

Esiselvitys tekoälyn hyödyntämisestä vesihuollossa

Vesilaitosyhdistyksen monistesarja nro 96

Helsinki 2024

Julkaisun jakelu:

Vesilaitosyhdistys
Ratamestarinkatu 7 B
00520 Helsinki

puh. (09) 868 9010
sähköposti: vvy@vvy.fi
kotisivu www.vvy.fi

ISSN-L 2242-7279
ISSN 2954-2014

ISBN 978-952-7545-19-5

Helsinki 2024

KUVAILULEHTI			
<i>Julkaisija</i>	Suomen Vesilaitosyhdistys ry		
<i>Tekijät</i>	Lepistö, Joonas; Rouhiainen, Anna; Pulkkinen, Jaana		
<i>Julkaisun nimi</i>	Esiselvitys tekoälyn hyödyntämisestä vesihuollossa		
<i>Julkaisusarjan nimi ja numero</i>	Vesilaitosyhdistyksen monistesarja nro 96		
<i>Julkaisun teema</i>			
<i>Saatavuus</i>	Julkaisu on saatavissa Vesilaitosyhdistyksen verkkosivuilta.		
<i>Tiivistelmä</i>	Tekoäly tai AI (Artificial Intelligence) on tietotekniikan ala, jolla pyritään luomaan koneita ja ohjelmistoja suorittamaan ihmismäistä älykkyyttä vaativia tehtäviä. Ratkaisuja etsitään muun muassa ongelmanratkaisuun, oppimiseen, suunnitteluun sekä kielen ymmärtämiseen ja tunnistamiseen. Tekoälyn avulla tavoitellaan koneiden kykyä suorittaa monimutkaisia toimintoja itsenäisesti ja jopa tehdä päätöksiä perustuen oppimaansa tietoon. Tekoälyratkaisu tulisi aina valita olemassa olevan tarpeen tai ongelman mukaisesti. Käytettävä lähtödata on ratkaisevassa roolissa, sillä tekoälyn tuottamat tulokset ovat aina niin laadukkaita kuin mitä lähtödata on. Tekoälymallin kouluttaminen on aikaa vievä prosessi, joka vaatii laadukkaan lähtödatan lisäksi substanssiosaamista sekä käsiteltävän ongelman että tekoälyn osalta. Tieto- ja kyberturvaan on myös kiinnitettävä huomiota, jotta prosessi pysyy luotettavana. Tähän tekoälyn hyödyntäminen vesihuollossa -julkaisuun on kerätty yleistietoa tekoälystä ja sen eri osa-alueista, sekä esitelly vesihuoltoalalla toteutettuja tekoälyratkaisuja esimerkiksi konkreettisina case-esimerkkeinä. Julkaisun laadinnassa on hyödynnetty kirjallisuuslähteitä, haastatteluja sekä työpajoja. Tekoälyratkaisut kehittyvät jatkuvasti, joten tämän julkaisun esittämät asiat tulevat päivittymään uusien tutkimustulosten ja teknologisten edistysaskelten myötä. Vesihuoltoalalla tekoälyä hyödynnetään tällä hetkellä lähinnä poikkeustilanteiden havainnointiin sekä ennusteiden laatimiseen. Tekoälyratkaisuiden potentiaalinen hyödyntäminen vaatii substanssiosaamista sekä vesihuollon että tekoälyratkaisuiden osalta.		
<i>Avainsanat</i>	AI, koneoppiminen, tekoäly		
<i>Rahoittaja/toimeksiantaja</i>	Suomen Vesilaitosyhdistys ry		
	ISBN 978-952-7545-19-5		ISSN 2954-2014
	<i>Sivuja</i> 48	<i>Kieli</i> suomi	<i>luottamuksellisuus</i> julkinen
<i>Julkaisun jakelu</i>	Vesilaitosyhdistys, www.vvy.fi		
	Tekijät vastaavat julkaisun sisällöstä eikä julkaisun sisältöä voida tulkita Vesilaitosyhdistyksen kannanotoksi.		

BESKRIVNINGSBLAG			
<i>Publicerat av</i>	Finlands Vattenverksförening r.f.		
<i>Författare</i>	Lepistö, Joonas; Rouhiainen, Anna; Pulkkinen, Jaana		
<i>Publikationens titel</i>	Förstudie om utnyttjande av artificiell intelligens inom vattenförsörjningen		
<i>Publikationsseriens titel och nummer</i>	Vattenverksföreningens duplikatserie nr 96		
<i>Publikationens tema</i>			
<i>Tillgänglighet</i>	Publikationen finns på Vattenverksföreningens webbsida.		
<i>Sammanfattning</i>	Artificiell intelligens eller AI (Artificial Intelligence) är ett område inom informationsteknik som syftar till att skapa lösningar för att utföra uppgifter som kräver mänsklig intelligens. Man söker lösningar för bland annat problemlösning, lärande, planering och språkförståelse och identifiering. Med hjälp av artificiell intelligens strävar man till att maskinerna kan utföra komplicerade funktioner självständigt och till och med fatta beslut utifrån den kunskap de lärt sig. En AI-lösning bör alltid väljas utifrån ett befintligt behov eller problem. Utgångsdatan som används spelar en avgörande roll, eftersom resultaten som genereras av AI, bestäms av kvaliteten på utgångsdatan. Skolning av AI-modellen är en tidskrävande process som utöver högklassiga utgångsdata kräver även substanskompetens både i fråga om det problem som behandlas och i fråga om AI. Informations- och cybersäkerhet måste också uppmärksammas för att hålla processen pålitlig. I denna publikation om utnyttjande av AI inom vattenförsörjningen, har man samlat allmän information om AI och dess olika delområden, samt presenterat AI-lösningar som förverkligats inom vattenförsörjningsbranschen, till exempel som konkreta case-exempel. Vid utarbetandet av publikationen har man utnyttjat litteraturkällor, intervjuer och workshoppar. AI-lösningarna utvecklas kontinuerligt, så de saker som denna publikation presenterar kommer att uppdateras i och med nya forskningsresultat och tekniska framsteg. Inom vattenförsörjningsbranschen utnyttjas artificiell intelligens för närvarande främst för observation av undantagstillstånd samt för att göra prognoser. Utnyttjandet av potentialen i AI-lösningar, kräver substanskompetens både vad gäller vattenförsörjning och AI-lösningar.		
<i>Nyckelord</i>	AI, maskininlärning, artificiell intelligens		
<i>Finansiär/uppdragsgivare</i>	Finlands Vattenverksförening r.f.		
	ISBN 978-952-7545-19-5		ISSN 2954-2014
	Sidantal 48	Språk finska	Konfidentialitet offentlig
<i>Distribution av publikationen</i>	Vattenverksföreningen, www.vvy.fi		
	Författarna är ensamt ansvariga för rapportens innehåll, varför detta ej kan åberopas såsom representerande Vattenverksföreningens ståndpunkt.		

Sisällysluettelo

1	Johdanto	2
1.1	Mitä on tekoäly?	4
1.2	Tekoälyn historia	5
1.3	Tekoälyn hyödyntäminen organisaatioissa	6
2	Data	8
2.1	Laatu ja määrä	8
	CASE – Laadukkaan datan merkitys ja sen hyödyntäminen liiketoiminnassa	9
2.2	Luokittelu	10
	CASE – Tekoälyn hyödyntäminen kasvihuonekaasupäästöjen vähentämisessä	11
2.3	Massadata (<i>Big Data</i>)	12
2.4	Algoritmit	13
	CASE – Poikkeamien havaitseminen ja ennustaminen datan avulla	14
	CASE – Ennusteita ihmisten hyödynnettäväksi	15
	CASE – Tekoälyn rooli vesihuollon tehostamisessa	16
	CASE – Koneoppivien mallien hyödyntäminen viemäriverkon saneerausajankohdan ennustamisessa	17
	Tärkeimmät nostot: Data	18
3	Koneoppiminen	19
	CASE – Tekoäly jätevedenpuhdistuksen tehostajana	21
	CASE – Algoritmit ja koneoppimismalli vesihuoltolaitoksen apuna	22
3.1	Neuroverkot ja syväoppiminen	23
3.2	Tärkeimmät nostot: Koneoppiminen	25
	CASE – Vesijohtoverkoston vuotojen havaitseminen Lontoossa	26
4	Luonnollisen kielen käsittely (NLP)	27
4.1	Suuret kielimallit (Large Language Models)	28
4.2	Arviointi ja validointi	31
	CASE – Tekoälyn hyödyntäminen operatiivisessa tehokkuudessa	32
	CASE – Asiakaspalvelun työtehtävien tehostaminen	33
4.3	Tärkeimmät nostot: Luonnollisen kielen käsittely (NLP)	34
5	Energiankulutus ja kestävyys	35
5.1	Tärkeimmät nostot: Energiankulutus ja kestävyys	36
6	Turvallisuus	37
6.1	Etiikka	38
6.2	Lainsäädäntö	38
6.3	Tietosuoja	39
6.4	Tietoturva	40
6.5	Kyberturva	41
	Tärkeimmät nostot: Turvallisuus	42
7	Yhteenveto: Tekoäly vesihuollossa	43
7.1	Tulevaisuus	45
	Liite 1 – AI-sanasto	46

SANASTO

Suomeksi	Englanniksi	Lyhenne	Kuvaus
Algoritmi	Algorithm		Askel askeleelta etenevä laskentamenetelmä
Alisovittaminen	Underfitting		Malli ei sovi tarpeeksi hyvin koulutusdataan
Hyperparametri	Hyperparameter		Parametri, joka asetetaan ennen koulutusta
Koneoppiminen	Machine Learning	ML	Algoritmit, jotka oppivat datasta
Koulutus	Training		Algoritmeja opetetaan tunnistamaan ja ennustamaan tiettyjä malleja tai käyttäytymistä syöttämällä niille suuria määriä dataa
Koulutusdata	Training Data		Data, jota käytetään mallin kouluttamiseen
Luokittelu	Classification		Datan luokittelu ennalta määriteltuihin luokkiin
Luonnollisen kielen käsittely	Natural Language Processing	NLP	Tietokoneet ymmärtävät ihmiskieltä
Lähtödata	Input Data		Tiedot, joita käytetään mallin kouluttamiseen ja testaamiseen
Malli	Model		Matemaattinen esitys prosessista
Neuroverkko	Neural Network	NN	Laskennallinen malli, joka jäljittelee ihmisaivoja
Neuroverkko-regressio	Neural Network Regression		Koneoppimisen menetelmä, jossa käytetään neuroverkkoja ennustamaan jatkuvia arvoja
Ohjaamaton oppiminen	Unsupervised Learning		Oppiminen ilman merkittäviä tietoja
Ohjattu oppiminen	Supervised Learning		Oppiminen merkittävien tietojen avulla
Ohjelmointirajapinta	Application Programming Interface	API	Ohjelmointirajapinta, jota käyttäen sovellukset kommunikoivat keskenään
Piirreteknikka	Feature Engineering		Piirteiden luominen raakadatan pohjalta
Ristiinvalidointi	Cross-validation		Tekniikka mallin suorituskyvyn arviointiin
Ryhmittely	Clustering		Datan ryhmittely samankaltaisuuden perusteella
Suositusalgoritmi	Recommendation Algorithm		Analysoi dataa tarjotakseen kohdennettuja suosituksia
Syväoppiminen	Deep Learning	DL	Edistynyt koneoppiminen neuroverkoilla
Tekoäly	Artificial Intelligence	AI	Simuloi ihmisen älykkyyttä
Testidata	Test Data		Data, jota käytetään mallin testaamiseen
Vahvistusoppiminen	Reinforcement Learning	RL	Oppiminen kokeilemalla ja palautteen avulla
Validointidata	Validation Data		Data, jota käytetään mallin validointiin
Ylisovittaminen	Overfitting		Malli sopii liian hyvin koulutusdataan

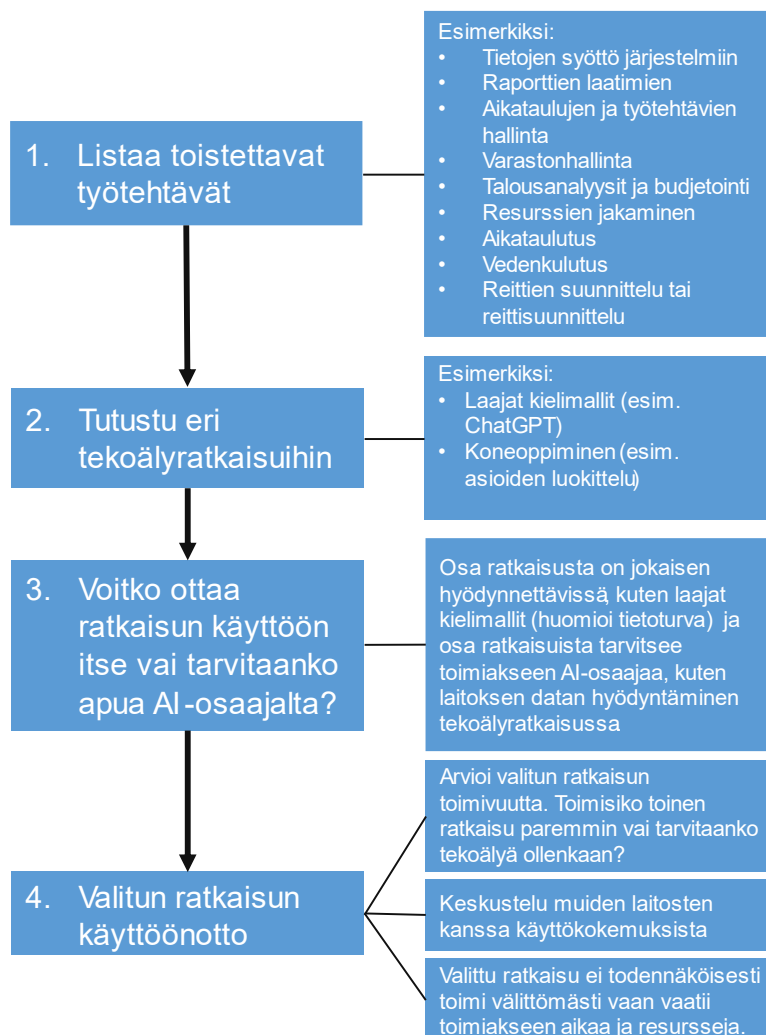
Liitteessä 1 on esitetty lisää AI-sanastoa.

1 JOHDANTO

Tekoälyn käyttö on yleistynyt ja monipuolistunut viime vuosina tarjoten lukuisia hyötyjä ja mahdollisuuksia eri aloilla. Tekoälyä hyödynnetään esimerkiksi asiakaspalvelun chatboteissa, etsittäessä karttasovelluksen avulla nopeinta kulkureittiä kohteesta toiseen, poikkeamia havaitsevista algoritmeista sekä laadittaessa tekstiä ChatGPT:llä. Tekoäly on kovassa nosteessa ja Microsoftin vuonna 2024 teettämässä tutkimuksessa Suomi listattiin tekoälyn käyttöönottamisessa edelläkävijäksi verrattuna muihin Pohjoismaihin¹. Tekoälyn hyödyntäminen edellyttää aina myös tietoturvanäkökulmien huomiointia sekä riskien arviointia.

Tekoälyä voisi verrata työkalupakkiin, josta löytyy erilaisia työkaluja, eli tekoälyratkaisuja. Tekoälyratkaisun valitsijan tulee tunnistaa nämä erilaiset ratkaisut sekä ymmärtää käyttökohdetta ja tarvetta, jotta osaa valita sopivimman työkalun. Usein tämä tarkoittaa vesihuoltoalalla sitä, että tarvitaan kaksi osaajaa: vesihuollon ja informaatioteknologian, tai tarkemmin tekoälyratkaisujen, substanssiosaajat. Kuvassa 1 on esitetty vaiheita tekoälyn hyödyntämiseen.

Tekoälyn tämänhetkisen hyödyntämispotentiaalin ymmärtääkseen on ensin tiedettävä, miten tekoäly toimii ja mitä tekoälyllä tarkalleen tarkoitetaan. Julkaisun tavoitteena on selkeyttää, millaisia työkaluja tekoäly nykyään tarjoaa. Julkaisussa esitellään käytännön esimerkkien avulla tekoälysovelluksia hyödynnettäviä ratkaisuja ja jokaisen pääkappaleen loppuun on koottu tärkeimmät nostot kyseisestä teemasta. Julkaisuun on kerätty case-esimerkkejä vesihuoltolaitoksilta, energia-alalta sekä vesihuollon palveluntarjoajilta. Julkaisussa käydään myös läpi, mitä näiden ratkaisujen käyttö edellyttää esimerkiksi datan laadun osalta. Lähtödatalla onkin suuri painoarvo, sillä tekoälyllä aikaansaatua tulos on aina juuri niin laadukas, kuin siihen syötetty lähtödata.



Kuva 1. Miten aloittaa tekoälyn hyödyntäminen.

¹ Isosomppi, P. 2024. Tutkimus: Suomi Pohjoismaiden kärjessä tekoälyn käyttöönotossa. Luettavissa: <https://news.microsoft.com/fi-fi/2024/04/16/tutkimus-suomi-pohjoismaiden-karjessa-tekoalyn-kayttoonotossa/>. Viitattu 30.7.2024.

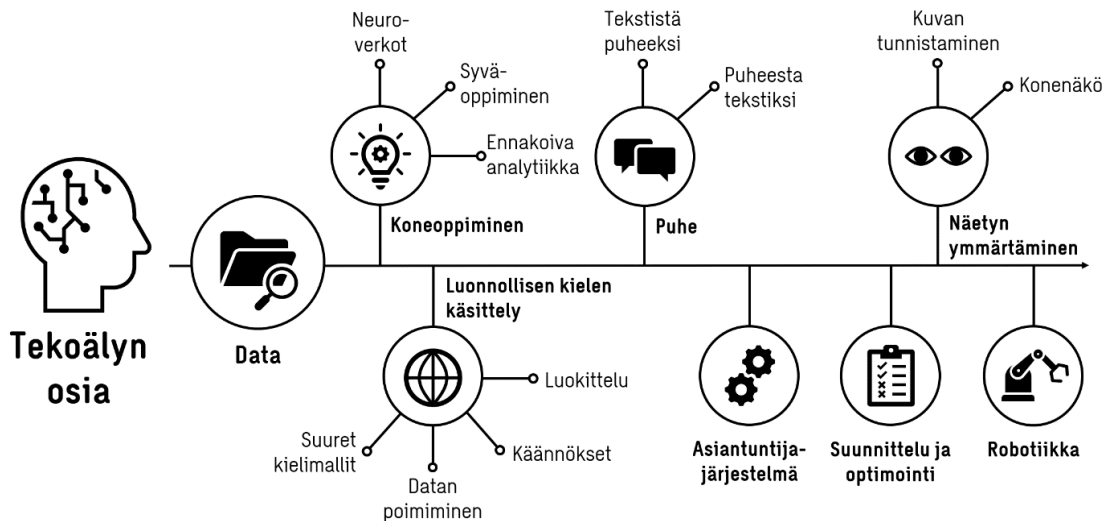
Laadukkaan ja kattavan lähtödatan lisäksi tekoäly vaatii energiaa toimiakseen. Esimerkiksi laaja kielimalli (ChatGPT) kuluttaa kysymykseen vastatessaan energiaa moninkertaisesti normaaliin hakukonehakuun verrattuna. Kestävä tekoälyratkaisujen käyttö edellyttää myös kestäväää energiantuotantoa. Tekoälymallin koulutus on aikaa vievä prosessi, joka kuluttaa energian lisäksi henkilöstöresursseja.

Tekoäly ei ole yleisvastaus kaikkiin haasteisiin, mutta väsymättömänä ja tarkkaavaisena puurtajana tarjoaa loistavat apukädet. Ihmisten luovuus ja kyky ymmärtää kokonaisuuksia ovat vielä toistaiseksi lyömättömiä. Tekoälyllä saatuihin vastauksiin ja tuloksiin on syytä suhtautua kriittisesti, sillä vastuu säilyy ihmisillä.

Julkaisun on laatinut Sweco Finland Oy Suomen Vesilaitosyhdistys ry:n kehittämisrahaston ja vesihuoltolaitoksista koostuva ohjausryhmän toimeksiannosta. Oman arvokkaan panoksensa julkaisuun ovat antaneet Aalto yliopiston työelämäprofessori Markus Sunela sekä Traficomin asiantuntijat Kimmo Saarni, Jukka Isosaari ja Aleksi Blomqvist. Hankkeen ohjausryhmään kuuluivat: Jarno Laine HS-Vesi Oy, Jani Väkevä ja Joonas Leinonen Kymen Vesi Oy, Kimmo Viinikka Kuusamon EVO, Sauli Pihamaa ja Janne Mäki-Petäjä Lahti Aqua Oy, Teemu Vuorma Neve Oy sekä Saijariina Toivikko Suomen Vesilaitosyhdistys ry.

