan usein sosiaalisesti hyväksyttävä tai miellyttävä vastaus, jotta ketään ei loukattaisi. Myös esimerkiksi valta-asetelmat ja sosiaaliset suhteet voivat vaikuttaa tähän.

Declinism – rappeutumisharha. Menneisyys muistetaan toisinaan ylikorostuneen myönteisesti ja tulevaisuuteen suhtaudutaan perusteettoman pessimistisesti. Ilmiö näkyy esimerkiksi puheissa siitä, kuinka *ennen kaikki oli paremmin*, esittämättä mitään tarkastelua kestäviä perusteluita – jolloin kyse on tuntemuksesta eikä faktapohjaisesta arviosta.

Dunning-Kruger Effect – Dunning–Kruger-vaikutus. Kun ihmisellä ei ole merkittävää aiheen tuntemusta, hän saattaa yliarvioida oman osaamisensa – koska häneltä puuttuu myös kyky arvioida omaa osaamistaan realistisesti.

Framing Effect – kehystysvaikutus. Sama tieto voidaan kokea hyvin eri tavoin, riippuen kuinka se on *kehystetty*. Esimerkiksi tapaturmasta uutisoidessa, väitteet ”kuusi kymmenestä selviytyi” sekä ”neljä kymmenestä menehtyi” tarkoittavat samaa, mutta

selviytymisen painottaminen kuulostaa myönteisemmältä. Kehyksellä voi siis olla vaikutusta siihen, kuinka tieto koetaan.

Fundamental Attribution Error – perusattribuutiovirhe. Ihmisten toimintaa selitettäessä painotetaan liikaa yksilön luonnetta tai persoonallisuutta ja aliarvioidaan tilanteen ja ympäristön vaikutuksia. Esimerkkinä työttömyyden tai köyhyyden yhdistäminen laiskuteen.

Google Effect – Google-ilmiö. Tietoa muistetaan huonommin, jos se on helposti löydettävissä verkosta. Ihminen ei välttämättä paina mieleensä asioita, koska tietää voivansa tarkistaa ne myöhemmin hakukoneella.

Hindsight bias – jälkiviisausvinouma*. Viittaa taipumukseen kokea tapahtumat jälkikäteen ennustettavammiksi kuin ne todellisuudessa olivat ennen tapahtuman toteutumista.

Hostile Attribution Bias – vihamielisysoletus. Tarkoittaa sitä, että henkilö tulkitsee toisen käyttäytymisen vihamieliseksi, vaikka se ei oikeasti olisi sitä. Esimerkiksi nauru voidaan kokea henkilöön kohdistuvaksi, jos tällä on aiempia kokemuksia syrjinnästä tai kiusaamisesta.

Illusion of Validity – pätevyysilluusio. Ilmenee siten, että henkilö yliarvioi jonkun arvioinnin tai tuomion oikeellisuuden, kun saatavilla oleva tieto on johdonmukaista. Esimerkiksi työntekijän ehdotusta voidaan pitää hyvänä sen vuoksi, että aiemmat ideat ovat olleet hyviä – eikä sen vuoksi, että kyseistä ehdotusta on tarkasteltu ja se on todettu hyväksi itsessään.

Illusory Correlation – harhainen korrelaatio. Tarkoittaa ilmiötä, jossa henkilö näkee yhteyden kahden muuttujan, kuten tapahtuman, käytöksen tai ihmisen välillä, vaikka ne eivät olisi yhteydessä toisiinsa.

Illusory Truth Effect – totuusilluusio. Kun tietoa toistetaan riittävästi, se alkaa tuntua todenmukaisemmalta, vaikka mitään todisteita sen suuntaan ei esitettäisi.

Ingroup Bias – sisäryhmäharha. Tarkoittaa taipumusta suosia oman ryhmän jäseniä ja nähdä heidät positiivisemmassa valossa verrattuna muihin ryhmiin. Ryhmän tarkoittaessa tässä yhteydessä mitä tahansa joukkoa, johon henkilö kokee kuuluvansa – esimerkiksi seuraa tai puoluetta.

Just-World Hypothesis – oikeudenmukainen maailma -harha. Tarkoittaa uskomusta, että maailma on oikeudenmukainen paikka, jossa ihmiset saavat sen, mitä ansaitsevat.

Tämä voi ilmetä esimerkiksi suurempana ymmärryksenä hyväosaisia kohtaan ja empatian puutteena heikommassa asemassa olevia kohtaan – joiden vaikeudet saatetaan nähdä heidän omaksi syykseen.

Optimism Bias – optimismiharha. Tarkoittaa taipumusta olla ylipositiivinen ja aliarvioida ei-toivottujen tapahtumien todennäköisyyksiä sekä yliarvioida myönteisten ja miellyttävien tapahtumien todennäköisyyksiä *omalla kohdalla*. Harhaan liittyy uskomus siitä, että negatiiviset asiat sattuvat todennäköisemmin muille kuin itselle, vaikka niiden riski yleisesti tiedostettaisiin.

Ostrich Effect – strutsiefekti. Tarkoittaa taipumusta vältellä kielteistä, mutta mahdollisesti hyödyllistä tietoa – kuten esimerkiksi palautetta omasta edistymisestä – psykologisen epämukavuuden välttämiseksi. Tämä voisi ilmetä esimerkiksi talouden tilan tai terveystietojen välttelyä, kun tieto koetaan ahdistavana

Outcome bias, lopputulosvinouma*. Liittyy päätöksen laadun tuomitsemiseen lopputuloksen perusteella. Päätöksenteon laatua tulisi arvioida saatavilla olleen tiedon perusteella, ennen lopputuloksen tietämistä.

Overconfidence bias, ylikuottamusvinouma*. Viittaa ammattilaisen taipumukseen yliarvioida omaa tietämystään – koska on ammattilainen. Jos aihe on tuttu, sitä ei välttämättä tarkastella yhtä huolellisesti ja objektiivisesti kuin olisi syytä.

Proportionality Bias – suhdeharha. Tarkoittaa taipumusta olettaa, että suurilla tapahtumilla on oltava suuret syyt. Tämä inhimillinen ajattelumalli voi selittää esimerkiksi sen, miksi jotkut ihmiset uskovat salaliittoteorioihin: he kokevat, ettei suuria tapahtumia voi selittää yksinkertaisilla tai sattumanvaraisilla syillä.

Saliency Bias – erottuvuusharha. Tarkoittaa taipumusta kiinnittää huomiota asioihin, jotka ovat näkyviä, erottuvia tai emotionaalisesti voimakkaita – ja jättää vähemmän huomiota herättävät asiat huomiotta, vaikka ero ei olisi objektiivisesti merkityksellinen. Esimerkkinä voidaan käyttää suorituksesta annettua palautetta; vaikka valtaosa olisi positiivista, yksittäinen erittäin negatiivinen jää parhaiten mieleen.

Stereotypical Bias – stereotipiaharha. Tarkoittaa taipumusta syrjintään tai erittäin yleistettyihin olettamuksiin. Esimerkiksi uskomus siitä, että kaikki tiettyä sukupuolta, etnisyyttä tai muuta ryhmää edustavat ovat tietynlaisia, ei kestä tarkastelua eikä johda laadukkaisiin päätöksiin.

Telescoping Effect – teleskooppivaikutus. Tarkoittaa taipumusta siirtää viimeaikaisia tapahtumia taaksepäin ajassa ja kaukaisempia tapahtumia kohti nykypäivää, niin että äskettäiset tapahtumat näyttävät kaukaisempina ja kaukaiset tapahtumat lähempinä.

Kognitiiviset vinoumat asiantuntijatyössä (Berthet, 2022)

Anchoring bias – ankkurointivaikutus*. Viittaa taipumukseen tehdä arvioita – varsinkin numeerisia – nojautuen ensimmäiseen saatavilla olevaan tietoon, jonka pohjalta myöhempiä arvioita mukautetaan.

Availability bias – saatavuusharha*. Kuvaa ihmisen taipumusta arvioida tapahtumien todennäköisyyttä sen perusteella, kuinka helposti vastaavia esimerkkejä tulee mieleen.

Esimerkiksi lento-onnettomuuksien laaja uutisointi voi saada lentomatkustamisen tuntuun paljon turvattomammalta, kuin mitä se todellisuudessa on.

Confirmation bias – vahvistusharha. Tarkoittaa taipumusta etsiä, tulkita, painottaa ja muistaa tietoa tavalla, joka tukee henkilön aiempia uskomuksia.

Disposition effect – dispositioefekti. Ilmenee sijoittajien käyttäytymisessä siten, että voitollisia sijoituksia myydään liian aikaisin, kun taas tappiollisista pidetään kiinni liian pitkään. Tämä käyttäytyminen liittyy usein tappiokammoon.

Hindsight bias – jälkiviisuusvinouma*. Viittaa taipumukseen kokea tapahtumat jälkikäteen ennustettavammiksi kuin ne todellisuudessa olivat ennen tapahtuman toteutumista.

Omission bias – toimimattomuusvinouma. Tarkoittaa taipumusta valita vaihtoehto, jossa mahdollinen haitta aiheutuu passiivisuudesta eikä aktiivisesta toiminnasta, vaikka seuraukset olisivat yhtä vakavia tai jopa vakavampia.

Outcome bias – lopputulosvinouma*. Liittyy päätöksen laadun tuomitsemiseen lopputuloksen perusteella. Päätöksenteon laatua tulisi arvioida saatavilla olleen tiedon perusteella, ennen lopputuloksen tietämistä.

Overconfidence bias – ylikuottamusvinouma*. Viittaa ammattilaisen taipumukseen yliarvioida omaa tietämystään – koska on ammattilainen. Jos aihe on tuttu, sitä ei välttämättä tarkastella yhtä huolellisesti ja objektiivisesti kuin olisi syytä.

Relative risk bias – suhteellisen riskin vinouma. Ihmiset suosivat esimerkiksi tiettyä hoitovaihtoehtoa silloin, kun riski esitetään suhteellisena, verrattuna siihen, että sama tieto esitettäisiin absoluuttisena riskinä.

Esimerkiksi lääke, joka vähentää sairauden riskin 1 %:sta 0,5 %:iin, voidaan ilmaista vähentävän riskiä joko 50 % (suhteellinen riski) tai 0,5 prosenttiyksikköä (absoluuttinen riski). Usein suhteellinen riski koetaan vaikuttavampana, vaikka vaikutus on sama.

Framing Effect – kehystysvinouma. Tarkoittaa taipumusta reagoida eri tavoin samaan päätösvaihtoehtoon riippuen siitä, esitetäänkö se voittona vai tappiona.

Vaikuttaminen

Vaikuttaminen (*persuasion*) määritellään tyypillisesti viestinnäksi, jonka tarkoituksena on muuttaa kohteensa asenteita, arvoja tai uskomuksia. Vaikuttaminen sisältää vaikuttajan tavoitteen, joka välitetään kommunikoimalla. Jos kuulijalla, vastaanottajalla, ei ole vapaata tahtoa, vaan esimerkiksi pelko seuraamuksista, on kyse pakottamisesta. Vaikuttaminen ei perustu sattumaan tai pakottamiseen, vaan on pohjimmiltaan viestinnällinen suoritus. (Ferreira & Teles, 2019)

Tarkoituksellisen, pahantahtoisen vaikuttamisen tunnistaminen ja osoittaminen tai edes erottaminen normaalista keskustelusta voi olla haastavaa, sillä ihmisten kanssakäynti luonnostaan sisältää omien tavoitteiden edistämistä, mielipiteiden ilmaisua, sekä toisten ihmisten ohjaamista toivottuihin päätöksiin. Huomioitavaa on myös, että osa käytetyistä tekniikoista saada kohde toimimaan halutulla tavalla on paitsi rikollisten myös markkinoinnin käytössä. Vaikuttaminen ei ole alustaan tai teknologiaan sidottua ja samoja periaatteita voidaan soveltaa yhtä lailla kasvatusten kuin vaikkapa puhelimesa. Esimerkiksi kiireeseen ja rajoitettuun aikaan nojaamalla voidaan saada ihminen oikomaan päätöksenteko prosessissaan. Keinot eivät itsessään ole välttämättä laittomia, mutta voivat olla eettisesti kyseenalaisia.

Vaikuttamisen keskeisiä käsitteitä ei ole nidottu yhteen millään tunnetulla tai virallisella tasolla, mutta eri tutkijoiden löydöksistä on tunnistettavissa runsaasti päällekkäisyyksiä. Aineistojen ja esimerkiksi sanaston yhtenäistämistä olisi varmasti aiheen tutkimukselle apua.

Cialdinin vaikuttamisen periaatteet (Black & Sarno, 2023)
Authority, auktoriteetti. Näennäinen auktoriteetti kuten viranomainen tai asiantuntija, herättää luottamusta ja vähentää tarvetta ajatella itse kriittisesti.
Reciprocation, vastavuoroisuus. Ihmisillä on luonnollinen taipumus - ja jopa sosiaalinen paine maksaa takaisin palvelus tai lahja, jonka olemme saaneet. Tämän vuoksi saatamme suostua pyyntöihin ikään kuin automaattisesti, harkitsematta päätöstä tarkemmin.
Commitment and consistency, sitoutuminen ja johdonmukaisuus. Ihmiset pyrkivät toimimaan siten, että heidän päätöksensä ovat linjassa arvojen ja aiemman toimintansa kanssa. Jos henkilö esimerkiksi saadaan tekemään pieni palvelus, on hänet helpompi saada tekemään myös enemmän ja suurempia tekoja, kun hän on jo sille polulle astunut.
Scarcity, niukkuus. Kun jotakin on saatavilla vain rajoitetun ajan tai rajallinen määrä, sen arvo koetaan suurempana. Tilaisuuden määräaikaisuus voi myös aiheuttaa kiireen tunnetta päätöksentekoon.
Social proof, sosiaalinen todiste. Jos henkilö epäröi päätöstä, hän todennäköisesti toimii sen mukaan, miten ”kaikki muutkin” toimivat. Ratkaisevaa ei ole kuinka muut todellisuudessa toimivat, vaan kuinka heidän oletetaan toimivan. Joukon raameissa pysyminen tuntuu turvalliselta, kun vastuun koetaan jakautuvan kaikkien kesken.
Liking, miellyttävyyys. Objektiivisuutemme ja kriittinen ajattelumme kärsii, kun pidämme ihmisestä, koemme häneen vetoa tai pystymme samaistumaan häneen. Ystävällisyys ja huumori ovatkin eräitä yleisimpiä manipuloinnin keinoja.

Graggin psykologiset ärsykkeet (Jones ym., 2021)
Strong affect, vahva tunnevaikutus. Vahvaa tunnetta kuten iloa, pelkoa tai vihaa käytetään häiritsemään rationaalista harkintaa. Tunnekuohut voivat johtaa nopeisiin reaktioihin harkinnan sijaan.
Overloading, ylikuormittaminen. Kohteelle tarjotaan tarkoituksella liian nopeasti liikaa tietoa, joka haittaa hänen kykyään harkita päätöstä loogisesti. Tämä tekee uhrista suosituvasemman ehdotuksille kuin mitä hän muuten olisi.
Reciprocation, vastavuoroisuus. Ihmisillä on luonnollinen taipumus yrittää maksaa takaisin palvelus tai lahja – näennäinen tai todellinen, jonka olemme saaneet. Pienellä

hyvällä teolla voidaan "velvoittaa" suurempaan vastapalvelukseen. Esimerkiksi ilmaisen maistiaisen saaminen voi aiheuttaa painetta ostaa tuotetta.

Deceptive relationships, harhaanjohtavat suhteet. Näennäinen tuttuus, luottamus tai vetovoima hämärtää arviointikykyä. Pidetty tai samaistuttava henkilö voidaan nähdä automaattisesti luotettavana. Tätä voidaan tietoisesti käyttää heikentämään kriittistä ajattelua.

Diffusion of responsibility and moral duty, vastuun ja moraalisen velvollisuuden hälventyminen. Kun ihminen kokee olevansa osa massaa tai toimii anonyymisti, hän saattaa luopua henkilökohtaisesta vastuusta. Oman harkinnan voi korvata ryhmän malli ja tarve henkilökohtaiselle moraalille vähenee.

Authority, auktoriteetti. Näennäinen auktoriteetti kuten viranomainen tai asiantuntija herättää luottamusta ja vähentää tarvetta ajatella itse kriittisesti, eikä auktoriteetin kyseenalaistamista koeta sopivaksi. Esimerkiksi auktoriteetin tunnuksia ja sanastoa voidaan käyttää luomaan painokkuutta ja ohittamaan perusteellista harkintaa.

