Suureen kielimallimullistukseen on herättävä

Erkki Mervaala

Suuret kielimallit eli LLM:t valtaavat alaa myös tieteen tekijöiden keskuudessa. Kun heinäkuussa 2024 uutisoitiin, että tieteellisen tutkimuksen kustantajajätti Taylor & Francis on myynyt omistamansa Routledge-kustantamon julkaisemat tutkimukset Microsoftille tekoälyn koulutukseen, moni tutkija maailmalla älähti. Myös Wiley on ilmoittanut tekoäly-yhteistyökuvioistaan. Yhdistävänä tekijänä on se, että datan omistaja on luovuttanut kirjoitukset koulutuskäyttöön kirjoittajia varoittamatta saati heiltä lupaa kysymättä. (Palmer, 2024.) Esimerkiksi koodausfoorumi Stack Overflow (2024) ilmoitti myyneensä oikeudet sisältöihinsä ChatGPT:n takana olevalle OpenAI:lle, mistä moni foorumille ratkaisujaan kirjannut koodari närkästyi. Suomessa kielimalleja on koulutettu muun muassa Suomi24- ja Reddit-keskustelusivustojen aineistoilla (Luukkonen ym., 2024.)

Vastaavanlaiset datain valtaukset ovat yleistyneet kielimallivillityksen vanavedessä, mutta ne edustavat laajempaa jo monennetta vuosikymmentä jatkuvaa datarohmuamista, jonka lähtöoletuksena on, että ”vielä yksi data” johtaa auvoiseen onneen ja ongelmain ratkaisemiseen. Jos plagiointi-, tekijänoikeus- ja identiteettinäkökulmat unohdetaan, tieteellisen tiedon kaataminen valtavaksi tiedevirraksi kuulostaa jopa kiehtovalta. Tieteelliseen tutkimukseen perustuen ja siitä ammentaen uutta toistettavaa tutkimusta tuottava koneapuri on kuitenkin toistaiseksi totta vain mainospuheissa.

Keväällä saimme lukea, kuinka moni julkaistu tutkimusartikkeli oli jouduttu poistamaan LLM-generoidun valetutkimuksen vuoksi. Jopa korkealle arvostetuista julkaisuista löytyi vertaisarvioituja artikkeleita, joiden sisältä paljastui ChatGPT:sta tuttu ”Certainly!”-aloitus. Scientific American -lehden mukaan kyseessä on vain jäävuoren huippu, sillä tekoälychattibotit ovat tunkeutuneet syvälle tieteellisen julkaisemisen maailmaan. (Stokel-Walker, 2024.)

Tutkimuksessa tekoälyä käyttävien ja siitä ”kiinni jääneiden” suhteesta ei voitane sanoa mitään varmaa, mutta voisi kuvitella jälkimmäisten olevan murto-osa edellisestä. Mitä generatiivisen tekoälyn tuominen tieteelliseen tutkimukseen tekee työllemme?

Osaamistyhjiötä rakentamassa

Yksi kliseisimmistä, ja sijoittajien ja yritysjohtajien näkökulmasta odotetuimmista, tulevaisuudenkuvista liittyy ihmistyön korvaamiseen tekoälyllä – ainakin osittain. Valtaosa suurten kielimallienkin ympärille rakentuneesta triljoonan dollarin sijoituspöhinästä liittyy tähän turbotehostamisen toiveeseen. Yritysjohtajain odotukset voivat kuitenkin olla ylimitoitettuja: tähtitaloustieteilijä Daron Acemoğlu ennustaa generatiivisen tekoälyn tehostavan tuottavuutta vain 0,9 prosenttia tulevan vuosikymmenen aikana. Tuottojen epävarmuus ei kuitenkaan toistaiseksi näy alaan sijoittamisessa. (Goldman Sachs, 2024; Nicoud, 2024.)

Generatiivinen tekoäly on jo vaikuttanut useisiin, hyvin erilaisiin aloihin, jopa perustavanlaatuisesti. Esimerkiksi käsikirjoittajat ja näyttelijät ovat ärähtäneet luovan työn ylenkatsomisesta, jota esiintyy Hollywood-studioiden lisäksi myös tekoälyjättien johdossa. (Anguiano & Beckett, 2023; Tangalakis-Lippert, 2024.)