1.1 MITÄ ON TEKOÄLY?

Tekoäly (kuva 2) tai AI (*Artificial Intelligence*), on tietotekniikan ala, jolla pyritään luomaan koneita ja ohjelmistoja suorittamaan ihmismäistä älykkyyttä vaativia tehtäviä. Ratkaisuja pyritään löytämään muun muassa ongelmanratkaisuun, oppimiseen, suunnitteluun sekä kielen ymmärtämiseen ja tunnistamiseen. Tekoälyn avulla tavoitellaan koneiden kykyä suorittaa monimutkaisia toimintoja itsenäisesti ja jopa tehdä päätöksiä perustuen oppimaansa tietoon.²



Kuva 2: Tekoälyn osia. (Mukaillen Execution Edge³ N.D.) Tämän julkaisun case-esimerkeissä on käytetty kuvan symboleita havainnollistamaan sitä, mihin tekoälyn osa-alueeseen case-esimerkki voidaan sijoittaa.

Tekoäly voidaan jakaa karkeasti kahteen päätyyppiin: kapea ja yleinen tekoäly. Kapea tekoäly, joka tunnetaan myös nimellä sovellettu tekoäly, on suunniteltu suorittamaan tiettyjä tehtäviä ja se on yleisin tekoälyn muoto. Esimerkkejä kapeasta tekoälystä ovat chatbotit, suositusalgoritmit ja puheentunnistusjärjestelmät. Yleinen tekoäly, joka on vielä suurelta osin teoreettinen käsite, viittaa älykkyyteen, joka pystyy suorittamaan minkä tahansa kognitiivisen tehtävän samalla tavalla kuin ihminen.²

Tekoäly on vain yksi prosessin osa, eikä se ole ratkaisu kaikkiin ongelmiin. Ennen kuin ryhdytään hyödyntämään tekoälyä, on ensisijaisen tärkeää varmistaa lähtötietojen olevan kunnossa ja että on selvitetty, millaisia toimintoja halutaan helpottaa. On myös tärkeää määrittää, missä prosessin vaiheessa tekoälyä käytetään, sekä mikä on sen todellinen tarve ja tarkoitus. Selkeä suunnittelu ja ymmärrys tavoitteista auttavat maksimoimaan tekoälyn hyödyt ja varmistamaan, että saatuja tuloksia voidaan hyödyntää tehokkaasti ja tarkoituksenmukaisesti.

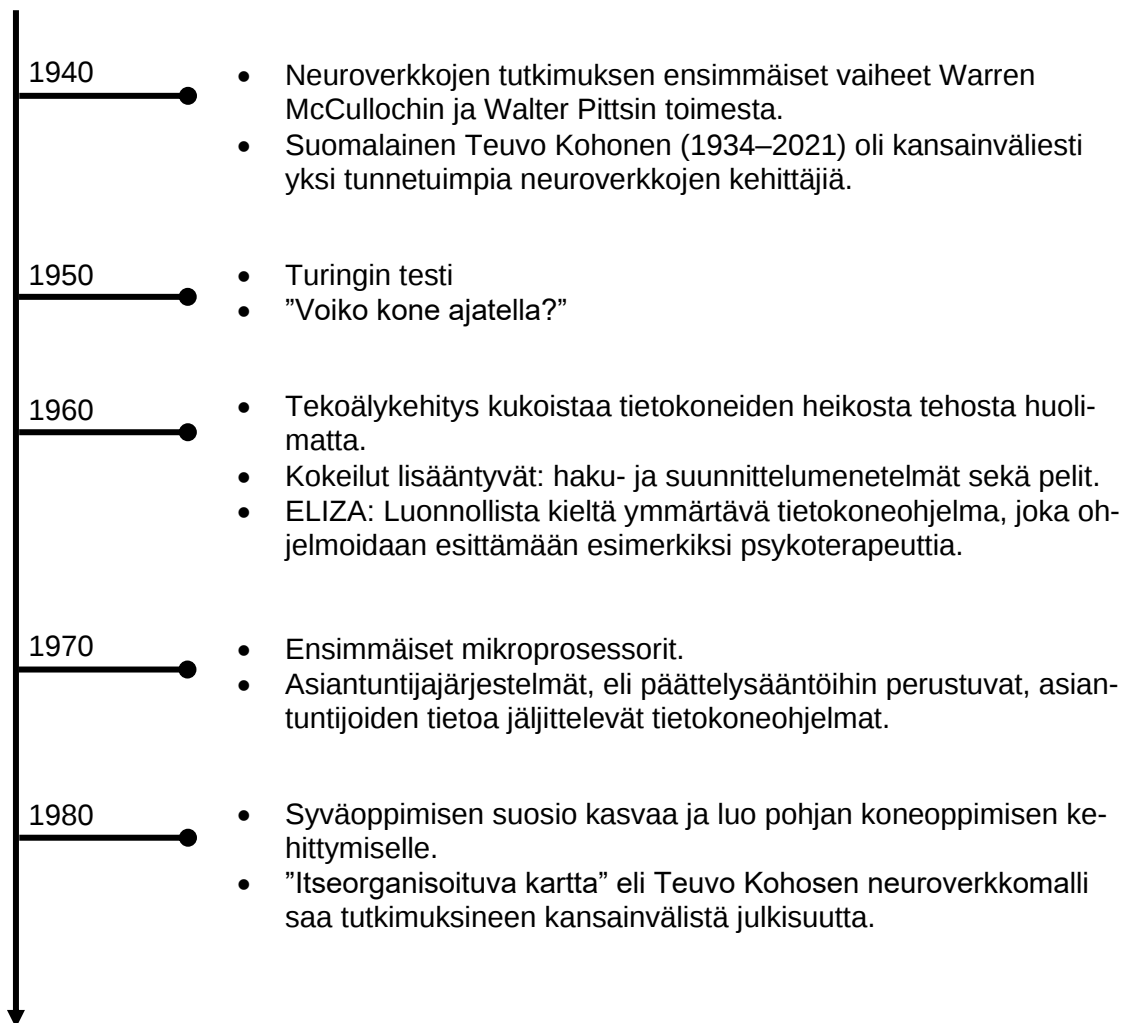
² Think Tank Mandag Morgen. 2023. Are Nordic organisations ready for AI? The effect of AI on how organisations think about their digital transition and skills. PDF-artikkeli. Viitattu 30.7.2024.

³ Execution Edge. N.D. Artificial Intelligence and the Future of Work. Verkkoartikkeli. Luettavissa: <https://www.executionedge.net/artificial-intelligence-and-the-future-of-work/>. Viitattu 21.8.2024.

1.2 TEKOÄLYN HISTORIA

Tekoälyn juuret löytyvät 1940-luvulta, kun ensimmäiset tietokoneet kehitettiin. Tekoälyn tutkimus alkoi virallisesti vuonna 1956, mutta sen kehitys on kohdannut monia haasteita ja rahoituksen vaihteluita. 2000-luvulla tekoäly koki uuden nousun, kun koneoppiminen alkoi tuottaa käytännön sovelluksia eri aloilla. Tämä jatkuva kehitys on muokannut tekoälyä sellaiseksi, kuin me sen tänään tunnemme.⁴

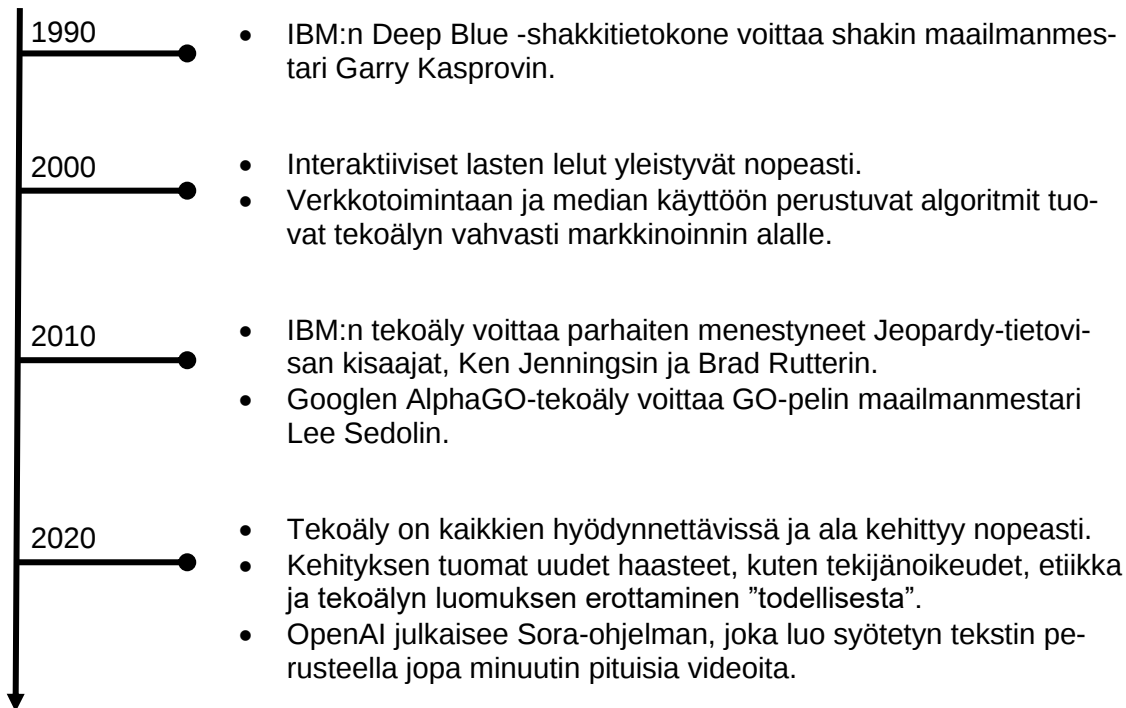
Tekoälyn modernin historian aikajana ^{5, 6}



⁴ Skycode Oy. N.D. Tekoälyn historia. Tekoäly.info. Verkkoartikkeli. Luettavissa: https://xn--tekoly-eua.info/tekoaly_historia/. Viitattu: 20.8.2024.

⁵ Hartikainen, M. 2022. Tekoälyn historia. Tampereen yliopisto. Verkkoartikkeli. Luettavissa: <https://projects.tuni.fi/kite/tekoalysta-yleisesti/tekoalyn-historia/>. Viitattu: 24.6.2024.

⁶ Wikipedia. 2024. Timeline of artificial intelligence. Verkkoartikkeli. Luettavissa: https://en.wikipedia.org/wiki/Timeline_of_artificial_intelligence. Viitattu: 24.6.2024.



Tekoälyn tulevaisuus näyttää haasteineen ja mahdollisuuksineen lupaavalta. Koneoppiminen ja neuroverkot jatkavat kehittymistään, ja ne tulevat yhä integroitumaan syvemmin päivittäiseen elämäämme. Tekoäly on jo nyt muuttamassa monia aloja, kuten terveydenhuoltoa, kuljetusta, koulutusta ja viihdettä. Sen uskotaan myös olevan avain monien suurten yhteiskunnallisten ongelmien, kuten ilmastonmuutoksen ja sairauksien, ratkaisemiseen. Tekoälyn tulevaisuus ei kuitenkaan ole vain teknologista edistystä. Se herättää myös kysymyksiä eettisistä ja yhteiskunnallisista vaikutuksista, kuten työpaikoihin, tietosuojaan ja yksityisyyteen liittyen. Tekoäly on inhimillisen älykkyyden ja luovuuden tuote, ja sen tulevaisuus on ihmisten käsissä. Vain aika näyttää, mitä uusia innovaatioita, haasteita ja mahdollisuuksia tekoälyn matka tuo mukanaan.

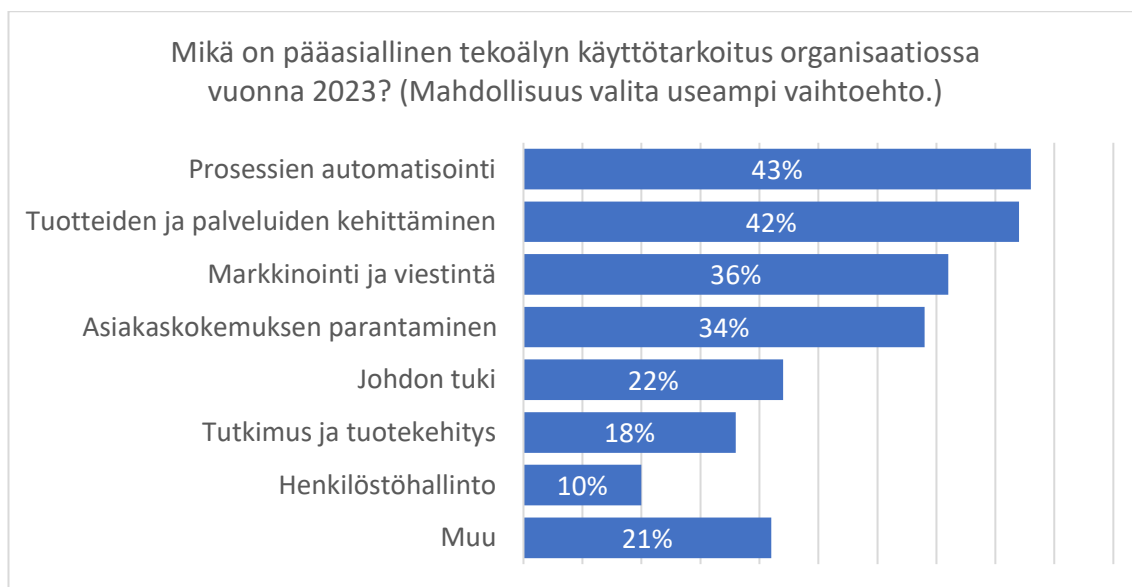
1.3 TEKÖÄLYN HYÖDYNTÄMINEN ORGANISAATIOISSA

Vuonna 2023 toteutetussa "Are Nordic organisations ready for AI?" -kyselytutkimuksessa haastateltiin 1211 pohjoismaalaista organisaatiota tekoälyn hyödyntämisestä, joista tuloksien mukaan puolet hyödyntävät tekoälyä. Vastaajina oli sekä yksityisen että julkisen sektorin organisaatioita. Mitä suurempi organisaatio on, sitä todennäköisemmin tekoälyä hyödynnetään suunnitelmallisesti. Syyt, miksi tekoälyä ei hyödynnetä, olivat useimmiten "epäselvä tai epävarma potentiaali" ja "ei johdon prioriteetti". Tämä osoittaa monilla organisaatioilla olevan epävarmuutta tekoälyn mahdollisuuksista ja sen tuomista hyödyistä.⁷

40 % vastaajista kertoi heidän organisaatiossaan olevan puutteita digitaadoissa. Tämä voi olla merkittävä este tekoälyn hyödyntämiselle, sillä perustason osaamisen puutteet nostavat kynnystä ottaa käyttöön uusia teknologioita. Vesihuoltolaitoksille tämä tarkoittaa, että panostamalla laadukkaaseen digikoulutukseen voidaan helpottaa arjen työkuormaa ja parantaa tehokkuutta sekä talouden hallintaa.⁷

⁷ Think Tank Mandag Morgen. 2023. Are Nordic organisations ready for AI? The effect of AI on how organisations think about their digital transition and skills. PDF-artikkeli. Viitattu 30.7.2024.

Suomalaiset vastaajat suhtautuvat tekoälyn tulevaisuuden mahdollisuuksiin hyvin myönteisesti. 70 % suomalaisista vastaajista uskoo, että tekoälyä tullaan hyödyntämään enemmän tulevan kolmen vuoden aikana. Kehitysmyönteisyys luo organisaatiolle hyvän pohjan lähteä kehittämään tekoälyn hyödyntämistä omassa toiminnassaan. Tällä hetkellä yleisimmät tekoälyn käyttötavat olivat tekstin tuottaminen ja datan analysointi. Käyttötarkoituksissa korostuivat prosessien automatisointi sekä tuotteiden ja palveluiden kehittäminen (kuva 2). Tämä voi vesihuollon alalla tarkoittaa esimerkiksi vuotoriskien ja vedenkulutuksen ennakointia, mekaanisten ja toistuvien kirjausten automaattisointia ja taloudellisten resurssien parempaa hallintaa.⁸



Kuva 3: Are Nordic organisations ready for AI? -kyselytutkimukseen vastanneiden organisaatioiden yleisimmät tekoälyn käyttötarkoitukset vuonna 2023. Muu-kohdassa kerrottuja esimerkkejä ovat mm. kielikäännökset, reformointi, inspiraatio sekä virtaviivaistaminen.⁸

Vaikka yhden kyselytutkimuksen perusteella ei voi vetää suoria johtopäätöksiä, voidaan tuloksien perusteella tulkita tekoälyn hyödyntämiselle löytyvän potentiaalia ja tarvetta. Panostamalla digikoulutukseen ja selkeyttämällä tekoälyn mahdollisuuksia voidaan madaltaa kynnystä tekoälyn käyttöönololle. Tämä voi merkittävästi helpottaa vesihuoltolaitosten arjen työkuormaa ja parantaa tehokkuutta, mikä on erityisen tärkeää taloudellisesti tiukkoina aikoina ja työtekijäpulalla keskellä.

Pienetkin tehokkuuden parannukset voivat tuoda organisaatiolle merkittäviä taloudellisia hyötyjä. Tekoäly voi esimerkiksi optimoida logistiikkaa, virtaviivaistaa toimitusketjuja ja automatisoida rutiinitehtäviä, vapauttaen henkilöstöä monimutkaisempiin ja strategisempiin tehtäviin. Lisäksi tekoälyn avulla voidaan helpottaa päätöksentekoa analysoimalla suuria tietomääriä nopeasti, mikä johtaa parempaan riskienhallintaan ja tarkempaan markkinatrendien ennustamiseen. Tekoälyllä voidaan myös parantaa tuotteita ja palveluita mahdollistamalla personoinnin ja innovaation, mikä parantaa asiakaskokemusta ja lisää asiakasuskollisuutta.⁸

⁸ Think Tank Mandag Morgen. 2023. Are Nordic organisations ready for AI? The effect of AI on how organisations think about their digital transition and skills. PDF-artikkeli. Viitattu 30.7.2024.

2 DATA

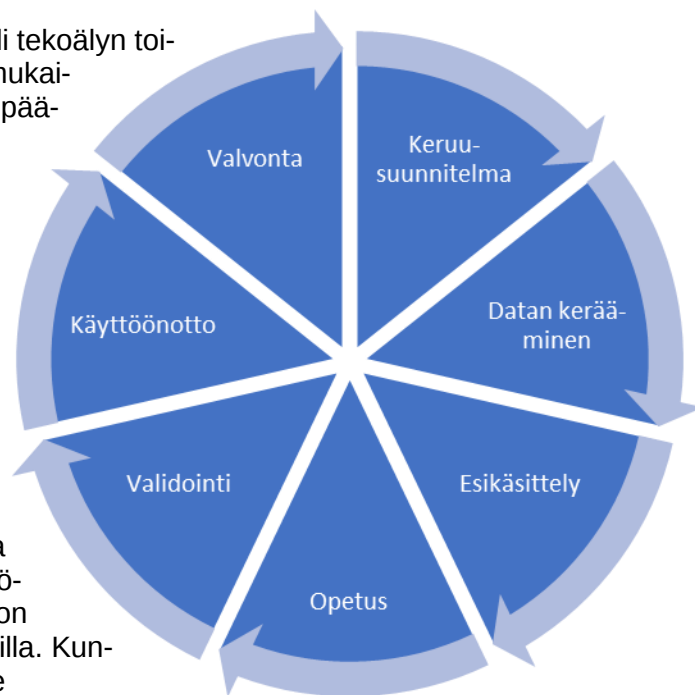
Tekoälysovellukset tarvitsevat lähtötiedoksi dataa. Digitaalisessa muodossa tallennettu tieto, kuten numerot, sanat, kuvat ja äänet, ovat datan eri muotoja, joita tekoäly hyödyntää toiminnoissaan. Tehokas tekoälyn hyödyntäminen vaatii organisaatiolta data edellä työskentelyä. Menestyvät yritykset ja organisaatiot painottavat datan saatavuutta ja käyttöä kaikessa toiminnassaan, mikä mahdollistaa tekoälyn tuomien hyötyjen pitkäaikaisen ja jatkuvan realisoitumisen.⁹ Tämä edellyttää, että tietojärjestelmiä käytetään niin kuin on suunniteltu, ja että dataa kerätään ja tallennetaan tavalla, joka mahdollistaa sen hyödyntämisen. Ilman tällaista lähestymistapaa tekoälyn hyödyt jäävät lyhytaikaisiksi ja pinnallisiksi.¹⁰

On tärkeää muistaa, että pienetkin virheet datan keruussa ja käsittelyssä voivat vaikuttaa merkittävästi tekoälyn toimintaan ja sen tuottamiin tuloksiin. Siksi datan laatuun, määrään, tasapuolisuuteen, mahdollisiin vinoumiin ja kaikkien näiden hallintaan tulee kiinnittää erityistä huomiota.¹¹

2.1 LAATU JA MÄÄRÄ

Koska datan laadulla on ratkaiseva rooli tekoälyn toiminnassa, voi epätarkkuus tai -johdonmukaisuus johtaa virheellisiin analyysihin ja päätöksiin. Laadukas data on esimerkiksi kattavaa, relevanttia ja virheetöntä. Myös määrällä on merkitystä, sillä suuret datamäärät mahdollistavat koneoppimismallien paremman kouluttamisen ja tarkemman mallintamisen. Tavoitteisiin pääseminen ja tiedolla johtaminen edellyttävät, että perusasiat tiedon keruusta ja hyödyntämisestä laitetaan kuntoon (kuva 4).¹¹

Kaikenlainen päätöksenteko, analyysi ja järjestelmien kehitys tekoälyn parissa perustuu datan saatavuuteen ja sen hyödyntämiseen. Kerätyn tiedon heikkous on yleinen haaste suomalaisilla vesilaitoksilla. Kunnollisen datan keruu on tärkeää, sillä se mahdollistaa luotettavien ja tarkkojen tekoälymallien rakentamisen.¹²



Kuva 4. Datan laadun varmistaminen.¹²

⁹ Taulli, T. 2019. Artificial Intelligence Basics – A Non-Technical Introduction. Apress Media LLC. Monrovia. California. USA. Luettavissa: <https://doi.org/10.1007/978-1-4842-5028-0>. Viitattu: 29.4.2024.

