



101 LAKES

Tekoäly juristin toiminnassa

Tkt, OTK, dos, Olli Pitkänen

Kuka puhuu?



TKT, OTK, dos OLLI PITKÄNEN

Olen koko urani työskennellyt tietotekniikkaan, dataan ja tekoälyyn liittyvien oikeudellisten kysymysten parissa

Perustin jo 1994 Opplex-nimisen lakitoimiston, sittemmin asiantajotoimiston, joka keskittyi atk-alaan.

Myöhemmin toimin vuosia opettajana ja tutkijana yliopistoissa sekä asiantuntijana eri yrityksissä.

Nyt 2019 alkaen 1001 Lakes Oy:n osakkaana ja vastannut yrityksen oikeudellisista palveluista.

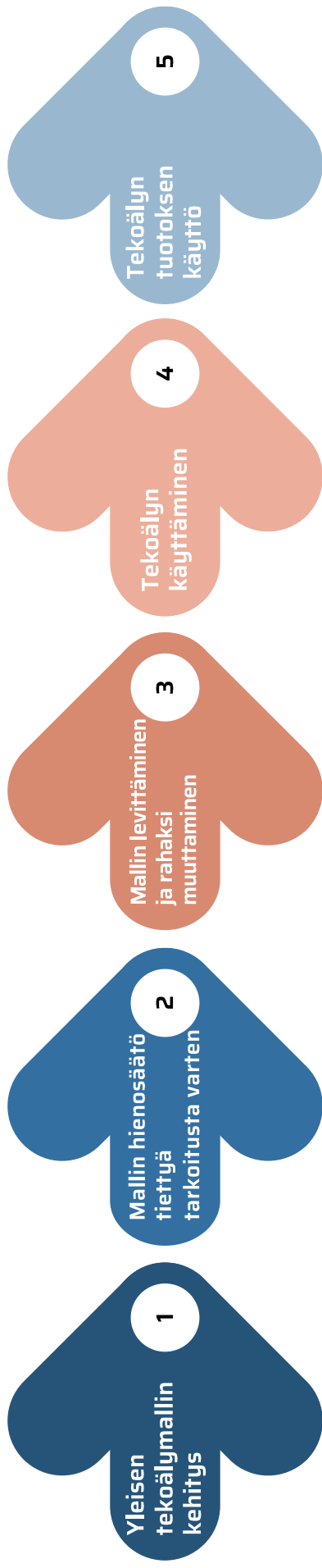
1001 Lakes tarjoaa monenlaisille yrityksille datan jakamiseen ja hyödyntämiseen sekä tekoälyn liittyviä palveluita. Kansainvälistä teknis-oikeudellista asiantuntemusta.

www.linkedin.com/in/pitkanenolli
www.1001lakes.com

101 LAKES



Tekoälyn elinkaari



LLM:n kaltaisen yleisen tekoälymallin kehittämiseen tarvitaan suuria määriä dataa eri lähteistä. Se voi sisältää esimerkiksi tekijänoikeuden suojaamia teoksia tai henkilötietoja, joiden käyttämiseen saatetaan tarvita lupa.

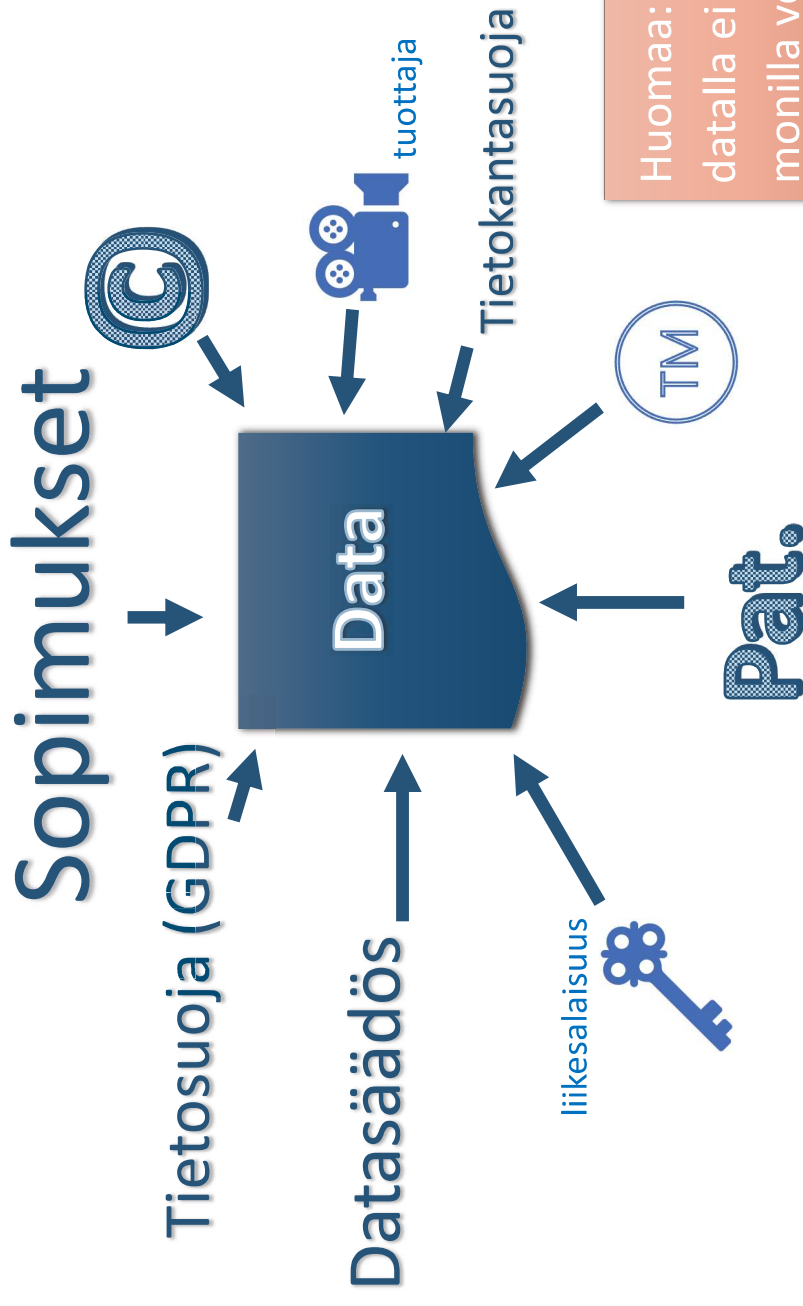
Yleisen mallin hienosäätämiseksi tiettyä tarkoitusta varten tarvitaan lisää dataa eri lähteistä. Niihin kohdostuu usein eri tahojen oikeuksia.

