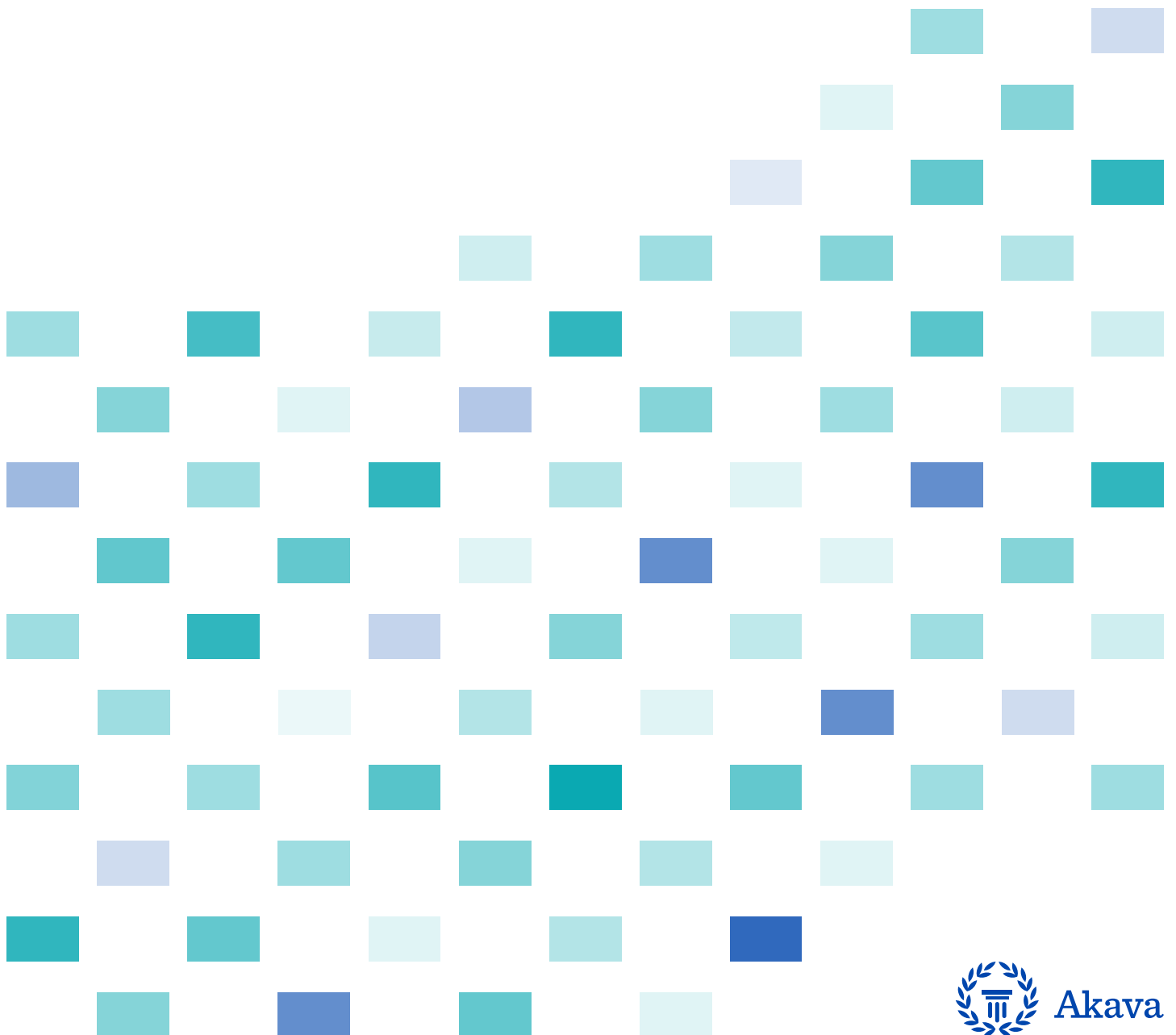


Miksi tekoäly oppi näkemään ja juttelemaan nyt ja mihin se riittää?

Artikkeli 1/2024



Artikkeli 1/2024

Kirjoittaja: Arto Klami, apulaisprofessori, Helsingin yliopisto

Avainsanat: tekoäly, ai

Päivämäärä: 14.2.2024

Kirjoittajan esittely

Arto Klami on tekniikan tohtori. Hän työskentelee tietojenkäsittelytieteen apulaisprofessorina Helsingin yliopistolla.

Yhteenveto

Artikkeli esittelee lyhyesti tekoälyn historiaa ja sen läpimurtoa. Siinä käydään läpi, tekoälyn toimintaperiaatteita ja miten tekoäly on kehittynyt vuosien varrella itseohjatusta oppijasta luovaksi sisällöntuottajaksi.

Artikkelin lopuksi tarkastellaan tekoälyn ja sen käytön haasteita ja pohditaan, miltä tekoälyn tulevaisuus näyttää.

Akava Works

Akavan tiedontuotanto toimii nimellä Akava Works.

Akava Works tarjoaa monipuolisesti tietosisältöjä raporteina, selvityksinä ja tutkimuksina. Tavoitteenamme on lisäksi herättää yhteiskunnallista keskustelua erityisesti akavalaisille tärkeistä ja ajankohtaisista aiheista. Akava Works -sisällöt eivät ole Akavan virallisia kannanottoja.

Akava Works -sivusto on osoitteessa www.akavaworks.fi.