¹⁰ Sunela, M. 17.1.2024. Työelämäprofessori. Aalto yliopisto. Etähaastattelu. Haastattelijana Joona Lepistö ja Jaana Pulkkinen.

¹¹ Sunela, M. 4.6.2024. Tekoälyn hyödyntäminen vesihuollossa. Työpaja 2 - Tekoälyn käytön edellytykset. Etätyöpajan luentosarja.

CASE – LAADUKKAAN DATAN MERKITYS JA SEN HYÖDYNTÄMINEN LIIKETOIMINNASSA



Data

*”Alvalla aloitettiin noin neljä ja puoli vuotta sitten työskentely data- ja tekoälystrategian parissa. Strategian keskiössä on ollut **datan laadun parantaminen ja sen hyödyntäminen liiketoiminnassa**. Tavoitteena on ollut poistaa haasteet, jotka liittyvät tekoälyn käyttöön ja varmistaa, että datan hyödyntäminen tuottaa liiketoimintahyötyä. Datan laatu ja oikeellisuus ovat avainasemassa, ja niiden varmistamiseksi on kehitetty prosesseja ja mittareita.*

*Strategiassa korostetaan, että datan keräämisen tulee perustua liiketoimintatarpeisiin ja että kerätyn datan tulee olla volyymihyötyä tuottavaa. Tämä tarkoittaa, että **kerätyn datan tulee olla sellaista, joka ei kuormita laitoksen taloutta negatiivisesti**. Ei siis kannata kerätä dataa mitä ei tulla hyödyntämään laitostyöskentelyssä tulevaisuudessa, joskin tämän määrittely ei aina ole yksiselitteistä.*

*Tärkeää datan keräämisessä ja säilyttämisessä on miettiä **kustannuksia, jotka aiheutuvat datan säilyttämisestä**. Siksi on tärkeää kerätä laadukasta dataa, jota laitos pystyy käyttämään esimerkiksi päätöksenteon tukena. Toinen tärkeä asia on **määrittää datalle käyttötarkoitus**. Kolmanneksi tärkeää on **merkitä data**, kun se otetaan sisään laitoksen järjestelmään ja määrittää minkä tiedon se korvaa järjestelmässä. Näin toimien laitos kerää laadukasta dataa ja välttää turhia kustannuksia ylimääräisen datan säilömisestä.*

*Hyödylliseksi on koettu, että laitoksella on ollut **omat ICT-osaajat, joiden avulla data- ja tekoälystrategiaa on voitu kehittää ja toteuttaa yhdessä verkostoista vastaavien henkilöiden kanssa**. Käytännössä tämä on esimerkiksi tarkoittanut, että jokaisesta uudesta tai saneeratusta putkimetristä on tallennettu tarvittavat tiedot laitoksen järjestelmiin ja samalla poistettu vanhentunut tarpeeton data, tätä kautta myös analyysiemme tarkkuus on parantunut.”*

-Alva-yhtiöt Oy

2.2 LUOKITTELU

Datan luokittelu mahdollistaa sen monipuolisen hyödyntämisen, eli kokonaisuuksien tarkastelun yksittäisten datapisteiden sijasta. Näin tarkasteltavasta datasta voidaan tehdä erilaisia johtopäätöksiä. Jokainen vesilaitos hyötyisi datasta ja sen käsittelystä vastaavasta henkilöstä, joka varmistaisi tiedon keräämisen ja analysoinnin prosessien olevan kunnossa, ja että dataa voidaan hyödyntää tehokkaasti ja vastuullisesti.¹²

Rakenteinen data (*Structured data*)

Rakenteinen data on järjestettyä ja muodoltaan määriteltyä tietoa, joka on tallennettu tietokantoihin. Se on helposti haettavissa ja käsiteltävissä, koska se noudattaa tiukkaa formaattia, kuten taulukoiden sarakkeita ja rivejä. Esimerkkejä rakenteisesta datasta ovat sääaseman lämpötilamittaukset ja pumppaamojen käyntiajat.¹³

Rakenteeton data (*Unstructured data*)

Rakenteettomalla datalla ei ole ennalta määrättyä muotoa tai rakennetta, mikä tekee sen analysoinnista ja käsittelystä haastavampaa. Tällaista dataa ovat esimerkiksi tekstidokumentit, kuvat, videot ja äänitiedostot. Suurin osa maailman datasta on rakenteetonta.¹³

Puolirakenteinen data (*Semi-structured data*)

Puolirakenteinen data on jotain rakenteisen ja rakenteettoman väliltä. Se sisältää sekä rakenteellisia elementtejä että sellaisia osia, jotka eivät ole tiukasti määriteltyjä. Esimerkkejä puolirakenteisesta datasta ovat sähköpostiviestit, joissa on sekä selkeästi määriteltyjä kenttiä (kuten lähettäjä, vastaanottaja ja päivämäärä) että vapaamuotoista tekstiä.¹³

Aikasarjadata (*Time-series data*)

Aikajärjestykseen sijoitettu data mahdollistaa muutosten, trendien ja kausivaihteluiden havainnoinnin ja ennustamisen, mikä on kriittistä monilla aloilla, kuten myös vesihuollossa. Dataa kerätään ja analysoidaan ajan kuluessa tapahtuvien muutosten ymmärtämiseksi.¹³

¹² Sunela, M. 4.6.2024. Tekoälyn hyödyntäminen vesihuollossa. Työpaja 2 - Tekoälyn käytön edellytykset. Etätyöpajan luentosarja.

¹³ Taulli, T. 2019. Artificial Intelligence Basics – A Non-Technical Introduction. Apress Media LLC. Monrovia, California, USA. Luettavissa: <https://doi.org/10.1007/978-1-4842-5028-0>. Viitattu: 29.4.2024.

CASE – TEKOÄLYN HYÖDYNTÄMINEN KASVIHUONEKAA- SUPÄÄSTÖJEN VÄHENTÄMISESSÄ



Koneoppiminen Data

"HSY:n sisäisessä hankkeessa on keskitytty kasvihuonekaasupäästöjen, erityisesti typpioksiduulin, vähentämiseen. Hankkeessa on hyödynnetty yli **kymmenen vuoden ajan kerättyä kasvihuonepäästödataa, joka on ollut perustana prosessien optimoinnille ja häiriöiden ennakkoinnille**. Tämä on ollut tärkeää, sillä prosessiongelmat voivat aiheuttaa merkittäviä kustannuksia.

Projektissa on ensin tunnistettu soveltuvat menetelmät ja sen jälkeen otettu ne käyttöön. Käytössä on ollut Pythonin kirjastoja koneoppimiseen, ja haasteena on ollut laboratoriodatan ja mitatun datan yhdistäminen. **Dataa on visualisoitu ja klusteroitu, jotta sitä voidaan paremmin hyödyntää prosessien hallinnassa**. Erityisesti neuroverkkoregressio on osoittautunut tehokkaaksi menetelmäksi tulosten mallintamisessa.

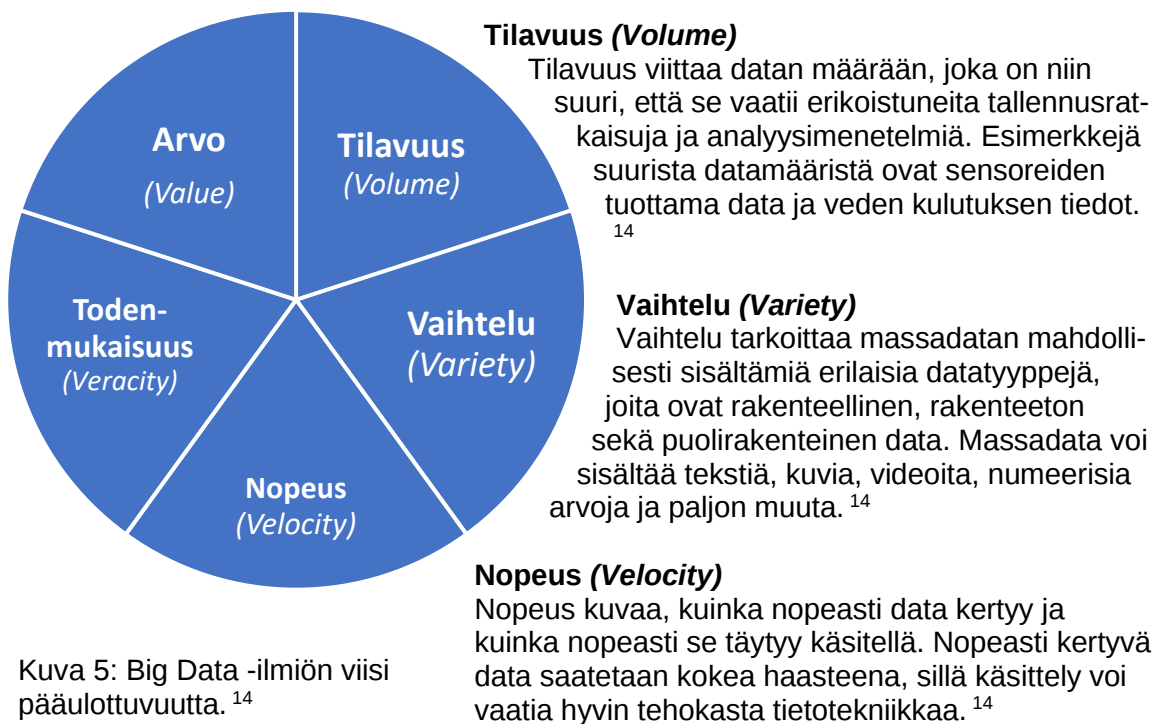
Projektitiimillä on ollut vahva tausta tutkimuksesta, koodaamisesta ja vesihuoltoalasta, mikä on mahdollistanut **syvällisen ymmärryksen laitoksen toiminnasta ja datan hyödyntämisestä**. Prosessihäiriöiden mallintamiseen on voitu käyttää resursseja, koska niiden ennaltaehkäisy säästää laitokselle rahaa. **Osaamisen kehittäminen on ollut keskeistä projektin onnistumiselle**.

Datan laadun varmistamiseksi on kiinnitetty huomiota **mittarivirheiden karsimiseen** ja mittareita on kalibroitu laboratoriodatan avulla. On tärkeää tarkastella datan käyttökelpoisuutta ja miettiä, **kuinka paljon labra- ja mittadataa tarvitaan ratkaisujen rakentamiseen**. Tämä kehitystyö on ollut intensiivistä ja vaatinut **paljon kokeilua ja virheistä oppimista**, mikä voi olla haastavaa konsulttivetoisesti toteutetuna."

-HSY

2.3 MASSADATA (BIG DATA)

Big Data, eli massadata, viittaa suureen ja monimutkaiseen tietomassaan, jota on vaikea käsitellä ja analysoida perinteisten tietojenkäsittelymenetelmien avulla. Jotta suuresta tietomassasta saadaan kaivettua esiin hyödyllisiä oivalluksia ja tietoa, tarvitaan edistyneitä työkaluja, teknologioita, sekä analyysimenetelmiä, kuten koneoppimista. Big Data on muuttanut tapaa, jolla yritykset ja organisaatiot pyrkivät ymmärtämään asiakkaitaan, kehittämään tuotteitaan, tekemään ennusteita ja tehostamaan toimintaansa.¹⁴ Big Data -ilmiöön liittyy yleensä kolme pääulottuvuutta, jotka tunnetaan "Three V's" -mallina (Kolmen V:n malli): Volume (tilavuus), Variety (vaihtelu) ja Velocity (nopeus). Myöhemmin tähän malliin on lisätty muitakin ulottuvuuksia, kuten Veracity (todenmukaisuus) ja Value (arvo) (kuva 5).¹⁴



Kuva 5: Big Data -ilmiön viisi pääulottuvuutta.¹⁴

Todenmukaisuus (Veracity)

Todenmukaisuus viittaa datan laatuun ja luotettavuuteen. Datan todenmukaisuutta on tärkeää pystyä arvioimaan, jotta analyysin perusteella voidaan tehdä oikeita päätöksiä.¹⁴

Arvo (Value)

Arvo kuvastaa datan hyödyllisyyttä. Suuren datamäärän kerääminen ja säilyttäminen ei itsessään tuo lisäarvoa, ellei datasta saada jalostettua merkityksellistä tietoa, joka auttaa päätöksenteossa tai toiminnan kehittämisessä.¹⁴

¹⁴ Taulli, T. 2019. Artificial Intelligence Basics – A Non-Technical Introduction. Apress Media LLC. Monrovia. California. USA. Luettavissa: <https://doi.org/10.1007/978-1-4842-5028-0>. Viitattu: 29.4.2024.

2.4 ALGORITMIT

Kuvitellaan, että vesilaitos haluaa ennustaa veden kulutusta. Avuksi voidaan kehittää tekoälypohjainen, koneoppiva algoritmi. Algoritmi voi käyttää vedenkulutuksen historiadataa ja muita vaikuttavia tekijöitä, kuten sääolosuhteita ja päivämäärää, oppiakseen ennustamaan tulevaa vedenkulutusta (kuva 6). Päättöpuun avulla algoritmi voi oppia tunnistamaan tiettyjä kuvioita ja tekemään ennusteita perustuen näihin havaintoihin.¹⁵

Kun malli on koulutettu, sitä voidaan käyttää reaaliaikaisesti verkoston datan analysointiin. Mallin tunnistamassa opitut tekijät, se voi antaa hälytyksen vesilaitoksen henkilökunnalle sekä tietojärjestelmään. Tämän kaltaisella tekoälyjärjestelmällä vesilaitos voi parantaa verkoston valvontaa ja ylläpitoa, tehostaa resurssien käyttöä ja tarjota parempaa palvelua.¹⁵

Algoritmit ovat siis ohjeiden tai sääntöjen joukkoja, jotka määrittävät tietyn ongelman tai tehtävän ratkaisun vaiheet. Algoritmeja voi verrata resepteihin tai kaavoihin, jotka kertovat tietokoneelle tai ihmiselle, mitä toimintoja suorittaa ja missä järjestyksessä, jotta haluttu lopputulos saavutetaan. Tekoälyn kontekstissa algoritmit ovat kriittisiä, sillä ne ovat koneoppimismallien perusta. Algoritmit voivat oppia historiallisesta datasta, tunnistaa trendejä ja suorittaa ennusteita tai päätöksiä uuden, näkemättömän datan perusteella.

Algoritmin kouluttamiseen ja oppimiseen liittyy yllättävä riski: mallista saattaa muodostua liian älykäs. Tätä kutsutaan ylisovittamiseksi, jolle koneoppimisalgoritmit ovat hyvin alttiita. Käytännössä se tarkoittaa, että malli on oppinut liikaa opetusdatan yksityiskohtia ja kohtinaa, eikä enää pysty yleistämään ja soveltamaan tietoa uusiin, näkymättömiin tietoihin. Tämä tapahtuu, kun malli on liian monimutkainen suhteessa käytettävissä olevan datan määrään, jolloin se mukautuu liiaksi opetusdatan erityispiirteisiin. Seurauksena malli saattaa näyttää erittäin tarkalta opetusdatalla, mutta sen suorituskyky heikkenee huomattavasti testidatalla. Ylisovittamisen välttämiseksi on tärkeää käyttää riittävästi dataa, jakaa data opetus- ja testijoukkoihin, ja valita sopivan monimutkainen tai yksinkertainen malli, joka ei ole liian joustava eikä liian jäykkä.¹⁶

Algoritmit voivat olla yksinkertaisia, kuten aakkosten mukainen lajittelu algoritmillä, tai erittäin monimutkaisia, kuten ne, joita käytetään syväoppimisessa ja neuroverkkojen kehittämisessä. Monimutkaisuus ei kuitenkaan aina ole välttämätöntä. Tärkeintä on, että algoritmi ratkaisee ongelman tehokkaasti ja tarkasti.



Kuva 6: Vedenkulutusta arvioivan ja ennustavan algoritmin vaiheet yksinkertaistettuna.

¹⁵ Alonso, E. G. - Jenny, H. - Minguez, R. & Wang, Y. 2020. Using Artificial Intelligence for Smart Water Management Systems. ADB Briefs. Asian Development Bank. Manila. Filippiinit. PDF-artikkeli. Luettavissa: <http://dx.doi.org/10.22617/BRF200191-2>. Viitattu: 29.4.2024

¹⁶ Helsingin yliopisto & MinnaLearn. 2024. Elements of Ai. Tekoälyn perusteet. Verkkokurssi. Käyntiosoite: <https://course.elementsofai.com/fi/>. Viitattu: 28.5.2024.

CASE – POIKKEAMIEN HAVAITSEMINEN JA ENNUSTAMINEN DATAN AVULLA



Koneoppiminen Data

”Algoritmit voidaan valjastaa havaitsemaan poikkeamia ja luomaan ennusteita eri lähteistä saatavan datan avulla. Hyödynnettäviä lähteitä ovat esimerkiksi automaatiojärjestelmistä saatava data (esim. pumppujen käyntiajat), verkkotietojärjestelmästä saatavat sijaintiedot, erilaisilta sensoreilta ja mittareilta saatava data sekä sadedata. Järjestelmä vastaanottaa dataa ketterästi API-rajapinnan kautta ja tarvittaessa voidaan toteuttaa myös muita integraatioita.

Smartvattenin järjestelmä käyttää koneoppimismalleja, jotka mahdollistavat dynaamisen oppimisen. Dynaaminen oppiminen tarkoittaa, että malli mukautuu jatkuvasti kerättävän datan perusteella. Tämä mahdollistaa ennakoivan analytiikan.

Käytännössä järjestelmän opetus alkaa syöttämällä siihen historiadata, yleensä vuoden ajalta. Historiadatan perusteella järjestelmä oppii toimintojen erityispiirteitä sekä toistuvia ja poikkeavia tapahtumia. Kun olemassa olevan historiadatan lisäksi järjestelmään virtaa rajapinnan kautta jatkuvasti myös uutta tietoa, pystyy järjestelmä tuottamaan ennusteita sekä havaitsemaan poikkeamia.

Jotta järjestelmästä saadaan irti kaikki hyöty, on sen käyttöön nimettävä vastuuhenkilöt ja varattava näille henkilöille työaikaa. Vaatii aikaa, osaamista ja mielenkiintoa selvittää esimerkiksi, mitkä ovat omalle laitokselle oleellisia hälytyksiä ja minkä tasoiset poikkeamat jätetään hälytysrajojen ulkopuolelle. Hälytysrajoja voidaan säätää joko laitosten tai Smartvattenin henkilöstön toimesta.

Järjestelmä on parhaimmillaan väsymätön vesilaitoksen apukäsi, jonka avulla voidaan ohjata kunnossapitoa, parantaa resurssienhallintaa sekä nopeuttaa ongelmatilanteisiin reagointia.

