

EUROOPAN KOMISSION KESÄKUUSSA 2018 PERUSTAMA
RIIPPUMATON

**TEKOÄLYÄ KÄSITTELEVÄ
KORKEAN TASON ASIANTUNTIJARYHMÄ**



**TEKOÄLYN MÄÄRITELMÄ:
TÄRKEIMMÄT VALMIUDET JA
TIETEENALAT**

Asiantuntijaryhmän työtä varten laadittu määritelmä

Tekoälyn määritelmä: tärkeimmät valmiudet ja tieteenalat


Tekoälyä käsittelevä korkean tason asiantuntijaryhmä

Vastuuvapauslauseke ja tämän asiakirjan käyttäminen: Seuraava kuvaus ja määritelmä tekoälyn valmiuksista ja tutkimusaloista on hyvin karkea yksinkertaistus kehityksen viimeisimmästä tasosta. Tämän asiakirjan tarkoituksena ei ole määritellä tarkasti ja kattavasti kaikkia tekoälytekniikoita ja -valmiuksia vaan kuvaila yhteenvedonomaaisesti se yhteinen käsitys tästä alasta, jota korkean tason asiantuntijaryhmä käyttää tuotoksissaan. Toivomme kuitenkin, että tämä asiakirja voi toimia hyödyllisenä lähtökohtana henkilöille, jotka eivät ole tekoälyn asiantuntijoita ja jotka sitten voivat jatkaa laajempiin ja syvällisempiin pohdintoihin tekoälystä saadakseen tarkempaa tietoa tästä alasta ja teknologiasta.

Tekoälyä käsittelevä korkean tason asiantuntijaryhmä on Euroopan komission kesäkuussa 2018 perustama riippumaton asiantuntijaryhmä.

Yhteyshenkilö Nathalie Smuha – koordinaattori, AI HLEG
Sähköposti CNECT-HLG-AI@ec.europa.eu

Euroopan komissio
B-1049 Bryssel

Asiakirja julkaistiin  huhtikuuta 2019.

Tämän asiakirjan ensimmäinen luonnos julkaistiin 18. joulukuuta 2018 yhdessä tekoälyä käsittelevän korkean tason asiantuntijaryhmän (AI HLEG) laatimien luotettavaa tekoälyä koskevien eettisten ohjeiden ensimmäisen luonnoksen kanssa. Sitä tarkistettiin eurooppalaisen tekoälyallianssin kautta ja ohjelunnoista järjestetyssä avoimessa kuulemisessa saatujen kommenttien perusteella. Haluamme kiittää nimenomaisesti ja lämpimästi kaikkia niitä, jotka antoivat palautetta asiakirjan ensimmäisestä luonnoksesta.

Euroopan komissio ja sen nimissä toimivat henkilöt eivät vastaa siitä, miten tässä asiakirjassa olevia tietoja käytetään. Tämän työasiakirjan sisältö on yksinomaan tekoälyä käsittelevän korkean tason asiantuntijaryhmän (AI HLEG) vastuulla. Vaikka komission henkilöstö on edesauttanut tämän asiakirjan laatimista, siinä esitetyt näkemykset edustavat AI HLEG:n kantaa eikä niitä voida missään olosuhteissa katsoa Euroopan komission viralliseksi kannanotoksi.

Lisätietoa tekoälyä käsittelevästä korkean tason asiantuntijaryhmästä löytyy verkosta (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).

Euroopan komission asiakirjojen uudelleenkäytöstä säädetään päätöksessä 2011/833/EU (EUVL L 330, 14.12.2011, s. 39). Sellaisten valokuvien tai materiaalin käyttöön, johon EU:lla ei ole tekijänoikeutta, on pyydettävä lupa suoraan tekijänoikeuden haltijalta.

TEKOÄLYN MÄÄRITELMÄ:

TÄRKEIMMÄT VALMIUDET JA TIETEENALAT

Käytämme lähtökohtana seuraavaa tekoälyn määritelmää, joka esitetään Euroopan komission tekoälyä koskevassa tiedonannossa¹:

”Tekoäly on järjestelmä, joka pyrkii älykkäästi saavuttamaan asetettuja tavoitteita analysoimalla ympäristöään ja toimimalla osittain itsenäisesti.

Tekoälyjärjestelmät voivat olla täysin ohjelmistopohjaisia ja toimia virtuaalimaailmassa (esimerkiksi äänikomennoilla ohjattavat digiapulaiset, kuva-analyysiohjelmistot, hakuohjelmistot, puheen- ja kasvojen tunnistusohjelmistot) tai laitteistojen osia (esimerkiksi edistyneet robotit, itsenäiset autot, nelikopterit (drone) tai esineiden internetin sovellukset).”