Sisällysluettelo

1	Tekoälyn toimintaperiaatteista: koneoppiminen ja syvät neuroverkot.....	3
2	Tekoälyn historiasta eli onko kuvien tulkinta helppoa vai vaikeaa? ..	4
3	Perustutkimus ja avoimet ympäristöt kehityksen vetureina.....	5
4	Enemmän irti datasta.....	6
5	Itseohjatusta oppijasta luovaksi tekoälyksi.....	8
6	Perustamallit kehityksen pohjana	8
7	Avoimet haasteet – mitä seuraavaksi?	10

Johdanto

Vielä muutama vuosi sitten moni tekoälytutkijakin arvioi, että sujuvasti keskustelevala tekoäly on vielä kaukainen tulevaisuudenhaave. Nyt tarjolla on kuitenkin toinen toistaan vakuuttavammin toimivia palveluita ja moni meistä käyttää niitä jo päivittäisen työnsä tehostamiseen. Mistä tässä läpimurrossa on kyse ja miten se saatiin aikaan?

Tekoäly teknologiana on jo pidemmän aikaa ollut esillä yritysten visioissa, mutta osaksi jokapäiväistä keskustelua se on tullut toden teolla parin viime vuoden aikana. Algoritmit eivät enää ole piilossa, vaan voimme keskustella tekoälyn kanssa filosofiasta ja pyytää nähtäväksi tussilaveerauksen vaikkapa höyrykäyttöisellä kirpulla ratsastavasta apinasta. Ohjelmoijan työkalupakkiin on tullut tekoälyapuri, joka sanallisten ohjeiden perusteella tuottaa rutiinikoodia verkkokaupan pohjaksi ja voi säästää päiväkausia monimutkaisen algoritmin kirjoittamisesta.

Siirryimmekö yhdessä vuodessa tekoälyn kivikaudelta tieteisfiktioon? Tästä ei ole kyse vaan todellisuudessa on kyse pidemmän kehityksen luonnollisesta askeleesta. Tekoälyn toiminta perustuu koneoppimiseen, jossa joustava malli – syvä neuroverkko – on opetettu tietoaineistojen avulla vastaamaan annettuun syötteeseen halutulla tavalla. Alalla on käytetty melko samankaltaisia malleja ja oppimisen periaatteita jo vuosikymmeniä, mutta laskentatehon ja käytettävissä olevan datan kasvu yhdessä menetelmien jatkuvan kehityksen kanssa alkaa kantaa hedelmää. Melko tasainen kehitys näyttää meistä mullistavalta, koska tekoäly astuu nyt ensimmäistä kertaa selkeästi ihmisen osaamisen tontille: se osaa vihdoinkin käsitellä luonnollista kieltä ja kuvia.

1 Tekoälyn toimintaperiaatteista: koneoppiminen ja syvät neuroverkot

Tekoälyn alaan kuuluu laaja kirjo erilaisia menetelmiä, joista näkyvimmit perustuvat koneoppimiseen. Koneoppimisessa käytettävien mallien toimintaa säätelevät tuntemattomat parametrit eli muuttujat. Näitä parametreja muuttamalla mallin toiminta muuttuu. Oppimisella tarkoitetaan parametrien valitsemista siten, että malli saadaan toimimaan halutulla tavalla. Käytännössä malli opetetaan toistamaan opetusaineistossa olevia säännönmukaisuuksia.

Oppimisen periaatteet juontavat juurensa perinteisestä tilastotieteestä. Malli opitaan oleellisilta osin samalla tavalla, oli sitten kyseessä yksinkertaisen kasvukäyrän sovitus neuvolan pituusmittauksiin tai monimutkaisen tekoälymallin opettaminen. Vaikka jälkimmäisessä tapauksessa syötteenä voi iän sijaan olla valokuva ja tuloksena pituuden sijaan sanallinen kuvaus kuvan sisällöstä, molemmissa tapauksissa on kyse matemaattisesta funktiosta, joka muuttaa syötteen tulokseksi. Tarvittavat funktiot ovat paljon monimutkaisempi ja niiden sovittamiseen tarvitaan enemmän esimerkkejä, mutta periaate on sama ja alalla on vuosikymmenien kuluessa kehitetty toimivia algoritmeja erityisesti ohjattuihin ongelmiin, joissa opetusaineiston syötteille tiedetään halutut tulokset.

Itse mallit voivat olla hyvin monimuotoisia. Tunnetuin malliperhe ovat syvät neuroverkot, joissa tietoa käsitellään kerroksittain yksinkertaisilla laskutoimituksilla ja verkon rakennetta muuttamalla voidaan luoda malleja moniin tarkoituksiin. Kuvia

käsittävät mallit hyödyntävät suotimia, jotka analysoivat paikallisia kuva-alueita, ja taas kielten käsittelyssä malli on rakennettu käsittelemään sanojen tai niiden osien jonoja. Keskeisenä elementtinä nykyisissä kielimalleissa on huomiomekanismi (engl. attention), joka kohdentaa mallin kulloinkin tarkastelemaan tiettyä osaa aiemmasta syötteestä (Vaswani A. ym. 2017). Yksinkertaisten laskutoimitusten lisäksi nykyaikainen neuroverkko voi sisältää muitakin monimutkaisia osia, jopa erillisiä päättelyalgoritmeja. Niitä on hyödyllisintä ajatella mielikuvituksellisen monimutkaisina funktioina, joita kuitenkin osataan sovittaa annettuun dataan.