Malli voi olla erittäin arvokas. Sen jakaminen ja rahaksi muuttaminen on usein kannattavaa. Mutta suojaavatko sitä mitkään oikeudet?

Tekoälyjärjestelmän käyttäminen edellyttää, että järjestelmään syötetään tietoja. Nämä tiedot voivat olla peräisin eri lähteistä ja niihin voi kohdistua oikeuksia.

Erityisesti generatiiviset tekoälyjärjestelmät, mutta myös muut tekoälyjärjestelmät voivat tuottaa arvokasta tietoa. Suojaako sitä mikään oikeus?

Mitä oikeuksia voi kohdistua dataan?



Huomaa: oikeudellisessa mielessä dataalla ei ole yhtä omistajaa, vaan monilla voi olla siihen erilaisia oikeuksia



Musiikin, kuvien, elokuvan, tv-ohjelmien, kaunokirjallisuuden, asiatekstien, arkkitehtuurin... **kaikenlaisen luovan työn oikeudellinen suoja**

Teos: mikä tahansa **luovan** työn tulos, **itsenäinen** ja **omaperäinen**

Automaattinen suoja: ei tarvitse rekisteröidä, merkitä ©-merkkiä tms.

Ei suojata aihetta, motiivia, ideaa, periaatteita, algoritmeja jne.

Taloudelliset oikeudet: kopiointi, levitys, taloudellinen hyödyntäminen

Moraaliset oikeudet: isyysoikeus, respektioikeus jne

Tekijä on se ihminen joka on luonut teoksen

Ei kone, yritys tai oppilaitos

Tekijöitä voi olla monta

Tekoäly ei ole tekijä – jos sen tuotos vaikuttaa luovalta, on kysyttävä, kenen luovuudesta on kyse

Tekijänoikeus on voimassa 70 vuotta tekijän kuolinvuodesta

Työsuhteessa luodun teoksen oikeudet siirtyvät usein automaattisesti työnantajalle

Tietokoneohjelmat ja tietokannat: Tekijäl 40 b §

Muut teokset: vakiintunut käytäntö, toimialan edellyttämä laajuus

Ei koske yliopistojen ja korkeakoulujen tutkijoita ja opettajia ellei erikseen sovita – yleensä sovitaan.

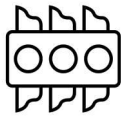
Arvioitava liittyykö teoksen luominen työtehtäviin

Esim. aika, paikka, välineet eivät ratkaise

Taloudellisista oikeuksista voi sopia,
moraalisista oikeuksista voi luopua vain rajoitetusti

- Tekijänoikeuslaissa, mutta monilta osin lähellä teollisoikeuksia
 - Investointien, kilpailun yms. suojaa
- Sekava kokoelma kapeita oikeuksia ilman selkeitä perusteita
 - Tietokanta- ja luettelosuojaa
 - Investointi suureen tietomäärään – tekoälymallit?
 - Tuottajat
 - Musiikki-, TV- ja elokuvatuottajat, ei esim. pelituottajat
 - Signaalin suoja
 - Radio- ja televisioyhteyksien lähettämä signaali on suojattu, muiden ei ole.
 - Esittävä taiteilija
 - Näyttelijät ja laulajat saavat suojaa, akrobaattit ja urheilijat eivät saa.
 - Valokuva
 - Kaikki valokuvat, jopa automaattisesti otetut, saavat suojaa, muut kuvat eivät saa.
- Digitaalisessa maailmassa epäjohdonmukaisesti joitain tietoja suojataan lähtökohteilla ja toisia ei.
- Korostamalla teollisoikeudellisia piirteitä ja esim. investointien suojaa, lähtökohteista voisi rakentaa järkevän suojan datalle.





