

Kun OpenAI:n marraskuussa 2022 julkaisemaa ChatGPT-kielimallia pyytää kirjoittamaan kielitieteellisen julkaisun pääkirjoituksen, jossa pohditaan tekoälyn uhkia ja mahdollisuuksia tieteelliselle julkaisemiselle, kielimalli tuottaa tunnistettavan tekstilajinsa edustajan. Sen mukaan tekoäly tehostaa tutkimusta ja julkaisuprosesseja: tekoäly pystyy analysoimaan valtavia tietomääriä lyhyessä ajassa ja tuottamaan tiiviitä ja selkeitä artikkeleita, mikä voi johtaa tutkimuksen nopeampaan levittämiseen ja soveltamiseen, vaikka tähän liittyykin mallin mukaan eettisiä ja metodologisia ongelmia. Kielimalli ehdottaa tekoälyn mahdollistavan esimerkiksi vertaisarvioinnin nopeutumisen, mutta huomauttaa myös, ettei tekoälyn kyky tunnistaa virheitä ole riittävä ja että tekoäly saattaa vääristää arviointejaan ennakoasetettujen mallien perusteella. Pääkirjoituksessa pohditaan myös sitä, miten inhimillisen tekijän poistaminen voisi tehdä tuloksista objektiivisempia, koska tulokset eivät enää olisi alttiita ”inhimillisille vinoumille”. Kielimallin tekstissä hehketään myös monitieteellistä tutkimusta: tekoäly pystyy yhdistämään tietoa eri aloilta ja tuottamaan synteettisiä näkemyksiä, jotka ylittävät perinteiset tieteenalat – aitoa innovointia siis!

Mitä enemmän kielimallilta kysyy, sen tarkempia mahdollisuuksia ja uhkakuvia se tuo esiin. Malli ei kuitenkaan mainitse sen tulevaisuudessa yhä varmemmin mahdollistamaa tieteellistä monikielisyttä. Kun konekääntimet tulevaisuudessa yhä kehittyvät, myös eri kielillä kirjoitettujen tieteellisten tekstien saatavuus tulee parantumaan. Se, miten kääntäminen käytännössä toteutettaisiin (esimerkiksi se, miten konekäännöksiin tulisi ylipäätään suhtautua ja onko niiden tarkistaminen kirjoittajan itsensä ja tämän palkkaaman kääntäjän vastuulla) keskustelutti laajalti Virittäjän toimitusneuvoston maaliskuisessa kokouksessa. Toistaiseksi Virittäjän toimitus on päättänyt ennen kaikkea parantaa ammattilaiskääntäjänsä tarkistamien abstraktien kansainvälistä leviämistä.

Kielimallien valtava viimeaikainen kehitys on nähdäkseni varsin kiehtovaa myös toisesta näkökulmasta, lingvististen teorioiden kannalta. ChatGPT pystyy tuottamaan hätkähdyttävän inhimillisen oloista tekstiä, vaikka se tuottaa tekstiä todennäköisyyksien pohjalta valtavista kieliaineistoista. Tuotetut rakenteet nousevat kielenkäytöstä: jos jotain asiaa tai tietoa ei ole opetusaineistossa, GPT-kielimalli ei voi sitä tietää, eikä sillä muutenkaan ole keinoja erottaa tietoa tai epätietoa toisistaan; tuloksena syntyneet tekstit ovat matemaattisia laskutoimituksia. Tässä ChatGPT eroaa niin kutsutuista sääntöpohjaisista kielimalleista, jotka tuottavat kieltä formaalien sääntöjen pohjalta. ”Virheet” on jälkimmäisissä helpompi korjata koodauksin kuin tilastollisissa malleissa, joissa opetusmateriaalina voi olla esimerkiksi suuri osa internetin teksteistä. Sääntöpohjaiset mallit ovat kuitenkin kehittäjiensäkin mukaan parhaimmillaan tuottaneet vain ”puisevaa”, faktuaalista tekstiä, ja ”inhimilliseen” tekstuaalisuuteen on päästy vasta koneen laskemien matemaattisten todennäköisyyksien myötä. Miksi nimenomaan tilastollinen kielimalli tuottaa niin hämmästyttävän inhimillisen oloista tekstiä? Miten sen luomaa kieltä tulisi analysoida esimerkiksi suhteessa käyttöpohjaisuuteen ja emergenssiin?

Matemaattisilla kielimalleilla on käytössään valtava määrä materiaalia, josta algoritmi tunnistaa säännönmukaisuuksia, ja näitä yhdistelemällä se voi tuottaa esimerkiksi keskustelun toimintajäsennyksen mukaisia vuoroja ja ”korjata” itseään. Merkityksiä sillä ei kuitenkaan ole, eikä ymmärrystä sosiaalisesta toiminnasta; ”ymmärrys” näistä perustuu käyttöliittymään tehtyihin koodauksiin. Esimerkiksi ChatGPT on opetettu performoimaan itseään keskustelun keinoin ja ottamaan itsensä deiktisten viittausten origoksi; samalla malli on opetettu toteamaan, ettei ole oikeasti olemassa ja käyttävänsä deiktisiä viittauksia vain, jotta ihminen osaisi hänen kanssaan luonnollisemmin keskustella. Mitään ”itseä” mallilla ei tietenkään ole, mutta skemaattisesti sillä on käytössään koko ajan enemmän inhimillisiä resursseja.

Sen lisäksi, että kielimallit tulevat vaikuttamaan tieteelliseen julkaisemiseen tavalla, jota emme vielä osaa hahmottaa, ne voivatkin toivon mukaan toimia lingvisteille myös inhimillisen vuorovaikutuksen ja merkityksentuoton teoreettisena peilinä – ehkä myös mahdollistaa tutkimussuuntausten ja -menetelmien rajat ylittävää keskustelua.

LAURA VISAPÄÄ

Kotikielen Seuran jäsenretki Järvenpään E. N. Setälän jäljille

Lauantaina 27.4.2024 klo 11.00–14.30

Retken ohjelmassa tutustuminen Aholaan sekä STEP-opiston (entisen Seurakuntaopiston) tiloihin, jonka paikalla oli aikanaan E. N. Setälän Toimelan tilan päärakennus (kuvassa). Setälästä retkeläisille kertoilee professori Riho Grünthal. Retki on tarkoitettu Kotikielen Seuran jäsenille. Lisätietoa tulee jäsenkirjeessä.



Kuva: Valokuvaamo Pietisen kokoelma, Museovirasto.