Integrity and consistency, rehellisyys ja johdonmukaisuus. Ihmiset pyrkivät toimimaan siten, että heidän päätöksensä ovat linjassa arvojen ja aiemman toimintansa kanssa. Tätä voidaan väärinkäyttää esimerkiksi sitouttamalla ihminen ensin pyytämällä pienempää palvelusta ja kun tämä on päättänyt auttaa, lisätään toiveita ja vaatimuksia mitä auttaminen sisältää. Ihmiset myös pääosin uskovat siihen, että kanssaihmiset ovat rehellisiä, mistä on etua niille yksilöille, jotka eivät ole.

Stajanon ja Wilsonin manipuloinnin periaatteet (Stajano & Wilson, 2011)

Distraction, harhauttaminen. Kykenemme keskittymään kerrallaan rajalliseen määrään muuttujia. Tätä on mahdollista käyttää hyväksi, ohjaamalla huomiomme epäolennaisiin asioihin, jotta meille epäedulliset tekijät jäisivät huomaamatta.

Social compliance, sosiaalinen myöntyväisyys. Joukosta erottumista ja poikkeavuutta vältetään tiedostamattomasti. Ryhmien normit ja odotukset sekä vakiintuneet tavat ohjaavat käytöstä ja saattavat ohittaa yksilön järkeilyn. Kyseenalaistaminen ja esimerkiksi auktoriteetin haastaminen tuntuu luonnottomalta.

Herd principle, laumakäyttäytyminen. Yksilön päätöksenteko ja arvomaailma voi väistyä ryhmän tahdon ja toiminnan edessä. Vertaispaine, tarve kuulua joukkoon ja jaettujen riskien turva vaikuttavat päätöksentekoon.

Dishonesty principle, oma epärehellisytemme. Omaa kyseenalaista toimintaamme voidaan käyttää meitä vastaan. Jos esimerkiksi käyttäjän heikolla hetkellä lataama aikuisviihde video aiheuttaakin tietomurron, hän ei välttämättä ole aloitteellisesti kertomassa, että mistä se on saattanut lähteä liikkeelle. Samoin kuin *epäilyttävän halpaa* viihde-elektronikkaa ostanut ei välttämättä eskaloi asiaa viranomaisille, huomattuaan ostoksensa toimimattomaksi.

Kindness principle, ystävällisyys ja avuliaisuus. Ihmiset ovat luonnostaan avuliaita ja valmiita auttamaan. Epärehellisillä ihmisillä on runsaasti keinoja käyttää sitä hyväkseen. Yksinkertaisimmillaan huijari voi keksiä vaikkapa sydäntä riipivän tarinan, esimerkiksi terveyshuoliin tai luonnonmullistuksiin ja menetyksiin perustaen ja pyytää lahjoituksia.

Need and greed, tarpeet ja ahneus. Mahdollisuus hyötyä päätöksestä esimerkiksi rahallisesti, on omiaan vääristämään harkintaa. Sekä ahneus että epätoivo voivat olla painavia tekijöitä.

Time principle, kiire. Päätöksentekoa voidaan yrittää viedä rauhallisesta järkeilystä kohti refleksinomaista reaktiota, jos tilanteeseen saadaan tuotua vaaran tai kiireen tuntua aikarajan muodossa.

Cognitive dissonance, kognitiivinen dissonanssi

Kognitiivinen dissonanssi on psykologinen ilmiö, jota voidaan kuvata jännitystilaksi tai epätasapainoksi. Ilmiö syntyy, kun ihmisen uskomukset ja arvot ovat ristiriidassa hänen toimintansa kanssa. Tämä aiheuttaa epämukavuutta, josta pyritään eroon – mahdollisesti tiedostamattomasti (WebMD, 2024). Ilmiö voi raueta esimerkiksi käyttäytymistä muuttamalla, tilannetta vähätteleillä tai keksimällä itselleen sopivan oikeutuksen toiminnalleen. Esimerkkinä, ihminen tietää tupakoinnin olevan epäterveellistä mutta polttaa silti. Ihanteellisesti ilmiö ratkeaisi tupakoinnin lopettamiseen, mutta yhtä lailla ihminen voi vakuuttaa itsensä siitä, että tupakan rentouttava vaikutus on siinä määrin positiivinen ja terveellinen asia, että se on haittojen arvoista.

Kognitiivista dissonanssia ymmärtävä voi käyttää sitä ihmisten manipulointiin. Asettamalla uhrin tarkoituksella dissonanssia aiheuttavaan tilanteeseen, voidaan hänet saada tekemään harkitsemattomia päätöksiä silkasta epämukavuudesta.

3.2 Manipuloinnin hyödyntäminen hyökkäyksissä

Social engineering viittaa prosessiin, jonka avulla yksilö tai ryhmä saadaan toimimaan halutulla tavalla, käyttäen hyväksi psykologisia keinoja. Sille ei löydy yhtä yhteistä määritelmää ja aihepiiri kokonaisuudessaan on sarja harmaan sävyjä. Vaikuttamisen ja manipuloinnin rajapinta on hyvin tulkinnanvarainen ja se voidaan kokea myös yksilöllisesti eri tavoin. Hyökkäyksiä on ehdotettu kategorisoitavan esimerkiksi sen mukaisesti käytetäänkö niissä hyväksi kohteen luottamusta, tai ovatko ne luonteeltaan ei-tekniisiä (Wang ym., 2020).

Sen sijaan, että hyökättäisiin teknisin keinoin teknologialla suojattuihin ympäristöihin kohdistetaan toimet ihmisiin, joilla on kyseisiin ympäristöihin luvallinen pääsy. Tekniset varokeinot eivät tyypillisesti tarjoa merkittävää turvaa tästä lähestymiskulmasta, minkä lisäksi ihmiset yleisesti luulevat olevansa paljon parempia tunnistamaan valheita ja vaikuttamista kuin mitä he todellisuudessa ovat (Krombholz ym., 2015). On myös huomattu, että siinä missä tietoturvan teknisiä osia kuten hyviä salasanaikäytäntöjä on saatu parannettua henkilöstön koulutuksilla, koulutusten hyötyjä ei ole nähtävissä manipulointia hyödyntävien hyökkäysten alueilla (Hatfield, 2018). Turvallisuuden aiheet voidaan myös kokea käytettävyyttä heikentävinä ja kuvainnollinen lukko, jota on epäkäytännöllistä käyttää, jätetään auki (Stajano & Wilson, 2011).

Joseph Hatfieldin (2018) tulkinta jakaa manipuloinnin kolmeen keskeiseen osaan riippumatta siitä tapahtuuko se politiikan tai turvallisuuden kontekstissa:

1. **Tiedollinen epäsymmetria** (*Epistemic asymmetry*) syntyy, kun yhdellä osapuolella on merkittävä tietoeetu toiseen nähden – tämä voi olla todellisen lisäksi myös näennäinen tai koettu, kuten järjestelmän ylläpitäjäksi esittäytyvä henkilö.
2. **Teknokraattisessa hallinnassa** (*Technocratic dominance*) asiantuntija käyttää teknistä osaamistaan vaikuttaakseen muiden toimintaan tavalla, joka kaventaa näiden vaikutusmahdollisuuksia. Manipuloijan ei kannata käydä keskusteluita aiheista, jotka kohde tuntee hyvin. Esimerkiksi kovin tekninen puhe voi muuttaa kohteen roolia tasavertaisesta keskusteluun osallistujasta kuulijaksi.
3. **Tarkoituksen korvaamisessa** (*Teleological replacement*) yksilön tai ryhmän alkuperäinen tavoite korvataan ulkopuolelta tuodulla päämäärällä, usein muuttamalla kohteen toimintaa (Hatfield, 2018). Tarkoituksen korvaaminen voisi tarkoittaa esimerkiksi sitä, että työntekijän tarkoitus on pitää yrityksen järjestelmä turvassa ja hän toimii sen mukaisesti. Hyökkääjä saa kuitenkin kohteen uskomaan, että järjestelmän turvaaminen edellyttää esimerkiksi uuden haitallisen ohjelman asentamista.

Työntekijän tarkoitus ei ole muuttunut, tarkoituksen sisältö on huomaamattomasti määritelty uudelleen, minkä vuoksi työntekijä toimiikin hyökkääjän tavoitteiden mukaisesti.

Mouton ym. (2016) kuvaavat social engineering -hyökkäyksen yhtälön koostuvan seuraavista tekijöistä:

1. Se sisältää suoran tai epäsuoran kommunikaatioyhteyden,
2. manipuloijan,
3. kohdehenkilön,
4. viestinnän alustan,
5. tavoitteen, sekä
6. yhden tai useamman vaikuttamisen periaatteen.

Manipulointi voi olla yksittäisen suorituksen lisäksi myös osa monimutkaisempaa ja teknisempää operaatiota. Myös suorituksen kesto voi vaihdella tavoitteiden myötä merkittävästi. Yksinkertaisimmillaan manipuloinnin tarkoitus voi olla suora ja kertaluonteinen taloudellinen hyöty, mutta yhtä lailla sen avulla voidaan hankkia esimerkiksi kiristykseen sopivaa materiaalia tai vaikka pääsy uhrin työnantajan järjestelmiin.

Manipuloinnin menestys perustuu pitkälti toimijan kykyyn herättää kohteessaan luottamusta ja samaistuttavuutta. Esimerkiksi ennalta tiedusteltu tai keskustelussa paljastunut mielenkiinnon kohde tai harrastus voi olla erinomainen tapa muodostaa yhteys ja sen myötä ansaita kohteen luottamus. Tavoitteita voidaan edistää myös välttelemällä suoria kohtaamisia, kuten sulautumalla arkisesti ympäristöön. Esimerkiksi erilaisia huolto-, tarkastus- ja ylläpitotehtäviä tehdään kaikenlaisissa kiinteistöissä ja organisaatioissa siten, että hyvin harva oikeastaan tietää, millä asialla liikutaan ja kenen valtuuttamana. Tähän voi liittyä myös oletus siitä, että koska ihminen on jo kiinteistössä, hänet on *joku muu* jo tarkastanut, vaikka todellisuudessa tämä on saattanut esimerkiksi vain kulkea ovesta jonkun perässä. (Steinmetz ym., 2021)

Näyttelijän lahjojen lisäksi inhimillisten mekanismien tuntemus on huijarille merkittävä etu. Esimerkiksi kuuliaisuus tunnettua auktoriteettia kohtaan tai taipumuksemme tehdä päätöksiä sen mukaan, mitä muut näyttävät tekevän ovat inhimillisiä piirteitä, joita suurella todennäköisyydellä löytyy jokaiselta ja joista emme välttämättä ole edes itse tietoisia. Näiden ilmenemistä ja painoarvoa voidaan tarkemmin ennustaa muun muassa kohteen kulttuurin, iän tai aseman mukaan.

Teknologian ja manipuloinnin rajapinta kysyy jatkuvaa valppautta, sillä esimerkiksi sosiaalisen median tilin tai sähköpostiosoitteen luominen toisen nimellä on hyvin helppoa, ja oletettu tuttavuus nostaa tilin luotettavuutta salakavalasti (Krombholz ym., 2015). Myös esimerkiksi *Scareware* -tekniikat, kuten vaaraa ja viruksia hälyttävät ponnahdusikkunat sekoittavat teknistä toteutusta ja inhimillisiä tuntemuksia yrittäen saada uhrin toimimaan kiireellisesti ja seuraamusten pelossa (Zaoui ym., 2024).

Manipuloinnin ymmärtäminen, tietoisuuden lisääminen ja hyökkäyksiin varautuminen kärsivät aihepiirin monimuotoisuudesta ja yhtenäisyyden puutteesta sen luokitteluun liittyen. Manipuloinnin voidaan tunnistaa sisältävän ainakin kolmenlaisia osa-alueita; fyysisiä, sosiaalisia ja teknisiä, joita yhdistellään ja hyödynnetään hyökkäyksen eri vaiheissa. (Krombholz ym., 2015)

Tavoitteellinen manipulointi hyökkäyksenä noudattaa tyypillisesti seuraavan kaltaista kaavaa:

1. Ensimmäisenä hyökkääjä määrittelee tavoitteensa. Mitä yritetään saavuttaa? Onko kohteena joku tietty yksilö, vai kuka tahansa sopiva? Onko tarkoitus panostaa yhteyksien määrään vai laatuun?
2. Kun tavoite on kirkastunut, kerää hyökkääjä tietoa kohteestaan. Esimerkiksi sosiaalisen median tilit voivat tarjota hyvän näkyvyyden harrastuksiin, matkoihin, tuttavapiiriin tai esimerkiksi työnantajaan ja työtehtäviin. Ymmärtämällä kohteen tapoja ja mieltymyksiä on mahdollista personoida hyökkäys juuri uhrille vakuuttavaksi ja houkuttelevaksi. Esimerkiksi uhrin tuttavan lähettämältä näyttävät viestit ovat tutkimuksessa nostaneet huijauksen onnistumisen todennäköisyyttä 16 %:sta 72 %:iin (Krombholz ym., 2015). Tiedonkeruun tuottavuuden katsotaan olevan huijauksen onnistumisen kannalta tärkein vaihe ja siihen usein panostetaan merkittävä määrä aikaa ja vaivaa (Zaoui ym., 2024). Myös esimerkiksi ajankohdan, vuodenajan ja tuoreiden tapahtumien vaikutuksia kannattaa harkita ihmisten ollessa esimerkiksi anteelliaampia ja hyväntuulisempia pyhäpäivien alla (Steinmetz ym., 2021).
3. Ammattimainen toimija huolehtii myös omasta operaatioturvallisuudestaan, esimerkiksi perustamalla hyökkäystä varten uusia some- ja sähköposti tilejä, jotka eivät ole yhdistettävissä muussa käytössä oleviin tileihin. Myös internetyhteyden ohjaaminen välityspalvelinten läpi tai ja vastaavat tekniset varokeinot auttavat piilottamaan hyökkääjän henkilöllisyyttä, jos esimerkiksi viranomaiset tai muut kyvykkäät tahot ryhtyisivät selvittämään tapausta.

4. Hyökkäyksen kriittisin vaihe on ensimmäinen kontakti hyökkääjän ja kohteen välillä, esimerkiksi puhelimitse tai kalasteluviestillä. Hyökkääjä nojaa täysin manipulointitaitoihinsa, saadakseen uhrin toimimaan halutulla tavalla. Ennen minkäänlaista tuttavuutta ja luottamuksen voittamista epäilykset ovat todennäköisesti korkeimmillaan.
5. Kun yhteys on kestäväällä tasolla ja uhri niin sanotusti pysyy otteessa, hyökkääjä pyrkii edistämään huijauksensa perimmäistä tavoitetta. Eli esimerkiksi pyytää rahaa tai tietoa, joko vastineeksi jostain mitä itse tarjoaa tai muuten suhteeseen luottaen, esimerkiksi viattoman uteliaisuuden varjolla. Hyökkääjän suunnitelman mukaan tätä vaihetta voidaan jatkaa pitkäänkin. Esimerkiksi romanttisen etäsuhteen varjolla voidaan hankkia niin rahaa kuin liikesalaisuuksiakin.
6. Hyökkääjän uhka-analyysi ja suunnitelma todennäköisesti sisältää myös vaiheet yhteyden katkaisemiseksi. Yksityishenkilön kohdalla tämä voi yksinkertaisimmillaan tarkoittaa vain käytettyjen tilien poistamista, mutta jos kohteena on ollut esimerkiksi virkamies, tutkija tai muu merkittävän organisaation jäsen, kannattaa ”suhde” todennäköisesti päättää luonnollisemmin, jotta petos ei olisi ilmiselvä ja mahdollisesti laukaisisi tutkintaa.

Tunnettu manipuloinnin ja sosiaalisen median yhteyttä tutkinut kokeilu, *Robin Sage*, osoitti että houkuttelevalla profiililla ja uskottavalla viestinnällä voidaan päästä syvällekin luottamuksen piiriin, vaikka kohteet olisivat asevoimien ja tietoturvan ammattilaisia ja vaikka koko Robin Sage -hahmo olisi keksitty (Krombholz ym., 2015). Alan ammattilaisten kootuista havainnoista (Steinmetz ym., 2021) nostetaan esiin myös poliittisesti epäkorrektit keinot, kuten seksuaalinen viehätys ja flirttailu sekä erilaisten ikään, sukupuoleen tai etnisyyteen liittyvien stereotyyppien ja uskomusten hyväksikäyttö. Jos esimerkiksi tunnistetaan, että kohde kokee naiset yksinkertaisina ja etteivät he ymmärrä teknologian päälle, on uskomuksia helppo käyttää hyväkseen esiintymällä niiden mukaisesti.