Tässä asiakirjassa tätä määritelmää laajennetaan, jotta voidaan selvittää eräitä tekoälyn näkökohtia tieteenalana ja teknologiana. Tarkoituksena on välttää väärinkäsityksiä, luoda tekoälyä koskeva yhteinen käsitys, jota myös muut kuin tekoälyasiantuntijat voivat käyttää hedelmällisesti, sekä antaa hyödyllisiä tietoja, joita voidaan käyttää keskustelussa sekä tekoälyä koskevista eettisistä ohjeista että siihen liittyvistä politiikkasuosituksista.

1. Tekoälyjärjestelmät

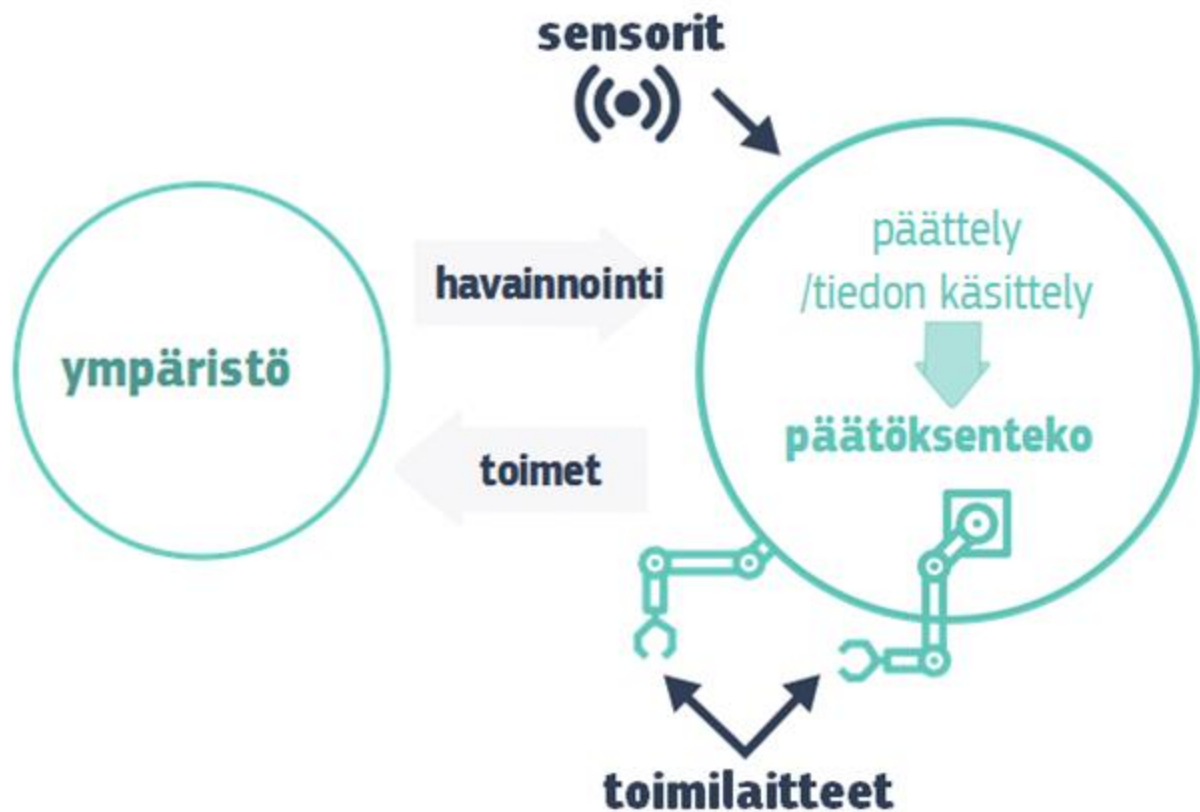
Termi 'tekoäly' sisältää selväsanaisen viittauksen älykkyyden käsitteeseen. Koska älykkyys (niin koneiden kuin ihmisten älykkyys) on kuitenkin epämääräinen käsite, vaikka psykologit, biologit ja neurotieteilijät ovat tutkineet sitä jo kauan, tekoälyn tutkijat käyttävät useimmiten rationaalisuuden käsitettä. Tällä tarkoitetaan kykyä valita paras toimintatapa tietyn tavoitteen saavuttamiseksi ottaen huomioon tietyt optimoitavat kriteerit ja käytettävissä olevat resurssit. Rationaalisuus ei luonnollisesti ole älykkyyden käsitteen ainoa ainesosa, mutta se muodostaa siitä merkittävän osan.

Seuraavassa termillä *tekoälyjärjestelmä* tarkoitetaan kaikkia tekoälyyn perustuvia komponentteja, ohjelmistoja ja/tai laitteistoja. Tekoälyjärjestelmät ovatkin yleensä suurempiin järjestelmiin *upotettuja* komponentteja pikemmin kuin erillisiä järjestelmiä.

