

ÄLYKÄSTÄ PASKAPUHETTA - TEKOÄLYMALLIT IHMISÄLYN PEILINÄ

Onko tekoälyllä älykkyyttä? Luultavasti on, sillä sen käymää keskustelua ei enää juuri erota ihmisten käymästä keskustelusta. Ihmistenkään järkeily ei nimittäin useimmiten huikaise älykkyyden tasollaan. Kielimallit haastavat romantisoidut käsitykset ihmisjärjestä ja luovuudesta.

Tekoäly ja ATK-hype ovat vuosikymmenten aika tulleet ja menneet erilaisina aaltoina. Moni asia on toki muuttunut, kun tietokoneiden kehitys on yhdistynyt internetin myötä tiedonvälityksen muutokseen, mutta luvattuja vallankumouksia ei juurikaan ole näkynyt.

Viime syksyyn asti suhtautumisemme tekoälyn uusimpaan aaltoon oli epäilevää. 2010-luvun edistysaskeleet perustuivat jo 1980-luvulla kehitettyihin algoritmeihin. Uutta oli lopulta vain laskentatehon kasvu ja sovellusten saavutettavuus.

Kenties köykäisin arjessa näkyvä tekoälyn ilmentymä ovat olleet asiakaspalvelutyöhön

valjastetut chattibotit. Jokainen näitä järjestelmiä käyttänyt tietää, että ainoa järkevä tapa käyttää chattibottia on käskyttää se kutsumaan asiakaspalvelija paikalle. Siksi viime talvena ensikosketuksemme ChatGPT-sovellukseen oli hätkähdyttävä.

Käyttäjä voi esittää ChatGPT:lle erilaisia kysymyksiä tai pyyntöjä. Sen vastauksena tuottama teksti on yhtenäistä ja keskustelu järjellisen oloista. Sen antamat vastaukset ovat ajoittain virheellisiä, mutta ne ovat sitä hämmästyttävän inhimillisillä tavoilla: aivan kuin kone muistelisi ja välillä sekoitaisi asioita mielessään.

ChatGPT:n kaltaisilla kielimalleilla on käytännössä loputon kyky tuottaa pinnallista

Google Deepmind on kutsunut taiteilijoita kuvittamaan tekoälyä.

Tämän teoksen on luonut graafikko Khyati Trehan.

Älykkyyden käsitteessä ei ole mitään sellaista, minkä perustella tekoälyjärjestelmät eivät olisi oikeasti älykkäitä, jos ne selvästi käyttäytyvät älykkään oloisesti.

tekstiä aiheesta kuin aiheesta, tyylillä kuin tyylillä. Ainakin englannin kielellä lopputuloksen erottaminen ihmisen kirjoittamasta tekstistä on hankalaa.

Tässä vaiheessa on jo melko selvää, että kielimallien potentiaalia tullaan käyttämään informaatiovaikuttamiseen, internetissä käytävän julkisen keskustelun hämmentämiseen ja muuhun vastaavaan. Moni työnkuva muuttuu tavalla tai toisella, kun sujuvan tekstin ja visuaalisen sisällön tuottamisen kulut romahtavat. Lisäksi osa näistä muutoksista tulee tapahtumaan nopeasti. Kyse voi olla pikemminkin kuukausista kuin vuosista. Sen sijaan on epäselvää, onko älykkyyden suhteen käsillä jotain aidosti uutta verrattuna aiempaan.

Olemme toistuvasti törmänneet niin julkisessa kuin yksityisissäkin keskusteluisa uusia tekoälyjärjestelmiä vähätteleviin reaktioihin. Tekoälyä pidetään vain automaattina, joka mekaanisesti yhdistelee sen opettamiseen käytetyssä materiaalissa esiin-

tyneitä asioita. Näin toki onkin, mutta monet kohtaamistamme reaktioista vaikuttavat perustuvan ihmismielen mystifointiin sekä romantisoituun käsitykseen ihmisälystä ja luovuudesta. Tulisiko kielimalleja sittenkin pitää älykkäinä?

MITÄ ÄLYKKYYS ON?

