Atte Kettunen

TEKOÄLYMALLIEN HYÖDYNTÄMINEN KALASTELUHUIJAUSSIVUSTOJEN TUNNISTAMISEN TYÖKALUISSA

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2024

TIIVISTELMÄ

Kettunen, Atte
Tekoälymallien hyödyntäminen kalasteluhuijaussivustojen tunnistamisen työkaluissa
Jyväskylä: Jyväskylän yliopisto, 2024, 30 s.
Tietojärjestelmätiede, Kandidaatintutkielma
Ohjaaja: Vuorinen, Jukka

Kalasteluhuijaukset ovat yksi nykyajan vaarallisimmista kyberrikollisuuden muodoista. Kalasteluhuijaus on matalan kynnyksen kyberhyökkäys, jolla kohteelta voidaan huijata arkaluontoisia ja luottamuksellisia tietoja. Kalasteluhuijausten tunnistamiseen on kehitetty erilaisia työkaluja, joiden avulla käyttäjä voi tunnistaa mahdollisia kalasteluhuijauksia. Tässä kandidaatintutkielmassa tutkitaan tekoälymallien hyödyntämistä kalasteluhuijausten tunnistamisessa. Tutkielmassa keskitytään erityisesti kalasteluhuijausten verkkosivustojen tunnistamiseen niiden URL-osoitteen ja sisällön perusteella. Tutkimuksen tarkoituksena on arvioida tekoälymallien tuomia hyötyjä, sekä niiden käyttöön liittyviä haasteita kalasteluhuijausten tunnistamisessa. Tutkielma on toteutettu kirjallisuuskatsauksena. Kirjallisuus koostuu vertaisarvioiduista tutkielmista, kirjoista ja artikkeleista. Tutkielman perusteella tekoälymallien avulla kalasteluhuijausten tunnistamisen työkaluista voidaan tehdä käyttäjien silmissä luotettavampia. Tekoälymalleja hyödyntämällä tunnistamisen työkaluista saadaan tarkempia ja laadukkaampia. Tekoälymallien avulla voidaan myös auttaa käyttäjää ymmärtämään työkalujen toimintaa paremmin, mikä parantaa niiden käytettävyyttä. Tekoälymallien suuret laskentakustannukset, epävarma tietoturvallisuus, sekä puutteellinen kyky sopeutua erilaisiin kalasteluhuijauksiin taas ovat esimerkkejä näihin malleihin liittyvistä ongelmista. Tutkimuksessa keskustellaan myös käyttäjän roolista kalasteluhuijausten torjumisessa. On huomattava, että pelkällä teknologialla ja tunnistamisen työkaluilla kalasteluhuijauksia ei koskaan voida välttää täydellisesti.

Asiasanat: kalastelu, tekoäly, koneoppiminen, syväoppiminen, kyberturvallisuus

ABSTRACT

Kettunen, Atte
Use of artificial intelligence in phishing site detection tools
Jyväskylä: University of Jyväskylä, 2024, 30 pp.
Information Systems Science, Bachelor´s Thesis
Supervisor: Vuorinen, Jukka

Phishing attacks are one of the most dangerous forms of cybercrime. Phishing is an easily executable attack aimed at deceiving the target into revealing sensitive or confidential information. Various tools have been developed for phishing detection, so that users can identify phishing attempts more easily. This bachelor's thesis explores the utilization of artificial intelligence models in phishing detection. The thesis particularly focuses on identifying phishing websites based on their URL-address and content. The aim of this thesis is to assess the benefits and challenges related to using artificial intelligence in phishing detection. The thesis is conducted as a literature review. The literature in this thesis is comprised of peer reviewed articles and theses, gathered from various databases. Based on this literature, artificial intelligence models can improve the reliability of phishing detection tools in the eyes of the users. Artificial intelligence models can enhance the accuracy and quality of these tools. Artificial intelligence can also be used to improve the usability of phishing detection tools, by providing insights on classifications made between phishing- and legitimate websites. This in part improves the user experience and reliability of phishing detection tools. On the other hand, the thesis explores potential downsides in using artificial intelligence for this purpose. The high computational costs, uncertainty around information security and limited capability to adapt to different phishing techniques are examples of these downsides. The study also discusses the role of users in combatting phishing attacks. It is essential to note that technology and detection tools by themselves will never be able to eliminate phishing.

Keywords: phishing, artificial intelligence, machine learning, deep learning, cybersecurity

KUVIOT

SISÄLLYS

1 JOHDANTO

Kalasteluhuijaukset ovat verkkoympäristössä erittäin laajalle levinnyt ongelma, jotka aiheuttavat suuria vahinkoja niin yksityishenkilöille, kuin organisaatioille. APWG:n (Anti-Phishing Working Group) raportin (2023) mukaan vuonna 2023 havaittiin ennätyksellisesti lähes viisi miljoonaa kalasteluhyökkäyksen yritystä. Kyseinen raportti kattaa vain APWG:n yhteistyökumppaneiden, sekä verkkosivuston kautta ilmoitetut yritykset, eli todellinen määrä on paljon suurempi. APWG on kuitenkin seurannut samalla menetelmällä kalasteluhuijausten trendejä jo vuodesta 2004, joten ennätysmäärä kertoo myös yleisesti kalasteluhuijausten määrän kasvusta.

Kalasteluhuijaukset ovat kehittyneet alkukantaisista huijausyrityksistä hienovaraisiksi hyökkäyksiksi, joita ihmisen on erittäin hankala tunnistaa (Stojnic, Vatsalan & Arachchilage, 2021). Kalasteluhuijauksella tarkoitetaan kyberhyökkäystä, jossa hyökkääjä pyrkii saamaan kohteen toimimaan haluamallaan tavalla, saadakseen haltuunsa tämän henkilökohtaisia tietoja (Alkhalil, Hewage, Nawaf & Khan, 2021). Kalasteluhuijaus eroaa monista kyberhyökkäyksistä, sillä siinä hyökkäys kohdistetaan järjestelmien ja teknologioiden sijasta käyttäjään (Wu, Miller, & Little, 2006). Käyttäjä nähdään usein tietoturvaketjun heikoimpana lenkkinä (Görling, 2006; Heartfield & Loukas, 2016), minkä takia hyökkäyksiä on tuottoisaa kohdistaa heihin järjestelmäheikkouksien sijaan. Kalasteluhuijaus on yksi vaarallisimmista kyberrikollisuuden muodoista (Stojnic ym., 2021). Kalasteluhuijauksen onnistuessa hyökkäyksen toteuttaja pääsee käsiksi kohteen henkilökohtaisiin, arkaluontoisiin tai kriittisiin tietoihin, kohteeksi valitun järjestelmän mukaan. Kalasteluhuijausten vaarallisuutta lisää myös käyttäjien puutteellinen kyky torjua kyseisiä hyökkäyksiä (Desolda, Ferro, Marrella, Costabile & Catarci, 2022).

Tekoälymallit ovat tuoneet merkittäviä edistysaskeleita useilla teknologian osa-alueilla. Niiden tuomat hyödyt ulottuvat myös kalasteluhuijausten tunnistamiseen liittyviin toimintoihin. Tässä tutkielmassa tutkitaan tekoälymallien tuomia mahdollisuuksia kalasteluhuijausten tunnistamisessa. Tutkimuksessa tarkastellaan erityisesti sitä, miten tekoälymallit voivat vastata nykyisten kalastelun tunnistusmenetelmien ongelmakohtiin ja heikkouksiin.

Aiemman kirjallisuuden perusteella tekoälymallien hyödyistä kalasteluhuijausten tunnistamisessa ei ole tehty kokoavaa tutkimusta. Kirjallisuutta erilaisista tekoälymalleista kalastamisen tunnistamiseksi on varsin kattavasti, mutta tarkempaa tarkastelua tekoälyn tuomista hyödyistä käytännön työkaluissa ei ole toteutettu ainakaan tekijän parhaan tiedon mukaan. Tutkielman taustalla on myös kasvava tarve kehittää parempia ja tehokkaampia menetelmiä kalasteluhuijausten tunnistamiseksi. Tutkimuksen tavoitteena on selvittää, millaisia hyötyjä tekoälymallien avulla voidaan saavuttaa kalasteluhuijausten tunnistamisen työkaluissa. Tekoälymallien tuomien hyötyjen lisäksi tutkielmassa tarkastellaan mahdollisia haasteita tekoälyn käytössä kalasteluhuijausten tunnistamisessa.

Tutkielman tutkimusmenetelmänä on kirjallisuuskatsaus, jossa kartoitetaan aiempia tutkimuksia kalastelusta, kalastelun tunnistamisen malleista ja tekoälyn hyödyntämisestä tässä käyttötarkoituksessa. Kirjallisuuskatsauksen lähdeaineisto on kerätty lukuisista tietokannoista, kuten JYKDOK, IEEE Explore, ACM digital library, Web of Science, ResearchGate ja ScienceDirect. Tutkielma pohjautuu lähteisiin, jotka koostuvat artikkeleista, kirjallisuuskatsauksista, tutkimuksista, sekä kirjoista ja APWG:n vuoden 2023 neljännen neljänneksen kalastelun trendeistä kertovasta raportista. Tämän kirjallisuuden pohjalta on pyritty vastaamaan tutkielman tutkimuskysymyksiin: ”Miten tekoälymalleja voidaan hyödyntää kalasteluhuijaukseen käytettyjen sivustojen tunnistamisen työkaluissa?” sekä ”Mitä haasteita liittyy tekoälymallien hyödyntämiseen kalasteluhuijaukseen käytettyjen sivustojen tunnistamisen työkaluissa?”.