Näin ollen tekoälyjärjestelmä on yhden yleisimmin käytetyn tekoälyä koskevan perusteoksen mukaan ensi kädessä rationaalinen.² Mutta miten tekoälyjärjestelmä saavuttaa rationaalisuuden? Kuten edellä esitetyn tekoälyn alustavan määritelmän ensimmäisessä virkkeessä todetaan, se saavuttaa sen havainnoimalla järjestelmän ympäristöä sensorien avulla ja keräämällä ja tulkitsemalla tällä tavoin tietoja, tekemällä päätelmiä havainnoistaan tai käsittelemällä tästä datasta saatuja tietoja ja päättämällä, mikä on paras toimintatapa, ja toimimalla sitten tämän mukaisesti toimilaitteiden avulla ja muuttaen näin mahdollisesti ympäristöä. Tekoälyjärjestelmät voivat joko käyttää symbolisia sääntöjä tai oppia numeerisen mallin, ja ne voivat myös mukauttaa käyttäytymistään analysoimalla, miten niiden aiemmat toimet vaikuttivat ympäristöön. Kaaviossa 1 oleva tekoälyjärjestelmän kuvaus havainnollistaa asiaa.

¹ Komission tiedonanto Euroopan parlamentille, neuvostolle, Euroopan talous- ja sosiaalikomitealle ja alueiden komitealle – Tekoäly Euroopassa, Bryssel, 25.4.2018, COM(2018) 237 final.

² ”Artificial Intelligence: A Modern Approach”, S. Russell ja P. Norvig, Prentice Hall, 3. painos, 2009.



Kaavio 1: Tekoälyjärjestelmän kaavakuva.

Sensorit ja havainnointi. Järjestelmän sensorit on kuvattu kaaviossa WiFi-symbolilla. Käytännössä ne voivat olla kameroita, mikrofoneja, näppäimistö, verkkosivusto tai muita syötelaitteita sekä fyysisten suureiden (kuten lämpötilan, paineen, etäisyyden, voiman/väännön) antureita tai kosketussensoreita. Yleisesti ottaen tekoälyjärjestelmä on varustettava sensoreilla, joilla se voi havainnoida riittävästi ympäristössä olevaa dataa, jolla on merkitystä sen ihmissuunnittelijan tekoälyjärjestelmälle antaman tavoitteen kannalta. Jos esimerkiksi halutaan rakentaa tekoälyjärjestelmä, joka automaattisesti siivoaa lattian, kun se on likainen, sensoreina voisi olla muun muassa kameroita, jotka ottavat kuvan lattiasta.

Kerätyn datan suhteen on usein hyödyllistä erottaa toisistaan jäsenneilty ja jäsentämätön data. *Jäsenneilty data* on järjestetty ennalta määriteltyjen mallien mukaan (kuten relaatiotietokanta), kun taas *jäsentämätön data* ei ole missään tunnetussa järjestyksessä (kuten kuvan tai tekstin sisältämä data).

Päättely/tiedonkäsittely ja päätöksenteko. Tekoälyjärjestelmän ytimen muodostaa sen päättely-/tietojenkäsittelymoduuli, joka käyttää syötteenä sensoreista tulevaa dataa ja ehdottaa toteutettavia toimia saavutettavan tavoitteen mukaisesti. Tämä tarkoittaa, että sensorien keräämä data on muunnettava tiedoksi, jonka päättely-/tietojenkäsittelymoduuli voi ymmärtää. Siivoavan tekoälyjärjestelmän esimerkissä kamera tuottaa päättely-/tietojenkäsittelymoduulille kuvan lattiasta ja moduulin on päätettävä, siivoaako se lattian vai ei (eli mikä on paras toimintatapa halutun tavoitteen saavuttamiseksi). Vaikka ihmisistä vaikuttaa helpolta ratkaista lattian kuvan perusteella, onko se siivottava, tämä ei ole yhtä helppoa koneelle, koska kuva on vain sarja nollia ja ykkösiä. Päättely-/tietojenkäsittelymoduulin on näin ollen

1. tulkittava kuva voidakseen ratkaista, onko lattia puhdas vai ei, mikä edellyttää yleensä, että sen on pystyttävä muuntamaan data tiedoksi ja mallintamaan tämä tieto suppeasti mutta kuitenkin siten, että kaikki merkityksellinen data otetaan mukaan (tässä tapauksessa lattian puhtaus tai likaisuus)
2. tehtävä tämän tietämyksen perusteella päätelmä tai käsiteltävä tämä tieto numeerisen mallin (eli matemaattisen kaavan) tuottamiseksi, jonka avulla se päättää, mikä on paras toimintatapa. Jos tässä esimerkissä

kuvasta johdettu tieto on se, että lattia on likainen, paras toimintatapa on aktivoida siivous, muussa tapauksessa paras toimintatapa on olla tekemättä mitään.

On syytä huomata, että termi 'päättös' tulee käsittää laajasti kaikenlaiseksi toteutettavan toimen valinnaksi eikä se välttämättä tarkoita, että tekoälyjärjestelmät ovat täysin autonomisia. Päättös voi olla myös jonkin ihmiselle – joka tekee lopullisen päätöksen – annettavan suosituksen valinta.