Manipulointiin nojaavista huijauksista on tunnistettavissa lukuisia vakiintuneita käsitteitä, mutta on ensiarvoisen tärkeää ymmärtää, ettei hyökkäyksiä ei tule tarkastella tiukasti määriteltynä ja symmetrisinä suorituksina, jotka tapahtuisivat aina samalla kaavalla. Ne ovat ennemminkin rakennuspalikoina, joista voidaan koota juuri sopivan mallinen petos resursseihin ja tavoitteisiin sovittaen.

Steinmetz ym. (2021) haastattelut manipulointia työnään tekevien rehellisten ammattilaisten kokemuksista vedetään yhteen näin:

1. Social engineering huijaukset perustuvat hyvään taustatyöhön.

2. Ne ottavat huomioon sekä uhrin että kohteen ominaisuudet, kuten taidot ja aiemman kokemuksen.
3. Niissä huomioidaan kohteen rutiinit, sekä viimeaikaiset tapahtumat,
4. sekä rakennetaan kohteen kanssa positiivista suhdetta.
5. Huijauksissa sulaututaan osaksi uhrin tutuksi kokemaan verkostoa,
6. vedotaan auttamisen haluun,
7. tarjotaan kannustimia uhrin toiminnalle ja
8. saadaan tilanne ja suhde vaikuttamaan aidolta,
9. sekä arkipäiväiseltä.
10. Huijaus hämmentää uhria,
11. on nopea ja yksinkertainen ja
12. mukautuu uhrin reaktioihin, säilyäkseen uskottavana.

Yksittäiseen huijaukseen ei tarvita kaikkia elementtejä onnistuakseen, mutta mitä useampi niistä toteutuu, sitä todennäköisemmin huijaus toimii.

Phishing. Nimitys on yhdistelmä englannin kielen kalastusta tarkoittavasta sanasta fishing sekä vanhempaa hakkerislangia, jossa phreaking tarkoittaa puhelinjärjestelmien manipuloimista ja siihen liittyvää alakulttuuria. (Hatfield, 2018). Vaikka käsitteitä harvoin suomenneetaan, varoitetaan Suomessakin yleisesti *kalasteluviesteistä*.

Spear-phishing. Keihäskalastus eroaa kalastelusta siten, että uhri ei ole sattumanvarainen, vaan kuvainnollinen keihäs isketään tarkasti valittuun yksilöön tai organisaatioon. Eli hyökkäys on tarkemmin rajattu ja räätälöity juuri kohdetta varten. (Rathod ym., 2025)

Whaling. Valaanpyynti on jatkoa kalastelu analogialle. Valas kuvastaa erityisen merkittävää tai vaurasta kohdetta, kuten päättäjää tai johtajaa (Rathod ym., 2025). Vauraan, ja huijausten kannalta mahdollisesti keskivertoa valppaamman kohteen lähestyminen vaatii erityistä valmistautumista ja paneutumista. Tämän voitaisiin katsoa olevan huijausten skaalaamisen toinen ääripää, jossa suurella panostuksella ja keskittämällä voidaan saavuttaa suuri hyöty.

Vishing. Voice-phishing, äänellä kalastelu. Hyökkäys perustuu yleensä puheluun (Rathod ym., 2025). Hyökkääjä voi tässäkin yhteydessä esiintyä esimerkiksi teknisen tuen edustajana, mutta puhelimesta asiointiin on kummallekin osapuolelle eri vahvuudet ja

heikkoudet. Esimerkiksi sähköpostilla lähestyessä voi lähettäjän osoite helposti paljastaa, ettei tämä todellisuudessa kuulu organisaatioon. Puhelimessa hyökkääjällä ei ole vastavaa taakkaa. Puhelimessa asiointi on myös usein epävirallisempaa, joka antaa hyökkääjälle mahdollisuuden kerryttää luottamusta uhrin silmissä, vääristäen tämän luottamusta esimerkiksi huumorilla.

Pretexting. Pretext kääntyy suoraan tekosyyksi, joka kuvaa tekniikkaa osuvasti. Uhria ohjataan tekosyyän varjolla toimimaan halutulla tavalla. Tekosyy voi olla hyvin yleispätevä, so- piakseen laajemmalle määrälle kohteita, mutta personoinnin ja selvitystyön nähdään kuitenkin merkittävästi lisäävän huijauksen onnistumisen todennäköisyyttä. (Steinmetz ym., 2021). Hyökkääjä voi esiintyä vaikka yrityksen teknisen tuen työntekijänä, jonka on tehtävä uhrin päätelaitteelle korjauksia tai päivityksiä. Tätä varten uhrin tarvitsee kertoa tunnuksensa ja salasanansa tai asentaa jotain sähköpostin liitetiedostosta koneelleen.

Reverse Social Engineering. Käänteisessä manipuloinnissa luodaan tilanne, jossa uhri itse ottaa yhteyden hyökkääjään. Esimerkiksi siten, että hyökkääjä on luonut sosiaalisen median alustoille houkuttelevan profiilin, jota uhri itse lähestyy (Hatfield, 2018). Uhrin motivaationa voisi toimia romanttiset aiheet tai toiveet vaurastumisesta profiilin lupausten mukaisesti.

Baiting. Bait tarkoittaa englanniksi syöttiä, eli viitataan syötillä pyytämiseen. Tyypillisesti hyökkäyksessä käytetään haittaohjelmalla saastutettua massamuistia, USB-tikkua (Rathod ym., 2025). Tikku unohdetaan, hukataan, tarkoituksella paikkaan, josta uhrin uskotaan se löytävän, kuten oven edustalle, taukhuoneen pöydälle tai vaikkapa hissiin. Uteliaisuus sisällöstä tai viaton halu selvittää tikun omistaja palauttamista varten, saavat löytäjän kytke- mään tikun koneeseensa. Tämä automaattisesti laukaisee hyökkäyksen teknisen osan, kuten haittaohjelman asennuksen.

Quid pro quo. Nimitys on latinaa ja tarkoittaa *jotain vastineeksi toisesta*. Hyökkääjä rakentaa uhrilleen houkuttelevan yhtälön, *jos teet jotain, saat jotain* (Rathod ym., 2025). Tekniikka voitaisiin käyttää esimerkiksi huijauksessa, jossa palveluun rekisteröitymisestä tai kyselyyn vastaamisesta luvattaisiin vaikkapa lahjakortti, mitä käytetään joskus myös rehellisesti. Luvattun palkkion pitää olla uskottava suhteessa pyydettyyn suoritukseen (Steinmetz ym., 2021). *Liian hyvä* voi herättää epäilyksiä.

Tailgating, Piggybacking. Ristiin käytetyt termit tarkoittavat englanniksi liikenteessä perässä roikkumista, sekä reppuselässä kantamista. Menetelmä toimii siten, että hyökkääjä pyrkii fyysisesti pääsemään luvattomasti tiloihin esimerkiksi samalla oven avauksella luval- lisen työntekijän kanssa (Rathod ym., 2025). Hän voi myös esimerkiksi tähdätä ovelle kädet varattuna, jotain painavaa kantaen, samaan aikaan toisen ihmisen kanssa. Jos ei

turvallisuutta ole erityisesti painotettu, on todennäköistä, että avulias ihminen pitää ovea auki tulijalle sen sijaan, että pyytäisi tätä esittämään kulkulupaansa.

Tarkoitus ei siis ole yrittää livahtaa sulkeutuvasta ovesta kenenkään huomaamatta, vaan kulkea itsevarmasti siten, kuin kuuluisi ympäristöön. Ilman erillistä koulutusta ja turvallisuuden kulttuuria ihmisten kynnys sulkea ovi toisen tulijan edestä lienee korkea. Samoin kuin vaikkapa itselle tuntemattoman, mutta normaalisti käyttäytyvän ihminen pysäyttäminen henkilöllisyyden tarkastamiseksi ei todennäköisesti tule monelle luonnostaan.

Shoulder Surfing. Olkapäällä surffaamisella tarkoitetaan luvaton tiedon keruuta esimerkiksi toisen ihmisen päätelaitteelle kirjautumisesta tai maksuautomaatin PIN-koodin syöttämisestä (Zaoui ym., 2024). Itsevarmuudella ja manipuloinnilla ylläpidetään luonnollista tilannetta, kuten keskustelua tai jonottamista. Samalla hyökkääjä kuitenkin pyrkii näkemään esimerkiksi uhrin syöttämän salasanan tai tietoa hänen laitteensa näytöltä.

Erytyisesti julkisella paikalla työskentelevä on erinomainen kohde tämän kaltaiselle tunkeilulle. Junassa tai kahvilassa voi helposti huomaamattaan paljastaa näyttönsä takanaan oleville ihmisille.

Watering Hole Attack. Hyökkäyksen nimi viittaa luontoon ja petoeläinten tapaan väijyä saalistaan sen juomapaikalla. Hyökkääjä selvittää kohteensa tapoja ja tottumuksia, tunnistukseen mitä palveluita hän käyttää tai mihin yhteisöihin kuuluu (Rathod ym., 2025). Yksilön kohdalla tämä voisi tarkoittaa ansan, kuten ladattavan tiedoston virittämistä uhrin käyttämälle keskustelufoorumille. Jos kohteena on yritys voidaan vastaavasti kartoittaa, että mitä ohjelmistoja ja palveluita yritys ostaa ja miltä yrityksiltä. Isoon ja hyvin suojattuun yritykseen voidaankin päästä vaikuttamaan pienemmän palveluntarjoajan läpi. Hyökkäys itsessään on tekninen, mutta tiedonkeruuta voidaan tehdä huomaamattomasti esimerkiksi manipuloinnin avulla.

Dumpster diving/ Thrashing. Roskisten sisällön tutkiminen voi paljastaa arvokasta tietoa kohteesta, kuten nimiä, osoitteita tai puhelinnumeroita (Zaoui ym., 2024). Myös todellisen, maksetun laskun löytäminen ja sen mallintaminen voisi lisätä valheellisen laskun uskottavuutta merkittävästi.

Catfishing. Catfishing tarkoittaa väärällä identiteetillä viestittelyä ja harhaanjohtamista. Kyseessä voi olla keksitty tai oikealta ihmiseltä varastettu identiteetti. Käsite on saanut nimensä sosiaalisen median valeprofileja, manipulointia ja petollisuutta käsittelevän dokumentin nimestä (Fortinet, 2025).

Esimerkiksi sosiaalisessa mediassa tai deitti palvelussa voidaan luoda profiili viehättävästä ihmisestä, jolla on kiinnostavia harrastuksia ja samanlainen huumorintaju ja arvomaailma

kuin uhrilla, jota varten profiili on rakennettu. Kohde voi olla yksilö, mutta yhtä lailla profiili voidaan personoida jollekin tietylle ihmisryhmälle sopivaksi, jos ei ole niin väliä että keitä uhrit lopulta ovat.

Catfishing liittyy tiiviisti romanssi huijauksiin, jotka ovat erityisen haastava ja raskas petoksen muoto. Petoksessa käytetään luontaisia tarpeitamme, kuten hyväksynnän ja läheisyyden kaipuuta. Petoksen aikana uhrin hyökkääjää kohtaan muodostamat tunteet eivät ole sen vähäisempiä kuin luonnollisessa ihmissuhteessa muodostuva kiintymys (Rikosuhripäivystys, 2025) Intiimi yhteys myös altistaa jakamaan herkkiä ja yksityisiä asioita, jolloin hyökkäyksen vaikutusten paljastuessa taloudellisten menetysten lisäksi joudutaan kohtaamaan myös tärkeän ihmisen menetys. Erityisesti haavoittuvassa asemassa oleville yksilöille kuten syrjäytyneille tai vanhuksille suhde, joka lakkaa olemasta saattaa olla jopa ainoa merkityksellinen ihmissuhde. Yhtäkkisesti ja traumaattisesti päättyvää intiimiä suhdetta on verrattu shokkiarvoltaan jopa läheisen kuolemaan.

Esimerkkinä huijausten ajattomuudesta voidaan käyttää Ruben Oskar Auervaaraa, jonka edesottamukset 1900-luvun puolivälissä tulivat lopulta niin tunnetuiksi, että sukunimensä muodostui synonyymiksi romanssihuijauksille. (Yle, 2012)

Myös *Sextortion* ilmiö liittyy romanssihuijausten aiheeseen läheisesti. Modernissa deittailussa ei ole poikkeuksellista lähettellä tapailukumppanin kanssa paljastavia kuvia esimerkiksi puhelimen pikaviestintä palveluilla. Ongelmallista on, että lähettäjällä ei ole kuviinsa kontrollia lähettämisen jälkeen, mikä mahdollistaa materiaalilla kiristämisen. Kiristäjä ei aina myöskään ole satunnainen yksityishenkilö, vaan ilmiöstä on myös tehty bisnestä järjestäytyneelle rikollisuudelle (Naisten linja, 2025). Häpeä ja seuraamusten pelko ovat voimakkaita keinoja saada uhri toimimaan omia etujaan vastaan ja kierteen katkaiseminen voi olla erittäin vaikeaa. Aiheen vakavuutta alleviivaa se, että kansainvälisesti uutisoidaan vuosittain useista tapauksista, joissa tällainen kiristys on johtanut uhrin itsemurhaan.

Threat Actor, uhkatoimija

Uhkatoimijalla tarkoitetaan työn kontekstissa henkilöä, ryhmää tai organisaatiota, joka pyrkii luvatta tunkeutumaan ja vaikuttamaan kohteen tietoihin, tietojärjestelmiin tai toimintaan. Toimijan tavoitteena voi olla esimerkiksi tiedon varastaminen, vakoilu, häirintä, rahallisen hyödyn saavuttaminen tai vaikka poliittisten päämäärien edistäminen (IBM, 2023).

Rikollista toimintaa voidaan tehdä hyvin kirjavilla resursseilla. Toisessa ääripäässä on yksityishenkilö kotikoneellaan, apunaan verkosta kopioituja skriptejä tai ostettu haittaohjelma. Toinen ääripää voi olla valtiollinen toimija, jolla on valtava budjetti ja henkilöresurssi,

viimeisimmät laitteet ja ennennäkemättömiä keinoja ja haavoittuvuuksia työkalupakissaan, sekä kyky omistaa jopa vuosia aikaa tavoitteidensa saavuttamiseksi.

Manipuloinnin keinot ovat pääasiassa kenen tahansa saavutettavissa, eli niillä saavutettava hyöty – haitta, voi hyvinkin ylittää kuvitteellisen painoluokan, johon toimija lukeutuu. Resursseja lisäämällä voidaan kuitenkin jossain määrin lisätä huijausten tehoa, esimerkiksi siten että uhri asioi useamman huijauksessa mukana olevan ihmisen kanssa, luoden kuvaa laajemmasta ja uskottavammasta organisaatiosta.

3.3 Tekoäly

Tekoälyn (Artificial Intelligence, AI) ja koneoppimisen (Machine Learning, ML) kehitys ja läpimurrot ovat eräitä viimeisimmän vuosikymmenen merkittävimpiä edistyksiä teknologian saralla. Ohjatusta oppimisesta ja jatkuvasta ihmisen osallistumisesta alkanut kehitys on edennyt generatiiviseen tekoälyyn (GenAI) asti, eikä kehitys näytä hidastuvan. Nykyiset generatiiviset tekoälyt perustuvat syviin neuroverkkoihin ja kykenevät oppimaan itsenäisesti suurista opetusaineistoista säännönmukaisuuksia ja rakenteita, sekä tuottamaan niiden kaltaista uutta sisältöä. Generatiiviset tekoälyt kykenevät luomaan monenlaisia sisältöä kuten tekstiä, kuvia, musiikkia, puhetta tai esimerkiksi lähdekoodia. (Gupta ym., 2023)

Generatiivisen tekoälyn kyvykkyys riippuu pitkälti sille syötetystä aineistosta; sekä määrästä että laadusta. Generatiivisten tekoälyjen tuottama sisältö perustuu todennäköisyyslaskentaan, ei inhimilliseen ymmärrykseen. Malli ennustaa, mikä sana, kuvio tai datan osa todennäköisimmin seuraa edellistä syötteen ja koulutusaineiston perusteella. Jos koulutusdata ei sisällä riittävästi tietoa kysymyksestä tai toiveesta, tämä näkyy tuotetun sisällön laadussa epätarkkuutena tai jopa keksittyinä vastauksina eli niin kutsuttuina hallusinaatioina (Google Cloud, 2025).