Tutkielman perusteella tekoälymallien avulla voidaan saavuttaa hyötyjä kalasteluhuijausten tunnistamisessa. Tekoälymallien avulla voidaan saavuttaa tarkempia ja laatutekijöiltään parempia tuloksia kalasteluhuijausten tunnistamisessa (Aljabri ym., 2022; Safi & Singh 2023; Orunsolu, Sodiya & Akinwale, 2019). Tekoälymallien avulla voidaan myös luoda käyttäjäystävällisempiä kalastelun tunnistamisen työkaluja (Schuetz ym., 2022; Bussone, Stumpf & O’Sullivan, 2015; Hadi ym., 2023). Tekoälymallien avulla pystytään siis toteuttamaan laadukkaampia, luotettavampia ja käyttäjäystävällisempiä ratkaisuja, kuin perinteisillä menetelmillä. Tekoälymallien käyttöön liittyy kuitenkin vielä tiettyjä haasteita. Näitä haasteita ovat mm. ominaisuuksien määrittelyn tarve (Saeed & Aghbari, 2023), laskentakustannukset (Basit ym., 2021; Thakur, Ali, Obaida & Kamruzzaman, 2023), sekä kyky sopeutua uudenlaisiin kalasteluhuijausten metodeihin (Thakur ym., 2023).

Tutkimuksen ensimmäisessä pääluvussa käsitellään kalastelua. Luvun alussa tarkastellaan kalastelun määritelmää, sekä rajataan käsite tämän tutkielman kontekstiin. Luvussa käsitellään kalasteluhuijauksissa käytettäviä metodeja, sekä niiden toteutuksen yleisimpiä vaiheita. Luvussa käsitellään myös kalasteluhuijauksen kohteen, eli käyttäjän roolia kalasteluhuijauksessa.

Toisessa pääluvussa tutkitaan kalasteluhuijauksen tunnistamisessa käytettyjä menetelmiä. Luvussa kerrotaan yleisimmistä malleista, joita on käytetty kalasteluhuijausten tunnistamisessa. Näiden menetelmien esittelyjen jälkeen tarkastellaan kyseisiin menetelmiin liittyviä haasteita.

Kolmannessa pääluvussa puolestaan käsitellään tekoälypohjaisia kalastelun tunnistusmenetelmiä. Luvun alussa tarkastellaan tekoälyn määritelmää ja rajataan se tämän tutkielman kontekstiin. Luvussa käsitellään tekoälymallien hyötyjä ja haasteita kalasteluhuijauksien tunnistamisen työkaluna. Luvun tarkoituksena on vastata tutkielman tutkimuskysymyksiin.

2 KALASTELU

Tässä luvussa käydään läpi kalasteluhuijausta kyberhyökkäyksenä. Luvussa tarkastellaan kalastelun määritelmää, sekä sen metodeja ja vaiheita. Luvun lopussa puhutaan käyttäjästä kalasteluhuijauksen kohteena. Tarkoituksena on avata kalasteluhuijausta käsitteenä ja kokonaisuutena.

2.1 Kalastelu käsitteenä

Kalastelu on kyberrikollisuuden muoto, jonka tavoitteena on huijata käyttäjältä henkilökohtaisia tietoja, kuten salasanoja, tai pankkitietoja. Kalastelua ei ole yksiselitteisesti määritelty terminä, sillä eri lähteet määrittelevät kalastelun hyvin vaihtelevasti. Kalastelu on jatkuvasti kehittyvä ja muuttuva kyberrikollisuuden muoto, mikä osaltaan vaikeuttaa käsitteen yksiselitteistä määrittelyä. Yhdistävänä tekijänä eri määritelmissä voidaan kuitenkin pitää kuvausta siitä, että hyökkääjän tavoitteena on huijata kalastelun kohde toimimaan hyökkääjän toivomalla tavalla (Alkhalil ym., 2021).

Yleisin keino näiden tietojen keräämiseksi on sähköpostin kautta tapahtuva huijaus, jossa hyökkääjä pyrkii manipuloimaan kohteen luovuttamaan tietojaan esittäen luotettavaa, tai tunnettua tahoa (Bhavsar, Kadlak, Sharma 2018; Nadeem ym., 2023). Huusko (2014) kertoo tutkielmassaan erilaisista kalasteluhuijauksen muodoista. Tutkielmassa listatuista menetelmistä nähdään, että käyttäjän arkaluontoista tietoa voidaan kerätä myös muilla tavoilla.

Koska keinoja tietojen keräämiseksi on useita, jotkin lähteet määrittelevät kalasteluhuijauksen tarkoittavan käyttäjän tietojen hankkimista rikollisin keinoin. Esimerkiksi Jakobsson ja Myers määrittelevät kalasteluhuijauksen kirjassaan (2006) sosiaalisen manipuloinnin muodoksi, jossa kalastelija pyrkii saamaan haltuunsa kohteen arkaluontoista, tai luottamuksellista tietoa. Tässä määritelmässä ei kiinnitetä huomiota kalastelun toteutustapaan, mutta kirjassa hyökkäyksen mainitaan useimmiten tapahtuvan sähköpostin välityksellä levitetyn huijaussivuston kautta.

Joissain julkaisuissa kalasteluhuijauksen määritelmä itsessään sisältää tarkennuksia huijauksen toteutustavasta. Esimerkiksi van der Merwe, Loock ja Dabrowski (2005) määrittelevät kalasteluhuijauksen valheellisen verkkosivuston luomisena, jotta hyökkääjä voi saavuttaa tavoitteensa. Heidän mukaansa valheellisen verkkosivuston osuus kalasteluhuijauksessa voidaan jopa nähdä kalasteluhuijauksen tärkeimpänä mahdollistajana, sekä erottavana tekijänä muihin kyberrikollisuuden muotoihin nähden (Aleroud & Zhou, 2017).

Lastdragerin kirjallisuuskatsauksessa (2014) pyrittiin tuottamaan yleispätevä määritelmä kalasteluhuijauksesta. Hänen mukaansa kalastelu voidaan aiempien määritelmien pohjalta yleisesti määritellä skaalattavana huijauksena, jossa jotain entiteettiä imitoimalla pyritään keräämään kohteen tietoja. On kuitenkin huomattava, että tämäkään määritelmä ei kata kaikkia huijauksia, jotka voitaisiin muiden määritelmien perusteella tulkita kalasteluhuijauksiksi. Esimerkiksi Abladi ja Weir (2020) mukaan kalastelua voi tapahtua myös ilman yksilön tai organisaation imitointia.

Kalasteluhuijauksen tarkkaa määrittelemistä varten olisi päästävä kollektiiviseen yhteisymmärrykseen siitä, mitkä hyökkäyksien muodot kuuluvat termin ”kalastelu” alle. Jos määritelmään ei sisällytetä huijausverkkosivustojen osuutta, voidaan kalasteluhuijauksina nähdä monenlaisia hyökkäyksiä, joissa on tavoitteena saada hyökkääjien haltuun kohteen tietoja. Tällöin tutkimuskenttä on erittäin laaja. Tässä tapauksessa kalasteluhuijauksilta puolustautuminen tarkoittaisi puolustautumista hyvin monenlaisilta hyökkäyksiltä. Käsitteen voitaisiin myös nähdä tarkoittavan erityisesti hyökkääjien luomien verkkosivustojen kautta tapahtuvaa tietojen kalastelua. Tällöin tutkimuskenttä olisi tarkempi, jolloin hyökkäyksien tutkiminen ja puolustautumiskeinojen kehittäminen voisi olla tehokkaampaa.

Tässä tutkimuksessa keskitytään erityisesti valheellisia verkkosivustoja hyödyntäviin hyökkäyksiin. Kalasteluhuijaus tarkoittaa siis tämän tutkielman kontekstissa huijausta, jossa hyökkääjä luo verkkosivuston, joka vaikuttaa luotettavan tahon alustalta, mutta onkin tarkoitettu pelkästään sisäänkirjautumis-, pankki-, tai muiden käyttäjätietojen keräämiseen. Useissa tutkielman lähteissä polku huijaussivustolle alkaa sähköpostin välityksellä välitetystä viestistä. On kuitenkin huomioitava, että kalastelusivusto voidaan jakaa mitä tahansa viestintäkanavaa hyödyntäen. Näin ollen ei ole aiheellista rajata tutkimusta pelkästään sähköpostin välityksellä tapahtuviin hyökkäyksiin.

2.2 Kalasteluhuijausten metodeja

Kuten aiemmin todettiin, kalasteluhuijausten toteuttamiseen on olemassa useita keinoja. Seuraavissa alaluvuissa tarkastellaan kalasteluhuijausten metodeja, joiden avulla kohteelta pyritään huijaamaan arkaluontoista tietoa. Listatut metodit mukailevat Nadeem ym. artikkelia (2023), sekä Huuskon tutkielmaa (2014) ja

Bhavsar ym. artikkelia (2018). Näiden artikkelien listaamista metodeista esitellään aiemmin rajattuun kontekstiin sopivat kalasteluhuijausten muodot.

2.2.1 Sähköpostikalastelu

Sähköpostikalastelulla tarkoitetaan sähköpostin välityksellä tapahtuvaa tietojen kalastelua. Nadeemin ym. artikkelin (2023) mukaan sähköpostikalastelussa hyökkääjät pyrkivät imitoimaan luotettavia organisaatioita tai yksityishenkilöitä saadakseen kohteen luottamuksen. Hyökkääjät luovat sähköpostiviestin, jossa pyritään sosiaalisen manipuloinnin avulla saada kohde toimimaan sähköpostiviestin mukaisesti (Koistinen, 2017). Toimiessaan hyökkääjien ohjeiden mukaan kohde päätyy lopulta luovuttamaan henkilökohtaisia tietojaan hyökkääjien käsiin. Usein sähköpostikalastelun päämääränä on ohjata kohde hyökkääjien luomalle kalastelusivustolle, jossa tietojen kerääminen tapahtuu (Alkhalil ym., 2021).