Toiminta. Kun toimesta on tehty päätös, tekoälyjärjestelmä on valmis toteuttamaan sen käytettävissään olevilla toimilaitteilla. Edellä olevassa kaaviossa toimilaitteet on kuvattu nivellettyinä jäseninä, mutta niiden ei tarvitse olla fyysisiä laitteita. Toimilaitteet voivat olla myös ohjelmistoja. Siivouksesimerkissä tekoälyjärjestelmä voisi tuottaa signaalin, joka käynnistää pölynimurin, jos valittu toimi on lattian siivous. Toinen esimerkki on keskustelujärjestelmä (niin kutsuttu chatbot), joka toimii tuottamalla tekstejä vastauksena käyttäjän lausumiin.

Suoritettu toimi saattaa muuttaa ympäristöä, joten seuraavalla kerralla järjestelmän on käytettävä taas sensoreitaan havainnoidakseen muuttuneesta ympäristöstä tulevaa mahdollisesti erilaista tietoa.

Rationaaliset tekoälyjärjestelmät eivät aina valitse tavoitteensa kannalta parasta toimintatapaa ja saavuttavat näin ollen vain *rajoitetun rationaalisuuden* ajan ja laskentatehon kaltaisten resurssien rajoitusten vuoksi.

Rationaaliset tekoälyjärjestelmät ovat hyvin yksinkertaisia tekoälyjärjestelmien versioita. Ne muuttavat ympäristöään mutta eivät mukauta käyttäytymistään ajan mittaan saavuttaakseen tavoitteensa paremmin. *Oppiva rationaalinen järjestelmä* on rationaalinen järjestelmä, joka toimen toteutettuaan arvioi ympäristön uuden tilan (havainnoimalla) määrittääkseen, kuinka menestyksellinen sen toimi oli, ja mukauttaa sitten päätelysääntöjään ja päätöksentekomenetelmiään.

2. Tekoäly tieteenalana

Edellä esitettiin hyvin yksinkertainen tekoälyjärjestelmän abstrakti kuvaus, joka perustuu kolmeen tärkeimpään valmiuteen: havainnointiin, päättelyyn/päätöksentekoon ja toimintaan. Se on kuitenkin riittävä, jotta voimme esitellä ja ymmärtää useimmat tekoälyjärjestelmien rakentamiseen nykyisin käytettävät tekoälytekniikat ja osat alueet, koska ne kaikki liittyvät järjestelmien eri valmiuksiin. Laajasti ottaen kaikki tällaiset tekniikat voidaan jakaa kahteen ryhmään, jotka perustuvat kykyyn *päätellä* ja *oppia*. Robotiikka on eräs toinen hyvin tärkeä ala.

Päättele ja päätöksenteko. Tämä tekniikoiden ryhmä sisältää tietämyksen esittämisen ja päättelyn, suunnittelun, vuoronnuksen, haun ja optimoinnin. Näiden tekniikoiden avulla sensoreista tulevasta datasta voidaan tehdä päätelmiä. Jotta tämä olisi mahdollista, data on muunnettava tietämykseksi, joten yksi tekoälyn ala käsittelee sitä, miten tällainen tietämys parhaiten mallinnetaan (*tietämyksen esittäminen*). Kun tietämys on mallinnettu, seuraavassa vaiheessa siitä tehdään päätelmiä (*tietämykseen perustuva päättely*), mihin sisältyy päätelmien tekeminen symbolisten sääntöjen avulla, *suunnittelua* ja *vuoronnusta* koskevia toimia, *haku* suuresta määrästä ratkaisuja ja *optimointi* ongelman kaikista mahdollisista ratkaisuista. Viimeinen vaihe on päätös toteutettavasta toimesta. Tekoälyjärjestelmään kuuluva päättely/päätöksenteko on tavallisesti hyvin monimutkaista ja edellyttää useiden edellä mainittujen tekniikoiden yhdistelmää.

Oppiminen. Tämä tekniikoiden ryhmä sisältää koneoppimisen, neuroverkot, syväoppimisen, päätöspuut ja monia muita oppimistekniikoita. Näiden tekniikoiden avulla tekoälyjärjestelmä voi oppia ratkaisemaan ongelmia, joita ei voida esittää täsmällisesti tai joiden ratkaisumenetelmää ei voida kuvata symbolisilla päätelysäännöillä. Esimerkkejä tällaisista ongelmista ovat ongelmat, jotka liittyvät havainnointikykyyn, kuten *puheen* ja *kielen ymmärtäminen* sekä *tietokonenäkö* tai *käyttäytymisen ennustaminen*. On syytä huomata, että nämä ongelmat vaikuttavat helpoilta, koska ne tosiaan ovat tavallisesti helppoja ihmisille. Ne eivät kuitenkaan ole helppoja tekoälyjärjestelmille, koska nämä eivät (ainakaan vielä) voi käyttää yleisjärjen päättelyä, ja ne ovat erityisen vaikeita, jos järjestelmän on tulkittava jäsentymätöntä tietoa. Tällaisissa tapauksissa *koneoppimista* noudattavat tekniikat ovat hyödyllisiä. Koneoppimistekniikoita voidaan kuitenkin käyttää moniin muihinkin tehtäviin kuin vain havainnointiin. Koneoppimistekniikat tuottavat numeerisen mallin (eli matemaattisen kaavan), jota käytetään päätöksen laskemiseksi datasta.