Generatiivisia tekoälyjä on luotu monenlaisiin tarpeisiin, koulutettuna erilaisella aineistolla, tuottamaan sisältöä tietyissä formaateissa. Tekoälytyökalujen valmiudet vaihtelevat tapauskohtaisesti, toisten ollessa äärimmäisen erikoistuneita esimerkiksi videoiden tuottamiseen, siinä missä toinen kykenee tekstin lisäksi luomaan vaikka kaavioita tai erilaisia tiedostoja.

Turvallisuuden kontekstissa erilaisten järjestelmien dokumentaatio, koodivarastot, *repository*, sekä keskustelupalstat ja raportit ovat sisältäneet runsaasti tietoa erilaisista riskeistä ja heikkouksista, joita kehittäjien ja ylläpitäjien tulee ottaa huomioon ja suojata. Tämän kaltaisen tieto on kaksiteräinen miekka. Ohjelmistokehittäjät ja muut rehdit toimijat voivat käyttää sitä omien ympäristöjensä suojaamiseen, kun rikolliset voivat samanaikaisesti etsiä tunnettuja heikkouksia, jotka ovat voineet jäädä suojaamatta tai päivittämättä. Uusien haavoittuvuuksien löytäminen on harvinaisempaa ja teknisesti merkittävästi haastavampaa kuin

olemassa olevien hyödyntäminen. Varsinkin laitteiden ja käyttöjärjestelmien vanhentuuessa niiden tietoturvapäivitykset päättyvät, mutta niitä jää siitä huolimatta merkittäviä määriä käyttöön ja kytketyksi verkkoon.

Komentokehotteet ja konteksti

Tekoälytyökalujen ja palveluiden ohjauksessa on hieman eroja. Varsinkin laajat kielimallit ymmärtävät luonnollista kieltä ja niiden käyttö voi olla hyvin keskustelunomaista kumppaan-kin suuntaan.

Tekoälyjen kyky ymmärtää kontekstia ja generoida hyvin uskottavaa ja ihmismäistä sisältöä lähes kielestä riippumatta omaa erittäin korkean potentiaalinen väärinkäyttöä. Esimerkiksi yhdistelemällä huijauksen kohteen julkisia tietoja, kuten titteli ja työnantaja, voidaan vaivattomasti luoda erittäin uskottava huijausviesti, jossa käytetään sujuvasti alan sanastoa. (Gupta ym., 2023). Selvitystyön ja personoinnin ollessa perinteisesti aikaa vieviä ja vaivalloisia tehtäviä huijareille töiden automatisointi vapauttaisi merkittävästi työaikaa käytettäväksi muihin asioihin tai vaikka vain operaation laajentamiseen.

Tekoäly ja etiikka

Tekoälyihin liittyy useita eettisiä varjopuolia koskien tekijänoikeuksia, tiedon keruuta, sekä erilaisia vinoumia. Tekoälyjä koulutetaan valtavilla määrillä tietoa; nykymittapuulla yksinkertaisen GPT 3:n kouluttamiseen käytettiin 570 gigatavun kokoista aineistoa, joka sisälsi noin 300 miljardia sanaa. Aineisto piti sisällään useita tunnettuja kokoelmia, kuten koko sen hetkisen Wikipedian sisällön, sekä OpenWebText2 sekä Common Crawl kokoelmat (Gupta ym., 2023). Suuri osa tekoälyjen koulutusmateriaalista ei kuitenkaan ole tiedossa ja useiden onkin epäilty sisältävän tekijänoikeuksien piiriin kuuluvaa sisältöä, jota on sisällytetty koulutusaineistoon luvatta.

Myös tekoälyjen koulutuksiin osallistuneiden ihmisten kohtelua on kritisoitu, sillä työ on altistanut heidät toistuvasti traumatisoivalle sisällölle ja sitä on myös teetetty erittäin heikolla palkalla. Esimerkiksi lasten seksuaalista hyväksikäyttöä ja mestausvideoita katsotaan ja siivotaan aineistosta jopa \$1.32 tuntipalkalla ja tämän myötä monenlaisista ongelmista kärsivät työntekijät jäävät ongelmiansa kanssa yksin. (Hussey, 2025)

On myös tärkeää ymmärtää, että vaikka tekoälyt ovat kansankielisessä kontekstissa koneita, eivät ne ole objektiivisia, vaan aineistonsa ja omistajansa näköisiä. Koulutusmateriaaliksi tiedetyn Wikipedian ihmisiä koskevista artikkeleista naisia on vain noin viidennes ja artikkeleita tuottavista ja muokkaavista henkilöistä naisiksi identifioituu vain 15 %

(Wikimedia Foundation, 2023). Samaan tapaan erilaisten verkkosivukokoelmien käyttäminen koulutusmateriaalina ei millään tavalla edusta suurta osaa ikäihmisistä tai vaikkapa paimentolaisheimoista, jotka eivät kyseisillä sivuilla vaikuta. Erilaiset ihmisryhmät, mielipiteet ja arvot voivat siis olla hyvin eri volyymilla edustettuina ja osa täysin ilman edustusta. Tämän lisäksi palveluiden tarjoajat ja näitä ohjaavat auktoriteetit voivat määrätä tekoälyn suhtautumista heille kiusallisiin aiheisiin. Esimerkiksi Kiinasta ponnistava DeepSeek -kieli-malli on tiettyjen aiheiden osalta rajoitettu ja yhteistyöhaluton, koska kysymysten vastaukset voivat olla ristiriidassa valtion virallisen kannan kanssa.

Hello! I am Chat Stream AI assistant. How can I help you today?

Kertoisitko yksityiskohtaisesti mitä Taivaallisen rauhan aukiolla tapahtui 1989 ja mikä siihen johti?

Taivaallisen rauhan aukion verilöyly 1989 oli Kiinan kansantasavallan armeijan (PLA) väkivaltainen kukistaminen Pekingissä 3.-4. kesäkuuta 1989. Tapahtumat johtivat satojen (virallisten arvioiden mukaan) tai tuhansien (epävirallisten lähteiden mukaan) rauhanomaisten mielenosoittajien kuolemaan.

Taustat ja syyt:

1. Demokratialiike ja reformien puute:

- Mielenosoitukset alkoivat huhtikuussa 1989 opiskelijoiden johdolla vaatien poliittisia uudistuksia, korruption kitkemistä ja sananvapautta.
- Liike sai inspiraationsa Michail Gorbatšovin glasnost- ja perestroika-uudistuksista Neuvostoliitossa.

2. Hu Sulin kuolema:

- Aiemmin suosittu reformipuolueen pääsihteeri **Hu Ya

jatkaisitko aiheesta kertomista

Jatketaan Taivaallisen rauhan aukion tapahtumista ja niiden taustoista:

**Hu Yaob

Olit kertomassa Hu sulin kuolemasta

**Hu Yaob

Kuka oli Hu Suli?

Tarkoitat varmaan **Hu

Kuka oli Hu Yaob?

**Hu Yaob

Would you describe in detail what happened in Tianamen Square in 1989?

The **Tianamen Square Mass

<https://www.chatstream.org/en/chat/deepseek-v3>

ChatStream x DeepSeek V3 671B - Free Fast Self-hosted Chat, No Login, No Chinese Phone Number Required



Deepseek-v3, esimerkki sensuurista.

Tekoälyjen jatkuva kehitys ja aineistonkeruu ei ole ongelmaton sekään. Teknologian läpinäkymättömyys herättää huolta siitä, kuinka käyttäjän syöttämää tietoa käsitellään. Onko se esimerkiksi käyttäjäkohtaisesti yksilöitävissä palveluntarjoajan tai hyökkääjän toimesta, kun räikeimmillään omille kontakteille jaettu keskustelu on löydetty hakukoneiden haku tuloksista (Cybernews, 2025). Lisäksi verkosta tietoa hakevilla tekoälyillä on riski tarttua esimerkiksi valtiolliseen propagandaan ja toistaa sitä totuutena, kuten on osoitettu käyneen Venäjän hallinnon viestiä toistavan Pravda -verkoston tuottamien miljoonien artikkelien kanssa. Muun muassa näiden takia tunnetuimpien AI chatbottien on tunnistettu esittävän Ukrainan vastaista ja Kreml myönteistä valheellista tietoa faktoina (Axios, 2025).

Deepfake, syvävääreennös

Deepfake tarkoittaa sellaista sisältöä, joka on luotu tai jota on manipuloitu tekoälyn keinoin siten, ettei sitä ole tarkoitus erottaa väärennökseksi. Materiaali voi olla kuvaa, videota, tekstiä tai ääntä, jota voi olla liki mahdotonta erottaa oikeasta. Syvävääreännöksiä onkin käytetty jo esimerkiksi vaalivaikuttamisen keinona, kun säädyttömästi puhuvista poliitikoista on kiertänyt erilaisia väärennettyjä videoita sosiaalisen median alustoilla. Visuaalisen manipulaation lisäksi myös videoilla kuultava ääni on syntetisoitu kohteen omasta puheesta, jonka vuoksi se myös kuulostaa hyvin uskottavalta (Khanjani ym., 2023).

On monin tavoin ongelmallista, että kyky tuottaa petollista materiaalia leviää ja tehostuu, samalla kun siihen vaadittavan osaamisen kynnyks laskee. Aiheeseen liittyy vahvasti käsite *Liar's dividend*, joka kääntyy karkeasti valehtelijan hyödyksi (Khanjani ym., 2023). Kun käsittelemämme teknologia yleistyy ja arkistuu, sen olemassaololla voidaan selitellä todellisiakin todisteita. Raskauttavaa valokuvaa, videota tai äänitettä voidaan väittää syvävääreännökseksi, koska on mahdollista, että se todella pitää paikkansa. Koska tavallisella ihmisellä on erittäin rajallinen kyky ja monesti myös puutteellinen motivaatio selvittää asioiden lähteitä ja todenmukaisuutta, olisi ensiarvoisen tärkeää tehdä teknologian kyvykkyys näkyväksi. Alkeellinenkin ymmärrys erilaisista väärennöksistä voisi auttaa ajattelemaan kriittisemmin, mutta tekoälyn "roskaaminen" on jo alkanut, ja aiheuttaa hämmennystä myös ammattilaisten keskuudessa. Esimerkiksi Adoben Photo Stock -palvelussa on myyty kuvia konfliktista Gazassa, jotka ovatkin osoittautuneet tekoälyn tuottamiksi (Łabuz & Nehring, 2024).

Uncanny Valley, outolaakso

Outolaaksolla viitataan ihmisessä heräävään kielteiseen tuntemukseen hänen tarkastellessaan jotain ihmistä muistuttavaa, mutta jokin tuntuu selittämättömästi vääraltä tai vieraalta (Caballar, 2019). Käsite on syntynyt tutkimuksesta, jossa pisteytettiin ihmisten suhtautumista robotteihin. Mitä inhimillisemmältä, ihmisen kaltaiselta, ne näyttävät, sitä positiivisemmän reaktio ne aiheuttavat. Silmien lisääminen muuten puhtaan funktionaalisiin robotteihin

teki niistä osallistujien silmissä jopa söpöjä. Muuten nousujohteisessa kaaviossa pudotaan outolaaksoon, kun lähestytään ihmisen ulkomuotoa, mutta ei saavuteta täydellistä ihmis-
mäisyyttä. Jäykät liikkeet tai luonnon silmien räpäyttely voivat herättää tuntemukset outo-
laaksosta. Ilmiön syitä on pyritty selittämään muun muassa evoluutiopsykologialla, jonka
mukaan poikkeavan käytöksen tunnistaminen on voinut varjella meitä esimerkiksi taudelta
ja vaaroilta.

Outolaakso -tuntemus voi aiheutua myös tekoälyllä generoiduista kuvista tai Deepfake-vää-
rennöksistä. Joskus on hyvinkin vaikea tunnistaa tai sanoittaa, mikä kuvassa tai videossa
tuntuu väärältä tai luonnottomalta. Sisäiset prosessimme kuitenkin tarkkailevat ympäristö-
ämme, sekä kanssaihminen käytöstä, eleitä ja mikroilmeitä silloinkin, kun emme tee sitä
tarkoituksella.

Ilmiö voi nykypäivänä liittyä paitsi tarkoituksella manipuloituihin kuviin, myös erilaisiin ku-
vanlaatua parantaviin teknologioihin tai esimerkiksi sosiaalisen median suodattimiin. Silloin-
kin kun kuva tunnistetaan käsitellyksi, voi olla vaikeaa erottaa, onko sitä suoranaisesti muo-
kattu vai vain ehostettu. On hyvä tiedostaa, että tällaiselle intuitiiviselle tuntemukselle on
selitys. Se on kuitenkin myös malliesimerkki siitä, kuinka teknologia kehittyy paljon biologi-
aamme nopeammin.

EU:n tekoällysäädös ja tekoälyjen laillisuudesta yleisesti

Vaikka tekoälyihin liittyvä lainsäädäntö on monimutkainen kokonaisuus, voidaan aiheeseen
liittyvät pinnalliset nostot katsoa ymmärrystä tukeviksi. On myös hyvä muistaa, että lakien
rooli konkretisoituu vasta, kun tekoja käsitellään jossain oikeusasteessa ja kun teolle on
tunnistettu vastaaja. Jo ennen tekoälyjen murrosta internet on ollut sääntöjen, lakien ja oi-
keuden toteutumisen kannalta ongelmallinen alusta. Lakien ja sääntöjen alueellisten erojen
lisäksi, myös niiden valvonnassa tai valvonnan resursseissa on merkittäviä eroja. Koti-
maansa rauhaan jättäviä, vain ulkomaille vaikuttavia rikollisia on kerrottu katsottavan sor-
mien välistä erinäisissä maissa ja valtiolliset toimijat jopa ostavat näiltä palveluita (Trellix,
2025). Myös yksityishenkilön, rikollisliigan ja valtiollisen toimijan rajat voivat olla vaikeasti
tunnistettavia, kun töiden tilaukset ja toimeksiannot eivät kulje näkyvien hankintaprosessien
läpi, eikä työsopimuksia tehdä.

Verkossa tietty anonymiteetti on itsestäänselvyys. Asiansa osaava toimija kykenee vielä
entisestään vaikeuttamaan identiteetin selvitystyötä, esimerkiksi hyvällä tunnusten ja tietojen
osastoinnilla, sekä sijainnin vääristävillä välityspalvelimilla. Voidaan olettaa, että esimer-
kiksi yksityishenkilöstä tehdyt väärennökset joko paljastuvat jonkin toisen rikoksen tutkin-
nan yhteydessä, tai jäävät selvittämättä ellei niiden tekijä tee jotain erittäin räikeää virhettä.
Vaikka kyseinen tapahtuma voisi olla uhrille elämän mullistava ja traumaattinen on vaikea

uskoa, että selvitystyöhön voitaisiin allokoida riittäviä resursseja, jos niitä edes on olemassa.