Saako teoksia käyttää tekoälyn kouluttamiseen?

- EU:n DSM-direktiivin 4 artiklassa (tekijänoikeuslaki 13 b §) säädetään rajoituksesta, joka koskee teosten kopiointia tekstin- ja tiedonlounhintaa varten, jollei tekijä ole nimenomaisesti ja asianmukaisella tavalla pidättänyt tätä oikeutta (**“TDM opt-out”**).
- Artiklan soveltuvuutta tekoälyn kouluttamiseen on sanamuodon perusteella epäilty, mutta esim tekoälyasetus näyttäisi vahvistavan soveltuvuuden.
- Siispä ilmeisesti se, jolla on laillinen pääsy teokseen, saa kopioida teoksen tekoälyn kouluttamista varten ja säilyttää kopioita tätä tarkoitusta varten, jollei tekijä ole nimenomaisesti ja asianmukaisella tavalla pidättänyt tätä oikeutta.

ODRL, ISCC ja muut välineet oikeudenhallintaan tekoälyn kouluttamisessa

Standardeja, kuten ODRL, TDM-AI, tunnisteita (esim. ISAN, ISNI ja ISCC) sekä muita teknisiä välineitä voidaan käyttää apuna oikeuksienhallinnassa.

- **Open Digital Rights Language (ODRL)** on yleisessä käytössä oleva kieli, jolla voidaan kuvata koneluettavassa muodossa erityisesti digitaalisten sisältöjen oikeuksia ja käyttöehtoja.
- **TDM-AI** on protokolla, jolla voidaan välittää oikeudenhallintatietoja järjestelmästä toiseen.
- **ISCC** on tunniste, joka voidaan automaattisesti laskea digitaalisesta sisällöstä ja joka kertoo myös kahden tiedoston samankaltaisuudesta.



Liikesalaisuudet

Liikesalaisuus on tietoa:

- joka ei ole yleisesti tunnettua tai helposti saatavissa selville ja
- jolla siksi on taloudellista arvoa elinkeinotoiminnassa ja
- jonka suojaamiseksi laillinen haltija on ryhtynyt kohtuullisiin toimenpiteisiin.

Tekoällyn käytössä on huolehdittava sekä koulutusvaiheessa että käytössä riittävästä tietoturvasta liikesalaisuuksien ja muiden luottamuksellisten tietojen suojaamiseksi!

Sääntelytrendit isossa kuvassa

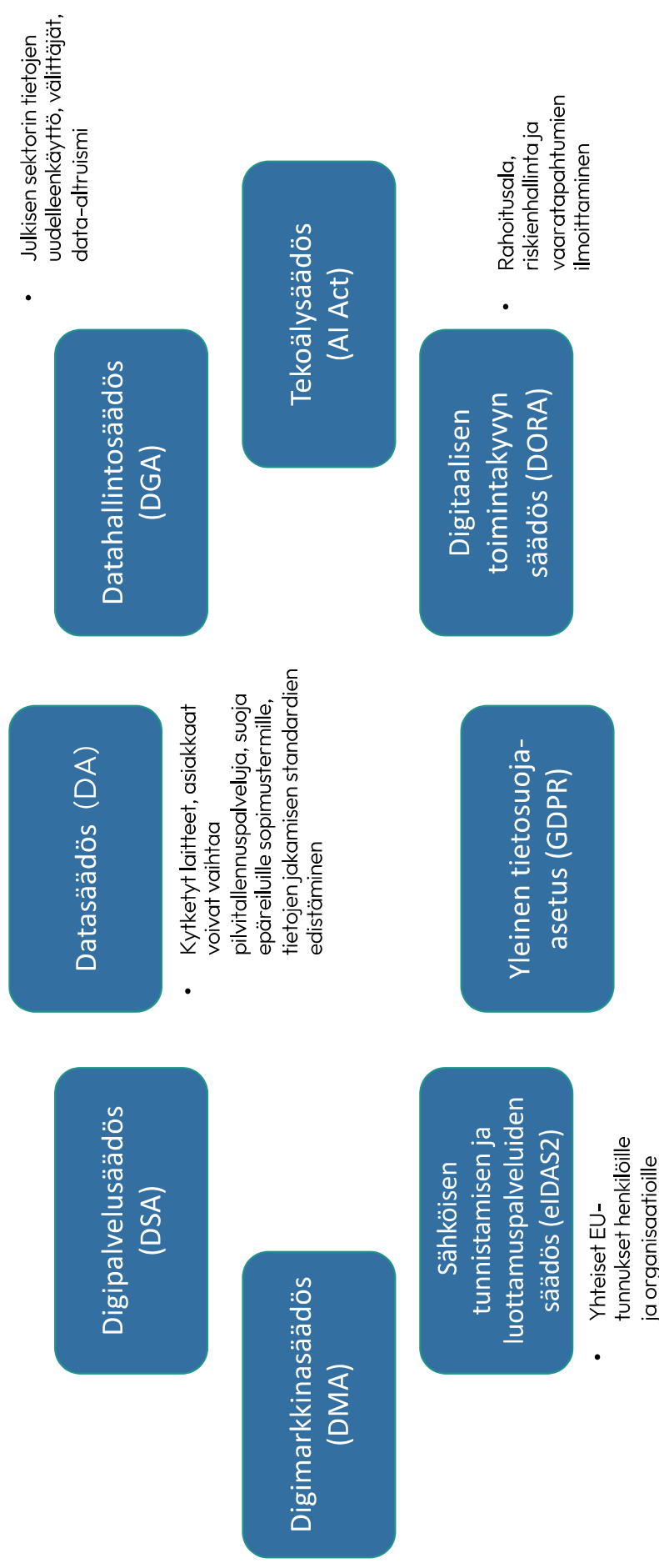
EU:n dataan liittyvä sääntely isossa kuvassa