On olemassa useita koneoppimisen muotoja. Yleisimmät menetelmät ovat *valvottu oppiminen*, *valvomaton oppiminen* ja *vahvistusoppiminen*.

Valvotussa koneoppimisessa järjestelmälle annetaan käyttäytymissääntöjen sijasta esimerkkejä syöte-tulos-käyttäytymisestä toivoen, että se pystyy tekemään esimerkeistä yleistyksiä (jotka tavallisesti kuvailevat menneitä tapahtumia) ja toimimaan hyvin myös tilanteissa, joita esimerkeissä ei esitetä (ja joita voidaan kohdata tulevaisuudessa). Käyttämässämme esimerkissä antaisimme järjestelmälle useita esimerkkikuvia lattiasta ja niiden vastaavat tulkinnat (eli onko lattia kuvassa likainen vai ei). Jos annamme riittävästi esimerkkejä, jotka ovat riittävän moninaisia ja kattavat riittävästi useimmat tilanteet, järjestelmä pystyy koneoppimisalgoritminsa avulla yleistämään ne siten, että se osaa tulkita hyvin myös sellaisia kuvia lattioista, joita se ei ole nähnyt. Joissain koneoppimisen menetelmissä käytetään algoritmeja, jotka perustuvat *neuroverkkojen* käsitteeseen. Niiden esikuvana ovat löyhästi ihmisaivot sikäli, että neuroverkko koostuu pienistä (ihmisen hermosoluihin rinnastettavista) tiedonkäsittely-yksiköistä, joiden välillä on paljon painotettuja kytkentöjä. Neuroverkon syötteenä on sensorien toimittama data (esimerkissämme kuva lattiasta) ja tuotoksena tulkinta kuvasta (onko lattia puhdas vai ei). Esimerkkien analysoinnin aikana (verkon *koulutusvaiheessa*) kytkentöjen painotuksia mukautetaan mahdollisimman paljon käytettävissä olevien esimerkkien antaman tiedon mukaiseksi (eli odotetun tuloksen ja verkon laskeman tuloksen välisen virheen minimoimiseksi). Koulutusvaiheen lopuksi testausvaiheella, jossa testataan neuroverkon käyttäytymistä aiemmin näkemättömillä esimerkeillä, tarkistetaan, että tehtävä on opittu hyvin.

On tärkeää huomata, että tähän menetelmään (kuten kaikkiin koneoppimistekniikoihin) liittyy aina tietty virheprosentti, joskin se on yleensä pieni. Näin ollen *tarkkuus*, joka mittaa sitä, kuinka suuri oikeiden vastausten prosenttiosuus on, on keskeinen käsite.

On olemassa useita neuroverkkoja ja koneoppimista koskevia lähestymistapoja, ja yksi menestyksekkäimmistä on *syväoppiminen*. Tämä menetelmä perustuu siihen, että neuroverkossa on syötteen ja tuloksen välillä useita kerroksia, joiden avulla se voi oppia yleisen syöte-tulos-suhteen vaiheittain. Tämä tekee menetelmästä kokonaisuutena tarkemman, ja se tarvitsee vähemmän ihmisen ohjausta.

Neuroverkot ovat vain yksi koneoppimisen väline, ja on olemassa monia muitakin välineitä, joilla on erilaisia ominaisuuksia: satunnaismetsät ja tehostetut puut, klusterointimenetelmät, matriisin faktorointi jne.

Toinen hyödyllinen koneoppimisen menetelmä on *vahvistusoppiminen*. Tässä menetelmässä tekoälyjärjestelmä voi tehdä päätöksensä vapaasti ja kustakin päätöksestä annetaan järjestelmälle palaute, joka kertoo sille, oliko päätös hyvä vai huono. Järjestelmän tavoitteena on ajan mittaan maksimoida saatu myönteinen palaute. Tätä menetelmää käytetään esimerkiksi suositusjärjestelmissä (esimerkiksi useat verkossa toimivat suositusjärjestelmät, jotka ehdottavat käyttäjille tuotteita, joita nämä saattavat haluta ostaa) ja myös markkinoinnissa.