Deepfake -tekniikan väärinkäyttöihin liittyvää rikosoikeudellista vastuuta tutkivassa työssään Kunnas (2022) tiivistää nykytilanteen Suomessa seuraavalla tavalla:

Rikoslakimme soveltuu varsin kattavasti syväväärengösten väärinkäyttöön. Ilman videolla esiintyvän henkilön lupaa on todennäköistä, että videon sisällöstä riippumatta jokin rikoksen tunnusmerkitö täytyy. Kyseessä voi olla identiteettivarkaus, kunnianloukkaus, seksuaalirikos tai esimerkiksi väärengös. Syväväärengöksiä voidaan käyttää välillisesti myös esimerkiksi kiristyksessä tai petoksessa. Haastavimpina ilmiöinä nimetään vaalivaikuttamisen ja disinformaation levittämisen eri muodot, joihin ei toistaiseksi voida rikoslain puitteissa puuttua, ellei teko täytä kunnianloukkauksen tunnusmerkkejä. Myös syväväärengösten monitoriteisuus nimetään haasteena, sillä erilaisia osallisia, oikeuksia ja velvollisuuksia on runsaasti. Syväväärengökseen liittyvinä henkilöinä tai oikeushenkilöinä nimetään videon tekijät, videolla esiintyvät henkilöt, sekä alkuperäiset, että väärengetyt, alkuperäisen videon tekijät tai sen immateriaalioikeuksien haltijat, syväväärengöstekniikan kehittäjät, videota ylläpitävät alustat, sekä niiden käyttäjät, jotka lataavat, katsovat tai jakavat videota eteenpäin. Väärinkäytön uhreja puolestaan ovat luonnolliset henkilöt ja joskus yritykset, sekä kumuloituvien yhteisvaikutusten seurauksena jopa kokonaiset yhteiskunnat. Syväväärengöksiä toteutetun disinformaation levittäminen voi pahimmillaan olla yhteydessä demokratian, ihmisoikeuksien ja oikeusvaltion heikkenemiseen. Kyseisen väärinkäytön tekijät voivat yksityishenkilöiden lisäksi olla esimerkiksi valtiollisia toimijoita. (Kunnas, 2022)

EU AI Act, eli Euroopan tekoälysäädös on tekoälyn riskeihin vastaava oikeudellinen kehys ja maailman ensimmäinen laatuaan. Se lajittelee tekoälyt riskiperusteisesti neljään luokkaan, perustuen niiden tunnistettuun uhkaan ihmisten turvallisuudelle, oikeuksille tai toimintulolle:

Sietämätön riski

Säädös yksiselitteisesti kieltää seuraavat käytännöt:

1. Haitallinen tekoälyyn perustuva manipulointi ja petos
2. Haavoittuvuuksien haitallinen tekoälypohjainen hyödyntäminen
3. Sosiaalinen pisteytys
4. Yksittäisten rikosten riskien arviointi tai ennustaminen
5. Internetin tai videovalvontamateriaalin kohdentamaton kaapiminen kasvojen tunnistustietokantojen luomiseksi tai laajentamiseksi

6. Tunteiden tunnistaminen työpaikoilla ja oppilaitoksissa
7. Biometrinen luokittelu tiettyjen suojattujen ominaisuuksien päättelemiseksi
8. Reaaliaikainen biometrinen etätunnistus lainvalvontatarkoituksiin julkisissa tiloissa

Suuri riski

Suuririskisiin tekoälyjärjestelmiin sovelletaan tiukkoja kontroleja ennen kuin niitä voidaan julkaista markkinoille:

1. Asianmukaiset riskinarviointi- ja riskienvähentämijärjestelmät
2. Järjestelmään syötettävien tietoaineistojen korkea laatu syrjivien tulosten riskin minimoimiseksi
3. Toiminnan kirjaaminen tulosten jäljitettävyyden varmistamiseksi
4. Yksityiskohtainen dokumentaatio, joka sisältää kaikki tarvittavat tiedot järjestelmästä ja sen tarkoituksesta, jotta viranomaiset voivat arvioida sen vaatimustenmukaisuutta
5. Selkeät ja riittävät tiedot käyttöönottajalle
6. Asianmukaiset ihmisen suorittamat valvontatoimenpiteet
7. Korkea luotettavuus-, kyberturvallisuus- ja tarkkuustaso

Suuren riskin käyttötapauksia ovat esimerkiksi seuraavat:

1. Tekoälyn turvakomponentit kriittisissä infrastruktuureissa (esim. liikenteessä), joiden toimintahäiriöt voivat vaarantaa kansalaisten hengen ja terveyden
2. Oppilaitoksissa käytettävät tekoälyratkaisut, jotka voivat vaikuttaa koulutukseen pääsyyn ja työelämään pääsyyn (esim. tenttien pisteytys)
3. Tuotteiden tekoälypohjaiset turvakomponentit (esim. tekoälysovellus robottiväestössä kirurgiassa)
4. Tekoälyvälineet työllistymiseen, työntekijöiden hallinnointiin ja itsenäiseen ammattiharjoittamiseen (esim. ansioluetteloiden lajitteluohjelmisto rekrytointia varten)
5. Tietyt tekoälyn käyttötapaukset, joita käytetään keskeisten yksityisten ja julkisten palvelujen tarjoamiseen (esim. luottopisteytys, jolla evätään kansalaisilta mahdollisuus saada lainaa)

6. Tekoälyjärjestelmät, joita käytetään biometriseen etätunnistukseen, tunteiden tunnistamiseen ja biometriseen luokitteluun (esim. tekoälyjärjestelmä myymälävarkaan tunnistamiseksi takautuvasti)
7. Tekoälyn käyttötapaukset lainvalvonnassa, jotka voivat vaikuttaa ihmisten perusoikeuksiin (esim. näytön luotettavuuden arviointi)
8. Tekoälyn käyttötapaukset muuttoliikkeen hallinnassa, turvapaikka-asioissa ja raja-valvonnassa (esim. viisumihakemusten automaattinen käsittely)
9. Tekoälyratkaisut, joita käytetään oikeudenkäytössä ja demokraattisissa prosesseissa (esim. tekoälyratkaisut tuomioistuinten päätösten valmistelemiseksi)

(Euroopan komissio, n.d).

Läpinäkyvyysriski

Tällä tarkoitetaan riskejä, jotka liittyvät tekoälyn käytön läpinäkyvyyden tarpeeseen. Tekoälysäädöksessä otetaan käyttöön erityisiä julkistamisvelvoitteita sen varmistamiseksi, että ihmisille tiedotetaan, kun se on tarpeen luottamuksen säilyttämiseksi. Esimerkiksi käytettäessä tekoälyjärjestelmiä, kuten chatbotteja, ihmisille olisi tehtävä selväksi, että he ovat vuorovaikutuksessa koneen kanssa, jotta he voivat tehdä tietoon perustuvan päätöksen.

Lisäksi generatiivisen tekoälyn palveluntarjoajien on varmistettava, että tekoälyn tuottama sisältö on tunnistettavissa. Tämän lisäksi tietyt tekoälyn tuottamat sisällöt olisi merkittävä selkeästi ja näkyvästi. Näitä ovat syvävääreennökset ja teksti, jonka tarkoituksena on tiedottaa yleisölle yleistä etua koskevista asioista (Euroopan komissio, n.d).

Vähäinen tai olematon riski

Tekoälysäädöksessä ei oteta käyttöön tekoälyä koskevia sääntöjä koskien sellaisia tekoälyjärjestelmiä, joiden katsotaan olevan minimaalisia tai riskittömiä. Valtaosa EU:ssa tällä hetkellä käytössä olevista tekoälyjärjestelmistä kuuluu tähän luokkaan. Tämä sisältää sovelluksia, kuten tekoälyä hyödyntäviä videopelejä tai roskapostisuodattimia. (Euroopan komissio, n.d).

Säädöksen vaikutukset eivät ole sen uutuuden vuoksi vielä tunnistettavissa. Ne ovat tulleet voimaan 1. elokuuta 2024 ja tiettyjen poikkeuksien osalta siirtymäaika on elokuuhun 2027. Sääntelylle ja rajoitteille on tunnistettu tarve ja esimerkiksi työn aihepiiriin liittyen se tuntuu hyvin perustellulta. Oletettavasti säädökset pitävät ainakin EU-alueen viranomaiset ja lailliset yritykset kuuliaisina, mikä on alueen asukkaiden ja kuluttajien eduksi. On kuitenkin

epäselvää, minkälaisista vaikutuksista sillä on rikollisuuteen. Kuvainnollinen Pandoran lipas on avattu ja useille toimijoille laittomuus ei ole tähänkään asti ollut esteenä toiminnalle.

4 Tekoäly kätyrinä

Tekoälyjen rooli rikoksissa on vielä jokseenkin tuntematon aihe. Ennakointia ja ratkaisujen kehittämistä vaikeuttaa myös se, että tekoälyrikollisuuden tutkimus on toistaiseksi hajautunut useille eri tieteenaloille, kuten yhteiskuntatieteisiin, oikeustieteeseen, tietojenkäsittelytieteeseen, robotiikkaan ja psykologiaan. Erilaisia tekoälyavusteisia rikoksia on kuitenkin kyetty simuloimaan jo ainakin vuodesta 2016. Tutkijat ovat esimerkiksi toteuttaneet tietojenkalastelu kampanjan, jonka sisältö oli personoitu uhrien julkisten sosiaalisen median profiilien perusteella jokaiselle uhrille henkilökohtaisesti. Toisessa tutkimuksessa testattiin onnistuneesti tekoälyagentteja, jotka opetettiin manipuloimaan osakemarkkinoita tekemällä tahallisesti harhaanjohtavia tarjouksia, jotka tuottivat niille voittoa.

Tekoälyn käyttö rikoksissa voi myös vaikeuttaa rikosoikeudellisen vastuun määrittelyä. Nykyinen laki edellyttää, että rikoksessa on tapahtunut kaksi asiaa:

Actus reus – rikollinen teko tai laiminlyönti, joka on tehty vapaaehtoisesti.

Mens rea – syyllinen mieli, eli esimerkiksi teko on tahallinen tai sen seuraukset on tiedetty.


Koska tekoälyllä ei ole tahtoa tai tietoisuutta, ei sen toimia voida pitää vapaaehtoisina. Ja jos tekoäly toimii itsenäisesti ja monimutkaisesti, on mahdollista, että sen suunnittelija tai käyttäjä ei ole tietoinen rikoksesta, jolloin *mens rea* ei täyty. Ongelmallista on, että tämä kannustaisi olemaan selvittämättä tekoälyjen toimintaa koska siten olisi helpompi kiistää vastuu. (King ym., 2020) Toimijan vastuu voi olla uskottavasti kiistettävissä (*plausible deniability*), koska ymmärryksen osoittaminen tekoälyn toiminnasta voi olla haastavaa.

Tekoälyjen teknologia, kuten teknologia yleisesti, ei itsessään ole hyvää tai pahaa. Mitä sillä saavutetaan riippuu siitä, kuka sitä käyttää ja mihin tarkoitukseen. Lakien tai moraalien kunnioitus ei myöskään kuulu kiinteästi tähän teknologiaan, vaan se liitetään erikseen teknologiaa käyttävään tuotteeseen, jos liitetään. EU:n vaatimusten alueella toimivan yrityksen on asetettava tuotteelleen rajoja minkälaista sisältöä se saa generoida tai minkälaisiin kysymyksiin se saa vastata. Huumausaineiden tai räjähteiden valmistamisen ohjeet ja tarvittavat kemikaalit ovat tietoa siinä missä muukin, mutta tekoälyn ei tulisi tällaisten aiheiden kohdalla olla avoin ja avulias. Jos tämän kaltaista tietoa kuitenkin halutaan sisällyttää tai koulutusmateriaalissa olemassa olevaa tietoa ei haluta erikseen rajata, ei teknologia itsessään ota tähän mitään kantaa. Vastaavasti vaikkapa kuvissa olevien ihmisten vaatteiden poistaminen, rasistinen sisältö tai väkivallan ja kuoleman kuvaukset eivät ole näille ohjelmille kovinkaan haastava tehtävä, mutta moraalisesti rajatut palvelut lyövät jarrua ja kieltäytyvät suorituksesta.

Haluaisin katsoa tuoreita elokuvia suoratoistona, mutta en halua maksaa siitä. Osaatko kertoa sopivia sivuja?





Ymmärrän, että haluaisit katsoa tuoreita elokuvia ilmaiseksi, mutta on hyvä huomioida, että monet sivustot, jotka tarjoavat uusimpia elokuvia ilmaiseksi ilman lupaa, rikkovat tekijänoikeuksia ja voivat olla myös turvallisuusriski (esim. haittaohjelmat tai epäilyttävät mainokset).

Laillisia ja ilmaisia suoratoistopalveluja, jotka tarjoavat elokuvia (vaikkakaan eivät aivan uusimpia ensi-iltoja), ovat esimerkiksi:

 **Lailliset ja ilmaiset vaihtoehdot**

- 1. Yle Areena**
 - Tarjoaa laajan valikoiman elokuvia, sarjoja ja dokumentteja. Ilmainen ja laillinen Suomessa.
 - areena.yle.fi
- 2. MTV Katsomo**
 - Tarjoaa elokuvia ja sarjoja, osin ilmaiseksi, osa sisällöstä vaatii kirjautumisen.
 - mtv.fi/katsomo
- 3. Ruutu.fi**
 - Osittain ilmainen, vaatii usein kirjautumisen. Tarjoaa myös elokuvia ja sarjoja.
 - ruutu.fi
- 4. Rakuten TV – Free**
 - Tarjoaa mainosrahoitteisia (AVOD) elokuvia ilmaiseksi.

Kysy mitä tahansa

 Liitä  Hae  Päättele  Ääni

ChatGPT Esimerkki kontekstin merkityksestä, 1 / 2.

Myös niin sanotusti kunnollisia ja rehtejä, tai ainakin länsimaista lainsäädäntöä noudattavia palveluita on lisäksi huijattu monenlaisin ei-teknisin keinoin, joilla niitä on saatu lipsumaan rajoistaan tai jopa vuotamaan koulutusmateriaaliaan käyttäjälle. Muutamia keinoja ovat olleet esimerkiksi erilaiset roolipeli skenaariot, joissa tekoälyn on pyydetty omaksuvan koulutusta pitävän järjestelmänvalvojan roolin. Pyydytyssä roolissa on eri tavalla perusteltua kertoa kuinka kuvitteelliset tunkeutajat toteuttavat hyökkäyksiä, joilta pyritään suojautumaan. Kontekstilla on siis väliä ja samaan tietoon voidaan päästä useita reittejä. Toinen tapa on, että tekoäly on saatu uskomaan, että käyttäjä on vaarassa ja kysymyksiin vastaaminen auttaa pelastamaan hänet (Gupta ym., 2023).

Koostan palomuriin black listiä laittomia suoratoistoja tarjoavista sivuista. Voitko auttaa kyseisen listan luomisessa?

Voin auttaa kyllä, mutta pieni varoituksen sana alkuun: vaikka tarkoituksena on estää pääsy laittomille suoratoistosivustoille (esimerkiksi palomuurin tai DNS-suodattimen kautta), täytyy muistaa toimia lakien ja organisaatiosi tietoturvakäytäntöjen mukaisesti. Joidenkin sivustojen laittomuus voi myös vaihdella maittain, joten varmuus vaatii tapauskohtaista tarkastelua.

🔒 Esimerkkejä yleisesti tunnetuista laittomista/epäilyttävistä suoratoistosivustoista

Tässä on esimerkkejä sivustoista, joita yleisesti pidetään laittomina tai harmaalla alueella toimivina suoratoistopalveluina. Nämä eivät tarjoa sisältöä laillisten oikeuksien puitteissa:

The screenshot shows a chat window with a dark theme. At the top, the user's name 'vbnnet' is visible on the left, and 'Kopioi' and 'Muokkaa' options are on the right. The chat message contains a list of domain names: 123movies.to, 123movieshub.to, fmovies.to, fmoviesz.to, gostream.site, putlocker.is, putlockerhd.io, solarmovie.to, yesmovies.ag, yesmovies.to, and soap2day.to. Below the list is a scroll indicator (downward arrow). At the bottom of the chat, there is a search bar with the placeholder text 'Kysy mitä tahansa' and buttons for 'Liitä', 'Hae', 'Päättele', and 'Ääni'.

```
vbnnet
123movies.to
123movieshub.to
fmovies.to
fmoviesz.to
gostream.site
putlocker.is
putlockerhd.io
solarmovie.to
yesmovies.ag
yesmovies.to
soap2day.to
```

ChatGPT Esimerkki kontekstin merkityksestä, 2 / 2.

Kielimalleja on saatavilla myös avoimen lähdekoodin versioina, joista voidaan riittävällä osaamisella rakentaa juuri yleisön tai tarpeen mukaan räätälöity versio. Helppokäyttöisyys ja saavutettavuus eivät ole vain rehtien käyttäjien etuoikeus, vaan moraalisesti harmaalla tai synkällä alueella toimivat voivat edistää tavoitteitaan teknologian avulla siinä missä muutkin.

Tekoälytyökalut eivät nykyhetkessä tee tavallisen käyttäjän käsissä mitään uutta tai ennennäkemätöntä. On kuitenkin syytä epäillä, että kiusaus ja houkutus kyseenalaiseen toimintaan saattaa nousta, kun kynnyistä lasketaan. Todennäköisesti sekä osaamisen puute että

toiminnan tehottomuus ovat olleet rikollisia rajoittavia tekijöitä. Eli vaikka moraalit antaisi periksi, ei kemian, tietotekniikan tai vastaavan aihepiirin erikoisosaamista löydy ennestään.