2.2.2 Kohdennettu kalastelu (Spear phishing / Whaling)

Kohdennettu kalastelu on vapaa suomennos, jonka voidaan tulkita tarkoittavan englanninkielisiä termejä ”spear phishing” ja ”whaling”. Kohdennetun kalastelun erottava tekijä muista kalasteluhuijauksen metodeista on sen kustomoinnin tasossa (Nadeem et al., 2023). Kohdennetussa kalastelussa kalasteluhuijaus kustomoidaan tarkempaa kohderyhmää tai yksilöä varten keräämällä etukäteen tietoa huijattavasta tahosta esimerkiksi sosiaalisen median ja internetin avulla, jotta huijauksen onnistumisen todennäköisyys kasvaa (Jagatic, Johnson, Jakobsson & Menczer 2007; Purhonen, 2023). Termit ”spear phishing” ja ”whaling” eroavat toisistaan kohteen statuksen mukaan. ”Whaling” menetelmässä kalasteluhuijaus kohdennetaan korkean statuksen ja arvon yksilöön, kun taas ”spear phishing” tarkoittaa yleisluontoisemmin kohdennettua kalasteluhuijausta kohteen statuksesta riippumatta (Bhavsar, Kadlak & Sharma, 2018)

2.2.3 Klikkaushuijaus

Huusko kertoo tutkielmassaan (2014) sosiaalisen median palveluissa tapahtuvasta klikkaushuijauksesta. Klikkaushuijaus (engl. clickjacking) on kalasteluhuijauksen muoto, jossa kohde pyritään saada klikkaamaan hyökkääjien luomaa linkkiä, mikä vie käyttäjän huijausverkkosivustolle (Huusko, 2014). Klikkaushuijauksen, tai klikkauksen kaappauksen esimerkkitapauksia ovat esimerkiksi tykkäyksen kaappaus (likejacking), sekä kursorin kaappaus (cursorjacking), joissa valheellisten verkkosivustojen avulla käyttäjä saadaan klikkaamaan hyökkääjän haitallista linkkiä (Huusko, 2014; Koskinen, 2016).

2.2.4 Kloonikalastelu

Nadeem ym. (2023) mukaan kloonikalastelussa hyökkääjät kopioivat täsmällisesti sähköpostiviestin, joka kohteelle on aikaisemmin varmuudella lähetetty. Heidän mukaansa tavoitteena on luoda vaikutelma siitä, että kyseessä on

jatkoviesti kyseiseen sähköpostiin liittyen. Klooni on parhaimmillaan niin tarkka kopio alkuperäisestä sähköpostiviestistä, että se eroaa ainoastaan sähköpostin lähettäjän- sekä klikattavien linkkien osoitteista (Nadeem ym., 2023). Kloonikalastelulla voidaan tarkoittaa myös sähköpostiviestin kopioimisen sijaan verkkosivuston kopioimista, jolloin hyökkääjä luo tarkan kopion tunnetusta verkkosivustosta ja pyrkii tätä kautta kalastelemaan kohteen tietoja (Banu & Banu, 2013).

2.2.5 Hakukonekalastelu

Hakukonekalastelussa hyödynnetään hakukoneoptimointia. Valheellinen verkkosivusto pyritään saamaan hakukoneoptimoinnin avulla hakukoneiden tuloksissa listojen kärkeen, jotta käyttäjät löytäisivät huijaussivustoille luonnollisesti, ilman erityistä sivuston levittämistä tai jakelua viestintävälineiden avulla (Nadeem ym., 2023).

2.2.6 Sosiaalisen median kalastelu

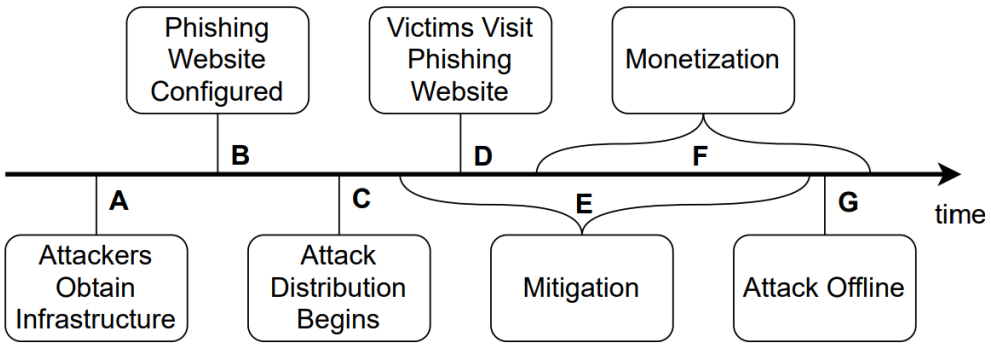
Kalastelua voidaan yrittää toteuttaa myös sosiaalisen median alustoilla. Huusko (2014) mukaan sosiaalisen median palveluita hyödynnetäänkin hyvin laajasti kalasteluhuijausten alustana. Hänen mukaansa sosiaalista mediaa hyödynnetään ihmisten tavoittamiseksi ja haitallisten linkkien levittämiseksi. Hän kertoo, että myös alustojen teknologisia haavoittuvuuksia voidaan hyödyntää tietojen kalastelussa. Käyttäjät voivat sosiaalisen median alustoilla olla jopa haavoittuvaisempia kalasteluhuijauksille, sillä he eivät välttämättä ole näillä alustoilla niin tarkkaavaisia mahdollisia kalasteluhuijauksia ajatellen (Huusko, 2014). Tätä haavoittuvuutta vahvistaa se, että sosiaalisessa mediassa hyökkääjien on helpompi kerätä tietoa kohteesta, sekä luoda kohdennettu hyökkäys tämän tiedon pohjalta (Desolda ym., 2022).

2.3 Kalasteluhuijauksen vaiheet

Kalasteluhuijauksen pääasiallisia vaiheita on yleisempien määritelmien mukaan kolme. Eri vaiheiden määrä on joissain tutkimuksissa suurempi, mutta näissä määritelmissä päävaiheet yleensä jaotellaan tarkemmiksi osavaiheiksi.

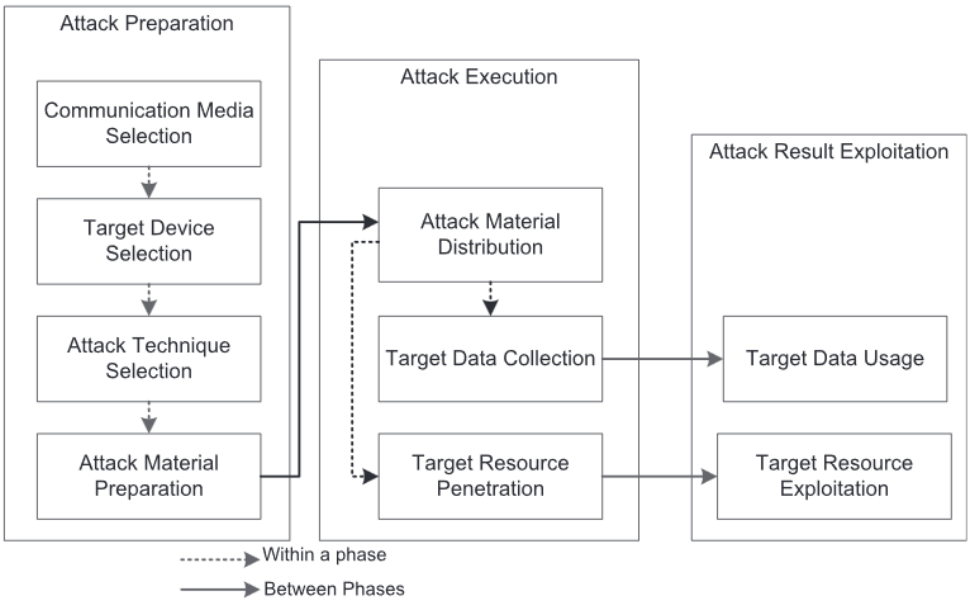
Esimerkki määritelmästä, jossa vaiheita on useampia, on Wetzelin (2005) artikkelin määritelmä kalasteluhuijauksen viidestä vaiheesta. Tässä artikkelissa vaiheiksi lasketaan hyökkäyksen suunnittelu, hyökkäyksen järjestely, hyökkäyksen toteutus, petos, sekä hyökkäyksen jälkitoimet. Tätäkin useampiin vaiheisiin kalasteluhuijauksen jakavat Jakobsson ja Myers (2006), sillä heidän mukaansa kalasteluhuijauksen vaiheita on yhteensä seitsemän. Heidän mukaansa omia vaiheitaan ovat hyökkäyksen valmistelu, tiedon tavoittelu jollain kalastelumetodilla, kohteen vasteen saaminen, arkaluontoisen tiedon kerääminen, tiedon murtaminen, kohteen imitointi, eli tiedon hyödyntäminen ja hyödyn saaminen tämän tiedon avulla. Myös Oest ym. (2020) jakavat kalasteluhuijauksen seitsemään eri

vaiheeseen (Kuvio 1). He määrittelevät kalasteluhuijauksen vaiheiksi infrastruktuurin hankkimisen, verkkosivuston luomisen, huijauksen levittämisen, kohteen tavoittamisen, hyökkäyksen lieventymisen, monetisoinnin ja hyökkäyksen loppumisen.