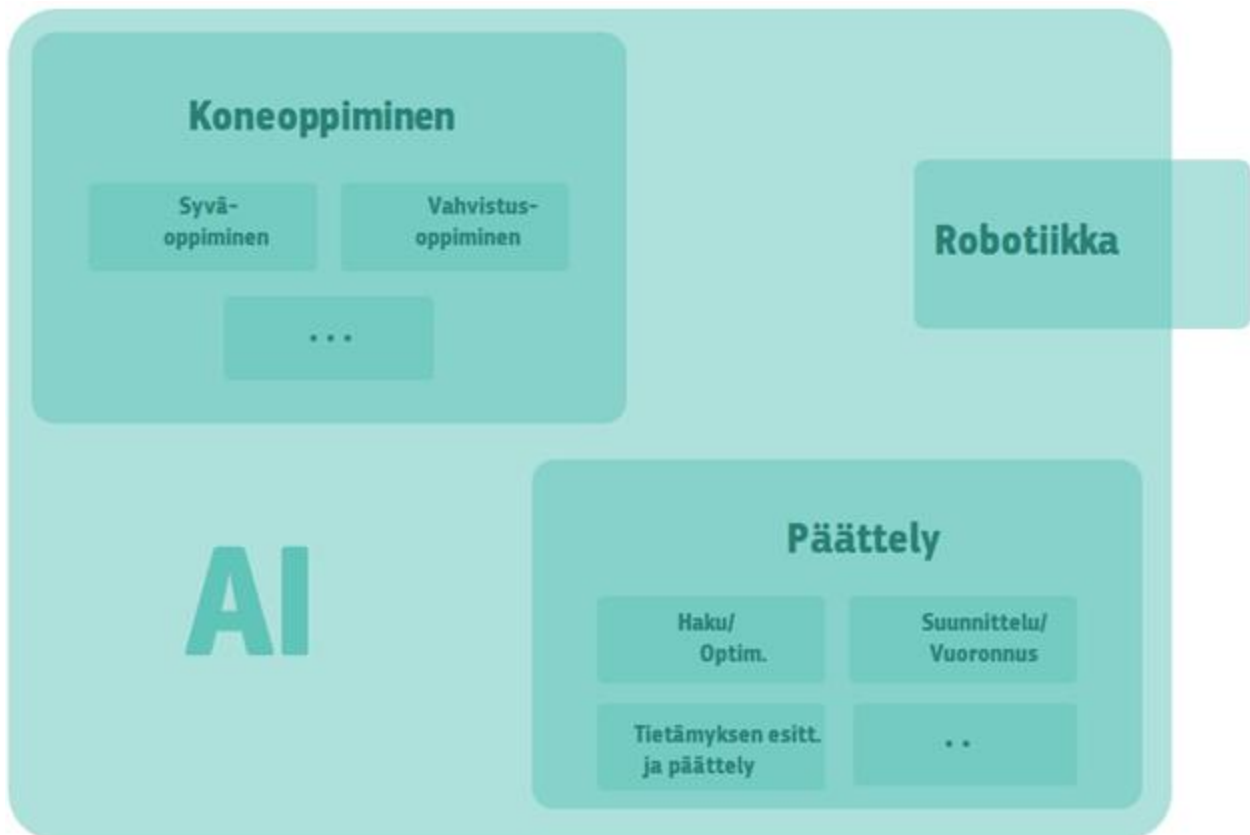
Koneoppimisen menetelmät ovat hyödyllisiä havainnointitehtävien, kuten näkemisen ja tekstin ymmärtämisen, lisäksi myös kaikissa sellaisissa tehtävissä, joita on vaikea määritellä ja joita ei voida kuvailla kattavasti symbolisilla käyttäytymissäännöillä.

On syytä pitää mielessä ero yhtäältä sellaisten koneoppimisen menetelmien, joissa opitaan uusi tehtävä, jota ei voida kuvailla hyvin symbolisesti, ja toisaalta sellaisten (edellisessä jaksossa mainittujen) rationaalisten toimijoiden oppimisen välillä, jotka mukauttavat käyttäytymistään ajan mittaan saavuttaakseen paremmin annetun tavoitteen. Nämä kaksi tekniikkaa saattavat toimia päällekkäin tai yhdessä, mutta ne eivät ole välttämättä sama asia.

Robottiikka. Robottiikka voidaan määritellä ”fyysisessä maailmassa toimivaksi tekoälyksi” (tätä kutsutaan myös *kehollistetuksi tekoälyksi*). Robotti on fyysinen kone, jonka on selviydyttävä fyysisen maailman dynamiikasta, epävarmuudesta ja monitahoisuudesta. Havainnointi-, päättely-, toiminta- ja oppimiskyky sekä kyky olla vuorovaikutuksessa muiden järjestelmien kanssa on yleensä integroitu robottiikkajärjestelmän hallintarakenteeseen. Robottien suunnittelussa ja käytössä ovat merkittäviä tekoälyn lisäksi myös muut alat, kuten koneenrakennus ja säätäteoria. Esimerkkejä roboteista ovat robottimanipulaattorit, autonomiset ajoneuvot (kuten autot, nelikopterit [drone], lentävät taksit) ihmisen kaltaiset robotit, robottipölynimurit jne.

