



# **Keinoäly ja koneoppiminen informaation visualisoinnissa**

Oulun yliopisto  
Tietojenkäsittelytiede  
LuK-tutkielma  
Esa Juujärvi  
2024

## Tiivistelmä

Olemme ihmisenä tottuneet keräämään tietoa talteen ja hyödyntämään sitä jokapäiväisessä elämässämme. Tiedon määrä on kasvanut ja viimeisten vuosikymmenien aikana sen määrä on suorastaan räjähtänyt uusien teknologisten palveluiden ja uusien älylaitteiden tultua markkinoille. Tuotamme päivittäin yhä enemmän paikkatietoa, kuvia, ääntä, videota sekä dataa sosiaalisesta mediasta, verkossa olevista kulkulaitteista, pyöristä, autoista ja muista kulkuvälineistä. Mittaamme atomien värähtelystä kaukaisimpia mustia aukkoja mitä moninaisimmilla laitteilla. Tiedon määrän kasvulle ei lähitulevaisuudessa näy loppua, mutta pystymmekö hyödyntää tietoa parhaalla mahdollisella tavalla päätöksen teossa ja ihmiskunnan kehityksen veturina. Oma ymmärryksemme saavuttaa rajat hyvin nopeasti kun puhumme tiedon määrästä, jota päivittäin keräämme ja tuotamme globaalisti ja reaaliaikaisesti.

Myös tapa käsitellä tietoa on kehittynyt. Konesaleissa ja laitteissa tapahtuu paljon, mitä sellaisenaan emme pysty hyödyntämään inhimillisessä ajattelussa. Miten voimme ymmärtää tietoa, joka ylittää ymmärryksemme rajan?

Tutkielmassani olen käynyt läpi 21 artikkelia ja tieteellistä julkaisua, joissa käsitellään tietoa, sen käsittelyä ja muokkaamista ihmiselle ymmärrettävään visuaaliseen muotoon. Pidän aihetta erittäin ajankohtaisena ja myös useissa julkaisuissa on ymmärretty sen keskeinen luonne. Tiedon ymmärtämisen problematiikka on läsnä lähes jokaisessa arkisessa askareessamme sekä maailman kannalta keskeisiä päätöksiä laatiessa. Ilman ymmärrystä emme voi toimia rationaalisesti parhaalla mahdollisella tavalla.

### **Avainsanat**

Keinöäly, koneoppiminen, big data, visualisointi

### **Ohjaaja**

FT, yliopistonlehtori Leena Arhippainen

## Lyhenteet

AI	Artificial Intelligence / keinoäly / tekoäly
DL	Deep learning / syväoppiminen
IoT	Internet of Things / esineiden internet
ML	Machine learning / koneoppiminen
NLP	Natural language processing / luonnollisen kielen käsittely
GPS	Global Positioning System / maailmanlaajuinen paikannusjärjestelmä
WiFi	Wireless Fidelity / langaton lähiverkko
ChatGPT	Generative Pre-trained Transformer / tekoälypohjainen kielimalli
VEVF	Video-based effective visualization framework / videopohjainen tehokas visualisointikehys
JSTOR	Journal Storage / akateeminen digitaalinen kirjasto
IEEE Xplorer	Institute of Electrical and Electronics Engineers (IEEE) provides a digital library called IEEE Xplore. / Institute of Electrical and Electronics Engineers (IEEE) -järjestön tarjoama digitaalinen kirjasto englanniksi
ACM	Association for Computing Machinery / tietojenkäsittelytieteiden järjestö

# Sisällysluettelo

Tiivistelmä .....	2
Lyhenteet .....	3
Sisällysluettelo .....	4
1. Johdanto .....	5
1.1. Tausta ja motivaatio .....	6
1.2. Tutkimuskysymys ja -menetelmä .....	7
1.3. Tutkielman rakenne .....	8
2. Keskeiset käsitteet .....	9
2.1. Massadata .....	9
2.2. Tietojärvi .....	10
2.3. Keinoäly .....	10
2.4. Koneoppiminen .....	10
2.5. Tiedon louhinta .....	11
2.6. Tiedon visualisointi .....	11
3. Tiedon visualisoinnin käyttötarkoituksia .....	12
3.1. Tiedon visualisointi päätöksenteon tukena .....	12
3.2. Tiedon visualisointi logistiikassa ja liikkumisessa .....	12
3.3. Visualisointi lääketieteessä ja vakuutusallalla .....	13
3.4. Urheiludatan visualisointi .....	13
3.5. Työkaluja tiedon visualisointiin .....	14
3.6. Digitaalisten kirjastojen ja tietovarastojen visualisointi .....	14
4. Löydökset ja pohdinta .....	16
5. Yhteenveto .....	18
Lähteet .....	19

# 1. Johdanto

Brittiläinen matemaatikko Clive Robert Humby totesi vuonna 2006 datan olevan uusi öljy (The Economist, 2017). Kuten öljyä pidetään teollisuuden kehityksen mahdollistajana, myös datalla on samanlaisia ominaisuuksia informaation sekä teknologian aikakauden voittokulun veturina. Vertaus viittaa siihen, että kuten öljyssä, myös datassa on suunnaton potentiaali, jonka voi valjastaa käyttöön jatkojalostamalla tuotetta kuhunkin käyttötarkoitukseen sopivaksi. Sekä öljynjalostus että datan analysointi vaativat erikoistaitoja ja tekniikoita. Datan analysoinnissa keinoäly ja datan visualisointi ovat keskeisiä tekniikoita, joilla datasta saadaan tietoa, ja sitä kautta ymmärrämme toimintaympäristöämme paremmin (Kuva 1).



**Kuva 1.** Tutkielman lähtökohtana on miten voisimme ymmärtää mitattua tietoa parhaalla mahdollisella tavalla visuaalisessa muodossa.