Kuvio 1 Kalasteluhuijauksen vaiheet 1 (Oest ym., 2020, s. 363)

Alkhalil ym. (2021) mukaan eri tutkimuksia tarkastellessa kalasteluhuijaus voidaan yhdistävien tekijöiden perusteella jakaa kolmeen päävaiheeseen. Heidän tutkimuksensa perusteella nämä päävaiheet ovat hyökkäyksen käynnistys, hyökkäyksen toteutuminen ja tulosten hyödyntäminen. Ajatusta hyökkäyksen kolmeen päävaiheeseen jakamisesta tukevat myös Aleroud ja Zhu (2017). Myös heidän mukaansa erilaisia tutkimuksia tarkastellessa löydetään yhdistäviä tekijöitä, joiden perusteella kalasteluhuijaus voidaan jakaa kolmeen merkittävimpään vaiheeseen. Nämä vaiheet ovat valmistelu, toteutus ja tulosten hyödyntäminen (Kuvio 2). Nämä päävaiheet voidaan taas jakaa sisältämään erilaiset kalasteluhuijauksen osavaiheet, kuten infrastruktuurin ja hyökkäysmetodin valmistelut, sekä hyökkäyksen levittämisen ja saatujen tietojen hyödyntämisen vaiheet (Aleroud & Zhou, 2017).



Kuvio 2 Kalasteluhuijauksen vaiheet 2 (Aleroud & Zhou, 2017, s. 163)

2.4 Käyttäjä huijauksen kohteena

Tietojärjestelmien tietoturvan heikoimpana lenkkinä pidetään yleisesti käyttäjää, eli ihmistä (Görling, 2006; Heartfield & Loukas, 2016; Desolda ym., 2022). Tietoturvan kannalta vahvimmatkin järjestelmät voidaan saada murrettua käyttäjän kautta, jos käyttäjä saadaan huijattua luovuttamaan arkaluontoista tietoa hyökkääjien käytettäväksi (Heartfield & Loukas, 2016). Desolda ym. (2022) mukaan tietoturvallinen järjestelmä yleensä olettaa ihmisen toimivan oikein tietoturvaprotokollien ja ohjeistusten mukaisesti. Kirjallisuuskatsauksen mukaan ihminen kuitenkin yleensä aiheuttaa tieturvauhkia muuten turvallisessa järjestelmässä. Tämän aiheuttavat esimerkiksi ihmisen kyvyttömyys arvioida ja tunnistaa mahdollisia tietoturvariskejä, sekä ihmisen rajallinen kapasiteetti käsitellä useita asioita samanaikaisesti (Desolda ym., 2022).

Hyökkäykset, jotka perustuvat inhimillisten heikkouksien hyödyntämiseen, kuten kalastelu, luokitellaan sosiaalisen manipuloinnin hyökkäyksiksi (Siddiqi, Pak & Siddiqi, 2022). Sosiaalinen manipulointi (engl. ”social engineering”) tarkoittaa tietoturvan kontekstissa metodeja, joilla hyökkääjä pyrkii hyödyntämään kohteen inhimillisiä heikkouksia saavuttaakseen tavoitteensa (Wang, Zhu & Sun, 2021). Sosiaalisen manipuloinnin keinoja ovat esimerkiksi samaistuttavuus, harhauttaminen, auktoriteetti, luottamuksen imitoiminen, suostuttelu ja lupaus palkinnosta tai negatiivisista seurauksista (Wang ym., 2021).

Käyttäjä on siis puutteellisten kyvykkyyksiensä, sekä inhimillisten heikkouksiensa takia kalasteluhuijauksien mahdollistaja. Kalasteluhuijausten torjumisessa onkin kiinnitettävä erityistä huomiota ihmisen avustamiseen kalasteluhuijauksien tunnistamisessa, sillä kalasteluhuijaukset eivät hyödynnä järjestelmien, vaan käyttäjien heikkouksia tavoitellessaan yksilön, tai organisaation

arvokasta tietoa. Ihmisen rooli kalasteluhuijauksen kohteena tekee kalastelusta kyberrikollisuuden muotona varsin poikkeavan teknologisia heikkouksia hyödyntäviin hyökkäyksiin nähden. Ihminen hyökkäyksen kohteena tuo kalasteluhuijaukseen monia ulottuvuuksia, joita muunlaisissa kyberhyökkäyksissä ei tarvitse huomioida. Jotta kalasteluhuijauksilta voidaan välttyä, on teknologisten näkökulmien lisäksi otettava huomioon inhimilliset näkökulmat sosiaalisen manipuloinnin takia.

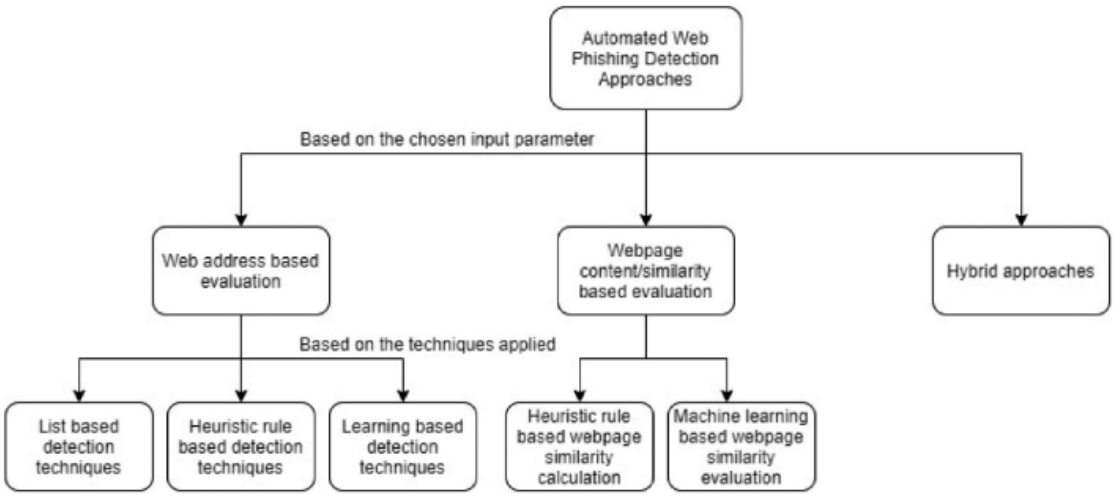
3 KALASTELUHUIJAUSTEN TUNNISTAMINEN

Tässä luvussa tarkastellaan kalasteluhuijausten tunnistamisen keinoja. Luvussa käsitellään yleisimpiä tunnistusmenetelmiä, sekä niiden käyttöä. Luvun tarkoituksena ei niinkään ole syventyä teknologioihin, vaan kalasteluhuijausten tunnistamiseen käytettävien menetelmien toimintamalleihin ja perusteisiin. Luvun lopussa tutkitaan näihin malleihin liittyviä ongelmia.

3.1 Kalastelun tunnistusmenetelmät

Vijayalakshmi, Mercy Shalinie, Yang ja Meenakshi kirjallisuuskatsauksen (2020) mukaan kalasteluhuijausten automaattiset tunnistusmenetelmät voidaan jakaa karkeasti kolmeen kategoriaan. He jakavat tunnistusmenetelmät verkkosivuston osoitetta, tai verkkosivuston sisältöä tarkasteleviin, sekä hybridimalleihin (Kuvio 3).

Vijayalakshmi ym. (2020) mukaan verkkosivuston osoitetta tarkastelevat tunnistusmallit voivat tunnistaa kalastelusivustoja kolmella erilaisella metodilla (Kuvio 3). Nämä metodit ovat listoihin perustuva tunnistaminen, heuristiset menetelmät sekä oppimisperusteinen tunnistaminen. Kirjallisuuskatsauksen perusteella verkkosivustojen sisältöä tarkastelevat tunnistusmenetelmät taas voivat tunnistaa kalasteluhuijaussivustoja kahdella menetelmällä (Kuvio 3). Nämä menetelmät puolestaan ovat heuristinen samankaltaisuuden laskelmointi, sekä koneoppimiseen perustuva samankaltaisuuden arviointi (Vijayalakshmi ym., 2020).



Kuvio 3 Kalastelun tunnistusmenetelmiä (Vijayalakshmi ym., 2020, s. 237)

3.1.1 URL-osoitteeseen perustuvat tunnistusmenetelmät

URL-osoitteen perusteella toimivat kalastelusivustojen tunnistusmenetelmät tarkastelevat verkkosivustojen URL-osoitetta luokitellessaan verkkosivustoja. URL-osoitteen perusteella määritellään, onko sivusto tunnistusmenetelmän perusteella haitallinen, vai luotettava. URL-osoitetta voidaan tarkastella vertaamalla sitä erilaisiin listoihin tunnetuista verkko-osoitteista, tai arvioimalla sitä erilaisilla heuristiikoilla. URL-osoitetta voidaan myös arvioida oppimisperusteisin tunnistusmenetelmin.

Safi ja Singh (2023) kertovat listaperustaisten tunnistusmenetelmien toiminnasta. Heidän mukaansa listaperusteiset tunnistusmenetelmät hyödyntävät joko mustia- tai valkoisia listoja. Mustat listat tarkoittavat kokoelmaa verkkosivustoista, joiden tiedetään olevan haitallisia, kun taas valkoiset listat tarkoittavat listausta tiedettävästi turvallisista verkkosivustoista (Safi & Singh, 2023). Kalasteluhuijausten tunnistamisessa yleisin tunnistusmenetelmä moderneissa selaimissa on listoihin perustuva tunnistaminen, etenkin mustia listoja hyödyntämällä (Roy, Awad, Amare, Erkihun & Anas, 2022).