Kaaviossa 2 esitetään useimmat edellä mainituista tekoälyn osa-alueista sekä niiden suhteet. On kuitenkin tärkeää huomata, että tekoäly on paljon monitahoisempi kuin tästä kaaviosta käy ilmi, sillä se sisältää monia muita osa-

alueita ja tekniikoista. Kuten edellä todettiin, robotiikassa käytetään myös tekniikoita, jotka eivät kuulu tekoälyn piiriin. Katsomme kuitenkin, että tämä riittää antamaan hedelmällisen panoksen tekoälyä, tekoälyn etiikkaa ja tekoälypolitiikkaa koskevaan kokemusten jakoon, tietoisuuteen ja keskusteluun, jota on käytävä hyvien erilaisia aloja ja sidosryhmiä edustavassa korkean tason asiantuntijaryhmässä.



Kaavio 2: Yksinkertaistettu yleiskuva tekoälyn osa-alueista ja niiden suhteista.

Sekä koneoppimiseen että -päättelyyn sisältyy monia muita tekniikoita, ja robotiikka sisältää tekoälyn ulkopuolisia tekniikoita. Tekoäly kokonaisuudessaan kuuluu tietojenkäsittelyopin piiriin.

3. Muita merkittäviä tekoälyn käsitteitä ja kysymyksiä

Kapea (tai heikko) ja yleinen (tai vahva) tekoäly. Yleisen tekoälyjärjestelmän on tarkoitus olla järjestelmä, joka voi suorittaa useimmat sellaiset toiminnot, joita ihmiset voivat suorittaa. Kapeat tekoälyjärjestelmät ovat sen sijaan järjestelmiä, jotka voivat suorittaa yhden tai muutaman rajoitetun tehtävän. Tällä hetkellä käyttöön otetut tekoälyjärjestelmät ovat esimerkkejä kapeasta tekoälystä. Tekoälyn alkuaikoina tutkijat käyttivät eri termejä (heikko ja vahva tekoäly). On yhä olemassa useita ratkaisemattomia eettisiä, tieteellisiä ja teknologisia vaikeuksia luoda valmiudet, jotka tarvittaisiin yleisen tekoälyn aikaan saamiseksi, kuten yleisjärkeen perustuva päättely, tietoisuus itsestä ja koneen kyky määritellä oma tarkoituksensa.

Dataa koskevat kysymykset ja vääristymät. Koska monet tekoälyjärjestelmät, kuten ne, joihin sisältyy valvotun koneoppimisen komponentteja, tarvitsevat valtavia määriä dataa toimiakseen hyvin, on tärkeää ymmärtää, miten data vaikuttaa tekoälyjärjestelmän käyttäytymiseen. Jos esimerkiksi koulutusdatassa on vääristymiä eli se ei ole riittävän tasapainoista tai kattavaa, tällaisella datalla koulutettu tekoälyjärjestelmä ei pysty yleistämään kunnolla ja se saattaa tehdä epäoikeudenmukaisia päätöksiä, jotka suosivat ehkä joitakin ryhmiä toisten kustannuksella. Tekoäly-yhteisö on viime aikoina pyrkinyt kehittämään menetelmiä, joilla havaitaan vääristymät koulutustietueissa ja myös muissa tekoälyjärjestelmän osissa ja lievennetään niitä.

Mustan laatikon tekoäly ja selitettävyyks. Jotkin koneoppimistekniikat ovat hyvin menestyksellisiä tarkkuuden kannalta, mutta on erittäin vaikeaa ymmärtää, miten ne tekevät päätöksiä. *Mustan laatikon tekoälyn* käsitteellä

viitataan tällaisiin tilanteisiin, joissa tiettyjä päätöksiä ei ole mahdollista jäljittää takaisin niiden perusteisiin. Selitettävyyden on yksi tällaisten tekoälyjärjestelmien ominaisuus, joka voi tämän sijaan tarjota jonkinlaisen selityksen niiden toiminnasta.

Tavoitesuuntautunut tekoäly. Nykyiset tekoälyjärjestelmät ovat tavoitesuuntautuneita, eli ihminen määrittelee niille saavutettavan tavoitteen ja ne käyttävät tiettyjä tekniikoita tämän tavoitteen saavuttamiseksi. Ne eivät määrittele omia tavoitteitaan. Joillakin tekoälyjärjestelmillä (kuten tiettyihin koneoppimistekniikoihin perustuvilla järjestelmillä) voi kuitenkin olla enemmän vapautta valita, millä toimintatavalla ne saavuttavat annetun tavoitteen.