Huomaa erot filosofiasa ja ideologiassa verrattuna Yhdysvaltoihin ja Kiinaan!

- Yleensä uusi EU:n datalainsäädäntö korostaa muutamaa lähestymistapaa:
 - Riskienhallinta
 - Erityisesti yksilöiden ja muiden kolmansien osapuolten näkökulmasta
 - Raportointi ja tietojen jakaminen
 - Raportointivaatimukset lisääntyvät, myös pienien alihankkijoiden ja muiden lähteiden puolesta, mikä edellyttää enemmän tietojen jakamista
 - Noudattaminen, säännöstenmukaisuus ja vastuu (compliance)
 - Lain noudattaminen ei riitä, vaan kyettävä myös todistamaan se: dokumentointi- ja tietojenjakovaatimukset lisääntyvät
- Sääntelyn keventäminen, Omnibus??

Sääntelytsunami

EU:n data- ja tekoälysäädäntö



- EU:n yleinen tietosuoja-asetus eli GDPR + kansallista lainsäädäntöä
- Laaja henkilötiedon käsite: tietosuoja koskee lähes kaikkea datan käsittelyä
 - Kannattaa selvittää, mitä henkilötietoja käsitellään
- Käyttötarkoitus määriteltävä etukäteen
 - Lähtökohtaisesti henkilötietoja ei saa käsitellä muihin tarkoituksiin
- Laillinen peruste oltava olemassa
 - Suostumus, sopimus, lakisääteinen velvoite, elintärkeä etu, yleinen etu / julkinen valta, oikeutettu etu
- Huomaa tietoturva ja kv-siirtoja koskevat säännökset
- Riskiarvio
 - Mitä vahinkoa henkilötietojen käsittelystä voi aiheutua ihmisille?
- Suunnittele jatkuva toimintatapa
- Dokumentoi

- Henkilötietojen käsittely tekoälyn kouluttamisessa ja käyttämisessä on lähtökohtaisesti henkilötietojen käsittelyä siinä missä muukin tietojen käsittely
 - Muista henkilötiedon laaja määritelmä: yleensä kannattaa lähteä siitä, että kaikkiin laajoihin aineistoihin sisältyy myös henkilötietoja
- Erityissäännöksiä, esim GDPR art 22
lähtökohtaisesti kieltää automatisoidut päätökset (ml profilointi), joilla on yksilöä koskevia oikeusvaikutuksia tms
- Maantieteellinen sijainti: ajetaanko tekoälyä paikallisesti vai pilvessä?

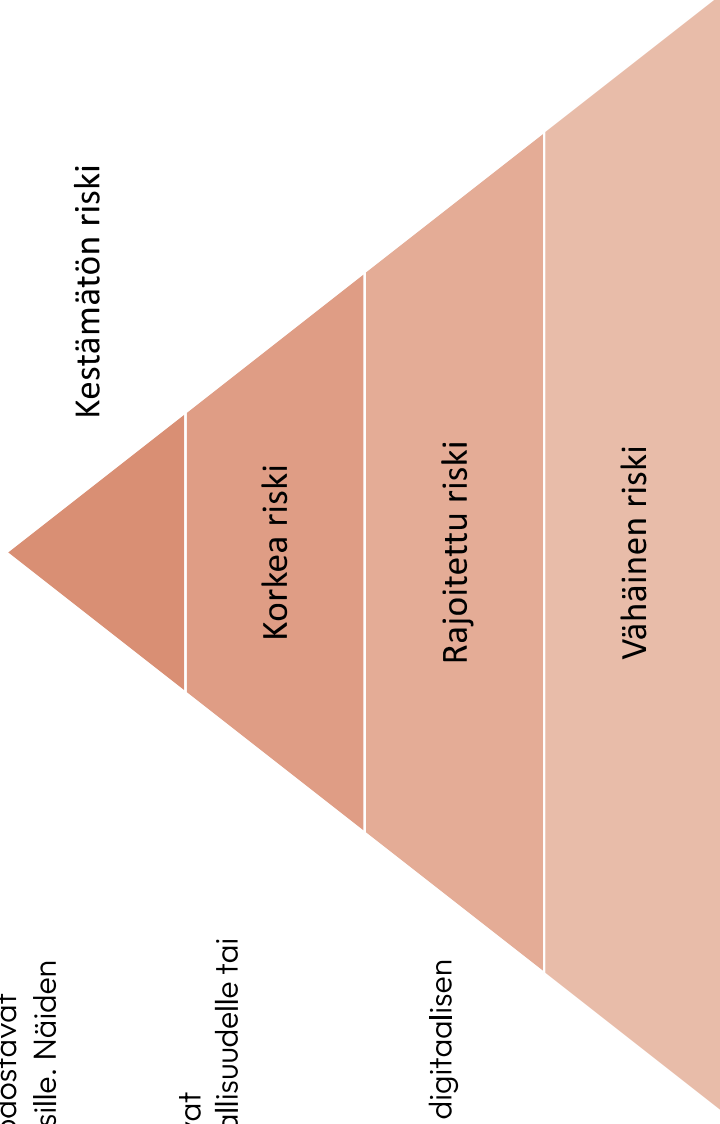
Tekoälylainsäädös (AI Act)