Prakash, Kumar, Kompella & Gupta kehittämä PhishNet tunnistusmenetelmä (2010) on esimerkki heuristiikkaan perustuvasta tunnistusmenetelmästä. Kuten PhishNet, heuristiikkaan perustuvat tunnistusmenetelmät vertaavat tunnettujen kalastelusivustojen verkko-osoitteiden ominaisuuksia annetun verkkosivuston osoitteen ominaisuuksiin. Tällä tavalla haitallisten linkkien tunnistamista voidaan toteuttaa yleisemmällä tasolla verrattuna listapohjaisiin tunnistusmenetelmiin, sillä annetun osoitteen ei tarvitse olla etukäteen luokiteltu haitalliseksi, tai hyväksytyksi (Prakash et al., 2010).

Oppimisperusteinen verkkosivuston URL-osoitteen luokittelu hyödyntää koneoppimista, tai muita tekoälymalleja tunnistaakseen haitallisia verkkosivustoja (Safi & Singh, 2023). Tekoälypohjaisia tunnistusmenetelmiä käsitellään tämän tutkielman luvussa 4.

3.1.2 Verkkosivuston sisältöön perustuvat tunnistusmenetelmät

Verkkosivuston sisältöä tarkastelevat tunnistusmenetelmät arvioivat verkkosivuston osoitteen sijaan sen sisältöä, eli itse verkkosivuston elementtejä. Verkkosivuston sisällön perusteella sivusto voidaan luokitella joko haitalliseksi, tai luotettavaksi.

Zhang, Hong & Cranor metodi (2007) on esimerkki kalasteluhuijaussivustojen tunnistamisesta verkkosivuston sisällön perusteella. Heidän menetelmänsä perustuu sivuston elementtien ja tekstisisällön analysoimiseen, valheellisten sivustojen tunnistamiseksi. Tässä menetelmässä verkkosivuston sisällön avulla sivustoa voidaan verrata muihin, tunnetusti luotettaviin sivustoihin ja tämän perusteella määritellä, onko sivusto luotettava.

Myös verkkosivuston sisältöä tarkastelevia tunnistusmenetelmiä voidaan toteuttaa hyödyntäen oppimiseen perustuvia tekoälymalleja (Vijayalakshmi ym., 2020). Palaamme näihin malleihin luvussa 4.

3.1.3 Hybridimallit

Kalasteluhuijaussivustojen tunnistamiseksi on luotu myös hybridimalleja, jotka yhdistävät URL-osoitteen ja sisällön perusteella toimivia tunnistusmenetelmiä. Vijayalakshmi ym. (2020) mukaan hybridimallit ovat yksittäisiä tunnistusmenetelmiä hyödyntäviä malleja tehokkaampia. He kertovat hybridimallien usein hyödyntävän malleja, jotka sisältävät useita tasoja, vaiheita, tai syötteitä kalastelusivustojen tunnistamiseksi. Hybridimallien avulla voidaan saada luotettavampia tuloksia kalasteluhuijauksia tunnistettaessa, kuin pelkästään yhtä keinoa hyödyntävillä työkaluilla (Vijayalakshmi ym., 2020).

3.2 Ongelmat kalastelun tunnistamisessa

kalastelun tunnistamiseksi on luotu lukuisia työkaluja, jotka eri metodein pyrkivät tunnistamaan haitallisia verkkosivustoja. Kalasteluhuijausten tunnistamisessa käytettävien työkalujen kanssa on kuitenkin omat ongelmansa.

Divakaran & Oest (2022) mukaan kalasteluhuijauksen tunnistamiseen käytettävä työkalu ei saa antaa liikaa vääriä positiivisia, tai negatiivisia tuloksia. Heidän mukaansa käyttäjä ei siedä liikaa vääriä tuloksia, tai hän päätyy ohittamaan työkalun suosituksia. Perinteisten huijausten tunnistamiseen käytettävien menetelmien tarkkuus haitallisten sivustojen tunnistamisessa usein laskee radikaalisti, kun väärien positiivisten raja asetetaan käyttäjälle siedettävälle tasolle (Divakran & Oest, 2022). Myös Orunsolu ym. (2022) kertovat tunnistusmenetelmien väärien positiivisten määrän olevan yksi mahdollinen rajoittava tekijä kalasteluhuijausten tunnistamisen työkalujen käyttöönotossa.

Schuetz, Steelman & Syler (2022) mukaan yksi vakavista ongelmista työkalujen käytössä liittyy työkalujen ja niiden käyttäjien väliseen vuorovaikutukseen.

Heidän tutkimuksensa mukaan käyttäjät ovat taipuvaisia luottamaan enemmän omaan päättelykykyynsä, kuin kalastelusivustojen tunnistamiseen tarkoitettuihin työkaluihin. Tutkimuksen mukaan käyttäjät päätyivät sivuuttamaan työkalujen suosituksia ja varoituksia, vaikka heillä oli käytössä työkaluja, jotka pystyivät tunnistamaan haitalliset sivustot yli 90 % tarkkuudella. Ongelmat kalastelusivustojen tunnistamisessa liittyvät siis osakseen käyttäjien toimintaan, eivätkä tunnistustyökalujen tarkkuuteen (Schuetz ym., 2022). Tutkimukset osoittavatkin, että käyttäjät eivät noudata työkalujen suosituksia tarvittavalla tasolla, jotta ne voisivat suojella käyttäjiä kalasteluhuijauksilta tehokkaasti (Abbasi, Zahedi, Zeng, Chen & Nunamaker 2015; Wu, Miller, & Garfinkel, 2006).

Jotta kalasteluhuijausten tunnistamiseen tarkoitettuja työkaluja käytettäisiin tarkoituksenmukaisella tavalla, niiden laatutekijöiden, kuten väärien positiivisten määrän olisi oltava käyttäjille siedettävällä tasolla. Tämän lisäksi työkalujen olisi oltava tarpeeksi tarkkoja tunnistamaan kalasteluhuijauksia. Nämä tekijät vaikuttavat käyttäjien kokemukseen työkalun luotettavuudesta ja ovat siksi kriittisen tärkeässä roolissa kalasteluhuijauksen tunnistamisen työkalujen näkökulmasta.

4 TEKOÄLYN HYÖDYNTÄMINEN KALASTELUHUIJAUSTEN TUNNISTAMISESSA

Tässä luvussa kerrotaan tekoälymallien hyödyntämisestä kalasteluhuijausten tunnistamisessa. Tarkoituksena on tutkia, miten tekoälymallien avulla voitaisiin vastata kalasteluhuijausten tunnistamiseen liittyviin ongelmiin. Aluksi käsitellään tekoälyn määritelmää, sekä sitä, mitä tekoälyllä tarkoitetaan tämän tutkielman kontekstissa. Seuraavaksi käydään läpi erilaisia malleja tekoälyn hyödyntämiseen tunnistamisessa. Lopuksi käsitellään näiden mallien mahdollisuuksia, sekä ongelmia suhteessa huijausten tunnistamiseen ja pohditaan tutkimuskentän tulevaisuuden suuntaa.

4.1 Tekoäly käsitteenä

Sheikh, Prins, Schrijvers (2023) mukaan tekoälylle ei ole yhtä yleisesti hyväksyttyä määritelmää. Heidän mukaansa tekoälyn voi määritellä eri tavoin, laajemmin tai suppeammin. Heidän mukaansa tekoäly voidaan laajalla tasolla määritellä vain joukoksi algoritmeja. Tämä määritelmä ei kuitenkaan heidän mukaansa kuvaile tekoälyn todellista luonnetta. Heidän tarkemman määritelmänsä mukaan tekoälyllä tarkoitetaan tietokoneiden pyrkimystä jäljitellä ihmisen kaltaista älykkyyttä. Jotkin määritelmät tarkentavat imitoidun älykkyyden luonnetta vielä tarkemmin, mutta tarkemmat määritelmät myös rajoittavat tekoälyn määritelmän yleistettävyyttä (Sheikh ym., 2023). Goyanes & Durotoye (2023) taas määrittelevät tekoälyn yleisellä tasolla koneiden tai tekoälyllisten olioiden kykynä suorittaa tehtäviä, ratkaista ongelmia, kommunikoida, vuorovaikuttaa tai käyttäytyä muilla tavoilla älykkäästi ja loogisesti, kuin biologinen ihminen. Tämän tutkielman kontekstissa tekoälyllä tarkoitetaan koneoppimisen ja syväoppimisen malleja, joiden avulla suoritetaan annettuja tehtäviä. Tämä määritelmä siis käsittää vain kaksi tekoälyn muotoa, joita tässä tutkielmassa käsitellään.

4.2 Tekoälymallit kalastelun tunnistamisessa

Kalasteluhuijausten tunnistamiseen käytettyjen perinteisten metodien lisäksi verkkosivustojen URL-osoitteita, sekä sisältöä voidaan tulkita myös tekoälymallien avulla. Tekoälymallit, joita kalastelun tunnistamisessa hyödynnetään, koostuvat enimmäkseen syväoppimisen, sekä koneoppimisen malleista (Basit ym., 2021). Koneoppimisen avulla tietokoneet voivat algoritmien ja matemaattisten mallien avulla suorittaa tehtäviä ilman nimenomaista ohjelmointia (Mahesh, 2019). Mishra, Reddy & Pathak (2021) mukaan syväoppiminen on koneoppimisen muoto, jossa tekoälyä koulutetaan neuroverkkojen avulla. Näihin neuroverkkoihin lukeutuvat mm. konvoluutioverkot, takaisinkytkeytyvät neuroverkot, sekä syvä neuroverkko (Mishra ym., 2021).