-Smartvatten

CASE – ENNUSTEITA IHMISTEN HYÖDYNNETTÄVÄKSI



Koneoppiminen



Data

*"Tekoälyn hyödyntäminen ennusteiden laadinnassa lähtee liikkeelle siitä, että **ymmärretään haaste tai ongelma, johon etsitään ratkaisua**. On myös määritettävä, minkä taseisia lähtötietoja tarvitaan ja kuinka tiheällä syklillä. Toisin sanoen lähtötietojen osalta on tiedettävä **mitä dataa kerätään, mistä, miten ja miksi**. Esimerkiksi vedentuotannon ennusteiden osalta riittää yleensä tuntitason data. Mikäli dataa tarvitaan useasta eri lähteestä tai palveluntarjoajalta, on **kokonaisarkkitehtuurin tunteminen** tärkeää.*

***Hälytysrajat voidaan asettaa tekoälyn avulla.** Tämä tarkoittaa sitä, että laskeaan **datan perusteella normaali vaihteluväli ja järjestelmä hälyttää, kun tapahtuu poikkeama**. Tämä on hyvä tapa äkillisten muutosten havaitsemiseen. Jos muutokset sen sijaan tapahtuu pikkuhiljaa, automaattinen hälytysraja voi olla siedäntynyt ja muutos jää havaitsematta. **Hitaasti kehittyvät ongelmat** sen sijaan **edellyttävät** toisenlaista lähestymistapaa: **erilaisia algoritmeja ja pitkän aikavälin tilannekuvia graafeineen**.*

*Tekoälyä hyödyntävien järjestelmiä hankittaessa on tärkeää huomioida, ettei pelkkä käyttöönotto riitä realisoimaan hyötyjä. Tekoälyä sisältävässä prosessissa hyödynnetään isoa datamäärää ja mallia täytyy hienosäätää, eli opettaa, useita kertoja. **Tekoälymallia hyödyntävän organisaation on tunnistettava, mitä tietoa tarvitaan ja mikä on oleellista. Henkilöstön on sitouduttava kehitystyöhön** ja ymmärrettävä, mitä "konepellin" alla tapahtuu. Esimerkiksi algoritmiperusteiset hälytykset on tarkasteltava vesilaitoksen henkilökunnan toimesta, jotta mallia voidaan säätää siten, että oleelliset hälytykset säilyvät ja epäolennaiset karsitaan pois.*

Tekoäly on hyvä työkalu, mutta se ei tuota tuloksia ilman, että ihmiset sitä käyttävät."

-Insta

CASE – TEKÖÄLYN ROOLI VESIHUOLLON TEHOSTAMISESSA



Data



Koneoppiminen

*"Kangasalan Vesi on ottanut käyttöön tekoälyä vesihuollon prosesseissaan, mikä on tuonut mukanaan merkittäviä hyötyjä. Aiemmin vesihuollon tilannekuvan ymmärtäminen ja poikkeamien havaitseminen vaati paljon ihmistyötä. **Nyt tekoäly auttaa analysoimaan kerättyä dataa ja tunnistamaan mahdollisia ongelmia, kuten vuotoja tai laitteiden toimintahäiriöitä.***

*Kangasalan Vedellä on käytössä Instan Focus AI -niminen tekoälysovellus, joka seuraa noin sataa puhtaan veden ja jäteveden kohdetta. Sovellus käy läpi automaatiojärjestelmästä kerättyä dataa ja tunnistaa poikkeamat, jotka se raportoi henkilökunnalle. Tämä mahdollistaa sen, että työntekijät voivat aloittaa päivänsä tarkastelemalla mahdolliset ongelmat ja **suunnata toimenpiteet oikeisiin kohteisiin.***

Tekoälyn avulla on löydetty esimerkiksi pumppujen tukoksia, jotka ovat ilmenneet pidentyneinä käyntiaikoina. Lisäksi on havaittu tilanteita, joissa viemäriverkkoon on päästetty sopimattomia vesiä, mikä on näkynyt kasvaneina pumppausmäärinä. Puhtaan veden puolella tekoäly onnistui paljastamaan virtausmittarin häiriön, joka ei olisi tavanomaisessa kaukovalvonnassa tullut ilmi. Lisäksi järjestelmän avulla on havaittu muutamia vuotokohteita puhtaan veden runko- ja talo johdoissa.

*Projektin haasteina on tunnistettu tekoälyosaajien resurssin puute ja nykyisten järjestelmien, kuten Kangasalan Veden palvelimien muutostyöt. Tavoitteena onkin tulevaisuudessa siirtää palvelimet pilveen ja kehittää järjestelmää entistä älykkäämmäksi. **Tämä tarkoittaa, että järjestelmä voisi tulevaisuudessa ilmoittaa itse huoltotarpeista, vähentäen kalenteripohjaista huoltoa.***

*Tekoälyn käyttöönotto on osoittautunut kustannustehokkaaksi, ja **Kangasalan Vesi on kokenut, että investointi on maksanut itsensä takaisin nopeasti.** Tämä on mahdollistanut myös henkilöstön vapauttamisen rutiinitehtävistä tärkeämpiin toimiin, kuten ennakoiwaan kunnossapitoon ja vikatilanteiden ratkaisuun.*

*Kangasalan Veden kokemukset osoittavat, että tekoälyn hyödyntäminen vesihuollossa **voi parantaa toiminnan tehokkuutta, nopeuttaa reagointia ongelmiin ja tuoda taloudellista säästöä.** Tekoäly on muuttanut tapaa, jolla vesihuoltoa hallitaan ja ylläpidetään, ja se on avannut uusia mahdollisuuksia toiminnan kehittämiseen."*

- Kangasalan Vesi

CASE – KONEOPPIVIEN MALLIEN HYÖDYNTÄMINEN VIEMÄRIVERKON SANEERAUSAJANKOHDAN ENNUSTAMISSESSA



Data



Koneoppiminen

"Aalto-yliopiston tohtorikoulutettava Sina Masoumzadeh tutkii, voiko viemäriverkoston saneeraustarpeen ajankohtaa määrittää erilaisia koneoppimismenetelmiä hyödyntämällä. **Tutkimuksen tavoitteena on kerätä tietoa ja selvittää, kuinka vesihuoltolaitosten keräämää dataa voidaan käyttää viemäriverkostojen saneerausajankohdan ennustamiseen koneoppimistekniikoita hyödyntäen.** Vaikka mallien kalibroiminen vaatii merkittävän määrän tietoa, tulokset osoittavat, että **pelkästään putkien fyysisten ominaisuuksien (esim. pituus, halkaisija, materiaali jne.) ja viemärikuvausten tulosten käyttäminen voi tuottaa kohtuullisen tarkan mallin.** Lisäksi, jos verkostosta saatavilla olevat tiedot putken kaltevuudesta, asennussyvyydestä tai virtaamasta voivat olla erittäin arvokkaita, sillä ne voivat vaikuttaa merkittävästi putkien vaurioitumiseen.

On olennaista, että vesihuoltolaitoksen keräämä data linkitetään olemassa olevaan tietoon verkkotietojärjestelmässä. Tämä tarkoittaa esimerkiksi sitä, että viemärintarkastusdatasta tehty kunnon luokittelu kohdistuu kuvattuun osioon, eikä se ole vain irrallinen asiakirja, jota ei voida yhdistää järjestelmässä oikealle putkelle. **Kun putkille on linkitetty riittävästi tiettyjen tapahtumien tietoja, voi koneoppimismalli ennustaa tarkemmin putkien käyttöikä.**

Koneoppimismenetelmien hyödyntämisessä on myös olennaista, että verkkotietojärjestelmien data on **helposti** saatavilla valitun menetelmän soveltamista varten. Tämä on tärkeää, koska malli on koulutettava uudelleen joka kerta, kun aineistoa päivitetään järjestelmään. Tämä tarkoittaa, että vesihuoltolaitosten ja järjestelmätoimittajien on panostettava yhteistyöhön ja siihen, että vesihuoltolaitoksen keräämä ja omistama data on helposti saatavilla siistissä ja hyödyllisessä muodossa."

**- Sina Masoumzadeh
(väitöskirjatutkija, Aalto yliopisto)**

TÄRKEIMMÄT NOSTOT: DATA



Datan laatu ja määrä

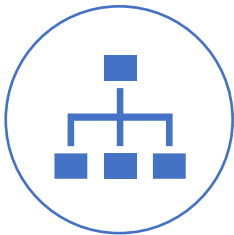
Laadukas data on ajantasaista, kattavaa, relevanttia ja virheetöntä. Suuret datamäärät mahdollistavat tarkemman mallintamisen ja luotettavien tekoälymallien rakentamisen. Esimerkkejä vesilaitoksella kerättävästä datasta:

- Veden laatu
- Kulutustiedot
- Paine- ja virtaustiedot
- Sään vaikutukset
- Asiakaspalaute
- Taloudelliset tiedot



Datan kerääminen ja säilyttäminen

- Keräämisen tulee perustua liiketoimintatarpeisiin ja olla taloudellisesti kannattavaa
- Säilyttämisen kustannukset on huomioitava, ja vanhentunut data poistetaan
- Tietosuoja ja tietoturva on huomioitava kerätessä ja säilytettäessä dataa
- Tallennusratkaisut valitaan käyttötarkoitukseen soveltuviksi
- Mittalaitteistojen laatu ja luotettavuus on tärkeä osa datan keruuta



Datan luokittelu

- Mahdollistaa datan tehokkaan hyödyntämisen, kokonaisuuksien tarkastelun ja johtopäätösten tekemisen
- Käyttötarkoituksen huomiointi dataa luokiteltaessa on tärkeää

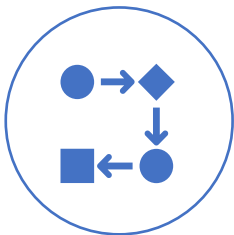


Datan hyödyntäminen

Datan analysointi ja jalostaminen merkitykselliseksi tiedoksi tukevat päätöksentekoa, resurssien optimointia ja palvelun parantamista sekä auttavat organisaatiota kehittämään toimintaansa tehokkaammin.

Esimerkkejä datan hyödyntämisestä:

- Kustannusanalyysit ja budjetointi
- Veden kulutuksen ennustaminen
- Ympäristövaikutusten arviointi
- Algoritmit



Algoritmit

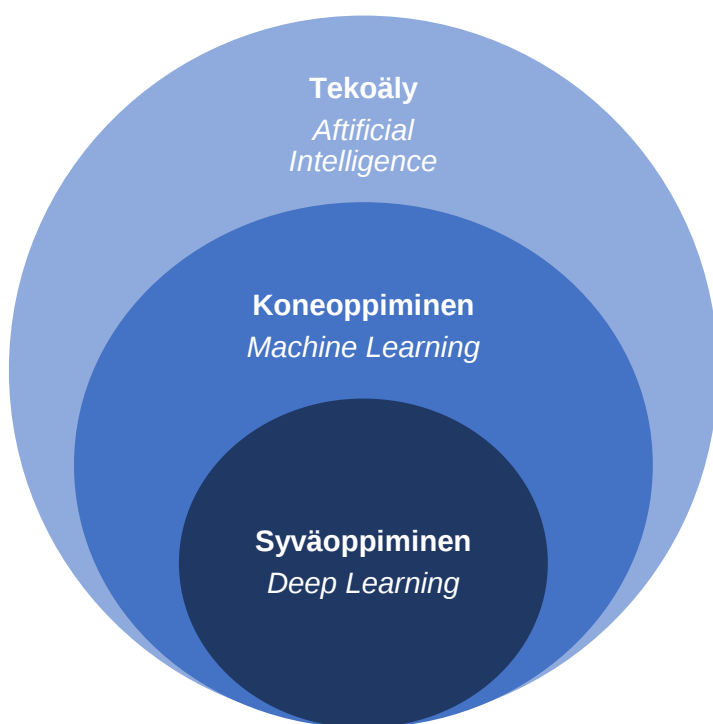
Algoritmeilla voidaan tehokkaasti luoda ennusteita tai muita johtopäätöksiä. Dataa on oltava riittävästi ja valittu malli ei saa olla liian yksinkertainen tai monimutkainen suhteessa käsiteltävään aiheeseen.

Esimerkkejä algoritmien hyödyntämisestä:

- Veden kulutuksen ennustaminen ja jakelun optimointi
- Vuotojen tunnistaminen
- Veden laadun ja sään seuranta sekä analysointi
- Häiriöiden ennakointi

3 KONEOPPIMINEN

Koneoppiminen on algoritmeihin perustuva tekoälyn ala (kuva 7).¹⁷ Sen avulla tietokoneet voivat oppia ja tehdä päätöksiä datasta itsenäisesti. Koneoppimisen soveltaminen ongelmanratkaisuun onnistuneesti edellyttää järjestelmällistä lähestymistapaa. Hyödynnettävän datan laatu on varmistettava, sillä tekoäly oppii, mitä sille syötetään. Jos annetaan huonoa tai laadutonta tietoa, tulee vastauskin olemaan huono. Käyttäjän ei silti tarvitse tietää, mitä kaikkea datasta halutaan saada irti. Esimerkiksi tässä tekoäly ja koneoppiva malli voi auttaa. Koneoppiminen on dynaaminen ja jatkuvasti kehittyvä ala, jolla on valtava potentiaali monilla eri toimialoilla.¹⁸



Koneoppimisprosessin tavoite on luoda yhteen tai useampaan algoritmiin perustuva malli, jota kehitetään kouluttamalla. Algoritmit koulutetaan datan avulla luomaan malleja, jotka mahdollistavat koneiden suorittavan tehtäviä. Koulutusdata opettaa malleja tekemään ennusteita ja testidataa käytetään mallin suorituskyvyn arviointiin. Mallin antamien ennusteiden tarkkuutta arvioidaan termien "true positive" ja "false positive" avulla. "True positive" tarkoittaa oikeaa ennustetta ja "false positive" viittaa tilanteeseen, jossa malli virheellisesti ennustaa tuloksen oikeaksi. Ajan myötä, kun koke-musta ja dataa kertyy lisää, järjestelmän suorituskky paranee.¹⁸

Kuva 7: Tekoälyn pääelementit.¹⁸

¹⁷ Taulli, T. 2019. Artificial Intelligence Basics – A Non-Technical Introduction. Apress Media LLC. Monrovia, California, USA. Luettavissa: <https://doi.org/10.1007/978-1-4842-5028-0>. Viitattu: 29.4.2024.

¹⁸ Sunela, M. 17.1.2024. Työelämäprofessori. Aalto yliopisto. Etähaastattelu. Haastattelijana Joona Lepistö ja Jaana Pulkkinen.

KONEOPPIMISEN NELJÄ PÄÄMENETELMÄÄ



Kuva 8: Koneoppimisen päämenetelmät.²⁰

Ohjattu koneoppiminen (*Supervised Learning*)

Verrataan koneoppimisen menetelmiä koulussa opiskeluun. Kuvittele opiskelijalla olevan joukko kysymyksiä, joista jokaiseen on liitetty oikea vastaus. Yksinkertaisimmillaan vastaus voi olla ”kyllä” tai ”ei”. Ohjatun koneoppimisen menetelmässä tietokoneelle näytetään näitä esimerkkejä oikeine vastauksineen. Tietokoneen tehtävä on opiskella ja ymmärtää, miten oikeisiin johtopäätöksiin päästään. Kun tietokone on opiskellut riittävän monta esimerkkiä, sen odotetaan pystyvän antamaan oikeita vastauksia uusiin, vastaavanlaisiin kysymyksiin, joita se ei ole ennen nähnyt.^{19, 20}

Ohjaamaton koneoppiminen (*Unsupervised Learning*)

Ohjaamaton koneoppiminen on kuin aiheeseen tutustumista oppikirjaa lukemalla. Tietokoneelle annetaan tietoja ilman etukäteen määriteltyjä vastauksia tai ohjeita. Se yrittää itse löytää ja tunnistaa datassa olevia piilotettuja ryhmiä, ”ryppäitä” (*clusters*) tai yhteyksiä, ja siten havaita toistuvuuksia. Ohjaamaton koneoppiminen voi auttaa tunnistamaan trendejä tai ongelmia, joita ei aiemmin ole huomattu.^{19, 20}

Puoliohjattu koneoppiminen (*Semi-supervised learning*)

Puoliohjattu koneoppiminen on termi vaikeasti lokeroitaville tai epämääräisille menetelmille, sillä se yhdistää ohjatun ja ohjaamattoman oppimisen piirteitä. Voit kuvitella opiskelijalle annettavan tehtäviä, joista osassa on vastaukset valmiina ja osassa ei. Opiskelijan tulee suorittaa oikein kaikki tehtävät, hyödyntäen niitä, joille on määritetty oikeat menetelmät, rutiinit tai muut vastaukset. Puoliohjaava koneoppiminen siis opettaa tietokonetta soveltamaan opittua. Tietokone voi oppia analysoimaan merkitsemätöntä dataa ja havaita esimerkiksi mahdollisia riskejä.^{19, 20}

Vahvistusoppiminen (*Reinforcement Learning*)

Vahvistusoppiminen on menetelmä, jossa tietokone oppii kokeilemalla ja saamalla palautetta suorituksistaan. Tyypillisesti tätä verrataan kirjallisuudessa koiran kouluttamiseen: kun koira tekee haluamasi tempun oikein, sille annetaan herkku, ja kun se ei tee, yritetään uudestaan tai koira ohjataan muuten oikeaan suuntaan. Tietokoneelle annetaan palkkioita hyvistä toiminnoista ja mahdollisesti rangaistuksia huonoista valinnoista. Ajan myötä tietokone oppii, mitkä toimet johtavat parhaisiin tuloksiin, ja tekee niitä päätöksiä yhä useammin itsestään.^{19, 20}

¹⁹ Taulli, T. 2019. Artificial Intelligence Basics – A Non-Technical Introduction. Apress Media LLC. Monrovia, California, USA. Luettavissa: <https://doi.org/10.1007/978-1-4842-5028-0>. Viitattu: 29.4.2024.

²⁰ Helsingin yliopisto & MinnaLearn. 2024. Elements of Ai. Tekoälyn perusteet. Verkkokurssi. Käyntiosioite: <https://course.elementsofai.com/fi/>. Viitattu: 28.5.2024.

CASE – TEKOÄLY JÄTEVEDENPUHDISTUKSEN TEHOSTAJANA



Data Koneoppiminen

”Kuusamon EVO on ottanut käyttöön tekoälyä hyödyntävän järjestelmän, joka parantaa jätevedenpuhdistuksen tehokkuutta ja optimoi energiankulutusta. **Tämä älykäs järjestelmä käyttää koneoppimista analysoidakseen puhdistusprosessia reaaliajassa ja tekee päätöksiä kahden minuutin välein.** Sen avulla voidaan varmistaa, että puhdistustulos täyttää asetetut vaatimukset.

Järjestelmä on suunniteltu huomioimaan puhdistamon kaksi erillistä linjaa ja se kykenee automatisoimaan prosessit. Tämä tarkoittaa, että järjestelmä voi itsenäisesti määritellä, miten puhdistusta tulisi säätää eri tilanteissa, kuten kun kapasiteetin tarve muuttuu. **Tällainen älykäs ohjaus auttaa pitämään puhdistamon toiminnan jatkuvasti optimaalisella tasolla.**

Evolla käytössä on Veolian Hubgrade-laitosmoduuleista Biologinen käsittely (Biological capacity). **Järjestelmä optimoi ilmastuksen energiankulutusta säätämällä ilmastuksen määrää eri ilmastuslohkoissa. Optimoinnin perusteena käytetään ilmastuksesta lähtevän veden ammonium- ja nitraattityypimittauksia.** Tavoitteena on vähentää ilmastuksen energiankulutusta samalla pitäen typen pitoisuudet asetusarvojen sisällä.

Molemmat laitokset ovat kaksilinjaisia, mikä mahdollistaa myös Standby-moduulin käytön. **Virtaaman ja kokonaistypen ollessa alle valittujen asetusarvojen, toinen ilmastuslinja voidaan poistaa käytöstä. Moduuli optimoi ja automatisoi myös linjojen vaihdon,** jolloin molempien linjojen biologiat saadaan pysymään aktiivisena. Virtaaman tai typen noustessa yli asetusarvojen, moduuli palauttaa molemmat ilmastuslinjat takaisin käyttöön.

Järjestelmän käyttöönotto on vaatinut aikaa ja yhteistyötä eri toimijoiden kanssa, sillä eri järjestelmien yhteensovittaminen on ollut haasteellista. Ennen järjestelmän käyttöönottoa valittiin ensin puhdistamon toimittaja ja sen jälkeen sopiva ohjausjärjestelmä.