Joitakin yksityiskohtia

- Tekoälylainsäädös sovellettavaksi 2.8.2026, joitakin poikkeuksia lukuun ottamatta.
 - Esim. kiellettyihin tekoälykäytäntöihin liittyvät säännökset ovat jo voimassa
- Tekoälyjärjestelmän laaja määritelmä:
Konepohjainen järjestelmä, joka on suunniteltu toimimaan käyttöönoton jälkeen vaihtelevilla autonomian tasoilla ja jossa voi ilmetä mukautuvuutta käyttöönoton jälkeen ja joka päättelee vastaanottamastaan syötteestä eksplisiittisiä tai implisiittisiä tavoitteita varten, miten tuottaa tuotteita, kuten ennusteita, sisältöä, suosituksia tai päätöksiä, jotka voivat vaikuttaa fyysisiin tai virtuaalisiin ympäristöihin.
- Koskee seuraavia osapuolia EU:n alueella:
 - Toimittajia tekoälyjärjestelmistä (+ heidän valtuutettuja edustajia)
 - Käyttäjiä tekoälyjärjestelmistä
 - Jakelijoita tekoälyjärjestelmille
 - Asianosaisia ("mikä tahansa luonnollinen henkilö tai henkilöiden ryhmä, joille tekoälyjärjestelmä kohdistuu tai jotka muuten vaikuttavat")
- Ei koske esim.: tieteellistä tutkimusta ja kehitystä
- Riskiperustainen lähestymistapa

Tekoälyasetuksessa määritellään neljä tekoälyjärjestelmien riskitasoa:

- Kestämätön riski: tekoälyjärjestelmät, jotka muodostavat selkeän uhan yksilöiden turvallisuudelle ja oikeuksille. Näiden käyttö on kielletty.
 - Esim. Sosiaalinen pisteytys
- Korkea riski: tekoälyjärjestelmiä, jotka aiheuttavat merkittävää haittaa yksilöiden terveydelle, turvallisuudelle tai perusoikeuksille, säännellään tiukasti.
 - edellyttää ihmisen suorittamaa valvontaa, riskienhallintaa, tiedonhallintaa ja dokumentointikäytäntöjä.
 - Esim. biometrinen tunnistaminen, kriittisen digitaalisen infrastruktuurin hallinnointi, rekrytointi.
- Rajoitettu riski: avoimuusvaatimukset.
 - Esim. Chatbotit and syväväärennykset.
- Vähäinen riski: ei asetettuja vaatimuksia.
 - Esim. roskapostisuodattimet, tekoälypohjaiset videopelit.



Datasäädös (Data Act)

Datasäädös (Data Act, DA) on osa EU:n isoa datastrategia- ja lainsäädäntöpakettia