Basit ym. (2021) mukaan koneoppimiseen perustuvia menetelmiä kalasteluhuijausten tunnistamiseen ovat mm. tukivektorikone, satunnaismetsä, C4.5, CART (classification and regression tree), päätöspuu, adaboost (adaptive boosting), sekä k-NN (k-nearest neighbor). Syväoppimisen malleiksi he luettelevat syvän-, eteenpäinsyöttävän- ja takaisinkytkeytyvän neuroverkon, sekä konvoluutioverkon, rajoitetun Boltzmannin koneen ja autoenkoodaajan. Metodeja, jotka hyödyntävät sekä koneoppimista, että syväoppimista kutsutaan hybridioppimisen malleiksi (Basit ym., 2021). Listattujen tekoälymallien tehokkuus kalasteluhuijausten tunnistamisessa riippuu suuresti käytetystä kalasteluhuijauksen metodista, joten näistä tekoälymalleista ei voida yksiselitteisesti määritellä parasta vaihtoehtoa (Abdillah, Shukur, Mohd & Murah, 2022).

4.3 Tekoälymallien hyödyt

Tekoälymallit ovat kokeellisten testien perusteella tarkimpia työkaluja kalasteluhuijausten tunnistamisessa (Basit ym., 2021). Tekoälymallien avulla voidaan luoda kalastelun tunnistustyökaluja, jotka osaavat luokitella haitalliset verkkosivustot jopa 99.98 % tarkkuudella (Aljabri ym., 2022; Safi & Singh 2023). Tekoälymallien avulla saadaan siis aikaan parhaat tulokset kalasteluhuijausten tunnistamisessa, kun tarkastellaan menetelmien tarkkuutta.

Hyvin tarkkoja kalastelun tunnistamisen työkaluja on kuitenkin pystytty luomaan jo aiemmin ilman tekoälymalleja. Esimerkiksi Zhang ym. (2007) esittelemä Cantina malli saavutti 97 % tarkkuuden kalastelusivustojen tunnistamisessa. Tekoälymallien tuomat edistysaskeleet kalasteluhuijausten tunnistamisessa eivät siis ole valtavan suuria työkalujen tarkkuutta ajatellen. Kalasteluhuijausten tunnistusmenetelmien laadusta kertovat kuitenkin muutkin mittarit.

Divakaran & Oest (2022) mukaan yksi tärkeistä mittareista tunnistustyökalujen arvioinnissa on väärien positiivisten ja negatiivisten luokittelujen määrä. Heidän mukaansa tunnistustyökalun FPR-arvon (väärien positiivisten määrä) on oltava 10^{-3} työkalun käyttöönoton mahdollistamiseksi, eli työkalu saa antaa väärän positiivisen tuloksen vain kerran tuhannesta. Heidän mukaansa liialliset

väärät positiiviset tulokset turvallisilla sivustoilla aiheuttavat liikaa keskeytyksiä käyttäjien toimintaan, jotta he käyttäisivät työkalua aktiivisesti. Tunnistustyökalun tarkkuus kuitenkin heikentyy FPR-arvon alentuessa (Divakaran & Oest, 2022). Tekoälymallien avulla on pystytty tuottamaan lupaavia työkaluja, joiden tarkkuus pystytään pitämään erittäin hyvänä samalla kun väärien positiivisten tulosten määrä pysyy alhaisena. Esimerkiksi Orunsolu ym. esittelemä malli (2019) saavutti 99,96 % tarkkuuden kalastelusivustojen tunnistamisessa, kun väärien positiivisten ja väärien negatiivisten tulosten osuus pysyi 0,04 prosentissa. Tekoälymallien avulla voidaan siis saavuttaa tarkempia tuloksia samalla, kun väärien positiivisten ja väärien negatiivisten tulosten määrä pysyy perinteisiin malleihin verrattuna alhaisempana.

Tunnistustyökalut ovat hyödyttömiä, jos niitä ei käytetä. Schuetz ym. (2022) mukaan käyttäjien liiallinen itsevarmuus, sekä epäluottamus työkaluihin tekee niiden käytön varmistamisesta hankalaa. Tutkimuksen mukaan käyttäjien luottamusta voidaan vahvistaa tarkemmilla tunnistustyökaluilla. Samassa tutkimuksessa havaittiin myös, että käyttäjien luottamukseen vaikuttaa työkalun läpinäkyvyys. Työkalun läpinäkyvyydellä on vaikutusta käyttäjän luottamukseen työkalua kohtaan, sekä näin ollen vaikutusta myös siihen, käyttääkö käyttäjä työkalua (Schuetz ym., 2022). Kalastelun tunnistamisen työkalujen läpinäkyvyyttä voitaisiin parantaa esimerkiksi selityksillä työkalun tekemistä luokitteluista (Bussone, Stumpf & O'Sullivan, 2015). Selityksien tuottamisessa voitaisiin hyödyntää laajoja kielimalleja, jotka voisivat luoda selityksen tekoälymalliin perustuvan työkalun tekemästä päätöksestä (Koide, Fukushi, Nakano & Chiba, 2024). Selityksien avulla käyttäjä ymmärtäisi paremmin työkalun tekemiä luokitteluja, mikä auttaisi käyttäjää luottamaan työkalun toimintaan paremmin (Schuetz ym., 2022). Tällä voitaisiin myös mahdollisesti estää työkalujen tarjoamien ohjeistusten ja suositusten ohittamista ja kyseenalaistamista.

4.4 Ongelmat tekoälymallien käytössä

Tekoälymallien hyödyntämiseen kalasteluhuijausten tunnistamisessa liittyy myös ongelmia. Ongelmat liittyvät käytettyihin teknologioihin, sekä käyttäjien ja tunnistustyökalujen väliseen vuorovaikutukseen.

Saeed & Aghbari (2023) mukaan tekoälymalleista koneoppimiseen perustuvien työkalujen ongelma on niiden vaatima ominaisuuksien määrittely. He tarkoittavat ominaisuuksien määrittelyllä työkalun vaatimaa määrittelyä siitä, mitä dataa tulisi tulkita, sekä miten kyseistä dataa tulisi tulkita. Heidän mukaansa koneoppimista hyödyntävät mallit vaativat liikaa ominaisuuksien määrittelyä, etenkin kun niitä verrataan syväoppimisen malleja hyödyntäviin työkaluihin. Syväoppimisen mallit eivät vaadi ominaisuuksien määrittelyä toimiakseen (Saeed & Aghbari, 2023).

Niin koneoppimisen, kuin syväoppimisenkin malleissa haasteeksi koituu myös niiden aiheuttamat laskentakustannukset (Basit ym., 2021; Thakur ym., 2023). Jotta tekoälymallit voisivat olla käytännöllisempiä oikeassa käytössä,

tutkijoiden olisi kehitettävä yksinkertaisempia malleja laskenta- sekä muiden kustannusten alentamiseksi (Thakur ym., 2023). Laskentakustannusten alentuminen ja tekoälyn käytön kustannukset voivat toki tippua myös ajan kuluessa ja yleisen tekoäly-, laskenta- ja prosessointiteknologian kehittyessä.

Tekoälymalleja hyödyntäessä potentiaaliseksi ongelmaksi voi myös koitua niiden kyky sopeutua uudenlaisiin kalastelumenetelmiin (Thakur ym., 2023). Mallit ja työkalut suunnitellaan tietynlaisten kalasteluhuijausten tunnistamiseen, mutta alati kehittyvänä kyberuhkana kalasteluhuijausten muodot voivat kehittyä nopeammin kuin niiden tunnistamiseen käytetyt teknologiat.

Tekoälymallien yleistyessä kalasteluhuijausten tunnistamisessa on otettava huomioon myös näiden mallien tietoturva (Thakur ym., 2023). On varmistettava, että luodut työkalut käsittelevät käyttäjien dataa ja tietoja tietoturvallisesti.

On myös muistettava, että vaikka teknologiat kalasteluhuijausten tunnistamiseksi kehittyisivät täydellisen tarkoiksi ja luotettaviksi, olemme siinä vaiheessa ratkaisseet vain osan vallitsevasta ongelmasta. Naqvi ym. (2023) mukaan teknologiapohjaiset ratkaisut eivät koskaan tule olemaan tarpeeksi kalasteluhuijausten torjumiseksi. Jotta kalasteluhuijauksilta pystytään suojautumaan, pitää huomiota kiinnittää myös käyttäjien kouluttamiseen kalasteluhuijausten tunnistamisessa ja niiden välttämisessä (Naqvi et al., 2023).

Ihmisen rooli kalasteluhuijausten kohteena tuo näiden huijausten torjumiseen erityisen ulottuvuuden. Ihmisen vuorovaikutus niin kalasteluhuijausyritysten, kuin niiden tunnistamiseen tarkoitettujen työkalujen kanssa on huomioitava aihetta tutkiessa. Schuetz ym. (2022) mukaan käyttäjien luottamus tunnistamiseen käytettävää työkalua kohtaan vaikuttaa merkittävästi siihen, miten paljon kyseistä työkalua käytetään. Tekoälymalleja hyödyntävät tunnistusmenetelmät eivät itsessään ole sen läpinäkyvämpiä, kuin muut työkalut, ellei niiden kehittämisessä ole erikseen panostettu käyttäjäystävällisyyteen. Voidaan siis jopa ajatella, ettei tunnistustyökalun käyttämällä teknologialla ole juurikaan merkitystä, jos työkalua ei ole suunniteltu tarpeeksi käyttäjäystävällisesti. Tämä tarkoittaisi, että tekoälypohjaisten ratkaisujen kehittäminen on osaltaan jopa tarpeetonta, jos niiden avulla ei voida parantaa teknologian lisäksi myös käyttäjän vuorovaikutusta tunnistamisen työkalujen kanssa.