Digitaalinen maailma, joka koostuu kaikista tietokoneiden, sensoriverkkojen, GPS-/WiFi-sijaintitietojen, verkkometatietojen, verkkolähteisten henkilötietojen, mobiililaitteiden, älykkäiden yhdistettyjen laitteiden ja seuraavan sukupolven sovellusten luomista tiedoista, muuttaa tapaamme havainnoida ympäröivää maailmaamme. Ennen näkemätön ja eksponentiaalinen tietojen kasvu tarjoaa yrityksille uusia ja ainutlaatuisia mahdollisuuksia ja haasteita. Tiedon määrän kasvaessa sen analysointi vaikeutuu ja kokonaisvaltaista kuvaa voi olla hankala hahmottaa. Tämä on ongelma niin markkinajohtajille kuin pienemmillekin toimijoille. (Zwolenski & Weatherill, 2014.)

On vahvaa näyttöä siitä, että liiketoiminnan suorituskykyä voidaan parantaa merkittävästi tietoon pohjautuvan päätöksenteon avulla. Tiedon hankkiminen, tallentaminen ja käsittely mahdollistaa päätösten tekemisen automaattisesti massiivisessa mittakaavassa (Provost & Fawcett, 2013). Ymmärrettävästi myös monessa muissa tapauksissa, kuin vain liiketoiminnassa, tiedon oikeanlainen hyödyntäminen helpottaa päätöksentekoa ja tuottaa parempia lopputuloksia. Tutkielmassa esitän kuinka datan käyttö ulottuu elämän joka osa-alueille kaupunkien ja valtioiden päätöksenteosta yksittäisten ihmisten käyttäytymiseen. Lääketieteestä kuntoiluun ja logistiikasta tuotteiden ja palveluiden kehittämiseen. Käyttömahdollisuudet ovat lukemattomia.

Tiedon ymmärtämisen tärkeys on nostettu esiin mm. Johannessen ja Berntzen tutkimuksessaan. Siinä esitetään lähtökohtaista mallia oivalluksen järjestelmästä, jossa voitaisiin perustaa data-analytiikan laboratorio tuomaan tietoa rekistereistä, avoimista tietolähteistä, osallistumistiedoista ja IoT:stä. Laboratoriossa käytettäisiin erilaisia koneoppimisanalyyseja ja tulokset esitettäisiin datan visualisointitekniikoilla, jotta päätöksentekijöillä olisi parempi pääsy rikkaampaan tietoon (Johannessen & Berntzen, 2021).

Ihmisen näköjärjestelmällä kestää alle 150 millisekuntia käsitellä monimutkaisia kuvia (Thorpe et al., 1996). Onkin siis erittäin loogista, että kaikki mahdollinen informaatio, mitä maailmassa tuotetaan, pitäisi voida esittää ihmisille nopeasti ymmärrettävässä visuaalisessa muodossa. On totta, että koneiden ja ihmisten välillä voidaan kommunikoida myös tekstuaalisesti tai esimerkiksi myös äänen välityksellä, mutta perehdyin tutkielmassani visuaaliseen kommunikaatioon sen merkittävyyden takia.

Datan visualisointi on kvantitatiivisen tiedon esittämistä graafisessa muodossa (Islam & Jin, 2019). Kaikki toimialat ja kaikenkokoiset yritykset hyötyvät datan visualisoinnista ja sen tuomasta liiketoiminta-avusta (Islam & Jin, 2019). Datan visualisointia voidaan hyödyntää myös kaikilla muilla elämän osa-alueilla. Esimerkiksi data-analyysissä, tieteellisessä tutkimuksessa, sosiaalipalveluissa, tiedon tunnistuksessa (Li & Qu, 2021).

## 1.1. Tausta ja motivaatio

Informaation visualisointia on tutkittu paljon. Onhan tietoa havainnollistettu kuvamateriaalilla ja grafiikoilla kivekautelta alkaen. Yksi tutkimuksissa esiin noussut kuuluisa kaavio on ranskalaisen insinöörin Charles Joseph Minardin luoma kaavio Napoleonin itärajan tappiosta vuonna 1821 (Li & Qu, 2021). Emme siis ole uuden asian äärellä, mutta tiedon digitalisoituminen ja valtava määrä asettaa tiedon kuvaamisen aivan uudenlaisen haasteen edelle. Tuottamaamme tietoon tulee myös tuottaa tapoja hyödyntää sitä keinoälyn ja uusien algoritmien avulla.

Perinteisesti informaatio on esitetty erilaisilla graafeilla ja kaavioilla esimerkiksi erilaisilla taulukoilla, kartoilla, viiva-, vesiputous-, pylväs-, piirakka-, mittari-, piste-, spiraali-, alue- sekä kuplakaavioilla (Mohammed et al., 2022). Tutkimuksissa löysin yllättävän vähän tietoa uudenlaisista malleista, joita nykyteknologioilla olisi mahdollista toteuttaa kuten animaatiot, videot tai interaktiivinen informaatiografiikka.

Koneoppimista ja keinoälyä enemmänkin on tutkittu analysoimaan videoa tai kuvaa kuin tuottamaan sitä informaation pohjalta.

Tietomäärän kasvaessa voimme valjastaa uusimmat teknologiset innovaatiot tuottamaan visualisointeja, joista voimme helposti havaita isoja kokonaisuuksia, trendejä sekä poikkeamia.

## 1.2. Tutkimuskysymys ja -menetelmä

Tutkielmassani käsitellään keinoälyn hyödyntämistä tiedon visualisoinnissa. Keinoälyä on suhteellisen uusi tutkimusalue sovelluksineen. Tiedon käsittelyssä sitä on käytetty jo vuosia ja monet ohjelmistot hyödyntävät sitä huomaamatta taustaprosesseissaan. Vaikka keinoälyllä voidaan havaita isosta tietomassasta toistuvia kuviota, niiden suhteita ja poikkeamia ohjelmistoille ymmärrettävässä muodossa, on laajan materiaalin kommunikointi ihmismielelle ymmärrettävässä muodossa haastavaa. Kokonaiskuvan hahmottaminen pelkkien loputtomalta tuntuvien listojen, taulukoiden tai isojen tietomassojen avulla on sellaisenaan lähes mahdotonta.