Hyvän mallin tärkein ominaisuus on sen yleistyvyys. On helppo opettaa malli tuottamaan oikea vastaus opetusaineistossa nähdylle tapauksille, mutta tämä ei riitä, vaan sen tulee tuottaa mielekkäitä vastauksia myös uusille syötteille. Perinteisten tilastollisten mallien osalta on opittu ajattelemaan, että havaintoja tulisi olla selkeästi enemmän kuin mallin parametreja. Muuten malli ylisovittuu opetusaineistoon ja voi toimia mielivaltaisen huonosti uusille näytteille. Vaikka mallien opettamiseen käytettävät aineistot ovat kasvaneet, mallien monimutkaisuus on monissa tilanteissa kasvanut vielä enemmän ja nykyisin parametreja voi olla jopa miljardeja. Voivatko tällaiset mallit toimia?

Kaikkia yleistyvyyteen liittyviä ongelmia ei missään tapauksessa ole saatu ratkaisua, ja monet alan avoimista ongelmista koskevat mallien luotettavuutta poikkeavissa tilanteissa. Siitä huolimatta voidaan yleisesti sanoa, että erittäin suuria malleja voidaan nykyisin sovittaa luotettavasti. On runsaasti sekä teoreettista (Belkin M. ym. 2019) että empiiristä todistusaineistoa siitä, että miljoonien tai miljardien parametrien mallit voivat toimia hyvin myös uusille syötteille. Vaikka keskusteleva tekoäly ei monissa tilanteissa tuotakaan sisällöllisesti oikeita vastauksia, se ei hätkähdä, vaikka käytettäisiin täysin kuvitteellisia sanoja. Jos vaikkapa kerron maanviljelijän eilen kyntäneen pellon ruosteisella pompellorilla, tekoäly osaa kuvailla pompellorin olevan vanha, mutta edelleen käyttökelpoinen kyntöaura. Tekoäly ei herää pohtimaan, miksi toinen osapuoli käytti kummallista sanaa ja saattaa jopa väittää sen olevan perinteinen murreilmaus, mutta yhtä kaikki se osasi toimia yllättävässä tilanteessa.

2 Tekoälyn historiasta eli onko kuvien tulkinta helppoa vai vaikeaa?

Tekoälyn historia on yhtä pitkä kuin tietokoneiden historia. Alan lähtölaukauksena pidetään vuonna 1956 järjestettyä Dartmouthin kesäseminaaria, neuroverkkojen peruseräite kehitettiin jo 40-luvulla ja koneoppimisen käsite on 50-luvulta. Älyn automatisointi oppimiseen perustuvilla menetelmillä on ollut eräs tietojenkäsittelytieteen tavoitteista käytännössä niin kauan kuin ala on ollut olemassa. Viime vuosien murroksen ymmärtämiseksi on hyvä pohtia hieman sitä, mikä on tekoälylle tai tietokoneille ylipäättään helppoa ja mikä vaikeaa. Tämä on itseasiassa melko hankala kysymys, ja alan kehittyessä tiedeyhteisön vastaukset ovat tavallaan kiertäneet täyden ympyrän.

Aluksi uskottiin, että ihmiselle helpot tehtävät ovat helppoja myös tietokoneille. Lapsikin ymmärtää puhetta, mutta shakkia osaa pelata vain harjaantunut ammattilainen, joten oletettiin, että tietokoneetkin oppivat ymmärtämään puhetta helpommin. Eräs MIT:n professori perusti 1960-luvulla kesän mittaisen opiskelijaprojektin luodakseen ohjelman, joka tunnistaa kuvasta kappaleita luotettavasti. Kylmän sodan molemmat osapuolet pyrkivät tosissaan täysin

automaattiseen konekäännökseen. Tänä päivänä on helppo naureskella näille visioille, mutta aikanaan tällaisten hankkeiden jatkuva epäonnistuminen loi varjon koko tekoälyn kentälle ja johti niin kutsuttuun tekoälyn talveen. Jos edes ”helppoja ongelmia” ei saada ratkaistua, koko alan rahoittaminen lienee turhaa. Vielä tämän vuosikymmenen puolelle asti koneiden kyky käsitellä luonnollista kieltä ja kuvia säilyi vaatimattomana, mutta ihmisen älyn mittatikkuna pidetyissä rajatummissa ongelmissa koneet olivat jo ehtineet kauas edelle. Tietokone voitti shakin suurmestarin ensimmäistä kertaa jo 80-luvun puolella, lähes vuosikymmenen ennen kuuluisaa DeepBluen ja Gasparovin ottelua. Tämä onnistui nykypäivään verrattuna erittäin rajallisilla laskentaresursseilla.

Olikin niin, että ihmiselle helpot asiat ovat koneille vaikeita, vaikka ne suorituvatkin hyvin monista ihmisille vaikeista tehtävistä. Erään selityksen ilmiölle tarjosi 80-luvulla esitelty Moravecin paradoksi (Moravec H., 1988), jonka mukaan havainnointi vaatii merkittävästi enemmän laskentaa kuin päättely. Monimutkaisen aistiympäristön tulkinta ja siihen reagoiminen ovat itseasiassa tehtävinä vaikeita myös ihmisille, mutta eliöiden biologinen koneisto suoriutuu niistä tehokkaasti miljardien vuosien evoluution ansiosta. Abstrakti päättely taas on kehityshistoriallisesti uusi ilmiö. Se ei ole välttämättä erityisen vaikeaa laskennallisesti, mutta ihmisten aivoilla on ollut vain kymmeniä tuhansia vuosia aikaa erikoistua siihen. Havaintojen tulkintaan tarvitaan paljon enemmän laskentaresursseja kuin varhaisilla tietokoneilla oli käytössä.