4.5 Pohdintaa

Käyttäjän ja tunnistustyökalun välinen vuorovaikutus vaikuttaa olevan yksi keskeisimmistä tekijöistä kalasteluhuijausten tunnistamiseen kehitettyjen työkalujen toimivuutta arvioitaessa (Schuetz ym., 2022). Tähän asti tuotettu tutkimus painottuu vahvasti kalasteluhuijausten tunnistamiseen käytetyn teknologian tarkkuuden kehittämiseen. Tulevaisuudessa aiheen tutkimuksen on keskityttävä entistä enemmän työkalujen laatutekijöiden, sekä ihmisen ja teknologian vuorovaikutukseen vaikuttavien tekijöiden määrittämiseen. Kun laatutekijät saadaan määriteltyä ihmisen ja työkalun välisen vuorovaikutuksen näkökulmasta, voidaan kehittää työkaluja, joita ihmiset käyttävät paremmalla todennäköisyydellä.

Tämän vuorovaikutuksen tutkiminen voisi parantaa käytettyjen työkalujen tuomia hyötyjä kalasteluhuijausten torjumisessa. Tähän ongelmaan voitaisiinkin mahdollisesti vastata muilla tekoälyn muodoilla, kuten laajoilla kielimalleilla. Laajat kielimallit ovat tekoälyn tyyppi, joiden avulla voidaan käsitellä ja tuottaa luonnollista kieltä (Hadi ym., 2023). Laajojen kielimallien avulla voitaisiin esimerkiksi tuottaa käyttäjäystävällisiä perusteluja työkalujen tekemistä ratkaisuista (Koide ym., 2024). Perustelujen avulla pystyttäisiin parantamaan työkalun läpinäkyvyyttä, joka voisi tehdä työkaluista käyttäjän silmissä luotettavampia.

Kalasteluhuijausten tunnistamisen työkalujen kehittäminen on erittäin tärkeää, jotta käyttäjiä voidaan auttaa välttämään kyseistä kyberrikollisuuden muotoa. Tutkimus vaikuttaa kuitenkin olevan liian teknologiapainotteista. Kalasteluhuijaus hyödyntää sosiaalisen manipuloinnin keinoja, joilta käyttäjää ei voida suojata pelkästään tarkoilla työkaluilla. Tulevaisuudessa kehitettyjä tunnistusteknologioita on tarkasteltava uusista näkökulmista, jotta niistä voidaan kehittää aidosti toimivia ratkaisuja.

5 YHTEENVETO

Kalasteluhuijaukset ovat matalan kynnyksen kyberhyökkäyksiä, joilla pyritään saamaan hyökkääjien haltuun kohteen arkaluontoisia ja luottamuksellisia tietoja. Kalasteluhuijauksen määritelmä ei vielä ole yksiselitteinen, sillä sen määrittäminen on vaikeaa kalasteluhuijauksen muuttuvan ja kehittyvän luonteen takia. Kalasteluhuijauksia on helppo toteuttaa, minkä takia ne ovat hyvin suosittu kyberrikollisuuden muoto. Kalasteluhuijaus luokitellaan yhdeksi vaarallisimmista kyberrikollisuuden muodoista. Käyttäjän on usein erittäin vaikea tunnistaa hyvin toteutettua kalasteluyritystä, minkä takia kalastelun tunnistamiseksi on kehitetty erilaisia työkaluja. Näiden työkalujen on tarkoitus auttaa käyttäjää havaitsemaan kalasteluhuijauksia. Tässä tutkielmassa tarkasteltiin kalasteluhuijausta, sen tunnistamiseen käytettäviä menetelmiä, sekä tekoälyn tuomia mahdollisuuksia ja haasteita kalasteluhuijausten tunnistamisessa. Tutkimuksessa tarkasteltiin kalasteluhuijauksiin käytettyjen sivustojen tunnistamista erityisesti niiden URL-osoitteen ja sisällön perusteella. Tavoitteena oli selvittää, mitä hyötyjä tekoälystä voisi olla kalasteluhuijaussivustojen tunnistamisen työkaluissa. Tavoitteena oli myös selvittää mahdollisia haasteita tekoälymallien hyödyntämisessä. Tutkimuskysymyksiä olivat: ”Miten tekoälymalleja voidaan hyödyntää kalasteluhuijaukseen käytettyjen sivustojen tunnistamisen työkaluissa?” sekä ”Mitä haasteita liittyy tekoälymallien hyödyntämiseen kalasteluhuijaukseen käytettyjen sivustojen tunnistamisen työkaluissa?”.

Tutkimus toteutettiin kirjallisuuskatsauksena. Kirjallisuutta haettiin lukuisista eri tietokannoista, kuten JYKDOK, IEEE Explore, ACM digital library, Web of Science, Scopus, ResearchGate ja ScienceDirect. Kirjallisuus koostuu vertaisarvioiduista lähteistä, jotka ovat artikkeleita, kirjallisuuskatsauksia, kirjoja ja tutkimuksia. Lähteisiin sisältyy myös APWG:n kalasteluhuijauksien trendeistä kertova raportti.

Tutkielman perusteella voidaan todeta, että tekoälymalleilla voidaan saavuttaa hyötyjä kalasteluhuijausten tunnistamisessa. Tekoälymallien avulla tunnistamisen työkaluista saadaan laadukkaampia ja tarkempia. Laatutekijät, kuten väärien positiivisten ja negatiivisten tulosten määrä, saadaan tekoälymallien avulla pidettyä paremmalla tasolla. Tämä tarkoittaa esimerkiksi väärien

positiivisten ja negatiivisten tulosten määrien tapauksessa sitä, että määrät pysyivät pienempänä, kuin perinteisillä menetelmillä. Tekoälyn avulla voidaan kehittää kalasteluhuijausten tunnistamisen työkaluja, jotka ovat tarkkoja erottelemaan kalastelusivustot luotettavista sivustoista, mutta eivät myöskään tuota liikaa vääriä positiivisia tai negatiivisia tuloksia. Tekoälymallien avulla voidaan myös parantaa käyttäjien luottamusta kalastelun tunnistamisen työkaluihin. Kun luottamusta saadaan parannettua, käyttäjät käyttävät työkaluja enemmän ja tarkoituksenmukaisesti. Tekoälymalleja voidaan myös hyödyntää työkalujen läpinäkyvyyden parantamisessa, mikä edelleen vaikuttaa käyttäjien luottamukseen näitä työkaluja kohtaan.

Tekoälymallien käytössä kalasteluhuijausten tunnistamisessa havaittiin myös haasteita. Tekoälyn laskentakustannukset vaikeuttavat niiden laajamittaista käyttöönottoa. Tekoälyn hyödyntämiseen liittyviä ongelmia ovat lisäksi epävarmuus niiden tietoturvaan liittyvissä seikoissa sekä mahdollinen kyvyttömyys sopeutua erilaisiin kalasteluhuijausten muotoihin. Myös käyttäjien vuorovaikutus tunnistustyökalujen kanssa aiheuttaa haasteita. Kalastelu on kehittyvä matalan kynnyksen kyberrikollisuuden muoto, jota toteutetaan jatkuvasti erilaisin menetelmin. Uudet ja muuttuvat hyökkäykset voivat aiheuttaa ongelmia tekoälymallien kehittämisessä. Kuten kaikessa tekoälyä hyödyntävässä teknologiassa, on myös huomioitava tekoälyn resurssien käytöstä aiheutuvat ympäristöön liittyvät kysymykset. Tässä tutkielmassa ei kuitenkaan paneuduttu tekoälymalleihin liittyvään ympäristökeskusteluun.

Tutkimus kalasteluhuijausten tunnistamisesta on hyvin teknologiakeskeistä. Tulevaisuudessa tutkimusta on suunnattava myös muihin tekijöihin, kuin teknologiaan. Kalasteluhuijausten tunnistamisen tutkimuksessa on tulevaisuudessa kiinnitettävä huomiota erityisesti siihen, miten voitaisiin kehittää työkaluja, joita käytetään tarkoituksenmukaisesti. Hyvien työkalujen kehittämiseksi tarvitaan tutkimusta käyttäjän ja työkalujen välisestä vuorovaikutuksesta. Jatkotutkimusta olisi erityisesti hyödyllistä toteuttaa työkalujen käyttäjäystävällisyyttä ajatellen. Työkalujen on oltava tarpeeksi läpinäkyviä, luotettavia ja helppokäyttöisiä niiden laajamittaisen käyttöönoton varmistamiseksi. Tulevaisuudessa olisi myös hyödyllistä pyrkiä tarkempaan kalasteluhuijauksen määrittelyyn. Toinen vaihtoehto on erotella erilaisia kalasteluhuijausten metodeja eri termistöjen alle tutkimuskentän selkeyttämiseksi.

LÄHTEET

Abbasi, A., Zahedi, F., Zeng, D., Chen, Y., & Nunamaker, J. (2015). Enhancing Predictive Analytics for Anti-Phishing by Exploiting Website Genre Information. *Journal of Management Information Systems*, *31*, 109–157. https://doi.org/10.1080/07421222.2014.1001260

Abdillah, R., Shukur, Z., Mohd, M., & Murah, Ts. M. Z. (2022). Phishing Classification Techniques: A Systematic Literature Review. *IEEE Access*, *10*, 41574–41591. https://doi.org/10.1109/ACCESS.2022.3166474

Albladi, S. M., & Weir, G. R. S. (2020). Predicting individuals' vulnerability to social engineering in social networks. *Cybersecurity*, *3*(1), 7. https://doi.org/10.1186/s42400-020-00047-5

Aleroud, A., & Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. *COMPUTERS & SECURITY*, *68*, 160–196. https://doi.org/10.1016/j.cose.2017.04.006

Aljabri, M., Altamimi, H. S., Albelali, S. A., Al-Harbi, M., Alhuraib, H. T., Alotaibi, N. K., Alahmadi, A. A., Alhaidari, F., Mohammad, R. M. A., & Salah, K. (2022). Detecting Malicious URLs Using Machine Learning Techniques: Review and Research Directions. *IEEE Access*, *10*, 121395–121417. https://doi.org/10.1109/ACCESS.2022.3222307

Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *Frontiers in Computer Science*, *3*, 563060.