Tutkielmassani on käytetty narratiivista kirjallisuuskatsausta (Denney & Tewksbury, 2013), jossa pyritään ymmärtämään tutkimuksen kohdetta keräämällä ja analysoimalla ei numeerista tietoa, kuten artikkeleita ja tieteellisistä julkaisuista. Käytin tutkimuksessa 21 lähdettä, joiden pohjalta pyrin ymmärtämään ja kirjoittamaan tämän päivän informaation visualisoinnista mahdollisimman kattavan kokonaiskuvan. Lähteiden rajallinen määrä teki kokonaiskuvan ymmärtämisestä hieman puutteellista, mutta löysin niistä kuitenkin samankaltaisia huomioita, joiden pohjalta pystyin tutkielmani kirjoittaa ja pohdintaa suorittaa.

Hain Scopuksesta yleisillä hakusanoilla kuten “data visualization” enintään kymmenen vuotta vanhoja artikkeleita. Kokeilin myös toisilla hauilla kuten “machine learning” OR “artificial intelligence” AND “information” OR “data” OR “big” AND “data”. Ensimmäisellä sain 34,210 tulosta ja toisella 260,945 tulosta. Lisäsin hakusanan “visualization or visualisation” hakuun ja rajasin tulokset koskemaan ainoastaan artikkeleiden ja tieteellisten julkaisujen otsikoita, sain tuloksen tarkentumaan 877 hakutulokseen. Näistä tuloksista löysin suurimman osan lähteistäni. Rajasin saamaani tulosmassaa vielä muutamilla eri hauilla löytääkseni tietoa yksittäisistä aiheista kuten “knowledge information”. Tällä haulilla löysin mielenkiintoisen artikkelin (Leung et al., 2022) aiheesta “Visualization and Visual Knowledge Discovery from Big Uncertain Data”, jossa keinoälyn avulla suuresta tietomassasta pyritään etsimään tietoa (Knowledge) pelkän informaation sijaan. Aineisto on kerätty Covid-19 liitetystä terveysdatasta (Leung et al., 2022). Tätä lähestymistä kaipasinkin enemmän löytämäni materiaaliin, miten informaatiosta voidaan teknologisesti tuottaa tietoa, jota visualisoimalla mahdollistetaan isompien kokonaisuuksien ymmärtämistä.

### 1.3. Tutkielman rakenne

Tutkielman toisessa luvussa esittelen keskeiset käsitteet, kolmannessa luvussa tiedon visualisoinnin käyttötarkoituksia. Kaikista käsitteistä ei ole vakiintunutta suomenkielistä käännöstä olemassa ja olen käyttänyt tekstissäni suomenkielisten käsitteiden lisäksi englanninkielisiä käsitteitä. Neljännessä luvussa käsittelen löydöksiä ja pohtinut, löytyykö niistä vastaus esitettyihin tutkimuskysymyksiin. Viides luku on yhteenveto.



## 2. Keskeiset käsitteet

Tähän lukuun olen kerännyt lähteissä käytettyjä keskeisiä käsitteitä. Monista käsitteistä ei ole vakiintuneita termejä suomen kielellä, siksi käytän käsitteissä myös paljon englannin kieltä.

### 2.1. Massadata

NASAn tutkijat käyttivät termiä "Big Data" ensimmäisen kerran vuonna 1997 kuvaamaan vaikeutta näyttää tietomassaa, joka on liian suurta mahtumaan tietokoneen päämuistiin. Muut organisaatiot ovat sittemmin omaksuneet termin (Mohammed et al., 2022). Käsite on sittemmin laajentunut ja sitä käytetään monella tapaa kuvaamaan yleensä ottaen tietoa, jonka prosessointi ja tarkempi analysointi sen eri ominaisuuksien johdosta on hankalaa. Tallennustilan ja muistin rajallisuus eivät ole enää keskeinen ongelma tiedon käsittelyssä. Mohammed et al. (2022) kuvaavat omassa tutkimuksessa Big Dataa neljällä ominaisuudella:

- Volume (tiedon määrä): Tuotetun tiedon kokonaismäärä on valtava.
- Velocity (tiedon nopeus): Tietoa syntyy ja kertyy suuria määriä erittäin nopeasti.
- Variety (tiedon monimuotoisuus): Tietotyyppejä on erittäin paljon.
- Vagueness (tiedon epämääräisyys): Tieto ei ole täsmällistä ja se sisältää paljon epäselvyyksiä.

Neljän V:n kategorisointia on laajennettu vastaamaan tarkempaa ymmärrystä Big Datan luonteesta ja määritelmät poikkeavat hieman eri tutkimuksissa. Esimerkiksi Rozić et al., (2023) jakaa Vagueness -käsitteen kahteen osaan: Veracity sekä Value.

- Veracity (tiedon luotettavuus): Tiedon on oltava luotettavaa ja tarkkaa, jotta sitä voidaan käyttää.
- Value (tiedon arvo): Kuinka arvokasta tieto on käytön kannalta. Tiedon tulee tukea käyttötarkoitusta.

Leung et al., (2022) lisää edelliseen jaotteluun vielä kaksi V:tä: Validity ja Visibility.

- Validity (tiedon kelvollisuus): Kuinka tarkasti tieto vastaa sitä mihin tarkoitukseen se on kerätty.
- Visibility (tiedon saatavuus): Kuinka helposti tieto on nähtävissä ja saatavilla.