4. Tekoälyn päivitetty määritelmä

Ehdotamme, että käytetään seuraavaa päivitettyä tekoälyn määritelmää:

"Tekoälyjärjestelmät ovat ihmisten suunnittelemaa³ ohjelmisto- ja mahdollisesti myös laitteistojärjestelmiä, joille annetaan monitahoinen tavoite ja jotka toimivat fyysisessä tai digitaalisessa ulottuvuudessa havainnoimalla ympäristöä hankkimalla dataa, tulkitsemalla kerätyn jäsenellisen tai jäsentämättömän datan, tekemällä päätelmiä tästä datasta johdetusta tietämyksestä tai käsittelemällä siitä johdettuja tietoja ja tekemällä päätöksen parhaasta toimesta tai toimista annetun tavoitteen saavuttamiseksi. Tekoälyjärjestelmät voivat joko käyttää symbolisia sääntöjä tai oppia numeerisen mallin, ja ne voivat myös mukauttaa käyttäytymistään analysoimalla, miten niiden aiemmat toimet vaikuttivat ympäristöön.

Tieteenalana tekoälyyn sisältyy useita lähestymistapoja ja tekniikoita, kuten koneoppiminen (josta syväoppiminen ja vahvistusoppiminen ovat erityisiä esimerkkejä), konepäättely (johon sisältyy suunnittelu, vuoronnus, tietämyksen esittäminen ja päättely, haku ja optimointi) ja robotiikka (johon sisältyy hallinta, havainnointi, sensorit ja toimilaitteet sekä kaikkien muiden tekniikoiden integrointi kyberfysiisiin järjestelmiin)."

Ehdotamme myös, että tätä asiakirjaa käytetään lisätiedon lähteenä tämän määritelmän tukemiseksi.

³ Ihmiset suunnittelevat tekoälyjärjestelmiä suorasti, mutta he voivat myös käyttää tekoälytekniikoita niiden suunnittelun optimointiin.

Tämän asiakirjan on laatinut tekoälyä käsittelevä korkean tason asiantuntijaryhmä (AI HLEG),

jonka jäsenet luetellaan seuraavassa aakkosjärjestyksessä:

Pekka Ala-Pietilä, AI HLEG:n puheenjohtaja AI Finland, Huhtamäki, Sanoma	Pierre Lucas Orgalim – Euroopan teknologiateollisuuden kattojärjestö
Wilhelm Bauer Fraunhofer	Ieva Martinkenaite Telenor
Urs Bergmann Zalando	Thomas Metzinger JGU Mainz, Euroopan yliopistojen liitto
Mária Bielíková Slovakian teknillinen korkeakoulu, Bratislava	Cateljne Muller ALLAI Netherlands, ETSK
Cecilia Bonefeld-Dahl DigitalEurope	Markus Noga SAP
Yann Bonnet ANSSI	Barry O’Sullivan, AI HLEG:n varapuheenjohtaja Corkin yliopisto
Loubna Bouarfa OKRA	Ursula Pahl BEUC
Stéphan Brunessaux Airbus	Nicolas Petit Liègen yliopisto
Raja Chatila IEEE Initiative Ethics of Intelligent/Autonomous Systems & Sorbonnen yliopisto	Christoph Peylo Bosch
Mark Coeckelbergh Wienin yliopisto	Iris Plöger BDI
Virginia Dignum Uumajan yliopisto	Stefano Quintarelli Garden Ventures
Luciano Floridi Oxfordin yliopisto	Andrea Renda College of Europe Faculty & CEPS
Jean-Francois Gagné Element AI	Francesca Rossi* IBM
Chiara Giovannini ANEC	Cristina San José Euroopan pankkiyhdistysten liitto
Joanna Goodey Perusoikeusvirasto	George Sharkov Digital SME Alliance
Sami Haddadin Münchenin teknisen yliopiston robotiikan ja koneälyn laitos	Philipp Slusallek Saksan tekoälyn tutkimuskeskus (DFKI)
Gry Hasselbalch The thinkdotank DataEthics, Kööpenhaminan yliopisto	Françoise Soulié Fogelman Tekoälykonsultti
Fredrik Heintz Linköpingin yliopisto	Saskia Steinacker Bayer
Fanny Hidvegi Access Now	Jaan Tallinn Ambient Sound Investment
Eric Hilgendorf Würzburgin yliopisto	Thierry Tingaud STMicroelectronics
Klaus Höckner Hilfsgemeinschaft der Blinden und Sehschwachen	Jakob Uszkoreit Google
Mari-Noëlle Jégo-Laveissière Orange	Aimee Van Wynsberghe Delftin tekninen yliopisto
Leo Kärkkäinen Nokia Bell Labs	Thiébaut Weber EAY
Sabine Theresia Köszegi Wienin tekninen yliopisto	Cecile Wendling AXA
Robert Kroplewski Asianajaja, Puolan hallituksen neuvonantaja	Karen Yeung Birminghamin yliopisto
Elisabeth Ling RELX	

*Francesca Rossi toimi tämän asiakirjan esittelijänä.