Anti-Phishing Working Group (2023). Phishing activity trends report, 4th Quarter, *APWG*

Banu, D. M. N., & Banu, S. M. (2013). A Comprehensive Study of Phishing Attacks. *International Journal of Computer Science and Information Technologies, Vol. 4*, 783-786

Basit, A., Zafar, M., Liu, X., Javed, A. R., Jalil, Z., & Kifayat, K. (2021). A comprehensive survey of AI-enabled phishing attacks detection techniques. *TELECOMMUNICATION SYSTEMS*, *76*(1), 139–154. https://doi.org/10.1007/s11235-020-00733-2

Bhavsar, V., Kadlak, A., & Sharma, S. (2018). Study on Phishing Attacks. *International Journal of Computer Applications*, *182*, 27–29. https://doi.org/10.5120/ijca2018918286

Bussone, A., Stumpf, S., & O'Sullivan, D. (2015). The Role of Explanations on Trust and Reliance in Clinical Decision Support Systems. *2015 International Conference on Healthcare Informatics,* 160-169. https://doi.org/10.1109/ICHI.2015.26

Desolda, G., Ferro, L., Marrella, A., Costabile, M., & Catarci, T. (2022). Human Factors in Phishing Attacks: A Systematic Literature Review. *ACM Computing Surveys*, *54*, 35. https://doi.org/10.1145/3469886

Divakaran, D. M., & Oest, A. (2022). Phishing Detection Leveraging Machine Learning and Deep Learning: A Review. *IEEE Security & Privacy*, *20*(5), 86–95. https://doi.org/10.1109/MSEC.2022.3175225

Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. (2024). A Scholarly Definition of Artificial Intelligence (AI): Advancing AI as a Conceptual Framework in Communication Research. *Political Communication*, *41*(2), 317–334. https://doi.org/10.1080/10584609.2023.2290497

Görling, S. (2006). THE MYTH OF USER EDUCATION. *Proceedings of the 16th Virus Bulletin International Conference*.

Hadi, M. U., Al-Tashi, Q., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M., Akhtar, N., Wu, J., & Mirjalili, S. (2023). *Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects*. https://doi.org/10.36227/techrxiv.23589741

Heartfield, R., & Loukas, G. (2016). A Taxonomy of Attacks and a Survey of Defence Mechanisms for Semantic Social Engineering Attacks. *ACM Computing Surveys*, *48*. https://doi.org/10.1145/2835375

Huusko, T. (2014). *Tietojen kalastelu sosiaalisen median palveluissa*. Jyväskylän Yliopisto

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, *50*(10), 94–100. https://doi.org/10.1145/1290958.1290968

Jakobsson, M., & Myers, S. (2006). *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft*. John Wiley & Sons.

Koide, T., Fukushi, N., Nakano, H., & Chiba, D. (2024). *ChatSpamDetector: Leveraging Large Language Models for Effective Phishing Email Detection* (arXiv:2402.18093). arXiv. https://doi.org/10.48550/arXiv.2402.18093

Koistinen, J. (2017). *SÄHKÖPOSTIN VÄLITYKSELLÄ TEHTÄVÄ TIETO-JENKALASTELU*. Jyväskylän yliopisto

Koskinen, I. (2016). *TIETOJENKALASTELUN TAVAT JA SUOJAUTUMIS-KEINOT*. Jyväskylän yliopisto.

Lastdrager, E. E. (2014). Achieving a consensual definition of phishing based on a systematic review of the literature. *Crime Science*, *3*(1), 9. https://doi.org/10.1186/s40163-014-0009-y

Mahesh, B. (2019). Machine Learning Algorithms -A Review. *International Journal of Science and Research, vol 9*, (381-386) https://doi.org/10.21275/ART20203995

Mao, J., Bian, J., Tian, W., Zhu, S., Wei, T., Li, A., & Liang, Z. (2018). Detecting Phishing Websites via Aggregation Analysis of Page Layouts. *Procedia Computer Science*, *129*, 224–230. https://doi.org/10.1016/j.procs.2018.03.053

Mishra, R. K., Reddy, G. Y. S., & Pathak, H. (2021). The Understanding of Deep Learning: A Comprehensive Review. *Mathematical Problems in Engineering*, *2021*, 1–15. https://doi.org/10.1155/2021/5548884

Nadeem, M., Zahra, S., Abbasi, M., Arshad, A., Riaz, S., & Ahmed, W. (2023). Phishing Attack, Its Detections and Prevention Techniques. *International Journal of Wireless Information Networks*, *12*, 13–25. https://doi.org/10.37591/IJWSN

Naqvi, B., Perova, K., Farooq, A., Makhdoom, I., Oyedeji, S., & Porras, J. (2023). Mitigation strategies against the phishing attacks: A systematic literature review. *Computers & Security*, *132*, 103387. https://doi.org/10.1016/j.cose.2023.103387

Oest, A., Zhang, P., Wardman, B., Nunes, E., Burgis, J., Zand, A., Thomas, K., Doupé, A., & Ahn, G.-J. (2020). Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale. *Proceedings of the 29th USENIX Conference on Security Symposium*, 361–377.

Orunsolu, A. A., Sodiya, A. S., & Akinwale, A. T. (2022). A predictive model for phishing detection. *Journal of King Saud University - Computer and Information Sciences*, *34*(2), 232–247. https://doi.org/10.1016/j.jksuci.2019.12.005

Prakash, P., Kumar, M., Kompella, R. R., & Gupta, M. (2010). PhishNet: Predictive Blacklisting to Detect Phishing Attacks. *2010 Proceedings IEEE INFOCOM*, 1–5. https://doi.org/10.1109/INFCOM.2010.5462216

Purhonen, T. (2023). *PHISHING SUSCEPTIBILITY RATE FOR MULTINATIONAL ORGANIZATIONS* (Pro gradu –tutkielma). Jyväskylän yliopisto.

Roy, S. S., Awad, A. I., Amare, L. A., Erkihun, M. T., & Anas, M. (2022). Multimodel Phishing URL Detection Using LSTM, Bidirectional LSTM, and GRU Models. *FUTURE INTERNET*, *14*(11), 340. https://doi.org/10.3390/fi14110340

Saeed, M., & Aghbari, Z. (2023). Survey on Deep Learning Approaches for Detection of Email Security Threat. *Computers, Materials & Continua*, *77*(1), 325–348. https://doi.org/10.32604/cmc.2023.036894

Safi, A., & Singh, S. (2023). A systematic literature review on phishing website detection techniques. *Journal of King Saud University - Computer and Information Sciences*, *35*(2), 590–611. https://doi.org/10.1016/j.jksuci.2023.01.004

Schuetz, S. W., Steelman, Z. R., & Syler, R. A. (2022). It's not just about accuracy: An investigation of the human factors in users' reliance on anti-phishing

tools. *Decision Support Systems*, *163*, 113846. https://doi.org/10.1016/j.dss.2022.113846

Sheikh, H., Prins, C., & Schrijvers, E. (2023). Artificial Intelligence: Definition and Background. In H. Sheikh, C. Prins, & E. Schrijvers (Eds.), *Mission AI: The New System Technology* (pp. 15–41). Springer International Publishing. https://doi.org/10.1007/978-3-031-21448-6_2

Stojnic, T., Vatsalan, D., & Arachchilage, N. A. G. (2021). Phishing email strategies: Understanding cybercriminals' strategies of crafting phishing emails. *SECURITY AND PRIVACY*, *4*(5), e165. https://doi.org/10.1002/spy2.165

Thakur, K., Ali, M. L., Obaidat, M. A., & Kamruzzaman, A. (2023). A Systematic Review on Deep-Learning-Based Phishing Email Detection. *Electronics*, *12*(21), Article 21. https://doi.org/10.3390/electronics12214545

van der Merwe, A., Loock, M., & Dabrowski, M. (2005). Characteristics and responsibilities involved in a Phishing attack. *Proceedings of the 4th International Symposium on Information and Communication Technologies*, 249–254.

Vijayalakshmi, M., Mercy Shalinie, S., Yang, M. H., & U., R. M. (2020). Web phishing detection techniques: A survey on the state-of-the-art, taxonomy and future directions. *IET Networks*, *9*(5), 235–246. https://doi.org/10.1049/iet-net.2020.0078

Wang, Z., Zhu, H., & Sun, L. (2021). Social Engineering in Cybersecurity: Effect Mechanisms, Human Vulnerabilities and Attack Methods. *IEEE Access*, *9*, 11895–11910. https://doi.org/10.1109/ACCESS.2021.3051633

Wetzel, R. (2005). Tackling phishing. *Business Communications Review*, *35*(2), 46-51.

Wu, M., Miller, R. C., & Little, G. (2006). Web wallet: Preventing phishing attacks by revealing user intentions. *Proceedings of the Second Symposium on Usable Privacy and Security - SOUPS '06*, 102. https://doi.org/10.1145/1143120.1143133

Zhang, Y., Hong, J. I., & Cranor, L. F. (2007). Cantina: A content-based approach to detecting phishing web sites. *Proceedings of the 16th International Conference on World Wide Web*, 639–648. https://doi.org/10.1145/1242572.1242659