## 2.2. Tietojärvi

Data Lake eli “tietojärvi” tai “datalampi” kuvaa tilannetta, jossa informaatiota on erittäin suuria määriä ja perinteisillä keinoilla sitä on erittäin vaikea analysoida. Viime vuosien nopea kehitys on johtanut suurien tietomassojen tuottamiseen ja siten datalampien syntyyn. Valtava määrä tietoa tuottaa haastetta kaikilla aloilla merkityksellisen tiedon erottamisessa. Keinoälyn ja koneoppimisen tekniikoiden avulla siitä kuitenkin voidaan tuottaa luotettavaa tietoa eri käyttötarkoituksiin. Visualisointi on noussut olennaiseksi työkaluksi monimutkaisen datan tehokkaassa esittämisessä. Tietojärvet ovat suosittu tapa hallita ja tallentaa laajaa määrää tietoa. Tieto on tietojärvissä erilaisissa muodoissa ja se tarjoaa joustavan tavan ympäristön tietojen käsittelyyn, yhdistämiseen, analysointiin ja visualisointiin (Rozić et al., 2023).

## 2.3. Keinoäly

Keinoäly on kauan kehitteillä ollut teknologia, joka viime aikoina on saanut suurta huomiota eri sovellusten kuten ChatGPT ja MidJorneyn ansiosta. Termi keinoäly keksittiin kuitenkin ensimmäisen kerran tietojenkäsittelytieteilijä John McCarthyn toimesta jo vuonna 1955 Dartmouth-konferenssissa, jossa hän ja muut tutkijat loivat pohjaa alan kehitykselle. McCarthy, yhdessä Marvin Minskyn, Nathaniel Rochesterin ja Claude Shannonin kanssa, katsotaan usein tekoälyn perustajaisiksi (McCarthy et al., 1955. s.13).

Vaikka keinoäly konseptina on jo vuosikymmeniä vanha, on viimeaikainen kehitys tehnyt siitä merkittävän teknologisen ratkaisun tämän päivän tietojenkäsittelylle, jonka haasteena on nimenomaan suurien tietomassojen käsittely. Esimerkiksi uudet keinoälyteknologiat, kuten luonnollisen kielen käsittely (NLP - Natural language processing), koneoppiminen (ML - Machine learning) ja syväoppiminen (Deep learning) voivat automaattisesti tunnistaa suuria tietomassoja koskevia kaavoja ja suhteita ja tuottaa niistä visualisointeja (Rozić, et al., 2023).

## 2.4. Koneoppiminen

Koneoppiminen on tekoälyn osa-alue, joka mahdollistaa tietokoneille oppimisen ja analyysin teon ilman nimenomaista ohjelmointia. Koneet parantavat suoritustaan kokemuksen kautta. Se pyrkii luomaan matemaattisia malleja ja mukauttamaan niiden parametreja käyttäen dataa. Se yhdistää tilastoteoriat ja tietokonealgoritmit (Järvinen et al., 2021. s.135).

Koneoppimista on käytetty menestyksekkäästi eri aloilla, kuten tiedonhaku, tietojen louhinta, puheentunnistus, tietokonegrafiikka, visualisointi ja ihmisen ja tietokoneen vuorovaikutus. Usein koneoppimismallit ovat kuitenkin käyttäjille mustia laatikoita, mikä hankaloittaa niiden ymmärtämistä ja kehittämistä. (Liu et al., 2017. s.4). Koneoppiminen jaetaan ohjaamattomaan ja ohjattuun oppimiseen, parametrisiin ja ei-parametrisiin menetelmiin sekä käyttää monipuolisesti syviä neuroverkkoja tietojen

analysoinnissa (Järvinen et al., 2021. s.135). Koneoppiminen on keskeisesti yhteydessä big dataan ja tiedon louhintaan.

## 2.5. Tiedon louhinta

Tietojen louhinnan tavoitteena on hyödyllisen tiedon hankkiminen suurista tietomassoista. Järvinen et al. (2021) ovat kategorisoineet tiedon louhinnan seuraavasti: Tutkivaan datan analysointiin, kuvaavaan mallintamiseen, ennustavaan mallintamiseen sekä kaavojen ja sääntöjen löytämiseen. Sen avulla voidaan paljastaa arvokasta tietoa, joka auttaa päätöksenteossa ja ennustamisessa eri aloilla, kuten liiketoiminta, terveydenhuolto ja tieteellinen tutkimus.

## 2.6. Tiedon visualisointi

Tiedon visualisoinnin tavoitteena on tehdä tiedosta helposti ymmärrettävää ja esittää monimutkaisia rakenteita selkeästi grafiikan avulla. Nykyään päätöksenteon edellytys on ymmärtää kerätystä aineistosta päätöksiin vaikuttavat seikat. Erilaisilla tiedon visualisointityökaluilla helpotetaan organisaatioita ja ihmisiä ymmärtämään isompia kokonaisuuksia ja tekemään parempia johtopäätöksiä tai luomaan uusia kysymyksiä. Uudet teknologiat ja käyttäjäystävälliset ohjelmistot ovat tehneet dataan pohjautuvien päätöksen teon helpommaksi (Islam & Jin, 2019).

Perinteisten graafien ja kaavioiden lisäksi tietoa voidaan visualisoida myös interaktiivisina kuvina, animaatioina tai videona. Interaktiivisuus mahdollistaa tiedon sisällä navigoimisen, avata yksittäisiä tietoelementtejä helposti näkyvämpään muotoon sekä muuttaa reaaliaikaisesti visualisointia arvoja muuttamalla. Ihmisäivot ymmärtävät grafiikkaa paremmin kuin monimutkaista numeraalista dataa (Islam & Jin, 2019). Visuaalisesti esitetyn datan pohjalta voidaan helpommin ja nopeammin tehdä johtopäätöksiä ja analyyskejä sekä omaksua tietoa. Tiedon esittäminen ymmärrettävästi myös demokratisoi tiedon.