Kuusamon EVO:lla on positiivisia kokemuksia järjestelmästä. **Järjestelmä ei ainoastaan optimoi energiaa, vaan myös vähentää kustannuksia pitkällä aikavälillä.** Lisäksi se sisältää instrumenttien valvonnan, mikä vapauttaa työntekijöiden aikaa muille tehtäville, sillä järjestelmä ilmoittaa automaattisesti, jos jokin laitteisto vaurioituu.

Vaikka Kuusamon EVO on vielä käyttöönottovaiheessa ja riippuvainen monista ulkopuolisista toimijoista, kuten koodauksen osaamisesta, **on tärkeää, että käyttäjät osallistuvat aktiivisesti järjestelmän kehitysprosessiin.** Tämä varmistaa, että järjestelmä palvelee parhaalla mahdollisella tavalla Kuusamon EVO:n tarpeita. Lisäksi laitos tutkii mahdollisuuksia laajentaa tekoälyn hyödyntämistä myös lämmöntuotannon optimointiin energiapuolella.”

-Kuusamon EVO

CASE – ALGORITMIT JA KONEOPPIMISMALLI VESIHUOLTOLAITOKSEN APUNA



Data



Koneoppiminen

*"Vesihuollosta vastaavat organisaatiot tarvitsevat tehokkaita työkaluja vedenjakelun hallintaan. Lining pyrkii vastaamaan tähän tarpeeseen kehittämällä **koneoppimiseen perustuvalla aluemittausjärjestelmällä**. Aluemittausjärjestelmä kykenee oppimaan ja ennakoimaan vedenkulutuksen malleja, mikä auttaa löytämään verkostossa olevia vuotoja.*

*Järjestelmä kerää kumulatiivista vesimäärätietoa eri lähteistä, kuten vedentuotantolaitoksilta, paineenkorottamoista, vesitorneista sekä erilaisista mitta-asemista. Etäluettavien vesimittareiden integrointi mahdollistaa aluekohtaisen vesitaseen seurannan. **Automaatiodata voi olla minuuttitasolla, tuntitasolla tai vuorokausitasolla, riippuen mittauspisteestä ja tarpeesta. Data validoidaan käyttämällä erilaisia algoritmeja, jotka tarkistavat datan eheyttä ja poikkeavuuksia.***

Järjestelmässä käytetään Singular Spectrum Analysis* koneoppimismallia. Ohjattu oppiminen auttaa ennustamaan vedenkulutusta historiatietojen perusteella. Algoritmi myös tunnistaa poikkeamat normaalista kulutuksesta, mikä voi viitata vuotoihin tai muihin ongelmiin verkostossa. Jos järjestelmä havaitsee, että tietyllä alueella vedenkulutus kasvaa yllättäen, se luo hälytyksen, jonka jälkeen huoltotiimi voi lähteä kentälle tarkastamaan tilanteen. Järjestelmän avulla parannetaan vesilaitoksen kykyä reagoida nopeasti ongelmiin, vähentää vedenhukkaa ja optimoida resurssien käyttöä.

- Lining

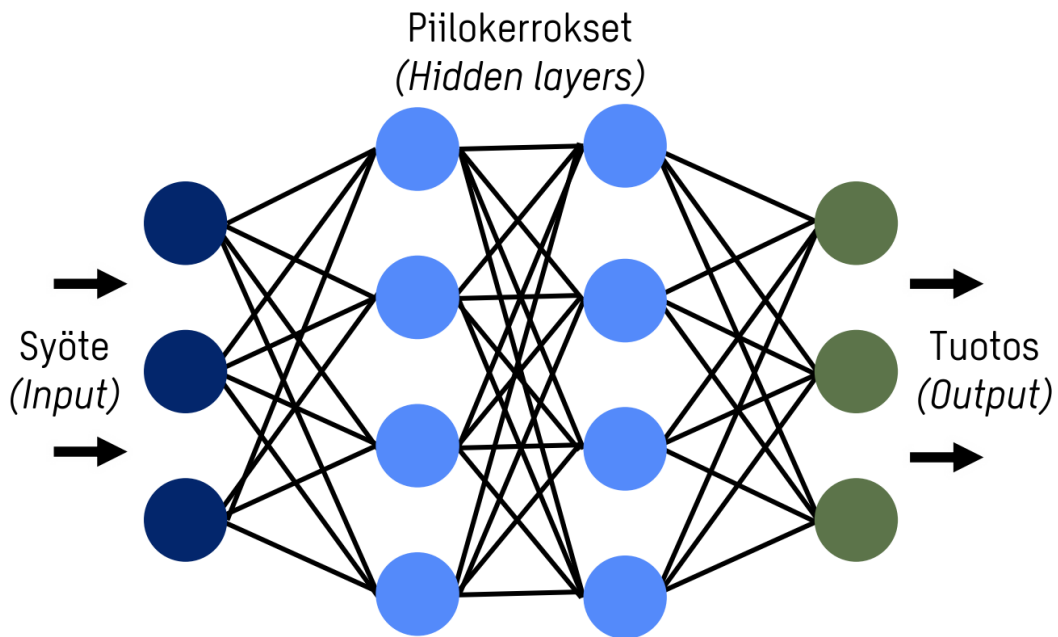
*Singular spectrum analysis (SSA) on tehokas aikasarjojen analysointimenetelmä, jota voidaan hyödyntää esimerkiksi erilaisten trendien löytämiseen, kausivaihteluiden komponenttien erottamiseen, samanaikaisten monimutkaisten trendien ja kausivaihteluiden erottamiseen, lyhyiden aikasarjojen rakenteen löytämiseen sekä muutospisteiden havaitsemiseen. SSA-menetelmä koostuu kahdesta vaiheesta: sarjan hajottamisesta ja uudelleenrakentamisesta. Uusien datapisteiden ennustamiseen käytetään uudelleenrakennettua sarjaa.^{21 22}

²¹ Hassani, H. 2007. Singular Spectrum Analysis: Methodology and Comparison. Luettavissa: [https://doi.org/10.6339/JDS.2007.05\(2\).396](https://doi.org/10.6339/JDS.2007.05(2).396)

²² Hassani, H. 2010. A Brief Introduction to Singular Spectrum Analysis. Luettavissa: https://ssa.cf.ac.uk/ssa2010/a_brief_introduction_to_ssa.pdf

3.1 NEUROVERKOT JA SYVÄOPPIMINEN

Syväoppiminen on koneoppimisen erikoistunut ala, joka jäljittelee ihmisaivojen toimintaa käyttäen hyväksi suuria verkostoja, joita kutsutaan neuroverkoiksi. Ne koostuvat kerroksista, joista jokainen käsittelee tietoa tietyllä tavalla, oppien malleja ja piirteitä datasta (kuva 9).²³



Kuva 9: Neuroverkon toiminta kuvattu yksinkertaistetusti. Neuroverkot jäljittelevät ihmisaivojen toimintaa käyttämällä kerroksittain järjestettyjä neuroneja, jotka käsittelevät ja välittävät tietoa. Ne oppivat tunnistamaan kuvioita ja tekemään ennusteita. Oppimisprosessi tapahtuu toistuvien iteraatioiden kautta, joissa verkko vertaa ennusteitaan todellisiin tuloksiin ja korjaa virheitään.²⁴

Neuroverkkojen avulla pyritään tunnistamaan malleja sekä suorittamaan monimutkaisia tehtäviä matemaattisesti. Tiedon käsittely perustuu suureen määrään neuroneiksi kutsuttuja yksiköitä. Neuronit on järjestetty kerroksiin, joiden välillä on painotettuja yhteyksiä, eli synapseja. Dataa syötetään neuroverkon ensimmäiseen kerrokseen, josta se etenee kerrosten läpi. Jokainen kerros suorittaa laskentoja ja lähettää tulokset seuraavalle kerrokselle. Neuroverkkojen ja syväoppimisen menetelmät ovat osoittautuneet erityisen tehokkaiksi kuvan- ja puheentunnistuksessa, luonnollisen kielen käsittelyssä sekä monissa muissa tehtävissä, joissa tarvitaan kykyä tunnistaa ja tulkita monimutkaisia malleja suurista datamääristä.²⁴

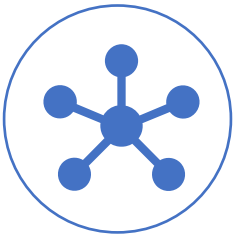
²³ Taulli, T. 2019. Artificial Intelligence Basics – A Non-Technical Introduction. Apress Media LLC. Monrovia. California. USA. Luettavissa: <https://doi.org/10.1007/978-1-4842-5028-0>. Viitattu: 29.4.2024.

²⁴ Chowdhary, K. R. 2020. Fundamentals of Artificial Intelligence. Springer Nature Private Limited. New Delhi. Intia. E-kirja. Luettavissa: <https://doi.org/10.1007/978-81-322-3972-7>. Viitattu: 29.4.2024

Vesihuollossa syväoppimista voidaan verrata esimerkiksi vedenpuhdistamoon. Kuvitellaan jokaisen puhdistamon osan ja puhdistusprosessin vaiheen edustavan jotakin tietojenkäsittelyprosessia. Puhdistettava vesi käy läpi prosessin, jonka jokaisessa vaiheessa vettä käsitellään ja suodatetaan, kunnes veden laatu on saavuttanut halutun tason. Samalla tavalla syväoppimisessa data kulkee kerroksesta toiseen, ja jokaisessa kerroksessa kone oppii jotain uutta datasta. Lopulta se pystyy tekemään tarkkoja ennusteita tai tunnistuksia, kuten esimerkiksi havaitsemaan häiriöitä biologisen jätevedenpuhdistusprosessin mikrobistossa.²⁵

²⁵ Alonso, E. G. - Jenny, H. - Minguez, R. & Wang, Y. 2020. Using Artificial Intelligence for Smart Water Management Systems. ADB Briefs. Asian Development Bank. Manila. Filippiinit. PDF-artikkeli. Luettavissa: <http://dx.doi.org/10.22617/BRF200191-2>. Viitattu: 29.4.2024

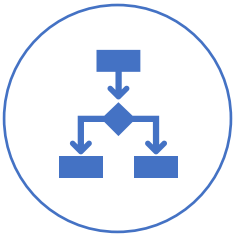
3.2 TÄRKEIMMÄT NOSTOT: KONEOPPIMINEN



Koneoppimisen perusperiaatteet

Koneoppiminen mahdollistaa tietokoneiden oppimisen ja päätöksenteon datan perusteella itsenäisesti. Datan laatu ja määrä ovat kriittisiä tekijöitä, sillä huono data johtaa huonoihin tuloksiin.

- Ratkaistavan ongelman ja mallin tarkoituksen määrittäminen
- Sopivat algoritmit – ei liian yksinkertaiset tai monimutkaiset
- Mallin koulutus ja testaus



Koneoppimisprosessin vaiheet

Mallia kehitetään kouluttamalla algoritmeja datan avulla. Koulutusdata opettaa mallia, ja testidataa käytetään suorituskyvyn arviointiin.

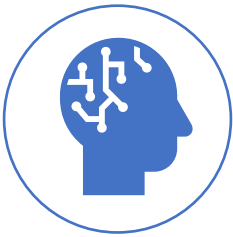
- Ongelman määrittely
- Datan keruu, esikäsittely ja jakaminen
- Mallin koulutus ja testaus
- Seuranta ja ylläpito



Koneoppimisen päämenetelmät

Ymmärtämällä koneoppimisen päämenetelmien eroavaisuudet voidaan valita oikea menetelmä eri ongelmiin ja datatyyppeihin.

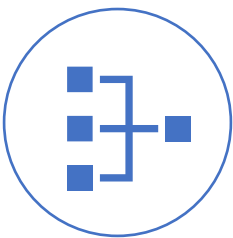
- Ohjattu koneoppiminen
- Ohjaamaton koneoppiminen
- Puoli ohjattu koneoppiminen
- Vahvistusoppiminen



Arviointi ja validointi

Mallin toiminnan luotettavuutta tarkastellaan arvioimalla ja validoimalla ennen käyttöönottoa ja päivitysten yhteydessä. Arviointi ja validointi auttavat luottamaan mallin kykyyn tehdä oikeita ennusteita ja päätöksiä.

- Testidatan käyttö
- Käytännön soveltaminen
- Päivitys ja optimointi
- Yhteistyö asiantuntijoiden kanssa



Syväoppiminen ja neuroverkot

Syväoppiminen ja neuroverkot ovat koneoppimisen kehittyneitä muotoja, jotka mahdollistavat monimutkaisten tehtävien suorittamisen. Ne ovat keskeisiä tekoälyn kehityksessä ja soveltamisessa.

- Voidaan hyödyntää suurten kielimallien kouluttamisessa, kuvantunnistuksessa ja puheen tunnistuksessa
- Jäljittelee ihmisaivojen toimintaa

CASE – VESIJOHTOVERKOSTON VUOTOJEN HAVAITSEMINEN LONTOOSSA



Data Koneoppiminen

Tausta: Syrinix on yritys, joka on erikoistunut automaatioon ja sensoritekologiaan vesijohtoverkoston vuotojen havaitsemiseksi. Yritys perustettiin yliopistoympäristössä vuonna 2004, ja se muuttui kaupalliseksi toimijaksi vuonna 2009. Syrinixin sensoriteknologiat keskittyvät pienten vuotojen tunnistamiseen akustisten sensoreiden avulla, kun taas paineantureita käytetään suurempien vuotojen havaitsemiseen.

Haaste: Lontoon vesijohtoverkosto on laaja ja monimutkainen, ja pienet ja suurten vuotojen havaitsemattomuus voi johtaa merkittäviin taloudellisiin ja ympäristöriskeihin. Perinteiset menetelmät vuotojen havaitsemiseksi olivat usein hitaita ja tehottomia, mikä aiheutti vesivuotojen kertymistä suuriksi ongelmiksi.

Ratkaisu: Syrinixin kehittämää ratkaisua käytetään esimerkiksi Lontoossa, jossa heidän akustiset sensorinsa ja paineanturinsa asennetaan strategisesti vesijohtoverkostoon. Pilvipohjainen ohjelmisto mahdollistaa datan reaaliaikaisen analysoinnin ja hälytysten hallinnan etänä, mikä parantaa reagoimiskykyä vuotojen havaitsemisessa. Sensoridatan perusteella koneoppivat algoritmit etsivät mahdollisia vuotokohtia verkostosta.

Tulokset: Syrinixin teknologian avulla Lontoon vesijohtoverkostossa on pystytty havaitsemaan pieniä vuotoja ajoissa, ennen kuin ne ovat kehittyneet suuriksi ongelmiksi. Tämä on johtanut verkoston vuotojen vähentämiseen ja säästänyt merkittäviä kustannuksia kunnallishallinnolle. Datan visualisointi ja analytiikka ovat myös parantaneet päätöksentekoa ja toimintaprosessien tehokkuutta.

Tulevaisuus: Vaikka Syrinix kehittää koneoppimista ja älykkäitä algoritmeja, teknologia vaatii edelleen insinöörin valvontaa. Yhtiö uskoo avoimen datan ja ymmärrettävien prosessien merkitykseen uusien teknologioiden käyttöön-otossa ja omaisuuden hallinnassa. Tämä sitoutuminen jatkuvaan kehitykseen ja oppimiseen näkyy Syrinixin tavoitteessa parantaa vesijohtoverkoston hallintaa globaalisti.

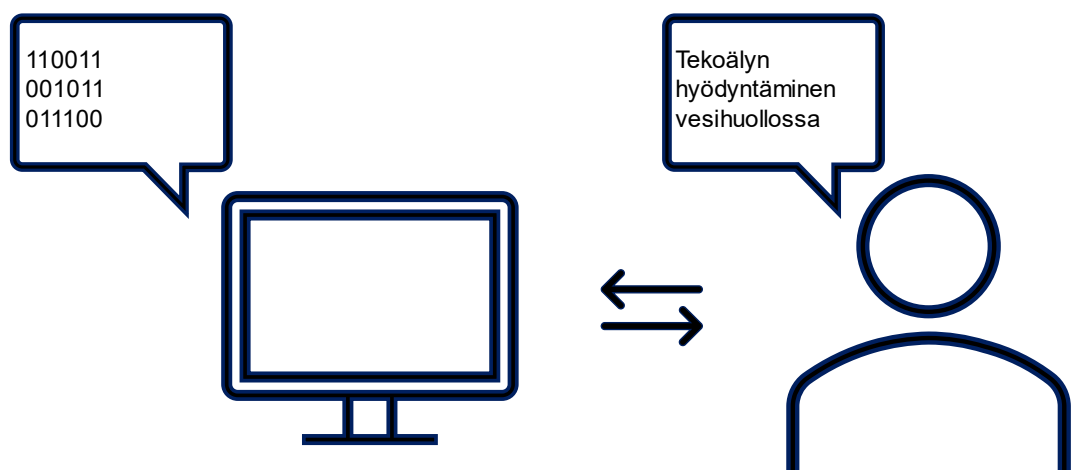
-Syrinix Ltd

4 LUONNOLLISEN KIELEN KÄSITTELY (NLP)

Luonnollisen kielen käsittely, lyhyemmin NLP (*Natural Language Processing*), on tekoälyn ja tietojenkäsittelytieteen ala, joka keskittyy siihen, miten tietokoneet ymmärtävät, tulkitsevat ja käsittelevät ihmisten käyttämää kieltä (kuva 9). Tavoitteena on, että koneet voivat lukea, ymmärtää ja jopa vastata kirjoitettuun ja puhuttuun kieleen ihmisen kaltaisesti.²⁶

Esimerkiksi käyttäessäsi älypuhelimien avustajaa kysyessäsi sääennustetta, NLP-prosessi käynnistyy. Puhelin tulkitsee puhettasi, ymmärtää kysymyksesi ja vastaa sinulle ymmärrettävästi. NLP mahdollistaa myös esimerkiksi sen, että voit kirjoittaa hakukoneeseen kysymyksiä luonnollisella kielellä ja saat relevantteja vastauksia.

Luonnollisen kielen käsittelyyn liittyy monia yllättäviäkin haasteita, koska ihmiskieli sisältää murteita, epäselvyyksiä, sarkasmia, idiomeja sekä monia eri merkityksiä yhdelle sanalle. Tekoälyn kehittäjät luovat malleja ja algoritmeja, jotta tietokoneet pystyisivät käsittelemään näitä haasteita sekä harvaan puhuttuja, esimerkiksi saamelaiskieliä.²⁷



Kuva 10. Esimerkki luonnollisen kielen käsittelystä.

²⁶ Helsingin yliopisto & MinnaLearn. 2024. Elements of Ai. Tekoälyn perusteet. Verkkokurssi. Käyntiosoite: <https://course.elementsofai.com/fi/>. Viitattu: 28.5.2024.