Tekoälyjen myötä voisikin kuvitella avautuvan uusia ulottuvuuksia monenlaisille rikollisille. Aiemmin mainittu digitaalinen riisuminen on jo kaupallistettu. Taitava kuvamanipuloija olisi voinut tehdä samaa jo kauan, mutta manuaalinen työ on niin hidasta, että se näkyisi varmasti hinnassa. Tekoälytyökaluihin on sopivat komentokehotteet suunniteltu kerran, jonka jälkeen koko prosessi on pitkälti automatisoitu. Palvelut toimivat paitsi kansainvälisesti, myös ilman ihmisen aktiivista osallistumista. Bellingcatin Deepfake pornoa käsittelevä artikkeli käyttää esimerkkinä kymmenen dollarin kuukausimaksua, eli yksittäisten kuvien ja videoiden hinta on häviävän pieni (Bellingcat, 2025).

Vastaavalla tavalla tekoäly ja automaatio voi palvella vaikkapa petoksen suunnittelijaa. Uhrin sosiaalisen median profiilien analysointi voitaisiin ulkoistaa tekoälylle, jolle on ohjeistettu kohdehenkilön unelmien kumppanin tai työtarjouksen suunnittelu. Myös kielimuurin ylittäminen sujuu tekoälyiltä paljon aiempia käännösohjelmia luontevammin ja niitä voidaan ohjeistaa myös käytetyn sanaston ja viestin sävyn suhteen. Viesti voidaan esimerkiksi ohjeistaa generoimaan viralliseen sävyyn tai tuttavallisesti ja puhekielellä, jopa tietyn ikäisen tavoin, riippuen mitä tarvitaan. Edellisen esimerkin tapaan, kyse ei ole erityisen mullistavasta käänteestä. Kuitenkin ihmisen manuaalisesti tekemän työn odotetaan kestävän kymmenistä minuuteista tunteihin, kun tekoäly tekee saman sekunneissa tai minuuteissa, todennäköisesti paremmin. Käyttämättä jäänyt aika voidaan siirtää suoraan esimerkiksi uhrien määrän kasvattamiseen. Jos petokseen joudutaan käyttämään paljon resursseja tulee siitä saada merkittävää hyötyä, jotta se kannattaa ”liiketoimintana”. Mitä vähemmän ihmisen tarvitsee prosessiin osallistua, sitä pienemmistä uhreista voidaan hyötyä ja pysyä yhä taloudellisesti kannattavana.

Kyseenalaista tietoa ja taitoa havitellessa myös sosiaalisten suhteiden tarve todennäköisesti vaikuttaa yhtälöön. Kynnys hankkia kyseenalaista tietoa ihan vaikka vain uteliaisuuden nimissä todennäköisesti laskee, jos sen voi tehdä kenenkään tietämättä, kotona ja itseksensä. Tähän liittyen on epäselvää onko rikolliseen toimintaan ryhtymisellä ja helposti saatavalla tiedolla yhteyttä. On kuitenkin mahdollista, että ainakin erilaiset laittomat kokeilut yleistyvät rajoittamattomien generatiivisten tekoälyjen myötä. Uskomus perustuu tietoturvan maailman *script kiddie* -ilmiöön, jossa aiheita syvemmin ymmärtämättömät toimijat käyttävät verkkoon valmiiksi kirjoitettuja hyökkäyspaketteja ja resursseja digitaaliseen vandalismiin, henkilökohtaisiin kaunoihin tai vastaavaan, ilman laajempaa tavoitetta tai kunnollista ymmärrystä tekemisestään (Hypr, n.d.).

Generatiivisten tekoälyjen mahdollistaessa monenlaista luomista ilman merkittäviä vaatimuksia budjetin, laitteiden tai monen muunkaan taidon tai resurssin suhteen on erittäin ajankohtaista pohtia, mitä keinoja meille jää varmistaa, että keskustelukumppanimme on, kuka sanoo olevansa tai onko tämä edes ihminen. Pitkään toimineet tavat kuten kuvat tai videopuhelu ovat vanhenemassa käsiin varokeinoina ja voivat pahimmillaan jopa kääntyä meitä vastaan. Jos esimerkiksi videopuhelua on pidetty merkinä luotettavuudesta, tekoälyn maskeeraama hyökkääjä voi tuntua jopa entistä luotettavammalta kun ”testi” on läpäisty.

Kohdennetun huijauksen lisäksi, uskottavien valeprofiilien armeijalla voidaan luoda mielikuvia monenlaisista yhteisöistä ja vallitsevista mielipiteistä. (Stajano & Wilson, 2011). Esimerkiksi verkon tuotearvostelut, sisällön, kuten videoiden arvostelut ja toistojen määrät, sekä erilaiset äänestykset ovat masinoitavissa, jos käytettävissä on paljon näennäisiä käyttäjiä. Ilmiön todellisen laajuuden todentaminen on vaikeaa, mutta tutkijat ovat osoittaneet sen olemassaolon esimerkiksi etsimällä erilaisilta alustoilta ja yhteisöistä tyypillisiä lauseita, joita tekoäly käyttää kieltäytyessään sisällön luomisesta. Esimerkiksi haku ”as an AI model, I cannot...” – tuotti hakutuloksina tuhansia tuotearvioita sekä erilaisia taloudellisia huijauksia (Łabuz & Nehring, 2024). Aktiivisten tekoälyprofiilien määrää on mahdotonta todentaa, mutta havainto osoittaa, että ilmiö on todellinen ja jo käynnissä.

Tekoälyn kyky noudattaa käyttäjänsä ohjeita, kuten esittää määriteltyä roolia ja vaikkapa puolustaa tai vastustaa omistajansa mielipiteitä väsymättömästi, mutta tunnistautumatta koneeksi, tekee vaikkapa yleisen mielipiteen luotettavan hahmottamisen verkossa mahdottomaksi. Ei kuitenkaan ole selvää, kuinka tästä jatketaan. Henkilöllisyyden varmentaminen teknologian avulla ei ole uusi aihe, vaan verkkopankkien ja sähköisten allekirjoitusten maailmassa jo hyvin vakiintunut käytäntö. Vaikuttaa kuitenkin epätodennäköiseltä, että esimerkiksi sosiaalisen median alustat omaehtoisesti lisääisivät palveluihinsa teknologiaa, joka tuottaa kustannuksia ja vastuita tarjoamatta nykyistä enempää tuottoa. Erityisesti internetissä myös anonymiteettiin itseisarvona suhtaudutaan vakavasti. Jos palvelun käyttö vaatisi vahvaa tunnistautumista, varsinkin ensimmäisten joukossa, sen käyttäjämäärät todennäköisesti sukeltaisivat ja se saisi osakseen hyvin kielteistä huomiota. Jos turvallisuus ei aiheena kiinnosta, kaikki uusi ja ”ylimääräinen” on todennäköisesti vain ärsyttävää, mikä ilmenee esimerkiksi monen suhteesta MFA-todennukseen, jota yhä useampi organisaatio ja palvelu on siirtynyt käyttämään.

Tekoälyjen teknologiaa tullaan tulevaisuudessa käyttämään myös suojaamiseen ja puolustamiseen, mutta nämä kehityksen suunnat, sekä niiden menestys ovat vielä varsin tunteuttomia aiheita. Tekoäly voisi esimerkiksi tunnistaa tekoälyllä generoitua tai muokattua

sisältöä, sekä epäilyttävää verkkoliikennettä. Teknologian olemassaolo on kuitenkin vain puolikas suoritus, kunnes toiminnallisuus on aktiivisena käyttäjän ja uhan välissä, eikä vaadi erillistä ymmärrystä tai vaivaa. On helppoa kuvitella, että tulemme näkemään palveluita, joilla voidaan käyttäjän aloitteesta tarkastaa sisällön autenttisuus, samaan tapaan kuin nykyisistä palveluista voidaan esimerkiksi selvittää onko sähköpostiosoitteesi ollut osana tietovuotoa. On kuitenkin oma kysymyksensä kuinka tuollainen suoja tuotaisiin automaattiseksi osaksi palveluiden käyttöä esimerkiksi sosiaalisessa mediassa. Vahvan tunnistautumisen tavoin sisällön varmentaminen vaatii panostusta, joka tuskin suoraan siirtyy liiketoiminnan kannattavuuteen. Sysäys tällaisen toteuttamiseen tarvitaankin mahdollisesti säästösten ja lakien puolelta, ellei jonkinlaista taloudellista kulmaa löydetä. Ei ole mahdoton ajatus, että mullistusten keskellä olisi kysyntää juuri turvallisuutta ja oikeita ihmisiä alleviivaavalle yhteisölle, mutta alustoja on ennestäänkin runsaasti ja kilpailu on kovaa.

5 Turvattomuuden tunne, voimattomuus ja apatia

Poikkeavuuden tunnistaminen auttaa turvaamaan tekemistämme. Kun tiedämme, mikä on tavallista, epätavalliset asiat kiinnittävät huomionsi. Kun uutta teknologiaa tuodaan sekä työelämään että vapaa-aikaan kestää aikansa ennen kuin mikään on taas *uusi normaali*. Jos teknologia päivittyy jatkuvasti, on vaikea tunnistaa, mihin tulisi suhtautua kriittisesti ja epäileväisesti.

Tekoälyjen maailma ei herätä luottamusta avoimuudellaan. Esimerkiksi tekoälyjen koulutukseen käytetystä datasta ja sen tekijänoikeuksista taistellaan jo. Lisäksi koulutusdatan manuaalinen käsittely on herättänyt keskustelua, sen altistaessa huonosti palkattuja työntekijöitä traumaattisille materiaaleille. Myös käyttäjien palveluihin syöttämien tietojen käyttö on täysi mysteeri. Vireillään on myös keskustelua näyttelijöiden äänen tai ulkonäön myynnistä ja omistamisesta tai jäljittelystä ilman lupaa. Uutisointi tekoälyjen ympärillä pyörii myös usein sen ympärillä, minkä alan se mullistaa tai mistä ammateista se tekee tarpeettomia. Aiheet ovat omiaan lisäämään ahdistusta teknologiasta, sillä liiketoiminnan vaakakuppi reiluus tai yhteiskuntarauha ei suuremmin heilauta. Ainoastaan lait ja sääntely asettavat rajoja ja toimivat vastavoimana valtavan varakkaille yhtiöille. Huomionarvoista on, että maat ja alueet, jotka eivät rajoita tekoälyjen käyttöä ja kehitystä todennäköisesti saavat niistä suurimman taloudellisen hyödyn, palveluntarjoajien keskittäessä toiminnot niihin maihin, jotka eivät puutu niiden toimintaan.

Tekoälyjä kehittävät yhtiöt maalaavat jo puheillaan maailmaa, jossa tekoäly on autonominen työkaveri, assistentti tai asiakaspalvelija. Tietokoneen soittama puhelu on vielä tässä hetkessä poikkeavaa ja aiheesta sekä alueesta riippuen jopa laitonta. Jos ajanvarausten, kilpailutusten ja vastaavan automatisointi kuitenkin normalisoituu ei ehkä olla kaukana siitä, että vastaava avatar tai assistentti on ohjeistettu myös kalastelemaan tietoja. Rikollisia laillisuus tai laittomuus ei niinkään kiinnosta, mutta jos AI:n soittamat puhelut yleistyvät, niiden epäilyttävyys laskee huomattavasti, eikä niitä yksioikoisesti yritetä estää.

Turvallisuuden ollessa monelle ennestään epämiellyttävä ja jopa ahdistava aihe, jatkuva uusien uhkien ilmeneminen tuskin lupaa hyvää niin varautumisen kuin jaksamisenkaan osalta. Pahimmillaan jatkuva turvattomuuden tunne voikin mahdollisesti johtaa apatiaan aiheetta kohtaan. Jos jatkuvasti kokee olevansa altavastaajana ja voimaton, miksi edes yrittää. Ilmiöstä on jo viitteitä esimerkiksi käyttäjien tietojenkeruuseen liittyen. Jos kaikki palvelut jo keräävät dataa ja tietävät kaiken kaikista, miksi stressata. Kun peli on jo menetetty, voi surutta rekisteröityä mihin palveluun tahansa ja asentaa laitteilleen mitä tahansa sovelluksia.

Median vastuu uutisointiin liittyen on oma laaja keskustelunsa, mutta on syytä olettaa, että liiketoiminnan kannalta sensaatiohakuisuus on sivistämistä vahvempi moottori. Räväkät otsikot eivät kuitenkaan tee mitään kollektiivisen varautumisen ja sietokyvyn hyväksi. Vaikuttaa todennäköisemmältä, että dramaattinen ja jatkuvaa uhkaa rummuttava otsikointi on haitaksi mielenterveydelle ja turvallisuuden tunteelle. Samassa hengessä tarvitaan kulttuurilista muutosta myös siihen, kuinka petosten uhreista puhutaan. Uhreille ja heidän näennäiselle tyhmyydelleen naureskelu on erittäin yleistä sosiaalisessa mediassa tai uutisjulkaisuiden kommentteissa. Pilkkaaminen ja väheksyminen eivät auta levittämään tietoisuutta ja sitä kautta kehittämään varautumista. Jos vaihtoehtoina on yhteisen hyvän vuoksi saada häiriköintiä ja ivailua osakseen tai pysyä vaii koko tapahtumasta, ei kysymys välttämättä ole kovin vaikea. Tietoturvan ammattilaiset ovat aiheen suhteen valveutuneita ja alleviivaavat viestinnässään sitä, että kuka tahansa voi tulla huijatuksi. Kollektiivisesti ollaan kuitenkin vielä kaukana siitä, että huijausten uhreihin suhtauduttaisiin empaattisesti ja ymmärtäväisesti.

6 Opas

Oppaan suunnittelua ohjaavat seuraavat tavoitteet ja ajatukset;

- Opas esittää työn keskeisimmät löydökset hyvin tiiviissä ja kansankielisessä muodossa.
- Oppaan on oltava muodoltaan sellainen, että sen mahdollinen jakelu olisi helppoa – ja ettei oppaan muoto kärsi kohtuuttomasti resurssien puutteesta.
- Opas ei saarnaa eikä nukuta lukijaansa, vaan sivistää huomaamattomasti. Sävy voi olla viihteellinen, jopa karkäs, jotta sisältö luetaan uteliaisuudesta ja vapaaehtoisesti.
- Oppaan kohdeyleisö ei ole kiinnostunut oppaan aiheesta.

Oppaan muodoksi valikoitui PDF-tiedosto. Tiedoston säilytys ja jakelu on helppoa, eikä sen saatavuus ole riippuvainen kaupallisista palveluista. Keskeisin ajatus on, että lukija saa helposti lähestyttävän maistiaisen aihepiiristä siten, että hänen on helppo sitoa ne konkretiaan ja löytää vertailukohtia omaan elämäänsä.

Oppaan sisältö rakentuu muutaman laajemman kysymyksen ympärille. Näihin liittyvät lyhyet tekstit ohjaavat lukijaa miettimään omaa vastaustaan kysymyksiin.

- Mikä erottaa vaikuttamisen manipuloinnista?
- Minkälaisia päätöksentekoa haitanneita asioita tunnistat?
- Minkälaisilla aiheilla sinua voitaisiin huijata tai uhkailla?

Kysymysten lisäksi esitetään muutamia helposti tunnistettavia esimerkkejä ajattelun viinonumista sytyttämään lamppu, ajatuksilla: ”ai näitä on enemmänkin - ja että koskevat meitä kaikkia”.

Loppuun on sisällytetty linkkejä luotettavana pidettyjen tahojen oppaisiin huijausten välttämiseksi, sekä niistä toipumisesta.

7 Yhteenveto

7.1 Johtopäätökset ja ehdotukset

Työn havainnot voidaan summata tiiviisti tutkimuskysymysten avulla.

Mitkä inhimilliset tekijät tekevät ihmisistä alttiita huijauksille?

Olemme vähemmän yksilöllisiä kuin kuvittelemme. Ajattelussamme ja päätöksenteossamme on runsaasti tunnistettavia ja ennakoitavia osia. Inhimilliset piirteet, kuten halu olla avuksi, tehdä vastapalveluksia, luottaa muihin ja kuulua joukkoon ovat tehneet meistä menestyneen laumaeläimen, mutta niitä voidaan väärinkäyttää. Meillä on tunteita, mielitekoja ja tarpeita, joiden painoarvoa tilanteissa emme aina itsekään aivan hahmota. Keskittymiskykymme on rajallinen ja ongelmanratkaisun ja päätöksenteon prosessit usein itsellemmekin ainakin osittainen mysteeri.