### 3. Tiedon visualisoinnin käyttötarkoituksia

Tiedon visualisoinnin käyttötarkoituksia on rajaton määrä. Olen kerännyt tähän lukuun keskeisiä käyttötarkoituksia.

#### 3.1. Tiedon visualisointi päätöksenteon tukena

Aiemmassa keräämissäni tutkimuksissa tiedon visualisointia oltiin tutkittu tarkemmin määrittelyissä ympäristöissä, ei niinkään yleisellä tasolla. Tutkimuksen kohteina olivat muun muassa kaupungit ja miten big dataa voidaan hyödyntää kaupunkien muuttuessa älykaupungeiksi sekä kestävässä kehityksessä (Lavalle et al., 2020). Tutkimuksessa käsitellään datan suurta ja vaihtelevaa määrää, jollaista kertyy eri antureiden ja laitteiden tuottamana niin kutsutussa esineiden internetissä (Internet of Things, IoT).

Tietojen erilainen luonne ja tiedon yhdistely tekee niiden käsittelystä monimutkaista, joka puolestaan vaatii erilaisten tekniikoiden soveltamista tietojen ymmärtämisessä. Uusien teknologioiden vallankumous on nousemassa tiedon käsittelyn haasteen ympärille ja se mahdollistaa uuden ja edistyksellisen teollisuuden synnyn sekä luo lukuisia liiketoimintamahdollisuuksia ja auttaa parantamaan kaupunkien kestävästä kehitystä tehokkuuden ja resurssienhallinnan kehittyessä. Kehitystä on kuitenkin vaikea saavuttaa elleivät päätöksen tekijät saa selkeää ja ymmärrettävää näkemystä tiedosta, jonka pohjalta tarvittavia päätöksiä kestävästä kehityksestä edesauttamiseksi tehdään. Lavalle ja hänen kollegansa (2023) ovat tutkimuksessaan ehdottaneet lähestymistapaa, joka parantaa tiedon keräämistä päätöksenteon tueksi esineiden internetin dataa hyödyntävien älykkäiden kaupunkien kontekstissa. Tutkimus auttaa käyttäjiä löytämään parhaan tiedon visualisoinnin äärelle, johon voidaan sisällyttää kaikki älykaupungin ja sen infrastruktuurin tuottama tieto reaaliaikaisesti strategisten ja taktisten päätösten tekemiseksi. Tutkimus auttaa myös ymmärtämään tekoälyalgoritmien tuloksia ja tarjoaa tarvittavan kontekstin luottamuksen synnyttämiseksi tuloksia kohtaan. Isosta tietomäärästä saadaan oikeilla menetelmillä prosessin kokonaiskuva, jota ilman käytäntöjä olisi vaikea ymmärtää. Käytännön data kerättiin palolaitoksen puhelinpalveluista sekä muusta kaupungin tuottamasta tiedosta.

#### 3.2. Tiedon visualisointi logistiikassa ja liikkumisessa

Logistiikka ja liikkuminen on keskeinen tutkimuksen kohde kaupunki-infrastruktuuria tutkittaessa. Lee ja kollegansa (2023) ovat tutkineet Big Dataa ja älykästä liikkumista omassa artikkelissaan. Erilaiset kuljetusmuodot ovat yhteydessä internetiin, ja uusia kokeiluja on lisääntynyt vastaavasti. Erityisesti älykkään liikkuvuuden ja IoT:n, tekoälyn ja suurten tietojen analyysitekniikoiden yhdistelmä muuttaa suurelta osin

liikkuvuuttamme. Artikkelissa analysoitiin ja arvioitiin erilaisia älykkään liikkuvuuden tuottamia tietoja hyödyntäviä viimeaikaisia tutkimuksia. Artikkelissa todettiin, että teknologisella kehityksellä on sekä positiivisia että negatiivisia puolia. Positiivisia puolia olivat liikkumisen mukavuus, turvallisuuden parantuminen, nopeat reittisuositukset ja saapumisaikojen tarkka ennustaminen. Negatiivisia puolia olivat puolestaan esimerkiksi työpaikkojen menetys kuljetusyrityksissä ja epäselvä vastuunjakaminen kuljettajien ja autonvalmistajien välillä. On tarpeellista käyttää uusimpia eksponentiaalisesti kehittyviä teknologioita asianmukaisesti ja miettiä sen vaikutusta kulttuurisen kehityksen muutoksiin.

Useat autovalmistajat ovat alkaneet tuottaa autoja, jotka ovat yhteydessä toisiinsa ja toimivat verkossa. Lähitulevaisuudessa kaikki logistiikka on liitetty verkkoon ja liikenteessä olevat laitteet jakavat sijaintinsa, määränpänsä, nopeutensa ja muuta tietoa eri tarkoituksiin, kuten liikennetiedon keräämiseen ja optimointiin. Tämä johtaa kasvavaan riskiin esimerkiksi henkilötietojen tietovuotoihin, kun ajoneuvot ovat kytkeytyneinä julkisiin verkkoihin. Autonomisten ajoneuvojen yleistyessä, tarvitaan ratkaisuja niiden hakkeroinnin, luvattoman toiminnan ja esimerkiksi avaimien kloonauksen estämiseksi. Verkossa olevien ajoneuvojen jakaman sijainnin avulla liikenteen sujuvuuden optimointi voi poistaa tarpeen liikennevaloille. Liikenteen sujuvuusalgoritmien lisäksi tarvitaan tarkempaa kuvankäsittelytekniikkaa ja nopeita viestintäjärjestelmiä. Suurten tietomäärien ja tekoälyn odotetaan tekevän merkittävän panoksen älykkään liikkuvuuden alalle. Älykkäiden liikkuvuuslaitteiden keräämien suurten tieto- ja informaatiomäärien käsittely tekee käyttäjien liikkumisesta kätevää ja turvallista sekä optimoi liikenteen hallinnan. Leen et al. (2023) tutkimuksessa ei käsitellä juurikaan tiedon visualisointia, mutta sisältää paljon muuta aihetta koskevaa tutkimusta.