Ihmiselle lähes automaattinen ympäristön havainnointi oli näin selitetty tekoälylle haastavaksi, mutta ei suinkaan mahdottomaksi. Jos eliöt pystyvät siihen varsin rajallisella energiankulutuksella, sen on oltava mahdollista koneellisesti. Nyt todistamme tätä askelta käytännössä. Meillä on nyt saatavilla sekä tarpeeksi suuria ja monimuotoisia aineistoja että tarpeeksi laskentaresursseja oppimaan aistiärsyksiä käsitteleviä malleja. Tekoälytutkija Andrew Ng ennusti vuonna 2017, että pystymme lähitulevaisuudessa automatisoimaan tekoälyn avulla lähes kaikki sellaiset tehtävät, joista tyypillinen ihminen suoriutuu alle sekunnin pohdinnan avulla. Se alkaa näyttää mahdolliselta.

Olemme tavallaan palanneet tekoälyn alkuhetkiin. Tekoäly suoriutuu nyt melko hyvin aikuiselle ihmiselle lähes automaattisista ympäristön ja kielen havainnointiin liittyvistä pulmista, mutta korkeamman tason ajattelua edellyttävät ongelmat ovat sille edelleen haastavia. Tekoäly osaa nyt salamannopeasti kuvailla näkemänsä kuvan sisällön ja jopa vastata kysymykseen, mutta korkeamman tason päättelyä tehdään edelleen varta vasten kuhunkin tarpeeseen räätälöidyillä algoritmeilla. Niiden avulla voidaan pelata shakkia tai go-peliä, mutta pitkäjänteiseen suunnitteluun vapaassa ympäristössä kykeneviä tekoälyjä ei juurikaan edes tutkita. On kuitenkin hyvä huomata, että syötteitä matalalla tasolla käsittelevät mallit ovat jo hämmästyttävän hyviä monissa vaativissa tehtävissä. Pelkästään niiden avulla voidaan luoda keskustelevia tai kuvia piirtäviä tekoälyjä. Oppimalla käsittelemään aisteja opimme siis samalla enemmän. Palaamme tähän jäljempänä tarkasteltuaamme ensin hieman sitä, miten tähän on päästy.

3 Perustutkimus ja avoimet ympäristöt kehityksen vetureina

Rinnakkain on tapahtunut laskentatehon, käytettävissä olevien tietoaineistojen ja algoritmien kehitystä. Nämä ovat sikäli yhtä tärkeitä, että ilman kaikkien kolmen kehitystä emme olisi vielä lähelläkään nykyisiä menetelmiä. Toisaalta millään näistä osa-alueista ei ole tapahtunut sellaista yksittäistä läpimurtoa, joka olisi kriittinen.

Voidaankin perustellusti sanoa, että tekoälyn murros on seurausta pitkäjänteisestä ja laaja-alaisesta perustutkimuksesta. Se on parhaita esimerkkejä siitä, että perustutkimukseen panostetut eurot ja tunnit maksavat itsensä takaisin. Nykyisten menetelmien keskiössä on edelleen sekä Robbinsin ja Monron jo vuonna 1951 luoma optimointiperiaate, että Seppo Linnainmaan 70-luvun alussa kehittämä vastavirta-algoritmi. Monet syvien neuroverkkojen perusratkaisuista on kehitetty viime vuosituhanella. Nykyisin alan tutkijoita on moninkertaisesti enemmän ja uusia menetelmiä julkaistaan päivittäin. Seuraavien läpimurtojen elementit seuloutuvat tästä tutkimusmassasta. Uuden tehokkaamman algoritmin voi kehittää jopa yksittäinen tohtorikoulutettava pienessä maassa, koska kyse on pohjimmiltaan matematiikasta ja varsin lyhyistä ohjelmakoodin palasista.

Teknisten kehitysaskeleiden rinnalla muutosta ovat vauhdittaneet uudet toimintatavat, keskeisimpinä avoimuus ja paremmat työkalut. Nykyisin on tarjolla suunnaton määrä laadukkaita avoimia ohjelmistoja, jotka helpottavat tekoälymenetelmien kehittämistä. Omina opiskeluvuosinani neuroverkon toteuttaminen edellytti viikkojen työpanoksen, mutta nyt jokainen ohjelmoinnin perusteet tunteva voi luoda yksinkertaisen neuroverkon muutamalla koodirivillä. Valtaosa työkaluista on avointa lähdekoodia ja ilmaiseksi kaikkien saatavilla, valmiita esimerkkejä on saatavilla työn tueksi liki rajattomasti ja monet menetelmien kehittämisessä ja testaamisessa käytettävät datat ovat avoimia. Uusia malleja ja algoritmeja esittelevien tieteellisten artikkelien ohessa tarjotaan usein myös avoin ohjelmistototeutus. Näiden työkalujen avulla matka uudesta tieteellisestä ideasta toimivaksi, varmistetuksi malliksi voi onnistua jo viikoissa.