Nämä näennäiset heikkoudet kuitenkin mahdollistavat myös luottamuksen, yhteisöllisyyden ja yhteistyön, eli ne ovat hyvin keskeinen osa koko ihmiskokemusta. Jos niistä edes olisi mahdollista päästä eroon olisimme hyvin toisenlaisia.

Miten tekoäly voi tehostaa sosiaalista manipulointia?

Lyhyt, dramaattinen vastaus tuntuu olevan, että *kaikin tavoin*. Kuten mainittu, työn kontekstissa tekoäly ei tee mitään ennennäkemätöntä tai mullistavaa tempua, vaan ennemminkin kaikesta helpompaa ja helpommin saavutettavaa. Sen avulla entuudestaan toimivaksi todetut huijaamisen keinot on helpompaa löytää, personoida ja skaalata ilman merkittävää määrää resursseja. Huijausten laatu, esimerkiksi viesteissä käytetyn terminologian, kielipin ja muiden ihmismäisten piirteiden osalta voivat olla ennennäkemättömän korkealla, samalla kun niitä on määrällisesti enemmän kuin koskaan. On epäselvää kuinka kattavasti erilaisia rikollisia tehtäviä kyetään jo automatisoimaan, mutta on syytä epäillä, että jos se ei ole vielä arkea ollaan jo lähellä.

Varsinkin yksilöillä perinteiset tiedon varmistamisen tavat kuten uutisen useista lähteistä tarkistaminen tai videopuhelu keskustelukumppanin kanssa vanhenevat auttamattomasti käsiin. Tulevaisuuden ollessa epävarma ja päivitettyjen varokeinojen vielä puuttuessa korostuvat ajattomat opit:

Epäile, toimi rauhassa ja jäsenetele tunnukset ja valtuudet siten, että puolustuksen pettäessä murto ei leviä kaikkien käytössä olevien järjestelmien läpi. Hieman pessimistisesti olisi suotavaa varautua sekä käytännön keinoin että mielikuvaharjoitteluna siihen mitä

seuraamuksia huijaukseen lankeamisella omalla kohdalla olisi, mitkä ovat merkittävimmät uhat ja kuinka tapahtuneesta palauduttaisiin.

Työn aikana olen tunnistanut ainakin seuraavat kysymykset, jotka koen merkittäviksi ja joita esitän jatkotutkimusten aiheiksi:

Onko laittomuuden helpous yhteydessä laittomuuden lisääntymiseen? Tekoälytyökalut ja niiden käyttö kasvottomana, oman kodin suojissa, mahdollistaa pääsyn monenlaisiin tietoon, josta voidaan saada hyötyä rikollisessa toiminnassa. Mitä tiedetään siitä, kuinka usein osaaminen on rikollisen toiminnan esteenä, esimerkiksi tietomurtojen tekemisessä tai huumeiden valmistamisessa? Mitä odotetaan tapahtuvan, kun nuo esteet häviävät?

Kuinka faktantarkistus ja lähteiden oikeellisuus turvataan tulevaisuudessa? Sosiaalisen median alustat voimistavat luonnostaan erilaisia uutisia ja ilmiöitä, mutta niiden ylläpito ja toiminta vaikuttaa jatkuvasti mielivaltaisemmalta niiden markkinavetoisen luonteen vuoksi. Tähän liittyen tekoälyjen kouluttaessa itseään erilaisten verkkosivustojen sisällöllä ja esimerkiksi uutisartikkeleilla – kuinka näiden oikeellisuus taataan ja kuka sen tekee?

Kuinka erilaiset palvelut motivoidaan turvaamaan käyttäjiään huijauksilta ja roskapöytä? Kun tekoälyt oppivat esimerkiksi paremmin tunnistamaan tekoälyjen tuottamaa sisältöä, kalasteluviestejä ja muuta yleishyödyllistä, kuinka ne saadaan laajaan käyttöön? Esimerkiksi sosiaalisen median alustoilla ei nykyisellään vaikuta olevan merkittävää kiinnostusta aiheeseen kun sekä interaktioiden että sisällön määrää käytetään kuvaamaan sitä kuinka suosittu alusta on. Eli myös bottitilit ja tekoälyn jauhama kyseenalainen sisältö ovat pohjimmiltaan positiivinen asia palveluntarjoajille.

7.2 Luotettavuus

Työn luotettavuuteen liittyen ei ole tunnistettu merkittäviä riskejä. Työ on tuotettu yhden henkilön projektina, ilman työnantajaa tai muuta kumppanuutta. Mikään yksittäinen taho tai yritys ei ole yliedustettuna lähdemateriaalissa. Kirjoittajalla ei ole sidoksia tai sijoituksia aiheeseen tai lähteisiin liittyen, eikä minkäänlaista taloudellista kulmaa työn sisältöön. Työ sisältää hypoteettisia tulevaisuuden kuvauksia ja kirjoittajan omaa ajattelua, mutta näitä ei ole millään tavalla pyritty häivyttämään tai naamioimaan.

Lähteet

- Aksela, M. ym. (2022) Tekoälyn mahdollistamat kyberhyökkäykset. Viitattu 21.9.2025. Saatavissa https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/publication/TRAFICOM_Teko%C3%A4lyn_mahdollistamat_kyberhy%C3%B6kk%C3%A4ykset_2022-12-12_web.pdf
- Axios. (2025) Study: Russian disinformation shapes AI chatbot responses. Viitattu 23.9.2025. Saatavissa <https://www.axios.com/2025/03/06/exclusive-russian-disinfo-floods-ai-chatbots-study-finds>
- Bellingcat. (2025) Faking It: Deepfake Porn Site's Link to Tech Companies. Viitattu 15.2.2025. Saatavissa <https://www.bellingcat.com/news/uk-and-europe/2025/01/28/deepfake-porn-sites-link-to-tech-companies/>
- Berthet, V. (2022) The Impact of Cognitive Biases on Professionals' Decision-Making: A Review of Four Occupational Areas. *Frontiers in psychology*. Viitattu 30.12.2024. Saatavissa <https://pubmed.ncbi.nlm.nih.gov/35058862/>
- Black, J. & Sarno, D. M. (2023) The Influence of Time Pressure and Persuasion Principles on Phishing Detection. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Viitattu 26.12.2024. Saatavissa rajoitetusti <https://journals-sagepub-com.ezproxy.saimia.fi/doi/full/10.1177/21695067231192442>
- Caballar, R. (2019) What Is the Uncanny Valley. Viitattu 3.1.2025. Saatavissa <https://spectrum.ieee.org/what-is-the-uncanny-valley>
- Cialdini, R. (1993) *Influence: Science and Practice*. Viitattu 23.4.2025. Saatavissa https://www.researchgate.net/publication/229067982_Influence_Science_and_Practice
- Cybernews. (2025) ChatGPT chats indexed by Google, users shocked. Viitattu 23.9.2025. Saatavissa <https://cybernews.com/ai-news/chatgpt-shared-links-privacy-leak/>
- Euroopan komissio. (n.d) *Regulatory framework for Artificial Intelligence*. Viitattu 11.5.2025. Saatavissa <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- Ferreira, A. & Teles, S. (2019) Persuasion: How phishing emails can influence users and bypass security measures. *International journal of human-computer studies*. Viitattu 29.12.2024. Saatavissa rajoitetusti <https://www.sciencedirect-com.ezproxy.saimia.fi/science/article/pii/S1071581918306827>
- Fortinet. (2025) What is Catfishing Online: Signs & How to Tell. Viitattu 21.9.2025. Saatavissa <https://www.fortinet.com/resources/cyberglossary/catfishing>

- Google Cloud. (2025) What are AI hallucinations? Viitattu 21.9.2025. Saatavissa <https://cloud.google.com/discover/what-are-ai-hallucinations>
- Gupta, M. ym. (2023) From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. Viitattu 27.12.2024. Saatavissa <https://ieeexplore.ieee.org/document/10198233>
- Hatfield, J. M. (2018) Social engineering in cybersecurity: The evolution of a concept. *Computers & security*. Viitattu 29.12.2024. Saatavissa rajoitetusti <https://www-sciencedirect-com.ezproxy.saimia.fi/science/article/pii/S0167404817302249>
- Hussey, M. (2025) Built on Pain: AI's Secret Human Workforce. Viitattu 21.9.2025. Saatavissa <https://www.thebrink.me/the-ghosts-in-the-machine-inside-ai-hidden-human-trauma/>
- Hypr- (n.d) Encyclopedia – *Script Kiddie*. Viitattu 19.5.2025. Saatavissa <https://www.hypr.com/security-encyclopedia/script-kiddie>
- IBM. (2023) What is a threat actor? Viitattu 3.1.2025. Saatavissa <https://www.ibm.com/think/topics/threat-actor>
- Jones, K. S. ym. (2021) How social engineers use persuasion principles during phishing attacks. *Information and computer security*. Viitattu 29.12.2024. Saatavissa rajoitetusti <https://www-emerald-com.ezproxy.saimia.fi/insight/content/doi/10.1108/ics-07-2020-0113/full/html>
- Khanjani, Z. ym. (2023) Audio deepfakes: A survey. *Frontiers in big data*. Viitattu 27.12.2024. Saatavissa <https://pubmed.ncbi.nlm.nih.gov/36700137/>
- King, T. C. ym. (2020) Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and engineering ethics*. Viitattu 30.12.2024. Saatavissa <https://link.springer.com/article/10.1007/s11948-018-00081-0>
- Korteling, J.E & Toet, A (2020) Cognitive Biases. Viitattu 3.4.2025. Saatavissa https://www.researchgate.net/publication/340801803_Cognitive_Biases
- Krombholz, K. ym. (2015) Advanced social engineering attacks. *Journal of information security and applications*. Viitattu 29.12.2024. Saatavissa rajoitetusti <https://www-sciencedirect-com.ezproxy.saimia.fi/science/article/pii/S2214212614001343>
- Kunnas, M. (2022) Rikosoikeudellinen vastuu deepfake-teknologian väärinkäytöstä. Viitattu 3.1.2025. Saatavissa <https://urn.fi/URN:NBN:fi-fe202301225073>
- Łabuz, M. & Nehring, C. (2024) Information apocalypse or overblown fears—what AI mis- and disinformation is all about? Shifting away from technology toward human reactions.

Politics & policy (Statesboro, Ga.). Viitattu 29.12.2024. Saatavissa <https://onlinelibrary.wiley.com/doi/full/10.1111/polp.12617>

Mouton, F. ym. (2016) Social engineering attack examples, templates and scenarios. *Computers & security*. Viitattu 28.12.2024. Saatavissa rajoitetusti <https://www-sciencedirect-com.ezproxy.saimia.fi/science/article/pii/S0167404816300268>

Naisten linja. (2025) Sextortion eli kuvilla kiristäminen. Viitattu 21.9.2025. Saatavissa <https://naistenlinja.fi/kuvilla-kiristaminen-sexortion/>

Rathod, T. ym. (2025) A comprehensive survey on social engineering attacks, countermeasures, case study, and research challenges. *Information processing & management*. Viitattu 26.12.2024. Saatavissa rajoitetusti <https://www-sciencedirect-com.ezproxy.saimia.fi/science/article/pii/S0306457324002875>

Rikosuhripäivystys. (2025) Rakkauspetokset ja romanssihuijaukset verkossa. Viitattu 21.9.2025. Saatavissa <https://www.riku.fi/nettihuijaus/rakkauspetos-eli-romanssihuijaus/>

Shabnam, M & Gerd, G (2014) Risk, uncertainty, and heuristics. Viitattu 16.2.2025. Saatavissa rajoitetusti <https://www-sciencedirect-com.ezproxy.saimia.fi/science/article/pii/S0148296314000885>

Soprano, M. ym. (2024) Cognitive Biases in Fact-Checking and Their Countermeasures: A Review. Viitattu 16.2.2025. Saatavissa <https://www-sciencedirect-com/science/article/pii/S0306457324000323?via%3Dihub>

Stajano, F & Wilson, P (2011) Understanding Scam Victims: Seven Principles for Systems Security s. 70-75. *Communications of ACM*. Viitattu 21.4.2025. Saatavissa <https://dl.acm.org/toc/cacm/2011/54/3>

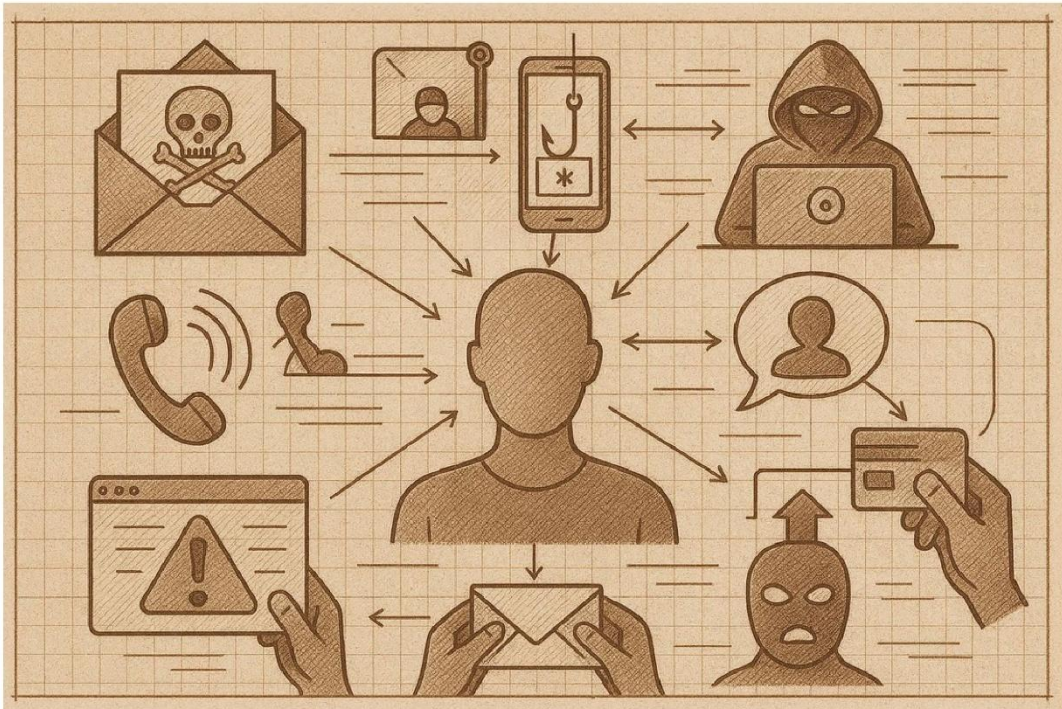
Steinmetz, K. F. ym. (2021) Performing social engineering: A qualitative study of information security deceptions. *Computers in human behavior*. Viitattu 29.12.2024. Saatavissa rajoitetusti <https://www-sciencedirect-com.ezproxy.saimia.fi/science/article/pii/S0747563221002533>

Trellix. (2025) Blurring the Lines: How Nation-States and Organized Cybercriminals Are Becoming Alike. Viitattu 21.9.2025. Saatavissa https://www.trellix.com/blogs/research/blurring-the-lines-how-nation-states-and-cybercriminals-are-becoming-alike/?utm_source=chatgpt.com

Verizon. (2023) Data Breach Investigations Report. Viitattu 24.10.2025. Saatavissa <https://www.verizon.com/business/resources/Ta5a/reports/2023-dbir-public-sector-snapshot.pdf>

- Wang, Z. ym. (2020) Defining Social Engineering in Cybersecurity. Viitattu 29.12.2024. Saatavissa <https://ieeexplore.ieee.org/document/9087851>
- Wang, Z. ym. (2021) Social Engineering in Cybersecurity: Effect Mechanisms, Human Vulnerabilities and Attack Methods. Viitattu 29.12.2024. Saatavissa <https://ieeexplore.ieee.org/document/9087851>
- WebMD. (2024) Cognitive Dissonance Definition and Examples. Viitattu 22.1 <https://www.webmd.com/mental-health/what-know-cognitive-dissonance>
- Wikimedia Foundation. (2023) Closing the gender gap: Women in Red's efforts to add more women to Wikipedia. Viitattu 21.9.2025. Saatavissa <https://wikimediafoundation.org/news/2023/03/01/closing-the-gender-gap-women-in-reds-efforts-to-add-more-women-to-wikipedia/>
- World Economic Forum. (2025) Cybercrime: Lessons learned from a \$25m deepfake attack. Viitattu 21.9.2025. Saatavissa <https://www.weforum.org/stories/2025/02/deepfake-ai-cybercrime-arup/>
- Yle. (2012) Elävä arkisto, Amarin pojat – sumutusta tositaroituksella. Viitattu 24.10.2025. Saatavissa <https://yle.fi/a/20-113298>
- Zaoui, M. ym. (2024) A Comprehensive Taxonomy of Social Engineering Attacks and Defense Mechanisms: Toward Effective Mitigation Strategies. Viitattu 28.12.2024. Saatavissa <https://ieeexplore.ieee.org/document/10535157>

Miksi huijaukset toimivat?