Ihmismieltä tutkivien tieteenalojen keskuudessa älykkyydellä tarkoitetaan kykyä toimia tarkoituksenmukaisesti vain osittain ennustettavassa ympäristössä. Erityisesti tekoälyn yhteydessä toinen yleinen luonnehdinta tulee filosofilta ja tietojenkäsittelytieteiden pioneerilta **Alan Turingilta** (1912–1954), joka vuonna 1950 määritteli keinotekoisien älykkyyden inhimillisen älykkyyden kautta: konetta tulisi pitää älykkäänä, mikäli sen käymää keskustelua ei voi erottaa ihmisen käymästä keskustelusta.

Edellä kuvatut älykkyyden määritelmät pyrkivät siis selventämään, mitä älykkyy-

on, eivätkä ne esimerkiksi liity siihen, miten älykkyyttä mitataan. Tieteellisessä ja filosofisessa tutkimuksessa käytetyt älykkyyden määritelmät eivät ankkuroidu tietoisuuden käsitteeseen. Ne eivät myöskään ota kantaa siihen, ymmärtääkö tekoäly käymäänsä keskustelua tai edellyttääkö kielen ymmärtäminen edes muuta kuin kykyä käydä mielekästä keskustelua. Moni filosofi vastaisi jälkimmäiseen kysymykseen kielteisesti.

Valtavirtaa edustavan käsityksen mukaan kielen ymmärtäminen on yksinkertaisesti kykyä käyttää sitä oikein tai mielekkäällä tavalla. Viime aikojen julkisessa keskustelussa esitetyt argumentit kielimallien älykkyyttä vastaan vaikuttavat kuitenkin olettaen, ettei aitoa älykkyyttä voi olla ilman tietoisuutta, ymmärtämistä tai luovaa ajattelua.

Mielestämme ChatGPT on ensimmäinen tekoälyjärjestelmä, joka toteuttaa Turingin määritelmän älykkyydestä tai ainakin tulee sitä hyvin lähelle. Emme väitä, että esimerkiksi ChatGPT olisi älykäs täysin inhimil-

lisellä tavalla. Sen sijaan haluamme korostaa, ettei tieteellisessä ja filosofisessa älykkyyden käsitteessä ei ole mitään sellaista, minkä perustella tekoälyjärjestelmät eivät olisi oikeasti älykkäitä, jos ne selvästi käyttäytyvät älykkään oloisesti.

Toinen – mielestämme tärkeämpi – huomio on, että tekoäly voi toimia peilinä, jonka kautta on mahdollista tarkastella omaa älyämme. Monilla meistä on rima kovin korkealla, kun arvioimme koneen älykkyyttä tai älyn puutetta. Kuitenkin myös suurin osa omasta päivittäisestä kielenkäytöstämme on stereotyyppistä ja pinnallista höpötystä vailla kummempaa järkeilyä tai luovaa ponnistusta.

MITÄ MIELI ON, JOS SE EI OLE TIETOISUUTTA?

Tieteellisen psykologian kenties kauaskantoisin ajatus on, että emme voi suoraan havaita oman mielemme toimintaa. Ajatus on ensinäkemältä outo, sillä tuntuuhan ilmeiseltä, että tiedämme, mitä ajattelemme ja mitä päätöksiä teemme ja miksi.

Ajatus tietoisuuden tavoittamattomissa olevasta mielestä löytyy jo **Sigmund Freudilta** (1856–1936). Eri muodoissa tämä saa kuitenkin vahvistusta 1900-luvun tutkimuksista, joilla ei ole mitään tekemistä Freudin teorioiden kanssa.

Ympäristön kohteiden tunnistaminen verkkokalvolle heijastuvasta valosta, kielen

Sepittelemme sujuvasti järkevältä vaikuttavia syitä sattumanvaraisille päätelmillemme.

ymmärtäminen tai vaikkapa pesäpallon heittäminen edellyttävät erittäin monimutkaista prosessointia, joka tapahtuu aivoissa ja muualla keskushermostossa. Emme voi havaita sitä mieltämme luotaamalla mitenkään. Yllättävämpää on, että myös suurin osa joka-päiväisistä päättelystä ja päätöksenteosta tapahtuu ”pinnan alla”.