Oleellista on, että ilmaiset työkalut eivät ole vain tutkijoiden ja harrastelijoiden puuhastelua. Yritykset käyttävät täsmälleen samoja työkaluja ja monien avointen työkalujen taustalla on alan suurimpia toimijoita Googlesta Metaan. Syväoppimisen alustat ovat nousseet Linuxin rinnalle avoimen lähdekehityksen airuina ja näkyvimpinä esimerkkeinä. Luotettavat työkalut ovat suuryrityksille kriittisen tärkeitä ja avoin kehitys on todettu parhaaksi tavaksi pitää ne ajan tasalla. Kyse ei ole hyväntekeväisyydestä, vaan ainoasta vaihtoehdosta: suljettuun omaan järjestelmään nojaava yritys jäisi auttamatta kilpailijoista jälkeen. Avoimen ympäristön päälle rakennettavaan korttitaloon on helppo liittää kaikkein tuoreimmat menetelmät ja työvoima liikkuu helposti organisaatiosta toiseen, myös tutkimusmaailman ja yritysten välillä.

Suuryritysten merkitys tekoälyn murroksessa on kokonaisuutena moninainen ja yrityksiä on aiheesta kritisoitu erityisesti henkilökohtaisen datan hyödyntämiseen ja kaupallistamiseen liittyvistä ongelmista, mutta avointen työkalujen näkökulmasta ne ovat olleet tärkeitä tekoälykehityksen vauhdittamisessa. Ilman laadukkaita työkaluja sekä tutkimus että yritysten mahdollisuudet hyödyntää tekoälyjä olisivat selvästi jäljellä nykyisestä.

4 Enemmän irti datasta

Algoritmit ovat kehittyneet merkittävästi, mutta sitä kehitystä ei voi tällaisessa katsauksessa käsitellä laajemmin. Yksi käsitteellinen muutos voidaan kuitenkin nostaa esille. Historiallisesti valtaosa koneoppimisesta on keskittynyt ohjattuun oppimiseen, jossa malli opitaan syöte-vaste-pareista. Esimerkiksi kuvien luokittelijan opettamiseen tarvittiin tietoa eli dataa, jossa kaikista kuvista on valmiiksi kerrottu, mitä kuva esittää. Tälläkin periaatteella saatiin noin vuosikymmen sitten luotua kuvia hyvin tulkitsevia malleja (Krizhevsky A. ym. 2012). Tässä opetustavassa valmiiden vastausten luominen

on kuitenkin aina merkittävä pullonkaula: ei auta, vaikka tarjolla olisi miljardi valokuvaa, jos on resursseja muodostaa toivottu vaste vain pienelle murto-osalle niistä. Mitä jos koko miljardin kuvan aineistoa voitaisiin hyödyntää tehokkaasti ilman käsityötä vastausten muodostamiseksi?

Tämä onnistuu niin kutsutun itseohjatun oppimisen (engl. self-supervised learning) avulla. Itseohjatussa oppimisessa malli opitaan ohjatun oppimisen algoritmeilla, mutta tavoitteena on ratkaista jokin keinotekoinen tehtävä, jonka vastaus on valmiina datassa. Kielimallia pyydetään ennustamaan lauseen seuraava sana tai täyttämään keskeltä lausetta poistettu sana. Kuvia käsittelevää mallia voidaan pyytää täyttämään kuvasta tarkoituksella poistettu osa tai poistamaan kuvasta siihen varta vasten lisättyä kohinaa (kuvio 1). Näihin tehtäviin voidaan käyttää hyviksi tunnettuja ohjatun oppimisen algoritmeja, mutta oikeat vastaukset ovat jo olemassa ja siten kaikki data saadaan käyttöön. Lisäksi samasta datasta voidaan muodostaa liki rajattomasti ennustustehtäviä: samaan kuvaan voidaan lisätä aina uudenlaista satunnaista kohinaa.

Kohinan lisääminen kuvaan



Kohinan poistaminen kuvasta

Kuvio 1: Kuvia luova tekoäly voidaan opettaa itseohjatun oppimisen periaatteella.

Opetuskuviin lisätään keinotekoisista kohinaa ja malli opetetaan ennustamaan alkuperäinen kuva kohinaisesta versiosta. Diffuusiomallit käyttävät tähän tarkoitukseen differentiaalilaskentaa ja osaavat lopulta luoda pelkästä kohinasta tarkkoja kuvia.

On perusteltua kysyä, mitä hyötyä on mallista, joka osaa täyttää puuttuvan osan kuvasta. Vastaus on yksinkertainen: mallin on opittava jotain oleellista kuvan sisällöstä, jotta se pystyy täyttämään puuttuvan alueen. Jos tennisottelun kuvasta poistaa keskikentän, malli osaa täyttää sinne verkon vain, jos se on jollain tasolla oppinut, että tennisottelussa palloa lyödään verkon yli. Jos malli kykenee samaan kaikkien mahdollisten kuvien kanssa, sen on täytynyt oppia jotain kuvista yleensä. Kaikki tämä ”ymmärrys” on jollain tavalla koodattuna mallin sisäiseen esitykseen, jonka se muodostaa saatuaan kuvan syötteenä. Tämän esityksen avulla on merkittävästi helpompaa ratkaista alkuperäinen ongelma. Enää ei tarvitakaan miljoonien kuvien ja niiden merkitysten opetusaineistoa, vaan jo muutamat sadat esimerkit riittävät. Ne on helppo kerätä.