### 3.3. Visualisointi lääketieteessä ja vakuutusallalla

Kaupungit ja ihmisten luomat logistiset kokonaisuudet ovat yksi merkittävimmistä aloista, jossa tiedon visualisoinnissa on hyötyä ymmärtämään isompia kokonaisuuksia ja luomaan toimivampia ympäristöjä ihmisten tarpeisiin. Lääketiede on myös erittäin merkittävä ala, josta vastaavaa tutkimusta ja käytännön sovelluksia on paljon. Lääketieteelliset tutkimukset kuitenkin painottuvat monesti enemmän lääketieteellisiin seikkoihin kuin tiedon visualisointiin niin ollen käyttänyt kyseisen alan tutkimuksia omassa tutkielmassani hieman vähemmän.

Myös valmiiksi datavetoisilla aloilla, kuten esimerkiksi vakuutusallalla on herätty miten ison datan ymmärtäminen ja analyysijärjestelmä tukee vakuutusallalla toimijoita automaattisesti havaitsemaan poikkeama suuressa tietomäärässä (Reis et al., 2022).

### 3.4. Urheiludatan visualisointi

Urheilussa keinoälyllä on analysoitu sekä visualisoitu urheiludataa. Dataa kertyy niin laskennallisista tiedoista kuin esimerkiksi urheilusuorituksen tarkkailusta kuvien ja videoiden avulla. Liu ja kollegat (2023) ovat omassa tutkimuksessaan käsitelleet

vuorovaikutteisen ja automatisoidun datan analyysin tärkeyttä monimutkaisen tiedon käsittelyssä ammattilaisurheilussa. Urheilijoiden seurantadatan kerääminen ja analysointi on yleistä, tarkoituksena arvioida pelaajien suoritusta, vähentää loukkaantumisriskiä ja analysoida sopeutumisreaktioita. Artikkelin ehdottaa videopohjaista tehokasta visualisointikehystä (VEVF), joka perustuu tekoälyyn ja suurten tietomäärien analysointiin. Artikkelin käyttää koneoppimismenetelmää luokitellakseen urheiluvideoita erottamalla sekä videoiden ajalliset että avaruudelliset piirteet. VEVF-malli, joka yhdistää konvoluutioneuroverkot ja ajalliset kokoavat kerrokset saavuttavat lupaavia tuloksia aikaan (Liu et al., 2023).

### 3.5. Työkaluja tiedon visualisointiin

Julkaisuissa ja artikkeleissa nousi esiin sovellusalueiden ohella erilaisia työkaluja datan visualisointiin. Ymmärryksen ja näkemyksen luominen grafiikoiden, kaavioiden tai animaatioiden perusteella riippuu paljon siitä, miten dataa on kerätty, prosessoitu ja visualisoitu. Kumar ja Shawkat (2022) antavat omassa tutkimuksessaan alustavan arvion tällä hetkellä saatavilla olevista suosituimmista Big Datan visualisointityökaluista. Tutkimuksessa vertailtiin kolmea eniten käytössä olevaan analytiikkatyökalua, Power BI, Tableau sekä Qlik View. Tutkimuksen vertailutestin perusteella näiden kolmen data-visualisointityökalun välillä Power BI saa parhaat pisteet liitettävyydestä, datanhallinnasta, upotetuissa analyyseistä, lisätystä analytiikasta, mobiili yhteensopivuudesta, käyttöjärjestelmästä, teknisestä tuesta sekä tietoturvasa. Tableau erottuu edukseen myös datanhallinnassa, datan visualisoinnissa, raportoinnissa, kyselyissä, käyttöjärjestelmätuessa, upotetuissa analyyseissä ja maantieteellisessä analytiikassa. Datahallinta, kyselyt, IoT ja mobiili analytiikka sekä tietoturva ovat kaikki alueita, joilla Qlik View erottuu edukseen. Kumarin ja Shawkatin (2022) tutkimus keskittyi ainoastaan eniten käytettyihin visualisointityökaluihin ja se rajoittuu hakutuloksiin Google trendien perusteella ja näin ollen joitain artikkeleita voi olla jäänyt huomioimatta. Työ kuitenkin tunnistaa joukon avoimia kysymyksiä, jotka voivat ohjata tulevaa tutkimusta data-visualisoinnin alalla.

Citespace on tietovarantojen visualisointiin tarkoitettu ohjelmisto, jota käytetään tieteellisen kirjallisuuden analysointiin ja yhteystietojen havainnollistamiseen. Se auttaa käyttäjiä tunnistamaan kirjallisuudesta keskeisiä suuntauksia, yhteyksiä ja tutkimushaaroja, ja se on erityisen hyödyllinen kirjallisuuskatsausten ja tutkimuksen suunnittelun yhteydessä. Citespace tarjoaa visuaalisia työkaluja ja graafisia esityksiä tiedon selkeyttämiseksi ja helpottamiseksi. Tämä ohjelmisto auttaa tutkijoita ja akateemisia ammattilaisia ymmärtämään paremmin, mitä aiheeseen liittyvä kirjallisuus sisältää ja miten tieteelliset käsitteet ovat vuorovaikutuksessa keskenään (Wang & Lu, 2019, s. 8173).