²⁷ Helakallio, A. 2024. Ruotsissa kehitteillä pohjoissaamea osaava tekoäly – Yle mukana kehityksessä. Tivi. Verkkolehtiartikkeli. Luettavissa: <https://www.tivi.fi/uutiset/ruotsissa-kehitteilla-pohjoissaamea-osaava-tekoaly-yle-mukana-kehityksessa/da76e5da-8f0d-4d08-a8bf-67ea25448c70>. Viitattu: 4.10.2024

4.1 SUURET KIELIMALLIT (LARGE LANGUAGE MODELS)

Suuret kielimallit (*LLM, Large Language Processing*) ovat edistyneitä tekoälyjärjestelmiä, jotka avaavat uusia mahdollisuuksia ihmisen kielen ymmärtämisessä ja tuottamisessa. Esimerkkeinä näistä malleista ovat OpenAI:n kehittämät GPT-3 ja GPT-4, jotka perustuvat syväoppimistekniikoihin ja ovat koulutettu valtavilla määrillä tekstidataa. Koulutusprosessi mahdollistaa LLM-mallien oppivan juuri mainituista kielen monimutkaisuuksista, kuten kielipöytä, kontekstista ja jopa hienovaraisista merkityksistä. Näin ollen LLM-mallit kykenevät suorittamaan erilaisia tehtäviä, kuten tekstin syöttöä, käännöstä, tiivistämistä ja kysymysten vastaamista, mikä tekee niistä äärimmäisen monipuolisia työkaluja luonnollisen kielen käsittelyssä. Esimerkki LLM-mallin toiminnasta esitetty kuvassa 11.^{28, 29}

LLM-mallien kehittäminen ja käyttöönotto herättävät myös eettisiä kysymyksiä. Yksi tärkeä näkökulma on puolueellisuuden välttäminen. Koska LLM-mallit perustuvat suureen määrään tekstidataa, ne voivat heijastella olemassa olevia kielenkäyttötrendejä ja sisältää vääristymiä tai ennakkoluuloja. Tämä voi johtaa epäoikeudenmukaisiin tai syrjiviin tuloksiin. Lisäksi LLM-mallit voivat levittää virheellistä tietoa, jos ne eivät kykene arvioimaan lähteiden luotettavuutta tai tulkita monimutkaista tietoa oikein. Väärinkäytön mahdollisuus on myös huomioitava, koska LLM-mallit voivat tuottaa uskottavan kuuloista mutta perusteetonta tietoa, jota voidaan hyödyntää esimerkiksi huijauksissa tai manipulaatioissa.^{28, 29}

Näin ollen suurten kielenmallien kehittämisessä ja käytössä on otettava huomioon eettiset näkökulmat, kuten puolueellisuuden välttäminen, virheellisen tiedon leviämisen ehkäiseminen ja väärinkäytön estäminen. On tärkeää kehittää vastuullisia käytäntöjä ja työkaluja, jotka auttavat valvomaan ja hallitsemaan LLM-mallien toimintaa.^{28, 29}

Tekstisyöte



1. Alkiointi

- Syöte jaetaan alkioiksi

2. Numerointi

- Alkiot muutetaan numeroiksi
- Vastaavat opetuksessa käytettyjä

3. Rinnakkaishuomiointi / Monikerroshavainnointi

- Jokainen huomiointipää tunnistaa ja pisteyttää alkioiden välisiä yhteyksiä eri näkökulmasta
- Verrataan miten läheisesti eri alkiot liittyvät toisiinsa ja miten relevantteja ne ovat lopputuloksen kannalta
- Parantaa ennustetta iteratiivisesti

4. Kielipää

- Valitsee todennäköisimmän seuraavan alkion ulostuloon



Tekstiulostulo

Kuva 11. LLM-mallin toiminta karkeasti.²⁹

²⁸ Taulli, T. 2019. Artificial Intelligence Basics – A Non-Technical Introduction. Apress Media LLC. Monrovia, California, USA. Luettavissa: <https://link.springer.com/content/pdf/10.1007/978-1-4842-5028-0.pdf>. Viitattu: 29.4.2024.

²⁹ Sunela, M. 3.5.2024. Tekoälyn hyödyntäminen vesihuollossa. Työpaja 1 - Tekoäly tutuksi. Etätyöpajan luentosarja. Viitattu: 21.8.2024.

Tärkeää LLM-mallien käytössä käyttäjän näkökulmasta on siirtyä hakukoneajattelusta keskustelupohjaiseen tiedonhakuun. Käytännössä tämä tarkoittaa sitä, että käyttäjän tarvitsee muotoilla hakusanansa siten, että ne muistuttavat enemmän keskustelua toisen henkilön kanssa, kuin se, että käyttäjä hakee tieto esimerkiksi tietoa selaimen hakukoneella. Tätä tapaa kutsutaan kehtesuunnitteluksi (Prompt engineering).³⁰ Tampereen yliopiston mukaan hyvä kehote sisältää seuraavia asioita:

- Tehtävänanto
- Aihe
- Kohderyhmä ja rooli
- Konteksti
- Tyyli ja muoto

Tehtävänanto

Määrittele aluksi selkeä tehtävä tekoälylle. Mitä haluat sen tuottavan? Pyydä tekoälyä esimerkiksi vastaamaan kysymykseen, tekemään tiivistelmän, kääntämään tekstiä, laatimaan kyselyn, tunnistamaan tekstistä osia. Jaa monimutkaiset tehtävät pienempiin osiin.³⁰

”Luo tiivistelmä alla olevasta tekstistä:”

”Korjaisitko muistiinpanoistani virheet ja kirjoittaisit niistä ehyen tekstin liitettäväksi raporttiin?”

”Luonnostele asiakkaille jaettava kirje. Kirje sisältää ... ja sen pituus tulee olla maksimissaan viisi lausetta.”

Aihe

Aiheen tai kysymysten määrittelyllä ohjataan oikean sisällön pariin. Jos etsit ideoita johonkin aiheeseen käsittelyyn voi hyvin yksinkertainen ja laajan aiheen antaminen olla perusteltua. Usein kuitenkin aiheen kuvaaminen mahdollisimman tarkasti antaa laadukkaamman vastauksen. Voit antaa teemoja tai alaotsikot, joita tekstissä käsitellään.³⁰

”Mikä oli laitoksen asiakkaiden keskimääräinen vuorokausikulutus vuonna 2023?

Perustuen lähtötietodataan.”

”Anna määritelmä sanalle vesihuolto”

³⁰ Tampereen yliopisto. 2.4.2024. Tekoälyä ohjaavien kehtteiden eli promptien muotoilu. Luettavissa <https://sites.tuni.fi/vinkkipankki/tekoaly/tekoaly-ohjaavien-kehotteiden-eli-promptien-muotoilu/>. Viitattu 14.10.2024

Kohderyhmä ja rooli

Kerro kehotteessa, että kenelle vastaus halutaan suunnata. Kuka on kohderyhmä ja lukija? Tämä määrittely kehotteessa auttaa tekoälyä muodostamaan sopivan tasoista tekstiä. Saman asian voi kirjoittaa hyvin eri tavoin riippuen siitä kirjoitetaanko aiheesta, vaikka laitoksen henkilökunnalle tai asiakkaille. Tekoälyä voi pyytää myös toimiaan tietyssä roolissa kirjoittaessaan tekstiä. Tämä asettaa usein myös tekstille halutun näkökulman.³¹

”Kirjoita vastaus kansantajuisella kielellä.”

”Kirjoita vastaus vesihuoltoalan asiantuntijoille”

”Olet asiantuntija vesihuoltolaitoksen viestinnässä. Autat minua kirjoittamaan kirjeen asiakkaille.”

Konteksti

Voit myös ohjata tarkemmin sisällön käsittelyä. Mitä asioita toivot, että vastaus sisältää. Ideoinnissa aihe voi olla laajempi, mutta jos haluat apua tarkkarajaiseen aiheeseen, voit ohjata aiheen käsittelyä, vaikka ranskalaisilla viivoilla tai kysymyksillä. Voit kertoa myös taustatietoja tilanteesta johon vastausta etsit, ja näin ohjata vastausta oikeaan suuntaan. Tarkempaa informaatiota voit antaa myös saatua ensimmäisen version vastauksesta. Kehotteeseen voi sisällyttää myös rajoituksia, esim. tekstin pituuteen liittyen. Voit myös rajata ulkopuolelle teemoja tai asioita.³¹

”Vastauksessa tulisi olla mukana sanat: putkirikko, häiriö ja vesihuoltolaitoksen henkilökunta.”

”Alueella xx on tapahtunut putkirikko. Kirjoittaisitko luonnoksen kirjeestä alueen asukkaille.”

Tyyli ja muoto

Minkälaisessa muodossa haluat vastauksen? Onko kyseessä asiateksti, kertomus, johdanto, lista vai taulukko? Kohderyhmä ohjaa tekstin tyyliä, mutta voit antaa myös tarkempia määrittelyjä tyylistä. Onko kyseessä asiallinen vai hauska teksti? Useissa tekoälysovelluksissa on näkyvillä myös painikkeet, joilla ohjaat tekstin tyyliä. Tekoäly toimii myös hyvänä kielenkääntäjänä, joten voit pyytää vastauksen haluamallasi kielellä tai vaikka useamman kieliversion samasta tekstistä.³¹

”Käännä teksti englanniksi ja ruotsiksi.”

³¹ Tampereen yliopisto. 2.4.2024. Tekoälyä ohjaavien kehotteiden eli prompttien muotoilu. Luettavissa <https://sites.tuni.fi/vinkkipankki/tekoaly/tekoalya-ohjaavien-kehotteiden-eli-prompttien-muotoilu/>. Viitattu 14.10.2024

4.2 ARVIOINTI JA VALIDOINTI

Koneoppimismallin toiminnan luotettavuutta tarkastellaan arvioimalla ja validoimalla (taulukko 1) ennen mallin varsinaista käyttöönottoa sekä mahdollisten päivitysten yhteydessä. Arvioinnissa käytetään koulutusdatan osaa, jota malli ei ole ennen nähnyt. Validointi on arvioinnin jälkeinen prosessi, jossa mallin suorituskykyä testataan useilla eri datanäytteillä. Näin toiminnan voidaan todeta olevan johdonmukaista, eikä vain satumalta tarkoituksenmukaista yhdellä datanäytteellä. Arviointi ja validointi auttavat ihmisiä luottamaan mallin kykyyn tehdä oikeita ennusteita ja päätöksiä todellisessa maailmassa.^{32, 33}

MALLIN LAADUN ARVIOINTI

OIKEELLISUUS	Kuinka todennäköisesti mallin antama tulos on oikein?
RELEVANTTIUS	Kuinka hyvin malli vastaa dataa?
ROBUSTISUUS	Kestääkö malli häiriöitä, kuten virheellistä tai myrkytettyä dataa?
TURVALLISUUS	Kuinka hyvin järjestelmä sietää ja suojaa mallia vaaralta sekä vahingolta?
YKSITYISYYS	Kuinka hyvin malli suojaa yksityisiä tietoja?
TEHOKKUUS	Kuinka nopeasti malli laskee tuloksen, tai kuinka nopeaa mallin rakentaminen on?
TASAPUOLISUUS	Tekeekö malli päätöksiä oikein ja oikeista syistä?
TULKITTAVUUS	Kuinka helposti ja mihin asti käyttäjä voi ymmärtää mallin tekemän päätöksen perusteet?

Taulukko 1: Mallin laadun arvioinnissa huomioitavia tekijöitä.³³

³² Chowdhary, K. R. 2020. Fundamentals of Artificial Intelligence. Springer Nature Private Limited. New Delhi. Intia. E-kirja. Luettavissa: <https://doi.org/10.1007/978-81-322-3972-7>. Viitattu: 29.4.2024

³³ Sunela, M. 4.6.2024. Tekoälyn hyödyntäminen vesihuollossa. Työpaja 2 - Tekoälyn käytön edellytykset. Etätyöpajan luentosarja.

CASE – TEKÖÄLYN HYÖDYNTÄMINEN OPERATIIVISESSA TEHOKKUUDESSA



Asiantuntijajärjestelmä Koneoppinen

UPM Energy on ottanut käyttöön tekoälyä vesivoimalaitosten toiminnan tehostamiseksi ja ympäristövaikutusten hallitsemiseksi. Yksi merkittävä esimerkki tästä on **tulovirtaaman ennustaminen**, joka on keskeinen osa vesivoimalaitosten operointia. Vesistöjen vedenkorkeuden hallinta on kriittistä, sillä liian korkea vedenpinta voi aiheuttaa tulvia, kun taas liian matala pinta vähentää voimalaitosten tuotantokykyä sekä aiheuttaa muita haittoja. Tähän liittyen tekoäly tarjoaa ratkaisuja.

Ratkaisuna vesivoimalaitosten yläpuolisiin vesistöihin on asennettu IoT-antureita, jotka keräävät tietoa muun muassa sateista, lumien sulamisesta ja muista fysikaalisista ilmiöistä, jotka vaikuttavat veden virtaamaan. **Kerätty data yhdistettynä muuhun saatavilla olevaan dataan, syötetään koneoppimismalleihin, jotka on koulutettu ennustamaan, miten nämä tekijät vaikuttavat tulovirtaamaan.**

Koneoppimismallit toimivat älykkäinä ennustajina, jotka analysoivat säädataa ja vesistön käyttäytymistä. Ne **oppivat historiallisista tiedoista ja pystyvät ennakkoimaan**, milloin kevättulvat saapuvat ja kuinka paljon vettä voidaan turvallisesti juoksuttaa voimalaitoksiin sähköntuotantoa varten. Tämän avulla vesivoimalaitos voi **optimoida sähköntuotantoaan ja varmistaa, että vedenpinta pysyy hallinnassa.**

Tekoälyn käyttö ei kuitenkaan ole pelkkää napin painamista. **Se edellyttää yhteistyötä matemaatikkojen ja substanssiasiantuntijoiden välillä.** On myös tärkeää, että **käytetyt mallit ovat sellaisia, että ihmiset voivat ymmärtää niiden toimintaa ja tuloksia.** Tämä tarkoittaa, että **joskus yksinkertaisempi malli voi olla parempi**, vaikka se ei olisikaan teknisesti kaikkein edistynein.

UPM Energyn tapauksessa **tekoäly on osoittautunut arvokkaaksi työkaluksi, joka auttaa yhtiötä käyttämään vesivaroja tehokkaasti ja ympäristöystävällisesti.** Mallit ovat olleet käytössä jo muutaman vuoden ajan ja ovat osoittautuneet luotettaviksi, mikä on mahdollistanut niiden laajentamisen myös muihin kohteisiin.

-UPM Energy

CASE – ASIAKASPALVELUN TYÖTEHTÄVIEN TEHOSTAMINEN



Luonnollisen kielen käsittely Puhe

”Lahti Aqua on tehostanut asiakaspalvelua tekoälyn sekä ohjelmistorobotin avulla. Laajaan kielimalliin perustuvaa tekoälysovellusta hyödynnetään Lahti Aquassa sähköposteihin vastaamisessa sekä ohjelmistorobottia vesimittarilukematietojen kirjaamisessa asiakaspalvelujärjestelmään.

Lahti Aqua on selvittänyt vuosittain asiakkailtaan mieluisimpia yhteydenottotapoja, joista sähköposti on Minun Aquani -palvelun rinnalla suosituin. Lahti Aquan asiakaspalvelun yhteiskäyttösähköpostiin saapuukin vuosittain noin 12 000 postia.

*Tekoälysovellus muodostaa saapuneisiin sähköposteihin vastausehdotuksen, jonka asiakaspalveluhenkilöstö tarkastaa ja tarvittaessa muokkaa ennen lähettämistä asiakkaalle. **Tekoälysovellus on suunniteltu vuorovaikutteiseen keskusteluun ja sen käyttämään äänensävyyn sekä vastausten pituuteen pystytään vaikuttamaan.** Tekoälysovelluksen kielimalli perustuu OpenAI ChatGPT:n. **Mallia tulee siis kouluttaa, jotta se vastaa vesilaitoksen tarpeisiin.** Mallin kouluttaminen vaatii työpanosta, mutta käyttöönotto on muuten helppoa.*

***Ohjelmistorobotti on tuonut apua vesimittarilukematietojen kirjaamisessa asiakaspalvelujärjestelmään.** Vesihuoltolaitoksen tulot ovat tiukasti sidoksissa vedenkulutustietoihin, joten henkilöasiakkaille on haluttu järjestää mahdollisimman monipuoliset tavat ilmoittaa vesimittarilukemansa. Vesimittarilukeman voikin ilmoittaa Lahti Aquan verkkosivuilta, Minun Aquani -portaalissa, sähköpostitse, puhelimitse sekä postitse lähetettävällä lukukortilla. Minun Aquani -portaalista tieto siirtyy automaattisesti asiakastietojärjestelmään, mutta verkkosivujen ja asiakastietojärjestelmän välinen rajapinta osoittautui haastavaksi, johon apua on saatu ohjelmistorobotilta.*

***Ohjelmistorobotin tehtävänä on kirjata verkkosivuilta sähköpostiin lähetetty lukematiieto asiakaspalvelujärjestelmään.** Kyseessä on opetetun rutiinin mekaanisesta suorittamisesta, jota ohjelmistorobotti hoitaa väsymättä ja virheettä aloittaen työt henkilöstön työajan päättymisen jälkeen ja saaden ne aamuun mennessä valmiiksi. Jatkossa ohjelmistorobotin työtehtäviin on tavoitteena lisätä muitakin asiakaspalveluun liittyviä perustehtäviä kuten omistajanvaihdosilmoitusten kirjaamisia.*

*Tekoälytyökalun sekä ohjelmistorobotin toteutuksessa on hyödynnetty ulkopuolista palveluntarjoajaa. **Ohjelmistorobotin sekä tekoälytyökalun toteuttaminen vaatii kattavaa IT-osaamista ja ymmärrystä eri kielimallien mahdollisuuksista sekä rajoituksista.** Ohjelmistorobotin käyttö edellyttää myös, että **käytössä on vakioidut lomakkeet, joissa tietosisältö on aina samassa paikassa ja sen siirtyminen noudattaa ennalta määriteltyä polkua.** Kun nämä asiat ovat kunnossa, on käyttö helppoa.*

***Asiakaspalvelussa ohjelmistorobotti ja tekoälyavusteiset sähköpostivastaukset vähentävät mekaanista työtä vapauttaen työaikaa muihin tehtäviin.** Lahti Aquan tulevaisuuden tavoitteena on saada tekoälyavusteinen haku myös verkkosivuille.”*

-Lahti Aqua

4.3 TÄRKEIMMÄT NOSTOT: LUONNOLLISEN KIELEN KÄSITTELY (NLP)



Luonnollisen kielen käsittelyn peruseriaatteen

Mahdollistaa tietokoneiden ymmärtävän ja käsittelevän ihmisten käyttämää kieltä. Esimerkkejä ovat älypuhelimien avustajat ja hakukoneet, jotka tulkitsevat ja vastaavat kysymyksiin luonnollisella kielellä.

- Kielimallin toiminta riippuu koulutusdatan laadusta
- Puolueellisuuden välttäminen on tärkeää
- Vääristymien ja ennakkoluulojen riskit huomioitava



Suuret kielimallit

Perustuvat syväoppimistekniikoihin ja valtaviin tekstidatamääriin. Kykenevät suorittamaan monipuolisia tehtäviä, kuten:

- Tekstinsyöttö
- Tiivistelmät
- Käännökset
- Kysymyksiin vastaaminen



Mahdollisuudet ja haasteet

Ihmiskieli sisältää murteita, epäselvyyksiä ja monia merkityksiä, mikä tekee kielen käsittelystä haastavaa. Kehittyneet mallit ja algoritmit pyrkivät käsittelemään näitä haasteita ja mahdollistavat harvinaistenkin kielten käsittelyn.

- Asiakaspalvelun tehostaminen
- Palautteen analysointi
- Automaattinen raportointi
- Tietojen järjestäminen
- Laatu riippuu datan laadusta ja monipuolisuudesta
- Eri kielten ja murteiden ymmärtäminen
- Yksityisyys ja tietosuojat



Prompt engineering

Tarkoittaa hakukoneajattelusta keskustelupohjaiseen tiedonhakuun siirtymistä, jossa käyttäjä muotoilee hakusanansa keskustelunomaisesti.

Hyvä kehote sisältää:

- Selkeän tehtävänannon
- Aiheen
- Kohderyhmän ja roolin
- Kontekstin
- Tyylin ja muodon

5 ENERGIANKULUTUS JA KESTÄVYYS

Vaikka tekoäly voi auttaa optimoimaan energiankäyttöä, sen omat operaatiot kuluttavat merkittäviä määriä energiaa. Tekoälyn hyödyntämisen suosion nopea kasvu on herättänyt huolenaiheita energiankulutuksen ja ympäristövaikutusten suhteen, eikä syyttä. Erityisesti suurikokoiset AI-mallit, kuten Chat GPT, tarvitsevat datakeskuksia valtavine laskentatehoineen. Suuria määriä sähköä kuluttavia datakeskuksia jäähdytetään vedellä, joten ympäristövaikutukset ovat monitahoisia. Pahimmillaan datakeskukset voivat lisätä kasvihuonekaasupäästöjä ja aiheuttaa vedenpuutostilanteita alueilla, joilla datakeskukset sijaitsevat.³⁴

Haasteiden ratkaiseminen vaatii monitahoista lähestymistapaa. Energiatehokkuutta on parannettava, jotta AI-mallit voivat suorittaa tehtäviä tehokkaammin ja samalla kuluttaa vähemmän energiaa. Uusiutuvia energialähteitä on hyödynnettävä datakeskusten energiantarpeen tyydyttämiseksi ja päästöjen vähentämiseksi.³⁴

Tekoälyä tulee hyödyntää kestävin tavoin, jotta sen mahdollistamat edut voidaan saavuttaa ilman haitallisia ympäristövaatimuksia. Tekoälyn avulla voidaan ratkaista monimutkaisia ongelmia, mutta on myös tilanteita, joissa tekoälyn käyttö ei ole välttämättöntä. On tärkeää arvioida, voitaisiinko käsillä oleva asia ratkaista tehokkaasti ilman tekoälyä.³⁵ Esimerkiksi yksi ChatGPT-haku vaatii noin 10-kertaisen määrän energiaa tavalliseen hakukonehakuun verrattuna, synnyttäen saman verran päästöjä kuin teekupillisen keittäminen (kuva 12).³⁴ On pohdittava, olisiko vastaus mahdollista löytää tavallisella verkko-
haulla tai muin perinteisemmin tavoin, ilman tekoälyn käyttöä. Yksinkertaiset haut kielimallista kuluttavat ympäristöä turhaan.^{34, 35}



Kuva 12: Yksi ChatGPT-haku vaatii noin 10-kertaisen määrän energiaa verrattuna tavalliseen hakukonehakuun.³⁴

Lisäksi on hyvä kiinnittää huomiota haun (prompt) asetteluun. Liian ympäripyöreeä haku kielimallilta todennäköisesti tuottaa epämääräisiä tuloksia. Siksi on tärkeää miettiä ja kuvata tarkasti, mitä tekoälyn halutaan tuottavan. Tällä tavoin tarpeettomia hakuja voidaan vähentää ja siten pienentää ympäristövaikutuksia.^{34, 35}

³⁴ A. Shaji George - A.S. Hovan George - A.S.Gabrio Martin. 2023. The Environmental Impact of AI: A Case Study of Water Consumption by Chat GPT. PDF-artikkeli. Luettavissa: <https://doi.org/10.3390/en16020745>. Viitattu: 21.8.2024.