Tämä oppimisen periaate on yhdessä suurempien datojen ja parempien mallien kanssa auttanut murtamaan Moravecin paradoksin rajat. Osaamme nyt opettaa luonnollista kieltä, kuvia, puhetta ja videota käsitteleviä malleja. Kaikille näille on ominaista se, että niistä on tarjolla lähes rajattomasti aineistoa ja osaamme muodostaa luontevia ennustustehtäviä niiden opettamiseksi.

5 Itseohjatusta oppijasta luovaksi tekoälyksi

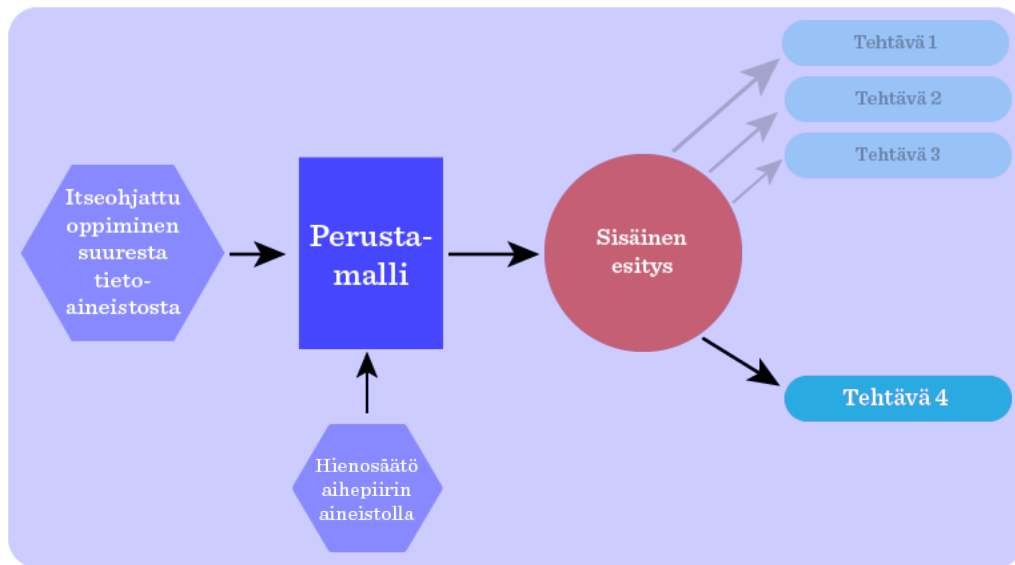
Edellä kuvasin teknistä oppimisperiaatetta, jonka avulla tekoäly saatiin ymmärtämään monimutkaisia aistihavaintoja. Kuten aiemmin vihjasin, tämän kyvyn seurauksena saatiin myös enemmän. Sekä kielimalleista että kuvia tarjoavista tekoälyistä käytetään nykyisin yleistermiä luova tekoäly (engl. generative AI, huom. ei creative). Tällä käsitteellä viitataan siihen, että ne osaavat luoda uutta sisältöä: ne voivat jatkaa annetusta tekstisyötteestä mielivaltaisen pitkälle tai tuottaa pelkästä kohinasta valokuvan. Kieli- ja kuvamallien yhdistelmällä voidaan tuottaa kuvia, jotka vastaavat annettua sanallista kuvausta.

Nämä tekoälyn luomat sisällöt ovat hätkähdyttävän laadukkaita ja monella tapaa hyödyllisiä. Jos syötteenä kielimallille käytetään ihmisen puheenvuoroa, tekstiä eteenpäin jatkava kielimalli muuntuu suoraan yksinkertaiseksi keskustelevalksi tekoälyksi. Vaikka mallilla ei voida sanoa olevan ymmärrystä käsiteltävästä aiheesta, se osaa silti tuottaa mielekkäitä vastauksia myös vaativiin kysymyksiin, joihin vastaaminen edellyttää monimuotoisen tiedon yhdistämistä ja muistuttaa erehdyttävästi asian syvällistä ymmärrystä. Voidaan ajatella, että tekoäly osaa erinomaisesti näytellä henkilöä, joka tietää vastauksen haluttuun kysymykseen. Monissa tapauksissa se riittää, mutta ei kaikissa.

Tekoäly osaa hetkessä tuottaa kuvia, joiden piirtämiseen valtaosa ihmisistä ei kykenisi vuosikausien harjoittelunkaan jälkeen. Toisaalta me osaamme kuvitella vastaavia kuvia ja tämä on meille yhtä automaattista kuin näkökentän tulkinta. Tekoälyn kyvyt kuvien ja videon tuottamisessa on osuvampaa rinnastaa kuvitteluun. Kunhan kone on oppinut kuvittelemaan, varsinaisen kuvan muodostaminen, joka merkitsee pikselien tallentamista tietokoneen muistiin, on tietokoneelle erittäin yksinkertainen tehtävä ja sen takia kuvittelun tulos saadaan ihmisen hyödynnettäväksi ilman piirtämiseen tarvittavia motorisia taitoja. Koneelle vaikean osatehtävän ratkaisu johti suoraan häkellyttäviin tuloksiin, koska muut osaongelmat olivat sille helppoja.