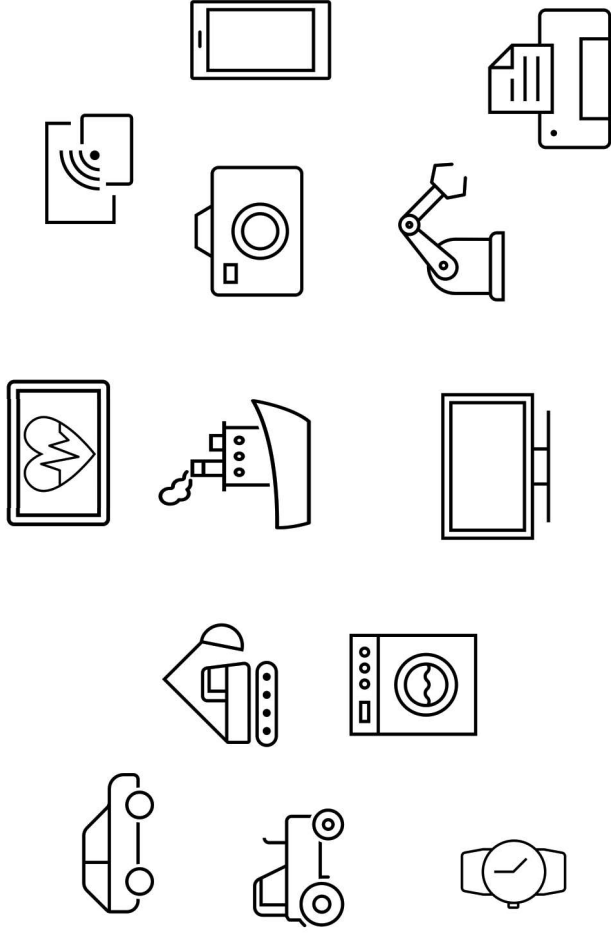
- Osa EU:n datastrategiaa ja ”säästösunamia”
- Pyrkii selventämään, kuka voi luoda arvoa datasta ja millä edellytyksillä
- Soveltaminen alkoi pääosin 12.9.2025
- Suomessa Traficom ensisijainen viranomainen
- Keskeinen sisältö:
 - Verkkoonliitetyn tuotteen käyttäjälle oikeus käyttödataan
 - Sopimusehtojen kohtuullisuus
 - Datan jakaminen poikkeustilanteessa
 - Palvelun vaihtaminen
 - Kansainväliset siirrot

Verkkoonliitetty tuote?

Todella laaja käsite: kaikki esineet, jotka kykenevät tuottamaan käyttöönsä liittyvää dataa

Verkkoonliitetty tuote:

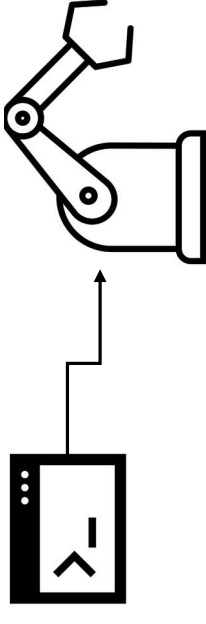
1. Saa, tuottaa tai kerää dataa käyttöstään ja ympäristöstään
2. Pystyy välittämään dataa sähköisen viestintäpalvelun, fyysisen yhteyden tai laiteyhteyden kautta (Huom! Ei tarvitse olla jatkuvasti verkossa!)
3. Pääasiallinen toiminto ei ole datan tallennus, käsittely tai siirto muun osapuolen kuin käyttäjän puolesta (esim pilvipalvelussa käytetty palvelinkone)



Liitännäispalvelu?

Kaikki olennaisesti tuotteeseen liittyvät palvelut

- Digitaalinen palvelu (ei sähköinen viestintäpalvelu), ml. ohjelmistot
- liitetty tuotteeseen osto-, vuokraus- tai liisaushetkellä siten, että tuote ei pystyisi ilman sitä suorittamaan jotain sen toiminnoista, tai
- valmistaja tai kolmas osapuoli myöhemmin liittää tuotteen toimintoja lisäksi, päivittäen tai mukauttaen
- Ei esim. konsultointi- ja rahoituspalvelut, säännöllinen korjaus ja huolto



Käytöstä syntyvä data?

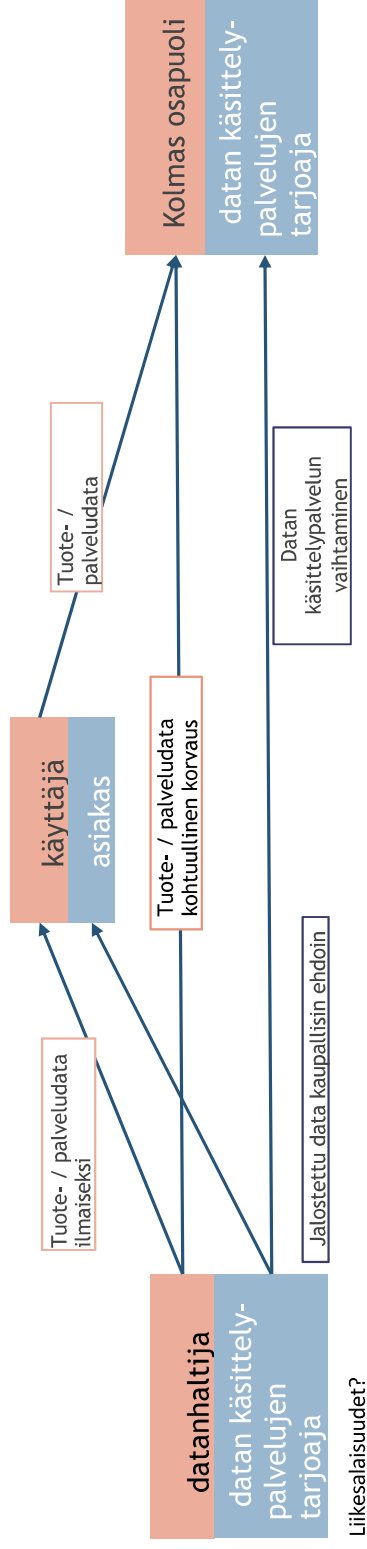
Tuotteen tai palvelun käytöstä syntyvä raakadata ja esikäsitelty data, muttei analysoitu ja johdettu data