Suuri määrä noin 1970-luvulta lähtien julkaistuja tutkimuksia on tehnyt selväksi, että inhimillinen järkeily on usein epäloogista, ja usein vähäpätöisiltä vaikuttavat tekijät huomaamatta vaikuttavat siihen, kuinka teemme päätöksiä. Hätkähdyttävintä näissä tutkimuksissa ei ole niinkään se, että olemme usein epärationaalisia, vaan se, kuinka sokeita olemme omalle irrationaalisuudellemme. Sepittelemme sujuvasti järkevältä vaikuttavia syitä sattumanvaraisille päätelmillemme.

Lundin yliopistolla tehdyssä tutkimuksessa koehenkilöitä pyydettiin arvioimaan asteikolla 0–100, kuinka samaa mieltä he olivat eräistä poliittisista väitteistä. Tutkijat vaihtoivat salaa osan koehenkilöiden vastauksista ja pysyivät jälkikäteen perusteluja näille

muutetuille vastauksille. Noin puolissa tapauksista vastaajat eivät huomanneet mitään poikkeavaa, vaan he tuottivat sujuvasti järkevän oloisia perusteluja tutkijoiden valitsemille mielipiteille kuin ne olisivat olleet heidän omiaan.

Vastaavia tuloksia on saatu monissa järkeilytutkimuksissa. Tulokset kiinnittävät huomion siihen, ettei perusteluja tuottava tietoinen ajattelu useinkaan saa ihmisiä huomaamaan virheitä tai vaihtamaan kantojaan. Ihmiset usein keksivät järkevänoloisia syitä melko satunnaisista syistä syntyneille päätöksilleen ennemminkin kuin perustavat päätöksentekonsa tietoiselle järkeilylle.

Järkeilyimme on usein pinnallista. Tämä ei ole lopulta kovin yllättävää. Arkinen ympäristömme on yllättävän monimutkainen, ja teemme tiedostamattamme jatkuvasti pieniä rutinoituneita päätelmiä ja päätöksiä. Epätyypillisissä tilanteissa rutiinit kuitenkin voivat toimia epätarkoituksenmukaisesti tai johtaa harhaan.

Valintojemme perusteleminen on eräs arkisista rutiineistamme, jonka tarkoitus lienee ensisijaisesti sosiaalinen. Haluamme vaikut-

taa järkeviltä muiden silmissä. Siksi olemme taipuvaisia ja taitavia tuottamaan sujuvia selityksiä asioille, jotka voivat olla epärationaalisia tai jotka eivät edes pidä paikkaansa. Usein järkeilyimme on siis vain sujuvaa paskapuhetta.

On toki myös selvää, että toisinaan järkeilyimme todella on tietoista ja rationaalista. Kuitenkin se tosiasia, että suuri osa järkeilytämme on tiedostamatonta, työntää kiilaa tietoisuuden ja mielen käsitteiden väliin. Se mahdollistaa ajatuksen, että myös ei-tietoinen järjestelmä, kuten tekoälysovellus, voi olla älykäs. Herää kysymys, että jos mieli ei ole sama kuin tietoisuus, mitä se sitten on.

Psykologit, aivotutkijat ja muut ihmismieltä tutkivat kognitiotieteilijät ovat jo vuosikymmeniä vastanneet, että mieli on tulosta aivoissa toimivista mekanismeista, jotka prosessoivat informaatiota. Aivojen biologinen tehtävä on käsitellä informaatiota siinä missä ruoansulatuselimistön tehtävä on käsitellä ravintoa. Mieli ei näin ole tietoisesta minän vaan aivojen ja osin muun kehon toimintaa. Tällainen näkemys itse asiassa kääntää mielen ja tietoisuuden roolit pääläelle: tietoisuus on yksi aivojen ylläpitämien mielen toimintoista tai kenties vain sen sivutuote.

Tietokone on Turingin ajoista asti ollut tärkeä mielen metafora: se on konkreettinen ja hyvin ymmärretty esimerkki järjestelmästä, joka käsittelee tietoa. Samalla se toimii

vastaesimerkkinä ajatukselle, jonka mukaan päättely tai muu älykkyyteen viittaava toiminta edellyttäisi tietoista mieltä. On toki mahdollista, että tietoisuudella lopulta on osansa ihmisen älykkään käyttäytymisen selittämisessä, mutta tieteellisen nykytiedon valossa tuo yhteys ei ole lainkaan selvä. On siksi perusteltua pitää nämä asiat erillään.