### 3.6. Digitaalisten kirjastojen ja tietovarastojen visualisointi

Kuten datan analysointi- sekä visualisointityökalut ovat kehittyneet isoin harppauksin, myös digitaaliset kirjastot ja tietovarastot, kuten JSTOR, IEEE Xplorer ja ACM:n

digitaalinen kirjasto ovat yleistyneet ja kehitys on tehnyt niistä yhä enemmän yleisölle saatavia. Asiakirjojen etsiminen avainsanojen perusteella on tehokas tapa saada kohdennettua tietoa. Suuren tietomäärän vuoksi käyttäjien on kuitenkin haasteellista oppia ja saada uutta tietoa. Viime vuosina paljon huomiota on kiinnitetty tekstiaineistojen louhintatekniikoihin, jotka analysoivat automaattisesti suuren määrän tekstidataa laskentatehon avulla. Aiheenmallinnus on yksi tällainen tekniikka. Pohjautuen ohjaamattomiin koneoppimismalleihin, aiheenmallinnus voi löytää teemallisen rakenteen kuvaamaan, mitä aiheet ovat ja miten aiheet ja asiakirjat liittyvät toisiinsa tilastollisesta näkökulmasta (Yang et al., 2017, s. 40).

Artikkelissaan Yang ja kollegat (2017) esittelevät visuaalisen analyysijärjestelmän VISTopic:n, joka hyödyntää vuorovaikutteisia visualisointitekniikoita suurten asiakirjakokoelmien ymmärtämiseen aiheenmallinnuksen avulla. Todennäköisyyden mukaan esitettyinä aiheen mallit ovat vähemmän ihmislukutaitoisia ja se tekee tiedon löytämisestä tehottomampaa, koska käyttäjien on tulkittava jokaista aihealuetta ja aiheasiakirjasuhdetta numeeristen arvojen perusteella koko mallin tulosten ajan. Usein myös tarvitaan linkitys aiheiden ja muiden tietojen, kuten aikojen, ihmisten ja paikkojen, välillä oivallusten löytämiseksi asiakirjakokoelmista. Kuitenkin näitä näkökohtia on saatavilla vain harvoissa tekstiaineiston kaivuutyökaluissa, mikä aiheuttaa eron tekstiaineiston kaivuutekniikoiden ja käytännön vaatimusten välillä.

## 4. Löydökset ja pohdinta

Tutkielman tarkoituksena oli saada kattava tieto siitä, miten edistynyttä teknologiaa hyödynnetään tiedon esittämisessä visuaalisessa muodossa nykyään, sekä minkälaisia työkaluja ja konsepteja tiedon visualisoinnin ympärille on luotu ja miten niitä on hyödynnetty eri elämän osa-alueilla.

Tutkimuksia yhdistää ajatus, jossa tietodon tiedetään olevan olemassa, sitä kerätään tehokkaasti, mutta sen hyödyntämiseen tarvitaan uusia tekniikoita kuten koneoppimista sekä keinoälyä tunnistamaan tiedosta isompia kokonaisuuksia, poikkeamia ja säännönmukaisuuksia. Tiedon ymmärtämiseen ja sen jatkohyödyntämiseen täytyy tieto esittää ihmiselle ymmärrettävässä muodossa. Johannessen ja Berntzen (2021) esittävätkin omassa tutkimuksessaan järjestelmää, joka mahdollistaisi pääsyn rikkaampaan teitoon ja sitä kautta parempaan ymmärrykseen.

Suurin osa tutkimuksista nostaa esille ongelmaksi datan suuren määrän ja sen nopean lisääntymisen (Zwolenski & Weatherill, 2014; Mohammed et al., 2022). Myös datan moninaisuus nousee esille ja se taas tuo oman haasteensa datan käsittelyyn. On siis selvää, että datan prosessointiin tarvitaan tehokkaita prosesseja, jotta saamme muokattua siitä ihmiskäyttöön sopivaa tietoa.

Monissa tutkimuksissa kuitenkin esitettiin datan visualisointiin perinteisiä työkaluja ja malleja (Mohammed et al., 2022; Islam & Jin, 2019). Suurin osa datan käytöstä yritysmaailmasta päätöksen tekoon käyttääkin jo olemassa olevia hyväksi havaittuja kaavioita ja graafeja (Mohammed et al., 2022; Islam & Jin, 2019). Näen, että nykyteknologialla voitaisiin pureutua tietoon vielä syvällisemmin ja esittämään se myös monipuolisemmin ja havainnollisemmin niin että kokonaiskuvan ymmärtäminen olisi entistä helpompaa. Joissain kohdin tutkimuksia mielestäni tehdään hieman keinotekoisia käsitteitä, esimerkiksi Big Datan V-alkuiset määritelmät (Mohammed et al., 2022; Rozić et al., 2023; Leung et al., 2022). Toki on ymmärrettävää, että teknologian kehittymisen myötä myös määrittelyitä pitää päivittää.

Olemme teknologisen harppauksen kynnyksellä. Monet esimerkit todistavat, miten uutta tietoisuutta voidaan tuottaa nykyteknologialla. Isot mallit tarkentuvat ja asioiden ymmärrettävyys ja ennustettavuus paranevat. Voimme havaita asioita, joiden olemassaolosta ei aikaisemmin ollut minkäänlaista viitettä. Ymmärrämme paremmin kvanttitaso tapahtumia ja toisaalta voimme mallintaa ympäröivää avaruuttamme ja sen tapahtumia ennen näkemättömällä tarkkuudella. Voimme tehdä ilmastomalleja, minkä soisi ohjaavaan ihmiskunnan kehitystä kestävämpään suuntaan. Uusien tautien ilmaantuvuuteen voimme reagoida nopeammin kuin koskaan. Tämä kaikki vaatii erittäin laajan tietomäärän ymmärtämistä ja vaikka valitettavan moni uusi teknologia ja sen sovellukset tähtäävätkin taloudelliseen voiton maksimointiin, uskon että tieto lisää tuskan lisäksi myös syvällisempää ymmärrystä maailma ja ihmiskunnan tilasta ja antaa meillä uusia eväitä kokonaisvaltaisempaan ajatteluun.