- 'Tuotteen data': verkkoon liitetyn tuotteen käytön tuloksena tuotettu data, jonka valmistaja on suunnitellut olevan saatavilla tuotteesta
 - Huom! Suunnitteluratkaisu, mikä kaikki data on saatavilla
 - 'Tuotteeseen liittyvän palvelun data': edustaa verkkoon liitettyyn tuotteeseen liittyvien käyttäjän toimien tai tapahtumien digitointia; käyttäjä on tallentanut tarkoituksella tai joka on tallentunut käyttäjän toiminnan sivutuotteena ja joka tuotetaan palveluntarjoajan tarjotessa tuotteeseen liittyvää palvelua
- Tuotteen dataa on
- Raakadata eli automaattisesti, ilman lisäkäsittelyä tuotettu data (sensorimittaukset)
 - Esikäsitelty data: käsitelty ymmärrettävyyden ja käytön mahdollistamiseksi ennen jatkokäsittelyä ja analysointia, esimerkiksi konversio dataformaattista toiseen.
 - Oleneminen metadata: tietoa datasta, sen sisällöstä ja käytöstä, helpottaa datan hakemista ja käyttämistä ja on tarpeellista datan tulkitsemiseksi ja käyttämiseksi
- Analysoitu ja johdettu data ei ole tuotteen dataa!

1010
1010

Verkkoon liitettyjen tuotteiden ja liitännäispalveluiden datan jakaminen

- Datasäädöksessä asetetaan säännöt esimerkiksi datanhaltijoille ja datan käsittelypalvelujen tarjoajille sekä velvollisuudet antaa tietoja ilmaiseksi tietyissä tapauksissa.
- Tuotteet ja palvelut suunniteltava niin, että data on oletusarvoisesti suoraan käyttäjän saatavilla



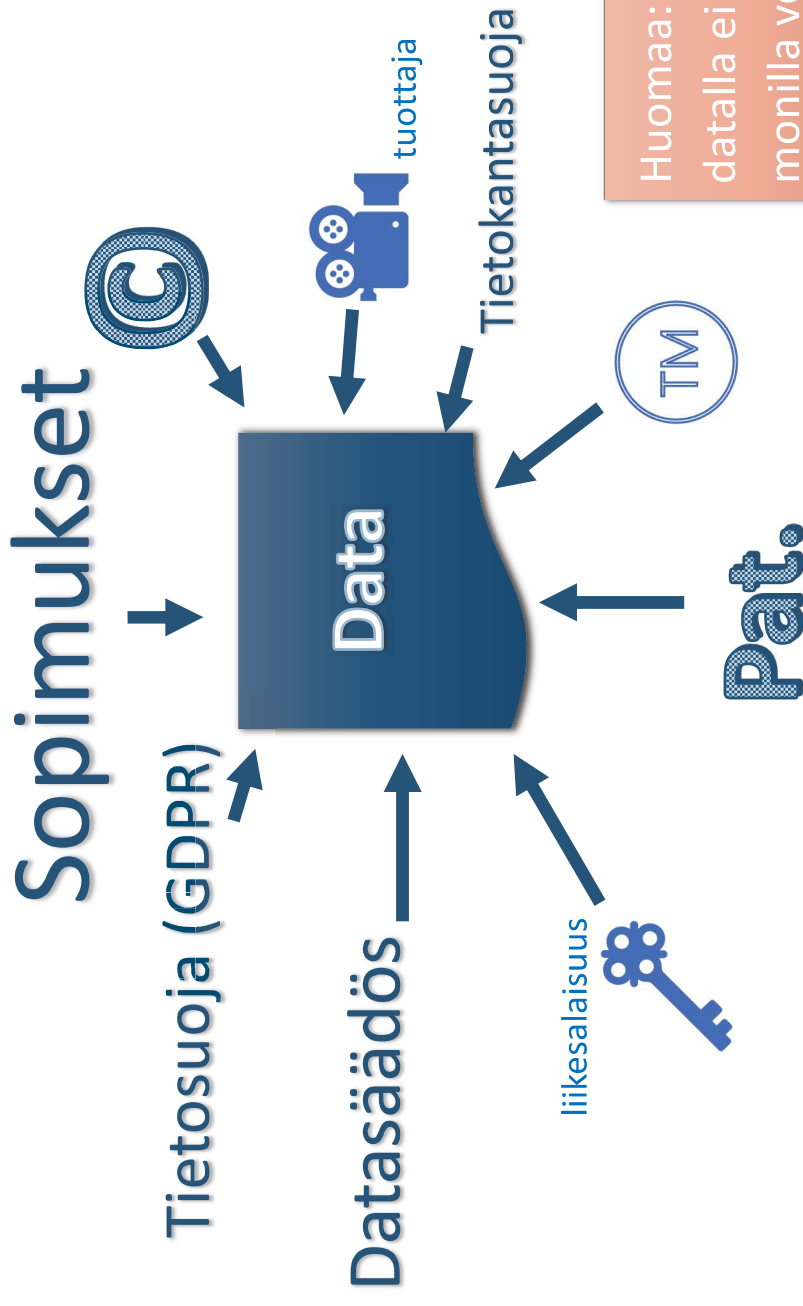
Datan jakamista ja käyttöä koskevat kohtuuttomat sopimusehdot