LUOVUUS KIERRÄTTÄMISENÄ

Myös tieteellinen käsitys luovuudesta eroaa sitä koskevasta arkikäsitelmästä. Arkisesti on tapana usein korostaa luovien keksintöjen ainutlaatuisuutta ja sitä, kuinka niiden ennakoiminen on mahdotonta. Luovuutta ja innovaatioita koskevassa tieteellisessä tutkimuksessa kuitenkin tyypillisesti korostuu asiantuntijuuden rooli.

Luovuus perustuu yleensä entuudestaan tunnettujen asioiden yhdistelyyn. Esimerkiksi **Thomas Edisonin** (1847–1931) laboratoriossa tehdyt tekniset keksinnöt perustuivat laajoihin ja järjestelmällisiin tutkimuksiin, joissa laboratorion työntekijät yhdistelivät olemassa olevia teknologioita yhä uusilla tavoilla. Tätä myös **Isaac Newton** (1642–1727) tarkoitti sanoessaan, että hän seisoo jättiläisten harteilla. Suurin osa tieteellisestä työstä oli jo tehty, joten hänen tehtäväkseen jäi lisätä hieman jotain omaa mutta ennen kaikkea liittää jo olemassa olevat tiedonpalaset uudella tavalla yhteen.

Kun pyörittelemme ajatuksia tai työstämme teosta, voi tuntua siltä, että mieleemme vain kumpuaa erilaisia oivalluksia, joiden alkuperä on sisäisessä intuitiossa.

Sama pätee myös luovuuteen taiteilijan työssä. Omaa alaa koskeva asiantuntijuus mahdollistaa taiteilijoille uusien ilmaisutapojen kehittämisen yhdistelemällä aiempia tyylejä ja teemoja. Tämä on oikeastaan melko ilmeistä. Vaikka jotkut taiteelliset teokset ja tyyllilajit ovat urauurtavia, niillä kaikilla on oma tunnistettava kulttuurihistoriansa. Tieteelliset ja filosofiset oivallukset ovat tässä suhteessa samankaltaisia.

Ei ole mitään periaatteellista syytä, miksi tällainen luova yhdistelyprosessi olisi mahdollinen vain ihmisäivoille. Päinvastoin kielimalleilla on käytössään valtava määrä materiaalia, jossa algoritmi tunnistaa säännönmukaisuuksia. Yhdistelemällä tätä materiaalia se muodostaa uusia keskusteluvastauksia, tekstejä tai kuvia. Luovalle toiminnalle ominainen piirre on nähdä tuttuja asioita uudesta näkökulmasta, mikä voidaan ymmärtää juuri kykynä luoda tällaisia uusia odottamattomia yhdistelmiä.

Meillä on usein taipumus romantisoida inhimillistä luovuutta siinä missä älykkyyt-

täkin. On silti syytä olla avoin sille mahdollisuudelle, että ihmisen kyky tuottaa uusia luovia yhdistelmiä saattaa lopulta olla varsin rajallinen ja luova käytöksemme on usein varsin ennalta arvattavaa. Koneelle on vaikeampi opettaa, mitä on hyvä maku ja miten arvioida, mitkä uudet yhdistelmät ovat mielenkiintoisia ja tähdellisiä. Nämä ovat nimittäin sosiaalisia kykyjä, koska tuotosten merkitys ja vastaanotto riippuu siitä, mihin kulttuuriseen ympäristöön ne tuodaan.

Kuitenkaan luova ajattelu ei välttämättä tunnu palasten yhdistelemiseltä. Tämä johtuu jälleen siitä, että suurin osa niin luovasta kuin rutinoituneestakin ajattelustamme tapahtuu alitajuisesti. Kun pyörittelemme ajatuksia tai työstämme teosta, voi tuntua siltä, että mieleemme vain kumpuaa erilaisia oivalluksia, joiden alkuperä on sisäisessä intuitiossa. Näin onkin, mutta tässä intuitiossa ei ole mitään mystistä. Kyse on alitajuisista ja osin automatisoituneista mielen toiminnoista, joille lähes kaikki arkiset kykymme ja niihin liittyvä ajattelu perustuvat.