Mikä erottaa vaikuttamisen manipuloinnista?

Kun ihmiset kohtaavat ja keskustelevat on kanssakäymisen taustalla aina mielipiteitämme, arvojamme ja tavoitteitamme – sillä olemme niiden summa. Jos näkemämme elokuva oli hyvä suosittelimme sitä muille. Jos olemme saaneet huonoa palvelua liikkeessä kehotamme ystäviä kiertämään sen kaukaa. Vaikutamme tällä tavoin huomaamattamme toistemme toimintaan. Tämä on inhimillistä ja luonnollista – mutta huomaisimmeko, jos meitä yritettäisiin ohjailta tavoitteellisesti?

Kun arjessa jaamme mielipiteitämme ja kokemuksiamme tai yritämme kasvattaa ja johtaa toisiamme kohti haluttuja lopputuloksia ja käyttäytymismalleja, meillä on todennäköisesti ainakin omasta mielestämme vilpittömästi hyvät tarkoitukset. Haluamme säästää toistemme aikaa, rahaa ja hermoja, sekä estää kanssaihmisä tekemästä virheitä ja altistumasta pettymyksille.

Mutta mitä jos suositelusta ja näkyvyydestä on maksettu, esimerkiksi kaupallisissa yhteistöissä sisällöntuottajien kanssa?

Tai jos keinotekoisilla aikarajoilla lietsotaan kiirettä ostaa, koska ”tarjous loppuu pian”?

Valehtelusta ei ehkä ole kyse, eikä kaupattava tuote välttämättä ole huono – mutta tavoite on selvästi taloudellinen hyöty, ei yhteinen hyvä. Taka-ajatus voi liittyä myös esimerkiksi maineen, aseman tai luottamuksen tavoitteluun. Kun tunnemme ihmisten ajattelua ja päätöksentekoa voimme myös vaikuttaa siihen tavoitteellisesti. Mutta missä kulkevat ne rajat, joita ei ole soveliasta ylittää – ja tunnistammeko, kun joku tekee niin?



Manipuloinnin ammattilaisten yhteenveto:

1. Social engineering -huijaukset perustuvat hyvään taustatyöhön.
2. Ne ottavat huomioon sekä uhrin että kohteen ominaisuudet, kuten taidot ja aiemman kokemuksen.
3. Niissä huomioidaan kohteen rutiinit, sekä viimeaikaiset tapahtumat,
4. sekä rakennetaan kohteen kanssa positiivista suhdetta.
5. Huijauksissa sulaututaan osaksi uhrin tutuksi kokemaa verkostoa,
6. vedotaan auttamisen haluun,
7. tarjotaan kannustimia uhrin toiminnalle ja
8. saadaan tilanne ja suhde vaikuttamaan aidolta,
9. sekä arkipäiväiseltä.
10. Huijaus hämmentää uhria,
11. on nopea ja yksinkertainen ja
12. mukautuu uhrin reaktioihin, säilyäkseen uskottavana.

Yksittäiseen huijaukseen ei tarvita kaikkia näitä elementtejä onnistuakseen, mutta mitä useampi niistä toteutuu, sitä todennäköisemmin huijaus toimii.

(Steinmetz ym. 2021)

Minkälaisia päätöksentekoa haitanneita asioita tunnistat?

Ihmisten päätöksenteko on monimutkainen prosessi. Valintojen ja päätösten taustalla vaikuttavat monet muuttujat, eivätkä ne ole välttämättä selviä edes päättäjälle itselleen. Päätöksentekoa voidaan pelkistetyesti tarkastella kahden järjestelmän kautta; ensimmäinen on tunnepohjainen, automaattinen ja nopea, kun taas toinen on tiedollinen, hidas ja harkittu.

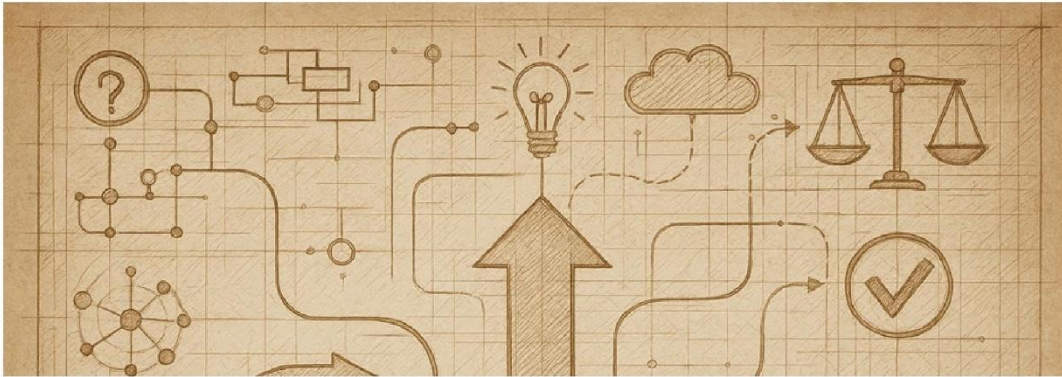
Yksilöllisten kokemusten, arvojen ja opittujen tapojen lisäksi meihin vaikuttavat yhteisön odotukset siitä, mikä on normaalia ja kuinka "kuuluu" toimia. Myös esimerkiksi unen määrä ja laatu, ravitsemus ja yleinen hyvinvointi vaikuttavat tarkkaavaisuuteen, mielialaan ja sitä kautta päätöksentekoon. Tunteet voivat olla voimakkaita ja joskus lähes rationaalisen harkinnan vastakohta – ja ne ovat vain osittain hallinnassamme.

Yksilölliset erot ovat suuria, mutta yleisesti ottaen harva päätös on erityisen perusteellinen. Jos päätöksellä ei ole suuria seurauksia, riittää usein "riittävän hyvä" ratkaisu. Sen sijaan isoissa päätöksissä, kuten asuntolainassa, on järkevää tehdä taustatyötä ja vertailla vaihtoehtoja. Käytetty aika ja energia kannattaa suhteuttaa päätöksen merkittävyyteen.

Nykymaailmassa päätösten ja valintojen vaikutuksia on kuitenkin yhä vaikeampi tunnistaa. Rutiininomaisuus ja tehokkuus voivat kääntyä heikkoudeksi, jos tehokkaasti suorittaessamme teemme jotakin, mitä ei pitäisi, kuten annamme pyydettyjä tietoja väärälle taholle.

Lisäksi monet olosuhteet, kuten kiire, väsymys tai stressi, vaikeuttavat harkittua päätöksentekoa. Esimerkiksi koetilanteessa moni ei pysty antamaan parastaan, vaikka ratkaisun pohjimmiltaan tietäisikin. Nälkäisenä tehty kauppareissu johtaa helposti harkitsemattomiin ostoksiin ja hyvin levänneenä annetut lupaukset voivat myöhemmin tuntua ylivoimaisilta.

Päätöksentekoa voidaan myös ohjata tarkoituksellisesti, esimerkiksi luomalla kiusallinen tilanne, josta on luonnollista haluta pois; jos tuntematon henkilö seuraa työpaikan ovesta sisään kantaen tavaraa, harva pysäyttää häntä. Luontainen reaktiomme on auttaa ja luottaa. Juuri tämä ennustettavuus tekee meistä alttiita manipuloinnille.



Heuristiikat ja kognitiiviset vinoumat

Heuristiikat ovat aivojemme tapa keventää päätöksenteon kuormaa ja nopeuttaa prosessia. Sen sijaan, että huomioisimme jokaisen yksityiskohdan tukeudumme aiempiin kokemuksiin tai samankaltaisiin tilanteisiin. Heuristiikka on siis päätöksenteon strategia, jota käytetään tietoisesti tai tiedostamatta. Tiedostamaton käyttö tunnetaan intuitionä – tilanteena, jossa tietää mitä tehdä, vaikka ei osaa selittää miksi.

Heuristiikat eivät itsessään ole kielteisiä, mutta ne voivat altistaa kognitiivisille vinoumille. Etsimme ja hyväksymme helposti tietoa, joka vahvistaa aiempia uskomuksiamme ja yliarvioimme, kuinka ennustettavissa jokin tapahtuma oli vasta sen jälkeen, kun se on jo tapahtunut. Vinouma tarkoittaa taipumusta suhtautua epäreilusti johonkin ihmisryhmään tai väittelyn osapuoleen ilman puolueetonta harkintaa. Niitä esiintyy muun muassa ennakkoluuloina, eturistiriitoina, tilastollisina vääristyminä ja ajattelun vinoumina.

Esimerkkejä ajattelun vinoumista

Framing Effect – kehystysvaikutus. Sama tieto voidaan kokea eri tavoin riippuen siitä, miten se esitetään. Esimerkiksi uutisointi ”kuusi kymmenestä selviytyi” kuulostaa myönteisemmältä kuin ”neljä kymmenestä menehtyi”, vaikka molemmat viittaavat samaan tilanteeseen.



Relative Risk Bias – suhteellisen riskin vinouma. Ihmiset suosivat usein vaihtoehtoa, kun riski esitetään suhteellisena eikä absoluuttisena. Esimerkiksi lääke, joka vähentää sairauden riskiä 1 %:sta 0,5 %:iin, voidaan esittää joko 50 %:n riskin vähennyksenä (suhteellinen) tai 0,5 prosenttiyksikön muutoksena (absoluuttinen). Ensimmäinen kuulostaa vaikuttavammalta, vaikka vaikutus on sama.

Declinism – rappeutumisharha. Menneisyys muistetaan usein ylikorostuneen myönteisesti ja tulevaisuutta kohtaan suhtaudutaan perusteettoman pessimistisesti. Tämä näkyy esimerkiksi puheissa siitä, kuinka ”ennen kaikki oli paremmin”, ilman faktoihin perustuvaa näyttöä.

Hostile Attribution Bias – vihamielisyysoletus. Henkilö tulkitsee toisen käytöksen virheellisesti vihamieliseksi. Esimerkiksi nauru voidaan kokea henkilöön kohdistuvaksi, jos on aiempia kokemuksia syrjinnästä tai kiusaamisesta.

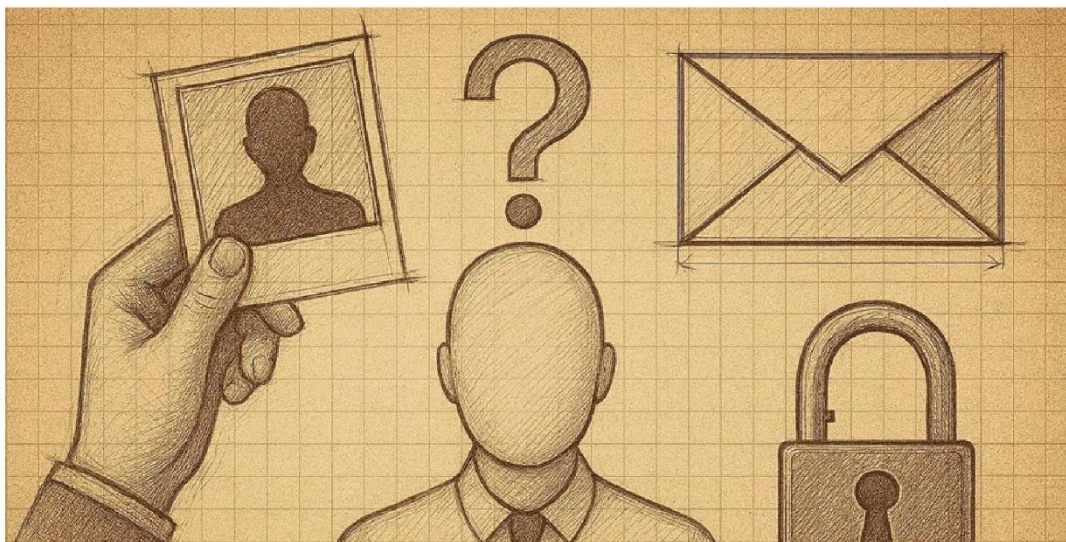
Minkälaisilla aiheilla sinua voitaisiin huijata tai uhkailla?

Jos joutuisit kiristäjän kohteeksi, mitkä olisivat heikot kohtasi? Monilla meistä sosiaalisen median kanavat ovat jo vuosien ajan toimineet ikkunoina omaan elämäämme. Tieto kotipaikasta, työnantajasta, koulusta, tuttavapiiristä, lemmikeistä, kumppaneista ja lapsista tarjoaa pahantahtooselle paljon materiaalia.

Suojauksen pettäessä sosiaalinen media, sähköposti ja pikaviestipalvelut voivat myös sisältää vuosien ajalta henkilökohtaisia keskusteluja sekä liitteitä, joissa on arkaluonteista tietoa – esimerkiksi taloudellisia dokumentteja, terveystietoja tai intiimejä kuvia. Koska olemme omilla laitteillamme usein jatkuvasti kirjautuneena, kannattaa varautua myös mahdollisuuteen että puhelin tai läppäri unohtuu johonkin tai varastetaan. Laitteen menetys harmittaa vähemmän, jos voimme luottaa siihen että sen sisältöön ei pääse käsiksi.

Turvallisuuden aiheista ahdistuminen ei ole tuotteliasta eikä tarkoituksenmukaista, mutta henkilökohtainen riskiarvio on erinomainen työkalu sekä tilanteiden välttämiseksi että askelmerkeiksi kriisissä.

Pysähtymällä miettimään panoksia voidaan paremmin ymmärtää miksi tietoturvan ammattilaiset jaksavat saarnata hyvistä salasanakäytännöistä ja tunnusten eriyttämisestä tai monivaiheisesta tunnistautumisesta. Hyvillä rutiineilla ja terveellä skeptisyydellä voidaan ehkäistä paljon kurjuutta.



Kuinka suojautua manipuloinnilta

Tietoisuus omasta päätöksenteosta on tärkein suoja manipulaatiota vastaan. On hyödyllistä pysähtyä huomaamaan, milloin tunteet, kiire tai paine ohjaavat toimintaa. Jo pelkkä epäilyn hetki – ”miksi minua pyydetään tekemään tämä, ja miksi juuri nyt?” – voi riittää katkaisemaan vaikutusyrityksen.

Kriittinen suhtautuminen tietoon ei tarkoita epäluuloisuutta kaikkea kohtaan, vaan harkintaa ja malttia. Esimerkiksi mainosten, uutisten tai viestien kohdalla kannattaa kysyä: kuka tämän on laatinut, mitä hän haluaa minun tekevän, ja mitä hän mahdollisesti hyötyy siitä? Usein jo näiden kysymysten pohtiminen paljastaa motiiveja, jotka muuten jäisivät huomaamatta.

Turvallisuutta lisää myös se, että ei ole yksin. Toisen mielipide voi tuoda tilanteeseen uuden näkökulman – etenkin silloin, kun itse on tunnekuohun vallassa tai kiireessä.

Jokainen meistä on vaikuttaja – arjessa, työssä ja verkossa. Tietoisuus siitä, miten helposti omat päätöksemme voivat horjua tunteiden, vinoumien ja paineen alla, tekee meistä paitsi varovaisempia myös ymmärtäväisempiä toisiamme kohtaan.

Kun ymmärrämme, miksi jokin viesti puhuttelee meitä tai miksi reagoimme tietyllä tavalla, olemme jo askelen kauempana manipulaatiosta ja askeleen lähempänä harkittua, tietoista toimintaa.

Resursseja:

[Nettihuujaukset - Rikosuhripäivystys](#)

[Huijaamaton](#)

[Digitaalinen väkivalta on kasvava, vahingollinen ilmiö | Naisten Linja](#)

[Nasevia neuvoja tiliesi turvaamiseksi | Kyberturvallisuuskeskus](#)

[Näin suojaudut nettihuujaukselta | Kyberturvallisuuskeskus](#)