Sopimisen vapautta on rajoitettu datan jakamista koskevissa sopimuksissa



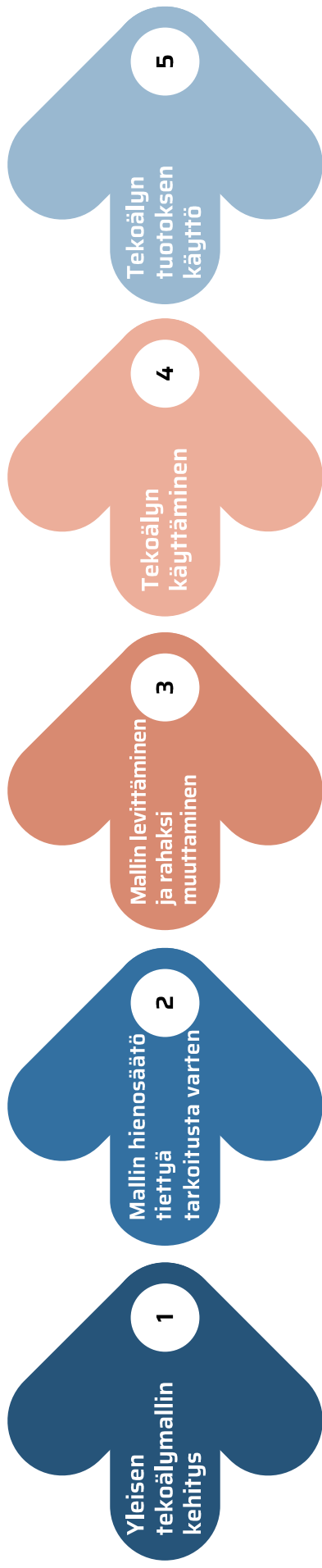
- Lähtökohtana sopimusvapaus: erityisesti yrityksillä on yleensä vapaus sopia mitä tahansa keskenään ja sopimukset sitovat osapuolia
- Datasäädös asettaa kuitenkin vaatimuksia dataa koskevien sopimusten ehdoille, muun muassa:
 - Data ilmaiseksi käyttäjälle
 - Käyttäjällä on oikeus jakaa dataa edelleen tietyissä rajoissa
 - Datan haltijalta kolmannelle osapuolelle käyttäjän sopimuksen täyttämiseksi kohtuullisella korvauksella
 - Tietyissä tilanteissa (datan vastaanottaja EU-alueella) vaaditaan oikeudenmukaiset, kohtuulliset ja syrjimättömät (FRAND) ehdot
 - komission mallilausekkeet?
 - Tietyistä asioista mainittava sopimuksissa
 - esim. toimenpiteet liikesalaisuuksien suojelemiseksi
- Datan haltija saa käyttää muuta dataa kuin henkilötietoja vain käyttäjän kanssa tehdyn sopimuksen perusteella.
 - Henkilötietojen osalta noudatettava GDPR:ää

Mitä oikeuksia voi kohdistua dataan?



Huomaa: oikeudellisessa mielessä dataalla ei ole yhtä omistajaa, vaan monilla voi olla siihen erilaisia oikeuksia

Tekoälyn elinkaari



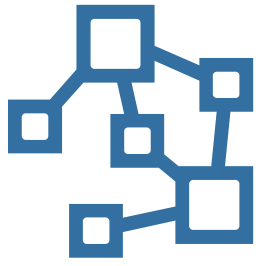
LLM:n kaltaisen yleisen tekoälymallin kehittämiseen tarvitaan suuria määriä dataa eri lähteistä. Se voi sisältää esimerkiksi tekijänoikeuden suojaamia teoksia tai henkilötietoja, joiden käyttämiseen saatetaan tarvita lupa.

Yleisen mallin hienosäätämiseksi tiettyä tarkoitusta varten tarvitaan lisää dataa eri lähteistä. Niihin kohdostuu usein eri tahojen oikeuksia.

Malli voi olla erittäin arvokas. Sen jakaminen ja rahaksi muuttaminen on usein kannattavaa. Mutta suojaavatko sitä mitkään oikeudet?

Tekoälyjärjestelmän käyttäminen edellyttää, että järjestelmään syötetään tietoja. Nämä tiedot voivat olla peräisin eri lähteistä ja niihin voi kohdistua oikeuksia.

Erityisesti generatiiviset tekoälyjärjestelmät, mutta myös muut tekoälyjärjestelmät voivat tuottaa arvokasta tietoa. Suojaako sitä mikään oikeus?



1001 LAKES

UNLOCKING DATA ECOSYSTEMS

Yhteystiedot

Olli Pitkänen

1001 Lakes Oy

olli.pitkanen@1001lakes.com

www.1001lakes.com