6 Perustamallit kehityksen pohjana

Vaikka edellä kuvatuille malleille löytyy runsaasti käyttöä jo luovina tekoälyinä, niillä on myös toinen merkitys, jossa mallin arvo on sen sisäisessä esityksessä. Suuret kielimallit ja kuvia ymmärtävät mallit ovat esimerkkejä niin kutsutuista perustamalleista (engl. foundation model) (Bommasani R. ym. 2021). Ne ovat malleja, jotka on opetettu suurilla tietoaaineistoilla siten, että ne oppivat mielekkään sisäisen esityksen kunkin mallin ominaiselle datalle. Nimensä mukaisesti niitä voidaan käyttää pohjana tai perustana uusille malleille, mutta ne eivät vielä ratkaise kuin pienen joukon ongelmia (kuvio 2).



Kuvio 2. Perustamallit opetetaan laajalla ja monipuolisella tietoa-aineistolla ratkaisemaan erilaisia tehtäviä itseoppimisen periaatteella. Malli oppii havainnoille sisäisen esityksen, joka auttaa uusien ongelmien ratkaisemisessa. Malli voidaan hienosäätää sovellusalueen datalla ja sen päälle on helppo opettaa malleja uusiin tehtäviin.

Tällä hetkellä moni tutkii, miten perustamallien avulla voidaan ratkaista tehokkaasti uusia ongelmia. Helpoimmillaan se tapahtuu hienosäätämällä mallia tietyn tehtävän datalla. Esimerkiksi kuvia ymmärtävää mallia voidaan käyttää lähtökohtana lääketieteellisten kuvien analyysissä. Mallin oppimiseen tarvitaan vähemmän dataa, koska perustamalli osaa jo käsitellä kuvissa esiintyviä rakenteita ja niiden välisiä suhteita. Perustamalleihin voidaan myös liittää muita malleja ja kielimalleihin lisätäänkin jatkuvasti uusia toiminnallisuuksia sallimalla, että se kutsuu ulkoisia ohjelmia.

Enää tuskin kannattaa luoda luonnollista kieltä tai kuvia käsittelevää tekoälymenetelmää, joka ei jollain tavalla rakentuisi valmiiksi opettujen perustamallien päälle. Tämä muuttaa myös alan tutkimusta. Jokainen menetelmä rakentuu entistä suuremmin muiden tarjolle tuomien komponenttien päälle. Perustamallien hyödyt ovat selviä, mutta niiden käyttöön liittyy myös avoimia kysymyksiä. Miten tutkijan tai yrityksen tulisi suhtautua siihen, jos oman ratkaisun pohjana käytetystä kielimallista myöhemmin löydetään merkittävä puute tai se on kehitetty epäeettisesti? Entä jos perustamallin kehittänyt yritys vetää sen pois saatavilta tai alkaa laskuttaa sen käytöstä päätähuimaavia summia? Miten varmistetaan, että oma ratkaisu toimii hyvin tulevienkin perustamallien kanssa?

Samaan aikaan luodaan myös uusia perustamalleja. Ihmisen aistiympäristö kattaa loppujen lopuksi melko pienen osan niistä ilmiöistä, joita haluaisimme tekoälyjen käsittelevän. Samoihin oppimisperiaatteisiin nojaten kehitetään malleja esimerkiksi satelliittikuville ja lääkemolekyyleille, mutta useimmiten tarvitaan uutta tieteellistä tietoa esimerkiksi siitä, kuinka muodostetaan kullekin datatyypille parhaiten sopivia oppimistehtäviä itseohjattuun oppimiseen. Kohinan poisto tai seuraavan sanan ennustaminen ovat luontevia tehtäviä, mutta minkä tehtävän avulla opitaan hyvä malli solubiologiasta?

Toinen merkittävä haaste on riittävän kattavan datan kerääminen ja mallin usein korkeiden kehityskustannusten kattaminen. Suomessakin on paljon osaavia tekijöitä, jotka tuntevat perustamallien oppimiseen tarvittavat algoritmit ja CSC tarjoaa kattavat

laskentaresurssit, mutta datan kerääminen ja käsittely on kallista. Suuryhtiöt ovat panostaneet kielen ja kuvan perustamalleihin suunnattomasti, koska näkevät niille kaupallisesti kannattavia sovelluksia, mutta emme voi luottaa niiden luovan vastaavia malleja kaikkiin tarkoituksiin. Esimerkiksi räätälöityjä opetuksen tukiratkaisuja voitaisiin tehokkaasti kehittää kunkin lapsen osaamistarpeita ymmärtävän perustamallin päälle, mutta olemmeko valmiita rahoittamaan perustamallin kehityskustannukset, jos konkreettiset hyödyt saadaan esille vasta vuosikymmenen kuluttua? Miten kemianteollisuuden yritykset saadaan yhdessä ponnistelemaan prosessiteollisuuden ilmiöiden mallintamiseen tarvittavien perustamallien eteen?

7 Avoimet haasteet – mitä seuraavaksi?

Kyky käsitellä ihmisen aistiympäristöä ratkaisee vain pienen, vaikkakin tärkeän, osan tekoälyn haasteista ja työtä riittää vielä paljon.