³⁵ Sunela, M. 3.5.2024. Tekoälyn hyödyntäminen vesihuollossa. Työpaja 1 - Tekoäly tutuksi. Etätyöpajan luentosarja. Viitattu: 21.8.2024.

5.1 TÄRKEIMMÄT NOSTOT: ENERGIANKULUTUS JA KESTÄVYYS



Tekoälyn energiankulutus

Tekoälymallit voivat kuluttaa merkittäviä määriä energiaa. Datakeskukset, joissa tekoälymalleja ajetaan, vaativat paljon sähköä ja jäähdytystä.

- On arvioitava, milloin tekoälyn käyttö on tarpeellista ja tuo todellista lisäarvoa
- Tekoäly voi auttaa optimoimaan energian käyttöä
- Mallien tehokkuuden ja toimivuuden kehittämisellä voidaan vähentää energian tarvetta



Ympäristövaikutukset

Datakeskusten energiankulutus voi pahimmillaan lisätä kasvihuonekaasupäästöjä ja aiheuttaa vedenpuutostilanteita.

- Tekoälyn käytön hiilijalanjäljen arviointi
- Uusiutuvien energianlähteiden käyttö
- Datakeskuksien sijoittaminen sopiviin kohteisiin



Energiatehokkuuden parantaminen

Energiatehokkuutta on parannettava, jotta tekoälymallit voivat suorittaa tehtäviä tehokkaammin ja kuluttaa vähemmän energiaa.

- Hyvä kehotteen (prompt) asettelu vähentää tarpeettomia hakuja
- Tekoälyn avulla voidaan ennakoida laitteiden huoltotarpeita ja näin parantaa energiatehokkuutta



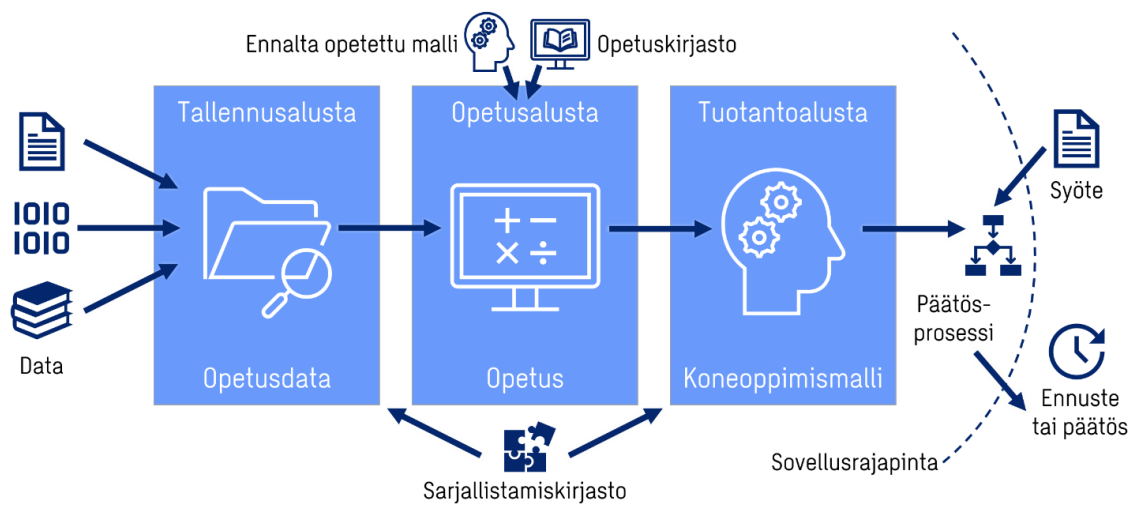
Kestävä tekoälyn käyttö

Tekoälyä tulee hyödyntää kestävin tavoin, jotta sen mahdollistamat edut voidaan saavuttaa ilman haitallisia ympäristövaikutuksia.

- Koulutus ja tietoisuus
- Voisiko käsillä olevan asian ratkaista tehokkaasti ilman tekoälyä?

6 TURVALLISUUS

Tekoälytekniikan nopea kehitys on tuonut mukanaan monia mahdollisuuksia, mutta myös uusia haasteita turvallisuuden, lainsäädännön ja eettisyyden näkökulmasta. Suomessa liikenne- ja viestintävirasto Traficomin Kyberturvallisuuskeskus tuottaa ajantasaista tilannekuvaa tietoturvallisuudesta sekä kehittää ja valvoo viestintäverkkojen turvallisuutta ja toimintavarmuutta. Tekoälyjärjestelmät, erityisesti koneoppimiseen perustuvat, voivat olla alttiita erilaisille hyökkäyksille (kuva 13), kuten mallin manipuloinnille väärällä datalla, mallin varastamiselle, vihamieliselle syönteelle ja sensoreiden sekä päätöksenteon välisen kommunikaation epäonnistumiselle. Nämä riskit voivat vaikuttaa järjestelmän luottamuksellisuuteen, eheyteen ja saatavuuteen.³⁶



Kuva 13: Yksinkertaistettu esimerkki koneoppimisjärjestelmän hyökkäyspinnasta. Mallin käyttökuntoon saamisessa on useita vaiheita. Kunkin vaiheen aikana malli voi altistua erilaisille riskeille.³⁶

³⁶ Aksela, M. - Marchal, S. & Vähä-Sipilä, A. 22.10.2021. Tekoälyn soveltamisen kyberturvallisuus ja riskienhallinta. Liikenne- ja viestintävirasto Traficom. Kyberturvallisuuskeskus. PDF-artikkeli. Luettavissa: <https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/publication/Teko%C3%A4lyn%20soveltamisen%20kyberturvallisuus%20ja%20riskienhallinta.pdf>. Viitattu: 24.6.2024

6.1 ETIIKKA

Eettinen kestävyys on tärkeää ottaa huomioon teknologian kehityksessä ja käytössä. Eettisten standardien noudattaminen voi vaikuttaa uuden teknologian sosiaaliseen hyväksyttävyyteen sekä käyttöönottoon, ja luoda luottamusta kansalaisten, julkisorganisaatioiden, kuluttajien sekä kaupallisten toimijoiden välillä. Teknologian eettisyys ei ole irrallinen yhteiskunnallisista ilmiöistä ja tapahtumista, vaan olosuhteiden muutokset voivat muuttaa ihmisten näkemystä siitä, mikä on hyväksyttävää tai toivottavaa. Yksi esimerkki lähihistoriastamme on COVID-19-pandemia, jonka myötä olemme saaneet lisää tietoa etätyöstä ja -opetusta tukevien sovellusten hyödyistä ja haitoista.³⁷

Lisäksi tekoälyn eettisillä kysymyksillä on poliittisia ulottuvuuksia. On tärkeää pohtia, kenellä on valta vaikuttaa tekoälyn kehitykseen ja käyttöön, kenen intressejä teknologia palvelee ja kuka hyötyy tekoälyn sääntelystä tai sen puutteesta. Näiden kysymysten tarkastelu auttaa ymmärtämään tekoälyn vaikutuksia yhteiskuntaan ja edistämään eettisesti kestävä teknologian kehitystä.^{37, 38}

6.2 LAINSÄÄDÄNTÖ

Tekoälysovelluksien käytön yleistymisen on kasvattanut tarvetta lainsäädännölle, joka varmistaa teknologian eettisen ja turvallisen käytön. EU:n perusoikeuskirja, perustuslait sekä eurooppalaiset ja kansalliset lait ja asetukset muodostavat perustan, jolle tekoälyn käyttöä säännellään. Nämä lainsäädännölliset instrumentit määrittelevät yksilöiden oikeudet ja vapaudet, jotka ovat olennaisia myös tekoälyn eettisessä arvioinnissa. Esimerkiksi yksityisyyden suoja ja tietosuojat ovat toistuvia teemoja tekoälyn etiikkaa käsittelevissä dokumenteissa, ja ne ovat keskeisiä myös lainsäädännössä.³⁸

Euroopan unioni on toiminut edelläkävijänä tässä kehityksessä luomalla EU AI Act -tekoälyasetuksen³⁹, joka asettaa vaatimuksia erityisesti korkean riskin tekoälysovelluksille, kuten kriittisen infrastruktuurin hallintaan ja lainvalvontaan liittyville järjestelmille. Asetus edellyttää riskienhallintaa, ihmisen tekemää valvontaa sekä vakauden ja turvallisuuden varmistamista. Yhdysvalloissa on vastaavasti ehdotettu AI Bill of Rights -lainsäädäntöä.³⁹

On tärkeää, että organisaatiot arvioivat huolellisesti tekoälyyn pohjautuvien ratkaisujen soveltuvuutta kriittisiin käyttötapauksiin ja ottavat huomioon datan ymmärtämisen, saatavuuden ja laadun. Tietoturvan ja tietosuojan näkökulmasta on tärkeää, että tekoälyjärjestelmät ovat läpinäkyviä, vikasietoisia ja tarkkailtavissa. Tämä edellyttää, että tekoälyn kehittäjät ja käyttäjät noudattavat voimassa olevaa lainsäädäntöä ja ottavat huomioon tekoälyn tuomat uudet riskit ja haasteet.³⁸

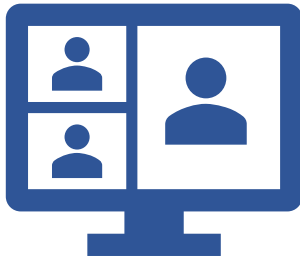
³⁷ Hartikainen, M. 2022. Tekoälyn etiikka. Tampereen yliopisto. Verkkoartikkeli. Luettavissa: <https://projects.tuni.fi/kite/tekoalyn-etiikka/>. Viitattu: 24.6.2024.

³⁸ Aksela, M. - Marchal, S. & Vähä-Sipilä, A. 22.10.2021. Tekoälyn soveltamisen kyberturvallisuus ja riskienhallinta. Liikenne- ja viestintävirasto Traficom. Kyberturvallisuuskeskus. PDF-artikkeli. Luettavissa: <https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/publication/Teko%C3%A4lyn%20soveltamisen%20kyberturvallisuus%20ja%20riskienhallinta.pdf>. Viitattu: 24.6.2024

³⁹ Euroopan Parlamentti. 2023. EU AI Act: first regulation on artificial intelligence. Verkkoartikkeli. Luettavissa: <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>. Viitattu: 20.8.2024.

6.3 TIETOSUOJA

Tietosuoja keskittyy henkilötietojen suojaamiseen ja yksityisyyden kunnioittamiseen. Monet tekoälyjärjestelmät käsittelevät suuria määriä henkilötietoja ja muita arkaluonteisia tietoja, joten tietosuoja on keskeinen huolenaihe.



- GDPR
- Henkilötietojen suojaus
- Tietojen keräämiseen liittyvä informointi
- Vastuu tietoa jaettaessa kolmansille osapuolille

EU:n yleinen tietosuoja-asetus⁴⁰ (GDPR) on merkittävä säädös, joka vaikuttaa tekoälyn käyttöön liittyvään tietosuojaan. GDPR:n 22. artikla antaa yksilöille oikeuksia automaattisen päätöksenteon yhteydessä, mikäli päätöksellä on heihin merkittäviä vaikutuksia. Tämä voi sisältää tekoälyyn perustuvat päätökset. Asetus edellyttää, että yksilöillä on oikeus saada ihmisen tekemä arviointi päätöksestä, ja heillä on oikeus esittää oma näkemyksensä.⁴¹

32. artikla puolestaan määrittelee vaatimukset henkilötietojen käsittelyn tietoturvalle, joka kattaa luottamuksellisuuden, eheyden, saatavuuden ja vikasietoisuuden. Tämä tarkoittaa, että tekoälyjärjestelmien suunnittelijoiden ja käyttäjien on varmistettava, että henkilötietoja käsitellään turvallisesti kaikissa järjestelmän vaiheissa, mukaan lukien opetus- ja validointidatan keruu, tallennus ja itse järjestelmän käyttö.⁴¹

Tekoälyjärjestelmät oppivat datasta, ja opetusdataan vaikuttamalla voidaan vaikuttaa järjestelmän toimintaan. Tämä korostaa opetusdatan hallinnan ja luottamuksellisuuden merkitystä. On tärkeää, että käytetty data on relevanttia, korkealaatuista ja kattavaa, ja että datan pääsynhallinta on selkeää ja oikeasuhteista. Lisäksi tekoälyjärjestelmien kehityksessä käytetään usein ulkopuolisia ohjelmistoja tai kirjastoja, mikä tuo mukanaan toimitusketjuhyökkäyksen riskit. Tämä voi altistaa datan luottamuksellisuudelle ja saatavuudelle riskejä, kuten kiristyshaittaohjelmat.⁴²

Organisaatioiden tulisi huolehtia, että tekoälyjärjestelmät on suunniteltu ja toteutettu tavalla, joka kunnioittaa yksityisyyttä ja suojaa henkilötietoja. Tämä tarkoittaa riittävän teknisen osaamisen tason varmistamista ja palvelumuotoilun huomioimista, jossa tekoälyjärjestelmät rakennetaan tarkasti valittujen syötteiden päälle, välttämällä tarpeetonta datan keruuta. Organisaatioiden on arvioitava huolellisesti tekoälyn pohjautuvien rat-

⁴⁰ Asetus (EU) 2016/679 luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta.

⁴¹ Euroopan Parlamentti. 2023. EU AI Act: first regulation on artificial intelligence. Verkkoartikkeli. Luettu-vissa: <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>. Viitattu: 20.8.2024.

⁴² Aksela, M. - Marchal, S. & Vähä-Sipilä, A. 22.10.2021. Tekoälyn soveltamisen kyberturvallisuus ja riskienhallinta. Liikenne- ja viestintävirasto Traficom. Kyberturvallisuuskeskus. PDF-artikkeli. Luettavissa: <https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/publication/Teko%C3%A4lyn%20soveltamisen%20kyberturvallisuus%20ja%20riskienhallinta.pdf>. Viitattu: 24.6.2024

kaisujen soveltuvuutta ja testattava prototyyppiä todellisissa ympäristöissä ennen laajamittaisen käyttöönoton harkitsemista. On myös tärkeää ymmärtää käytettävän datan laatua ja saatavuutta, sekä ottaa huomioon mahdolliset eettiset kysymykset ja sääntelyyn liittyvät seikat.⁴³

6.4 TIETOTURVA

Tietoturva suojaa tietojärjestelmiä ja verkkoja erilaisilta uhilta, kuten haittaohjelmilta ja tietomurroilta.



- Verkon ja järjestelmien suojaaminen
- Pääsynhallinta ja tietojen suojaus
- Varmuuskopiointi
- Tietoturva-auditoinnit

Riskienhallinta on olennainen osa tekoälysovellusten tietoturvaa. Riskien vaikutus riippuu käyttötarkoituksesta, ja ne ovat usein sidoksissa muihin riskeihin. Organisaatioiden tulee suorittaa säännöllisiä riskianalyyskejä, uhkamallinnusta ja tietoturvan suunnitteluperiaatteiden katselmointia. Tämä auttaa tunnistamaan potentiaaliset riskit ja hallitsemaan niitä tuotekehitysprosessin eri vaiheissa. Tietoturvariskit voidaan jakaa kahteen pääluokkaan: systeemisiin riskeihin ja toteutukseen liittyviin riskeihin. Systeemiset riskit liittyvät tekoälyn toimintaan osana laajempia järjestelmiä ja vuorovaikutuksessa ihmisten kanssa.⁴³

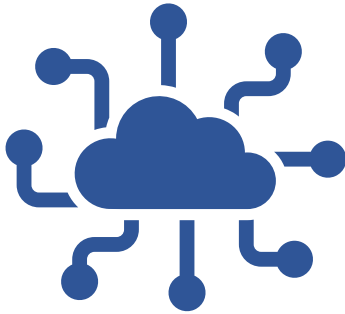
Tietoturvaominaisuudet, kuten selitettävyys, vikasietoisuus ja tarkkailtavuus, ovat keskeisiä tekoälyjärjestelmien turvallisuuden kannalta. Lisäksi on tärkeää huolehtia rajapintojen turvallisuudesta, mukaan lukien tiedonsiirron salaus, palvelunestohyökkäysten torjunta ja rajapintojen valvonta.⁴³

Organisaation sisäisen osaamisen varmistaminen ja selkeän suhteen määrittäminen palvelumuotoiluun ovat avainasemassa. Palvelumuotoilussa tulee keskittyä tarkasti valittuihin syötteisiin ja välttää tarpeetonta datan keruuta. Luottamuksen rakentaminen tekoälyyn perustuville tuotteille ja palveluille on tärkeää, jotta käyttäjät voivat luottaa järjestelmien turvallisuuteen ja eettisyyteen.⁴³

⁴³ Aksela, M. - Marchal, S. & Vähä-Sipilä, A. 22.10.2021. Tekoälyn soveltamisen kyberturvallisuus ja riskienhallinta. Liikenne- ja viestintävirasto Traficom. Kyberturvallisuuskeskus. PDF-artikkeli. Luettavissa: <https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/publication/Teko%C3%A4lyn%20soveltamisen%20kyberturvallisuus%20ja%20riskienhallinta.pdf>. Viitattu: 24.6.2024

6.5 KYBERTURVA

Kyberturvallisuus on laaja-alainen kokonaisuus, joka yhdistää tietotekniikan, viestinnän ja tietoturvan. Tekoälymallien on oltava vastustuskykyisiä, sillä ne voivat joutua kohteeksi erilaisille hyökkäyksille, jotka pyrkivät kiertämään tai vaarantamaan puolustusjärjestelmiä.



- Haavoittuvuuksien hallinta
- Varautuminen ja palautumissuunnitelmat
- Henkilöstön koulutus kyberturvallisuuteen liittyen
- Yhteistyö ja raportointi

On tärkeää huolellisesti arvioida tekoälyyn pohjautuvien ratkaisujen soveltuvuutta kriittisiin käyttötapauksiin, varmistaa datan ymmärtäminen, saatavuus ja laatu sekä testata prototyyppejä todellisissa ympäristöissä ennen laajamittaisen käyttöönoton harkitsemista. Dataan liittyvät haasteet ovat merkittävä tekijä myös kyberturvallisuudessa. Tietoaineistot ovat usein epätasapainossa, ja poikkeamien havaitseminen voi olla vaikeaa.⁴⁴

Tekoälymallien hyökkäyspinta on laajempi kuin perinteisissä järjestelmissä, sillä ne sisältävät opetusdatan tulkin, mallien sisältämät tiedot ja opetusprosessit. Keskeisiä teknisiä riskejä ovat mallin myrkyttäminen, mallin varastaminen, mallin väistämisen ja sensoreiden sekä päätöksenteon välisen kommunikaation epäonnistuminen.⁴⁴

⁴⁴ Aksela, M. - Marchal, S. & Vähä-Sipilä, A. 22.10.2021. Tekoälyn soveltamisen kyberturvallisuus ja riskienhallinta. Liikenne- ja viestintävirasto Traficom. Kyberturvallisuuskeskus. PDF-artikkeli. Luettavissa: <https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/publication/Teko%C3%A4lyn%20soveltamisen%20kyberturvallisuus%20ja%20riskienhallinta.pdf>. Viitattu: 24.6.2024

TÄRKEIMMÄT NOSTOT: TURVALLISUUS



Etiikka

Eettinen kestävyys on tärkeää teknologian kehityksessä ja käytössä. Eettisten standardien noudattaminen luo luottamusta ja vaikuttaa teknologian sosiaaliseen hyväksyttävyyteen.