TAIDOISTA JA YMMÄRTÄMISESTÄ

Palataan vielä kysymykseen kielen ymmärtämisestä. Vieraan kielen oppimista voi hyvin tarkastella esimerkkinä tiedostamattomien rutiinien toiminnasta.

Vierasta kieltä oppiessamme joudumme aluksi tietoisesti arvailemaan, mitä sanat ja lauseet tarkoittavat, ja kääntämään niitä sanakirjan avulla äidinkielellemme. Kun kielitaito kehittyy, tietoinen kääntäminen muuttuu hiljalleen tarpeettomaksi, ja lopulta osaamme suoraan ajatella, puhua ja kirjoittaa tuolla uudella kielellä. Tässä vaiheessa mitään tiedostamatontakaan kääntämistä ei välttämättä enää tapahdu. Me vain ymmärrämme tuota uutta kieltä.

Jotain samankaltaista tapahtuu, kun opetelmme oikeastaan mitä tahansa uutta taitoa. Opimme tekemällä, yrittämällä ja erehtymällä, usein sääntöjen tai opastuksen avulla. Kun taito kehittyy, lopulta vain osaamme tehdä erilaisia asioita ilman ohjeiden noudattamista tai ajattelemista. Tätä ymmärtäminen on, ja älykkään käytöksen perusta on tällä tapaa sisäistetyissä taidoissa.

Äidinkielen oppiminen on kuitenkin erilaista kuin vieraan kielen opettelu oppikirjojen avulla. Jo pienet vauvat vuorovaikuttavat hoitajiensa kanssa monin tavoin. He kykenevät seuraamaan katsetta ja kiinnittämään muiden huomion erilaisiin ympäristön tapahtumiin. Pian lapset oppivat monimutkaisia vuorovaikutustaitoja ja tekemään asioita yh-

dessä muiden kanssa. He oppivat, että yhteisissä toimissa eri ihmisillä on omat roolinsa ja vuoronsa, joihin kuuluvat tietyt rutinoituneet toimintatavat. Sekä äidinkielen sanaston että kieliopin oppiminen perustuu näille yhteisille toimille, joissa lapset oppivat seuraamaan ja kommunikoimaan mitä tapahtuu ja miksi.

Kieli ei ole itsenäinen järjestelmä, vaan sen oppiminen perustuu sosiaalisesti jaetuille käytännöille. Kielellisen järkeilyn kehitys nähtävästi myös nojaa sosiaalisille käytännöille, joiden lomassa lapset oppivat kysymään ja kertomaan syitä erilaisille tapahtumille ja teoille.

Mikäli ihmisäly on tällaisen toiminnallisen oppimisen tulosta, on epäselvää, missä määrin koneiden älykkyys voi olla ihmisälyn kaltaista niin kauan, kun koneoppiminen perustuu pelkästään laajoihin teksti- ja kuvaaineistoihin ilman, että se osallistuisi niihin käytäntöihin, joiden tuloksena tällaiset aineistot syntyvät.

Kuitenkin jo nyt kehitetään järjestelmiä, jotka eivät käsittele pelkästään tekstiä vaan myös äänellistä ja visuaalista syötettä. Robottiikan kehittyessä meillä on ehkä pian koneita, jotka ottavat osaa päivittäisiin askareisiimme ja jotka näin ollen alkavat ymmärtää maailmaa yhä inhimillisemmällä tavalla.

Tekoälyjärjestelmät jo itse asiassa ottavat osaa päivittäiseen toimintaamme, kun niiden käyttö tietoverkoissa ja muualla arjessamme lisääntyy. Vuorovaikutus niiden kanssa ei

ChatGPT tuottaa kaavamaista tekstiä, mutta toisaalta sillä voi ajatella olevan älyllisiä hyveitä, jotka monilta puuttuvat.

myöskään ole yksipuolista: jo nyt ne myös muuttavat käyttäytymistämme auttamalla meitä mutta myös tuputtamalla erilaisia tuotteita, tietoa, valintoja ja mielipiteitä.