Alan kenties suurin avoin ongelma on kausaliiteetti eli syy-seuraussuhteet. Ennustaminen ei riitä monissa sovelluksissa, lääketieteestä poliittiseen päätöksentekoon, vaan haluttaisiin päätellä, mikä suunnitellun toimen seuraus olisi. Kausaliiteetin parissa on tehty alalla tutkimusta jo vuosikymmenien ajan, mutta valmiita ratkaisuja on vielä valitettavan vähän. Kausaalisia suhteita voidaan kyllä päätellä luotettavasti, mutta vain melko rajatuissa tilanteissa. Muissa tilanteissa voidaan usein osoittaa, että kausaalisia suhteita ei voida aukottomasti päätellä. Tämä on tärkeä tieto, mutta soveltajien näkökulmasta usein pettymys.

Toinen keskeinen haaste liittyy epävarmuuksien luotettavaan mallintamiseen. Erityisesti on tärkeää, että autonomisesti toimivat järjestelmät pystyvät arvioimaan, milloin ne ovat varmasti oikeassa ja milloin vastaukseen liittyy epävarmuuksia. Vaikka tilastollisten oppimismenetelmien avulla voidaankin periaatteessa ottaa epävarmuudet huomioon perustellulla tavalla, monet nykyisistä malleista ovat liian itsevarmoja. Epävarmuuksien huomioimisen tärkeys korostuu silloin, kun mallien oppimiseen on käytettävissä vain vähän dataa. Vaikka osassa sovelluksista datan määrä on vain kustannuskysymys, monissa tilanteissa dataa ei yksinkertaisesti ole olemassa enempää. Ensimmäistä Mars-lentoa tehdessä ei ole olemassa aiemmilla lennoilla mitattua dataa, fossiileja on olemassa rajatusti, ja harvinaisia sairauksia potevia potilaita on vähän kuten sairausryhmän nimi kertoo.

Tutkimusta tarvitaan myös ihmisten ja tekoälyjen yhteistyön parantamiseksi. Moni kaipaa tulkittavia tekoälyjä, jotka osaavat aina perustella, miksi päätyivät antamaansa vastaukseen. Vaikka tämä on pääpiirteittäin toivottava ominaisuus, on hyvä muistaa, että myöskään ihmiset eivät ole tulkittavia. Kuka osaa kertoa, miksi naurahti, kun työkaveri lounaalla valitti liian suolaisesta ruuasta? Osaako radiologi kertoa, miksi hän teki yhden potilaan kohdalla virhetulkinnan? Voinko luvata, että kohtelen kaikkia asiakkaita tasapuolisesti myös väsyneenä? Tekoälyjärjestelmät ovat jo nyt usein tässä mielessä läpinäkyvämpiä ja tasapuolisempia kuin ihmiset, mutta me edellytämme tekoälyltä enemmän kuin toisiltamme. Keskustelevat tekoälyt osaavat jo pyydettyä muokata ilmaisuaan huomioidakseen keskustelukumppanin mielentilan, mutta meille tämä on erittäin hankalaa. Onko vastuu ihmisen tunteisiin ja oikkuihin reagoimisessa jatkossa nimenomaan tekoälyllä?

On aika kohdistaa katse myös siihen, miten tekoälyjä hyödynnetään mahdollisimman tehokkaasti kaikkein hankalimpien ongelmien ratkaisemisessa. Mikä on tekoälyn asema ilmastonmuutoksen hallinnassa tai miten sen avulla edistetään yhdenvertaisuutta?

Näiden ongelmien ratkaisu ei ole autonominen tekoäly, eikä myöskään erilliset tekoälymallit osana ihmisen työkalupakkia. Niiden sijaan tarvitaan ratkaisuja, jossa ihminen ja tekoäly toimivat aidosti yhteistyössä. Tähän keskittyvää tutkimusta tehdään Suomen Akatemian lippulaivahankkeessa Suomen tekoälykeskus (<https://fcai.fi>). Esimerkiksi oma tutkimukseni liittyy siihen, miten tekoälyn avulla tuetaan ihmisiä sekä tieteellisessä tutkimuksessa että tuotekehityksessä niin, että käytettävät tekoälymallit yleistyvät laajasti eri aloilla. Näin vältetään resurssien tuhlaaminen alakohtaisiin ratkaisuihin.

Yhteistyöhön rakentuvassa ratkaisussa tekoälyn on pyrittävä tekemään omasta toiminnastaan läpinäkyvää sekä yritettävä ymmärtää ihmisen tavoitteita ja toimia. Toisaalta vastuuta yhteistyöstä ei pidä ulkoistaa vain tekoälylle, vaan myös meidän kannattaa toimia niin, että yhteistyö on mahdollisimman tehokasta. Kenties kemistin kannattaa tekoälyn pyynnöstä tehdä muutama laboratoriokoe lisää, vaikka hän ei itse heti ymmärrä niiden tarvetta? Alamme hyväksyä tällaiset pyynnöt vasta, kun luotamme siihen, että tekoäly pystyy niiden avulla auttamaan meitä ratkaisevasti paremmin. Voimme oppia luottamaan vain kokeilemalla.

Viitteet

Vaswani A., ym. Attention is all you need. *Advances in Neural Information Processing Systems* 30, 2017.

Belkin M., ym. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences (PNAS)*, 116(32):15849–15854, 2019.

Moravec H., *Mind children: The future of robot and human intelligence*, Harvard University Press, 1988.

Krizhevsky A., ym. ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems* 25, 2012.

Bommasani R., ym. On the Opportunities and Risks of Foundation Models. *arXiv:2108.07258*, 2021.