Kuten tutkielmassa on yleisesti havaittu, tiedon määrä, sen paikkansapitävyys ja vaihtelevuus ovat rajoittaneet osaltaan myös tutkielmaani. Tietoa on valtavasti tarjolla, ja tutkielman suppeus antaa vain pintapuolisen näkemyksen aiheesta. Jo muutaman vuoden vanha tutkimus voi olla tänään vanhentunutta tietoa teknologian kehittyessä nopeutuvalla tahdilla. Materiaalin paljouden takia voisi olla viisampaa keskittyä jatkotutkimuksissa pienempiin osa-alueisiin. Keinoälyn merkitys sekä tiedon käsittelyssä että informaation visualisoinnissa ovat jo itsessään niin laajoja alueita, että molemmista aiheista voisi kirjoittaa useammankin erillisen tutkielman.

## 5. Yhteenveto

Tärkeimmät tulokset ovat tuoda esiin datan visualisoinnin tärkeys nykymaailmassa. Ilman datan jalostamista ymmärrettävään muotoon, sillä ei ole inhimilliseen toimintaan kontribuutiota.

Kuten monessa muussakin asiassa uskon, että poikkitieteellisyys liittyen tutkittuihin teknologioihin olisi suotavaa. Monissa tutkimuksissa voidaan todeta, että tutkimus on siiloutunutta, eikä esimerkiksi teknologisesta lähtökohdasta olla ymmärretty viestinnällistä tarvetta sekä viestinnällisessä tutkimuksessa teknologista. Asiaa voidaankin siis lähestyä monesta suunnasta. Viestinnällisestä sekä teknologisesta. Näiden leikkauspisteessä muodostuu kattava ymmärrys nykyisen teknologian mahdollistamasta tiedosta, joka voidaan pukea niin että se hyödyntää ihmistä ymmärtämään paremmin ympäristönsä ja toimimaan enemmän tietopohjaisesti suurten kysymysten äärellä. Keinoäly ja sen ympärille kasvaneet teknologiset innovaatiot tulevat vaikuttamaan ihmiskunnan tulevaisuuteen monella muullakin tapaa ja ongelmat sekä sen hyödyt eivät liity ainoastaan teknologioihin eikä viestintään.

Käsitlemme aihetta joka läpäisee koko yhteiskunnan toimintaa. Tutkimusta voidaan syventää aina kun uutta tietoa ja uusia tapoja käsitellä tietoa syntyy. Nykytutkimuksen valossa tiedon määrän kasvu ei lopu – päin vastoin, sen määrän lisääntyminen tulee kasvamaan näköpiirissämme olevaan tulevaisuuteen saakka.

## Lähteet

- Denney, A.S. & Tewksbury, R. (2013) How to Write a Literature Review. *Journal of Criminal Justice Education*, 24:2, 218-234, DOI: 10.1080/10511253.2012.730617
- Islam, M. & Jin, S. (2019). An Overview of Data Visualization.
- Johannessen, M. R. & Berntzen, L. (2021). Smart City - smart data? Towards a holistic system of insight for data analytics in smart cities.
- Järvinen, P., Siltanen, P. & Kirschenbaum, A. (2021). *Data Analytics and Machine Learning*. s. 129-146.
- Kumar, A & Shawkat Ali, A. B. M.(2022). Big Data Visualization Tools, Challenges and Web Search Popularity - An Update till Today. *Big Data Intelligence and Computing, International Conference, DataCom (2022)*, 305-315
- Lavalle, A., Teruel M. A., Maté, A. & Trujillo J. (2020). Improving sustainability of smart cities through visualization techniques for Big Data from iot devices.
- Lee, D., Camacho, D. & Jung, J. J. (2023). Smart Mobility with Big Data: Approaches, Applications, and Challenges.
- Leung, C. K. Madill, E. W. R. & Pazdor, A. (2022). Visualization and Visual Knowledge Discovery from Big Uncertain Data. *2022 26th International Conference Information Visualisation (IV)* s. 330-335.
- Li, X., Qu, M. (2021). Product Information Visualization Systems in the Era of Intelligence. *3rd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM)*. s. 514-517.
- Liu, A., Mahapatra, R. P. & Mayuri, A. V. R. (2023). Hybrid design for sports data visualization using AI and big data analytics. *Complex & Intelligent Systems (2023)* 9:2969–2980.
- Liu, S, Wang, X., Liu, M. & Zhu, J. (2017). Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics 1* 48–56.
- McCarthy, J., Minsky, M. L., Rochester, N. & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine Volume 27 Number 4 (2006)*
- Mohammed, L. T., AlHabshy, A. A. & ElDahshan, K. A. (2022). Big Data Visualization: A Survey. *4th International Congress on Human-Computer Interaction, Optimization and Robotic Applications, Proceedings*
- Provost, F. & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making.

- Reis, T., Kreibich, A., Bruchhaus, S., Krause, T., Freund, F., Bornschlegl, M. X. & Hemmje, M. L. (2022). An Information System Supporting Insurance Use Cases by Automated Anomaly Detection. *Faculty of Mathematics and Computer Science, University of Hagen*
- Rozić, R., Slišković, R. & Rosić, M. (2023). Artificial Intelligence for Knowledge Visualization: An Overview. *Digital Transformation in Education and Artificial Intelligence Application, MoStart 2023, CCIS 1827. pp. 118–131.*
- The Economist. (2017, May 6). The world's most valuable resource is no longer oil, but data. <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>
- Thorpe, S., Fize, D. & Marlot, C. (1996) Speed of processing in the human visual system. *Nature (London) 1996-06, Vol.381 (6582), p.520*
- Wang, W., Lu, C. (2019). Visualization analysis of big data research based on Citespace. *Soft Computing (2020) 24:8173–8186.*
- Yang, Y., Yao, Q., Qu, H. (2017) VISTopic: A visual analytics system for making sense of large document collections using hierarchical topic modeling. *Visual Informatics 1 (2017) 40–47.*
- Zwolenski, M. & Weatherill, L. 2014. The digital universe: Rich data and the increasing value of the internet of things.