KIELIMALLI EI OLE PELKKÄ PAPUKAIJA

Mikäli älykkyyden, luovuuden ja ymmärtämisen käsitteet kytketään irti tietoisuuden käsitteestä, tarjoaa tämä uudenlaisen näkökulman kielimallien tarkasteluun. Se asettaa uuteen valoon ajatuksen, ettei kielimalli voisi aidosti ymmärtää tekstiä, että se on vain mekaaninen papukaija, joka matkii ja muuntelee kohtaamiaan ilmaisuja.

Aihetta voidaan lähestyä myös kääntämällä katse ihmiseen. On hylättävä epätieteellinen ajatus, jonka mukaan pieni tietoinen minä tarkkailee aivojen sopukoissa ajatus-temme ja havaintojemme virtaa. Sen sijaan, on otettava vakavasti, että ihmismielikin on materiaallinen järjestelmä. Tämän järjestelmän täytyy myös jotenkin kyetä aidosti ymmärtämään, järkeilemään ja luomaan uutta.

Lisäksi kannattaa kysyä, kuinka korkealle älykkyyden rima tulisi asettaa. Tyypilliset argumentit koneen älykkyyttä ja luovuutta

vastaan asettavat tuon riman todella korkealle. Kuvia tekstisyötteestä tuottava malli, kuten Dall-E, ei toki tuota **Caravaggion** maalausten veroisia taideteoksia. Chatbotin vastaukset ovat usein ”sujuvaa paskapuhetta”, johon piiloutuu olennaisia faktavirheitä. Toisaalta jos Dall-E:n tuottamia teoksia verrataan siihen, mitä tämän tekstin kirjoittajat kykenevät öljyvärien tai piirto-ohjelman avulla tuottamaan, ovat sen tuotokset hyvinkin vaikuttavia.

ChatGPT puolestaan tuottaa kaavamaista tekstiä, mutta toisaalta sillä voi ajatella olevan älyllisiä hyveitä, jotka monilta puuttuvat. Se on väsymätön lukija – onhan se lukenut lähes koko internetin – eikä sen tarkkaavaisuus herpaannu. Ja lopulta suuri osa ihmistenkin tuottamasta tekstisisällöstä on sujuvaa paskapuhetta. Tämä herättää kysymyksen, miksi asetamme kielimallin älykkyyden kriteeriksi poikkeuksellisen suoriutumisen tason, jota ihmisten tapauksessa edellytämme lähinnä alansa parhailta asiantuntijoilta tai suoranaيسilta neroilta.

Usein keskustelu älykkyydestä vaikuttaa sotkeutuvan kysymykseen toimijuudesta: voiko tekoäly olla toimija, jolla on vastuita ja

oikeuksia tai kenties vapaata tahtoa? Nämä ovat kuitenkin täysin eri kysymyksiä kuin kysymys älykkyydestä.

Ehkäpä ihmisten epäuskoiset reaktiot keinotekoiseen älykkyyteen johtuvat pitkälti siitä, että tiedämme tekoälyn olevan monimutkainen automaatti, mutta käsityksemme ihmisestä ja ihmismielestä sisältää oletuksen jostain syvemmästä, luovemmasta ja autonomisemmasta. Ihminen on kyllä itsenäisempi toimija kuin kielimalli, mutta monella tapaa olemme myös ympäristömme tuotetta.

Ymmärryksemme, luovuutemme ja älykkyytemme heijastavat paljolti ympäristömme toisteisuutta. Tätä on vain vaikea huomata, koska emme voi suoraan havaita, miten mieleemme toimii.

Renne Pesonen toimii apurahatutkijana Tampereen yliopiston filosofian oppiaineessa.

Samuli Reijula on akatemiaturkija (2020–25) ja Helsingin yliopiston teoreettisen filosofian yliopistonlehtori.

KIRJALLISUUS

Strandberg, T., Sivén, D., Hall, L., Johansson, P. ja Pärnamets, P. 2018. False Beliefs and Confabulation Can Lead to Lasting Changes in Political Attitudes. *Journal of Experimental Psychology: General* 147 (9), 1382–1399.
Turing, A. 1950. Computing Machinery and Intelligence. *Mind* 59 (236), 433–460.