Pohdittavaa ja huomioitavaa:

- Milloin tekoälyä on sopivaa hyödyntää päätöksenteossa ja esimerkiksi asiakaspalvelussa?
- Kuka päättää tekoälyn käytöstä organisaatiossa?
- Läpinäkyvyys tekoälyn käytössä



Lainsäädäntö

Tekoälysovellusten käytön yleistymisen on lisännyt tarvetta lainsäädännölle, joka varmistaa teknologian eettisen ja turvallisen käytön.

- GDPR
- EU AI Act
- AI Bill of Rights (USA)



Tietosuoja

Keskittyy henkilötietojen suojaamiseen ja yksityisyyden kunnioittamiseen. GDPR antaa yksilöille oikeuksia automaattisen päätöksenteon yhteydessä ja edellyttää ihmisen tekemää arviointia merkittävissä päätöksissä.

Keskeisiä toimenpiteitä:

- Henkilötietojen suojaaminen
- Käyttäjien oikeuksien varmistaminen
- Datan jakaminen ja hyödyntäminen



Tietoturva

Suojaa tietojärjestelmiä ja verkkoja erilaisilta uhilta, kuten haittaohjelmilta ja tietomurroilta.

Keskeisiä toimenpiteitä:

- Verkon ja järjestelmien suojaaminen sopivilla turvatoimilla
- Pääsynhallinta
- Varmuuskopiointi
- Säännölliset tietoturva-auditoinnit



Kyberturva

Yhdistää tietotekniikan, viestinnän ja tietoturvan. Tekoälymallien on oltava vastustuskykyisiä hyökkäyksille.

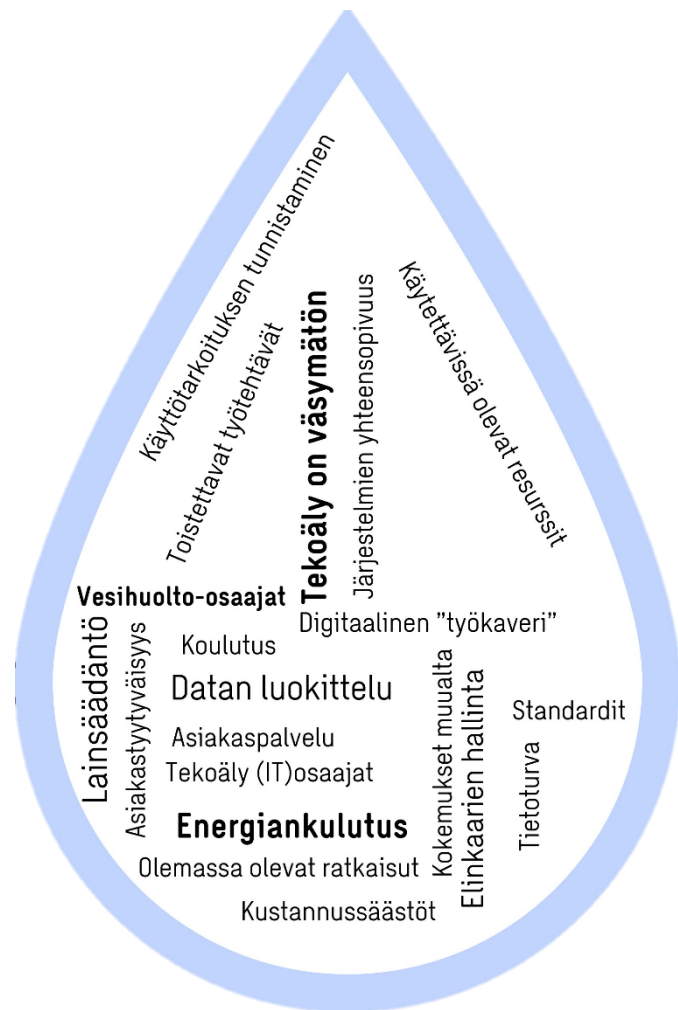
Keskeisiä toimenpiteitä:

- Haavoittuvuuksien hallinta
- Varautuminen, uhka-analyysit ja mahdollisten kyberuhkien tunnistaminen
- Henkilöstön koulutus
- Yhteistyö asiantuntijoiden kanssa

7 YHTEENVETO: TEKOÄLY VESIHUOLLOSSA

Teknologian kehittyessä myös vesihuoltoalalla on alettu hyödyntää yhä enemmän tekoälyä ja koneoppimista, joiden avulla pyritään tehostamaan toimintaa, ennakoimaan ongelmia ja optimoimaan resurssien käyttöä. Tämän julkaisun yhteydessä järjestettiin kolme työpajaa ja toteutettiin tutkimus, jossa haastateltiin useita vesihuoltoalan asiantuntijoita. Haastattelujen tarkoituksena oli selvittää, miten tekoälyä ja koneoppimista voidaan hyödyntää vesihuollossa, mitkä ovat alan nykyiset haasteet ja mahdollisuudet sekä millaista osaamista alalla tarvitaan tulevaisuudessa.

Haastateltaviksi valittiin vesihuoltolaitosten edustajia sekä vesihuoltoalan palveluntarjoajien asiantuntijoita, jotka ovat työskennelleet tekoälyn ja koneoppimisen parissa. Haastatteluiden perusteella tekoälyllä on potentiaalia parantaa vesihuollon tehokkuutta ja reagointikykyä. Lisäksi haastatteluissa selvisi, että alalla on vielä monia haasteita, kuten datan laadun ja keruun ongelmat, jotka on ratkaistava ennen kuin tekoälyn hyödyt voidaan täysin realisoida.



Yksi tärkeimmistä tekoälyyn ja sen hyödyntämiseen liittyvistä tekijöistä on data ja sen laadukkuus. Laadukas ja runsas data mahdollistaa eri tekoälysovellusten hyödyntämisen datan käsittelyssä, kun taas laaduton ja vähäinen data on heikosti tekoälysovellusten hyödynnettävissä. Siksi on tärkeää, että datan keräämisen edellytyksen ovat laitoksella kunnossa ja verkostoista sekä varusteista kerätään dataa toimivilla mittalaitteilla, jotka ovat sijoitettuna oikein verkostoon. Esimerkiksi alueellisessa vesitaselaskennassa etäluettavat vesimittarit sekä jakeluverkoston mittarit antavat yhdessä lähtötietoja laskenta-algoritmeille verkostosta. Datan määrän kasvaessa, olisi laitoksella hyvä olla henkilö, jonka vastuutehtävänä on datan laadusta, tallentamisesta ja säilyttämisestä vastaaminen.

Haastatteluissa ja työpajoissa tekoälyn vahvuudeksi havaittiin tekoälyn hyödyntämispotentiali erityisesti pienemmillä vesihuoltolaitoksilla (taulukko 2). Edellyttäen, että laitoksella on toimivat prosessit datan keräämiseen, voi laitos säästää henkilöstöresursseja hyödyntämällä tekoälyä arkisissa työtehtävissään. Esimerkiksi oikein käytetty ja koulutettu algoritmi oppii tunnistamaan verkostosta epätyypillisiä tapahtumia, kuten putkirik-koja alueella.

Haasteellista tekoälyn hyödyntämisessä on ihmisresurssien puute sekä koulutuksen että käytön näkökulmista. Vaikka tekoälyn on todettu säästävän henkilöstöresurssia esimerkiksi toistettavien työtehtävien automatisoinnissa, on myös havaittu, että tekoälyn käyttöönotto vaatii henkilöstön kouluttamista sekä järjestelmien kehittämistä. Teknologia-alalla onkin havaittu, että monet AI-hankkeet ovat kaatuneet kohonneiden kustannusten vuoksi, koska henkilöstön ja algoritmien koulutus ovat kuluttaneet hankkeissa resursseja. AI-mallien käyttöönotossa onkin huomioitava kustannukset mallien käyttöönotolle, ylläpidolle ja koulutukselle. Nämä kaikki vaiheet kuluttavat sekä henkilöstö- että energioresursseja.

Tekoälyn vahvuus	Käyttötapa	Mahdollisia haasteita
Väsymättömyys ja tarkkaavaisuus	- Suurien tietomäärien jatkuva seuranta ja analysointi - Eroavaisuuksien ja virheiden löytäminen - Jatkuva valvonta - Tietoturva- ja valvonta	- Datan hallinta ja tallennus voi olla kallista - Tieto siitä, miten tekoäly sai vastauksen, puuttuu - Tieto- ja kyberturvariskit - Väärät hälytykset voivat johtaa tarpeettomiin toimiin
Nopeus	- Nopea reagointi ongelmiin, kuten vuotoihin tai laitteiden häiriöihin - Reaaliaikainen asiakaspalvelu chatbotien avulla	Tekoäly muodostaa aina vastauksen, mutta vastaus perustuu todennäköisyyslaskelmiin eikä sillä ole takeita oikeellisuudesta
Havainnointi ihmisaistien ulkopuolelta	- Mittalaitteistojen ja data-analytiikan hyödyntäminen - Sään ja veden laadun seuranta sekä ennustaminen	Laitteiden ja datan laatu voi vaihdella, mikä voi johtaa virheellisiin havaintoihin tai tulkituksiin
Rutiinitehtävien toteutus	- Tietojen syötön automatisointi - Raporttien laatiminen - Aikataulujen ja työtehtävien hallinta - Varaston hallinta	- Liiallinen riippuvuus automaatiosta voi vähentää työntekijöiden osaamista - Muutokset prosesseissa voivat vaatia jatkuvaa koulutusta
Laskeminen	- Talousanalyysit ja budjetointi - Resurssien jakaminen - Aikataulutus - Veden kulutus - Reittisuunnittelu	- Virheelliset algoritmit tai datan syöttö voivat johtaa väärin laskettuihin tuloksiin - Koneoppimismallin yli- ja alioptimointi
Lyhytaikainen muisti	- Väliaikaisen tiedon säilytys ja käyttö - Reaaliaikainen päätöksenteko esim. sensoridatan perusteella	- Rajoitettu muisti voi estää tekoälyä oppimasta pitkäaikaisia kaavoja tai käyttäytymismalleja - Tiedon häviäminen lyhytaikaisessa muistissa voi johtaa epä johdonmukaisuuksiin
Yhtäaikaisten asioiden tekeminen	- Asiakaspalvelussa useiden asiakkaiden käsittely yhtä aikaa - Em. toimien suorittaminen samanaikaisesti vapauttaa henkilöstön muihin tehtäviin	Monimutkaisten tehtävien samanaikainen käsittely voi johtaa resurssien ylikuormitukseen ja heikentää suorituskykyä

Taulukko 2: Hankkeen työpajoissa tunnistettuja tekoälyn vahvuuksia, käyttötapoja vesihuollon alalla sekä huomioitavia mahdollisia haasteita.

Haastatteluissa ja työpajoissa kerättiin kokemuksia ja käytännön ratkaisuja vesihuoltoalan palvelun tarjoajilta Suomessa. Kokemusten perusteella palveluntarjoajilla on ratkaisuja laitoksille liittyen tekoälyn hyödyntämiseen vesihuollossa. Monet ratkaisut hyödyntävät tekoälyn osa-alueita ja esimerkiksi etäluettavat mittarit, data-analytiikka, avoimet järjestelmät, koneoppiminen ja automaatio ovat esimerkkejä käytössä olevista ratkaisuista.

7.1 TULEVAISUUS

Tekoälyn tulevaisuus vesihuollossa näyttää lupaavalta, mutta sen hyödyntäminen vaatii selkeää visiota ja tavoitteita. Organisaatioiden on nähtävä tie, jota pitkin haluavat edetä, ja ymmärrettävä tekniset mahdollisuudet sekä organisaation tarpeet. Tekoälyn avulla voidaan helpottaa ja tehostaa monia työtehtäviä, mutta sen käyttöönotto edellyttää harkintaa ja strategista lähestymistapaa. Tärkeintä on lähteä liikkeelle tarveperusteisesti, eli on ymmärrettävä käsillä oleva ongelma tai tehtävä, jota tekoälyllä halutaan helpottaa. Tämän jälkeen valitaan sopivin työkalu, joka voi perustua esimerkiksi koneoppimiseen.

Tehdyn esiselvityksen perusteella tekoälyä kannattaa hyödyntää erityisesti ennakoivassa kunnossapidossa, verkostojen vuotojen havaitsemisessa ja energiankulutuksen optimoinnissa, missä sen kyky käsitellä suuria datamääriä ja tunnistaa poikkeamia voi parhaimmillaan johtaa huomattaviin säästöihin. Sen sijaan tekoälyn käyttöä ei välttämättä ole järkevää laajentaa alueille, joissa ihmisen asiantuntemus ja paikallistuntemus ovat korvaamattomia.

Tekoäly muodostaa tuloksensa analysoimalla kerättyä dataa tai tunnistamalla kuvioita. Vastaukset perustuvat siis todennäköisyyksiin. Tämä prosessi edellyttää laadukasta dataa, sillä tekoälyn tuottamien tulosten laatu on suoraan verrannollinen lähtödatan laatuun. On siis kriittistä varmistaa, että datan keruu, tallennus ja analysointi ovat korkealaatuisia ja että tiedon laatu on jatkuvan tarkkailun ja parantamisen kohteena. Laadukkaan ja relevantin datan keräämiseen olisi kiinnitettävä entistä enemmän huomioita, jotta tulevaisuudessa pystyttäisiin hyödyntämään tekoälyn potentiaalia ja saamaan esimerkiksi aiempaa tarkempia analyyskejä verkoston todellisesta saneeraustarpeesta.

Tekoälyratkaisuja hyödynnettäessä on aina kiinnitettävä huomioita myös kyber- ja tietoturvallisuuteen. On hyvä muistaa, että tekoälymalleilla hyökkäyspinta on laajempi kuin perinteisissä järjestelmissä. Tekoälyn optimaalinen hyödyntäminen edellyttääkin vesihuollon sekä informaatioteknologian osaajien yhteistyötä.

Tekoäly ei ainakaan lähitulevaisuudessa ole korvaamassa ihmisiä, sillä ihmiset ovat vastuussa tekoälyn ohjaamisesta, sen tulosten tulkinnasta ja päätöksenteosta. Tekoäly on työkalu, joka vaatii asiantuntevaa käyttöä ja jatkuvaa valvontaa. Henkilöstön on sitouduttava kehitystyöhön ja ymmärrettävä, mitä "konepellin" alla tapahtuu, jotta voidaan varmistaa, että tekoäly palvelee vesihuoltoa tarkoituksenmukaisesti ja kestävästi. Inhimillinen älykkyys ja luovuus ovat tulevaisuudessakin avainasemassa, kun tekoälyä sovelletaan vesihuollon tarpeisiin.

LIITE 1 – AI-SANASTO

Suomeksi	Englanniksi	Lyhenne	Kuvaus
Aikasarja-analyysi	Time Series Analysis	TSA	Aikasarjojen analysointi ja ennustaminen
Aikasarjadata	Time Series Data		Aikajärjestykseen sijoitettu data
Algoritmi	Algorithm		Askel askeleelta etenevä laskentamenetelmä
Alisovittaminen	Underfitting		Malli ei sovi tarpeeksi hyvin koulutusdataan
Arviointi	Evaluation		Prosessi, jossa arvioidaan mallin suorituskykyä
Asiantuntija-järjestelmä	Expert System		Järjestelmä, joka jäljittelee ihmisen asiantuntijan päätöksentekoa
Datan louhinta	Data Mining		Hyödyllisen tiedon etsiminen datasta
Datan luokittelu	Data Classification		Prosessi, jossa data jaetaan luokkiin
Datan merkitseminen	Data Labeling		Prosessi, jossa dataan lisätään merkintöjä
Ennakoiva analytiikka	Predictive Analytics		Analytiikka, joka ennustaa tulevia tapahtumia datan perusteella
Ennuste	Prediction		Arvio tulevasta tapahtumasta
Hyperparametri	Hyperparameter		Parametri, joka asetetaan ennen koulutusta
Kapea tekoäly	Narrow AI		Tekoäly, joka on suunniteltu suorittamaan tiettyjä tehtäviä
Kehotesuunnittelu	Prompt Engineering		Hakukoneajattelusta keskustelupohjaiseen tiedonhakuun siirtyminen
Konenäkö	Computer Vision	CV	Tietokoneiden kyky tulkita ja ymmärtää visuaalista tietoa
Koneoppiminen	Machine Learning	ML	Algoritmit, jotka oppivat datasta
Koulutus	Training		Algoritmeja opetetaan tunnistamaan ja ennustamaan tiettyjä malleja tai käyttäytymistä syöttämällä niille suuria määriä dataa
Koulutusdata	Training Data		Data, jota käytetään mallin kouluttamiseen
Koulutusdata	Training Data		Data, jota käytetään mallin kouluttamiseen
Kuvantunnistus	Image Recognition		Tekniikka, joka tunnistaa ja luokittelee kuvia

Luokittelu	Classification		Datan luokittelu ennalta määriteltuihin luokkiin
Luonnollisen kielen käsittely	Natural Language Processing	NLP	Tietokoneet ymmärtävät ihmiskieltä
Malli	Model		Matemaattinen esitys prosessista
Massadata	Big Data		Suuret ja monimutkaiset tietomassat
Neuroverkko	Neural Network	NN	Laskennallinen malli, joka jäljittelee ihmisaivoja
Neuroverkko-regressio	Neural Network Regression		Koneoppimisen menetelmä, jossa käytetään neuroverkkoja ennustamaan jatkuvia arvoja
Ohjaamaton koneoppiminen	Unsupervised Learning		Oppiminen ilman merkittäviä tietoja
Ohjattu koneoppiminen	Supervised Learning		Oppiminen merkittävien tietojen avulla
Ohjelmointirajapinta	Application Programming Interface	API	Ohjelmointirajapinta, jota käyttäen sovellukset kommunikoivat keskenään.
Oletus	Assumption		Oletus tai ennakoarvio
Piilokerrokset	Hidden Layers		Neuroverkon kerrokset, jotka eivät ole suoraan yhteydessä syötteeseen tai tulokseen
Poikkeavuuksien havaitseminen	Anomaly Detection		Poikkeavien arvojen tunnistaminen
Puoliohjattu koneoppiminen	Semi-supervised Learning		Yhdistää ohjatun ja ohjaamattoman oppimisen piirteitä
Puolirakenteinen data	Semi-structured Data		Data, joka sisältää sekä rakenteellisia että vapaamuotoisia osia
Päätöspuu	Decision Tree	DT	Päätöksentekoa tukeva puumalli
Rakenteeton data	Unstructured Data		Data, jolla ei ole ennalta määrättyä rakennetta
Rakenteinen data	Structured Data		Data, joka on järjestetty tiettyyn muotoon
Regressio	Regression		Suhteen mallintaminen muuttujien välillä
Ristiinvalidointi	Cross-validation		Tekniikka mallin suorituskyvyn arviointiin
Robottiikka	Robotics		Teknologia, joka käsittelee robottien suunnittelua ja käyttöä
Ryhmittely (klusterointi)	Clustering		Datan ryhmittely samankaltaisuuden perusteella
Satunnaismetsä	Random Forest	RF	Useiden päätöspuiden yhdistelmä

Suositusalgoritmi	Recommendation Algorithm		Analysoi dataa tarjotakseen kohdennettuja suosituksia
Suuret kielimallit	Large Language Models	LLM	Syväoppimistekniikoilla koulutettuja suuria kielimalleja, jotka ymmärtävät ja tuottavat ihmiskieltä
Synapsi	Synapse		Neuronien väliset yhteydet
Syväoppiminen	Deep Learning	DL	Edistynyt koneoppiminen neuroverkoilla
Syöte	Input		Tieto, joka syötetään järjestelmään
Tekoäly	Artificial Intelligence	AI	Simuloi ihmisen älykkyyttä
Testidata	Test Data		Data, jota käytetään mallin testaamiseen
Tukivektori-kone	Support Vector Machine	SVM	Luokittelualgoritmi, joka erottaa luokat
Tuotos	Output		Tieto, joka saadaan järjestelmästä
Ulottuvuuksien vähentäminen	Dimensionality Reduction	DR	Datan ulottuvuuksien vähentäminen
Vahvistusoppiminen	Reinforcement Learning	RL	Tekoälymallin oppiminen kokeilemalla ja palautteen avulla
Validointi	Validation		Prosessi, jossa varmistetaan mallin toimivuus
Validointidata	Validation Data		Data, jota käytetään mallin validointiin
Yhdistelmä-oppiminen	Ensemble Learning	EL	Useiden mallien yhdistäminen
Yleinen tekoäly	General AI	AGI	Tekoäly, joka kykenee suorittamaan mitä tahansa inhimillistä tehtävää
Ylisovittaminen	Overfitting		Malli sopii liian hyvin koulutusdataan