Ohjelmointiala on toinen, missä tekoälyn vaikutuksia on jo ehditty havaita. ChatGPT, Copilot ja vastaavat ovat jo pystyneet tekemään junioritason koodaajan työtehtäviä. Yleisesti ottaen LLM-ihmiskorvikkeet ovat pystyneet korvaamaan juuri uransa alkuvaiheessa olevien työntekijöiden työt.

Ajatellaanpa tämän kehityksen vaikutuksia: Jo nyt on havaittu, että ohjelmointiyrityksissä ollaan huolissaan kokeneempien koodaajien tulevasta tekijäkadosta. Seniorikoodaajia kun ei synny, jos juniorit korvaa tekoäly. Muun muassa Googlella ja Amazonilla 30-vuotisen uransa aikana kehittäjänä toimineen Steve Yeggen mukaan seniorikoodaajien eläköityessä yritykset voivat päätyä niin kutsuttuun COBOL-tilanteeseen, jossa juniorikoodaajia ei vain enää ole (Yegge, 2024).

Akateeminen maailma avaa pikkuhiljaa ovensa LLM:lle myös tutkimuksen ulkopuolella. Atlantan Morehousesta on syksyllä 2024 tulossa ensimmäinen yliopisto, joka

käyttää virtuaalisia tekoälyavustajia opetuksessa ja akateemisen työn tukena (Morehouse College, 2024). Yliopiston professorin mukaan jokaisella professorilla tulee jatkossa olemaan 3–5 tekoälyavustajaa. Professorit tuottavat 3D-ympäristöön virtuaaliluentoja, joita 24/7-saatavilla olevat virtuaaliavustajat sitten opiskelijoille opettavat. Aika näyttää, miten paljon tekoälyn tuottamia virheitä, vääriä viitteitä ja stokastista mis- ja disinformaatiota opiskelijat saavat kokea osana opetustaan. Perinteisesti assistentin työ on ollut varsin tärkeäkin vaihe aloittelevan akateemikon uralla.

Ja vaikka koko työpaikka ei tulisi tekoälyn korvaamaksi, suuri osa niin sanotuista ”akateemisista paskaduuneista” sillä saatetaan hyvinkin korvata. Kun aiemmin korkeakouluharjoittelijat ovat saaneet työtehtävikseen esimerkiksi tutkimushaastattelujen litterointia, jossa toimessa saattaisi sujua vaikkapa koko kesätyörupeama, nyt tekoälylitterointityökalut voivat tehdä viikkojen homman päivässä. Toimittajanakin työskennelleenä tiedän tasan tuskan, joka tunti tunnilta turruttavammaksi käyvät haastattelulitteroinnit tuottavat.

Tekoälyturbulenssi on myös johtamisen kysymys. The Upwork Research Instituten tekemän tutkimuksen (Monahan & Burlacu, 2024) mukaan pomot kyllä odottavat tekoälyn tekevän työntekijöistään tuottavampia, mutta liki puolet työntekijöistä ei tiedä, miten tekoälyn odotetaan kasvattavan heidän tuottavuuttaan, ja jopa 77 prosenttia työntekijöistä koki tekoälyn heikentävän tuottavuuttaan. Tehokkuutta ja tuottavuutta työhönsä tahtovan on tiedettävä, mitä tekee.

Edellä kuvattujen esimerkkien kautta lienee ihan hyvä pohtia, missä ja miten tulevaisuuden tekijät kokemuksensa kartuttavat.

Läpinäkymättömän luotettava

Toisin kuin tutkimusten tulokset taikka vaikka Wikipedia eivät LLM:t eikä varsinkaan niiden kautta hankittu ”tieto” pysy muuttumattomina. Generatiivisten mallien DNA:han on kovakoodattuna tuotosten muuntuvaisuus, mistä johtuen myöskään hallusinoinnista eli tekoälyn omiaan satuilemisesta ei päästä 100 % varmuudella eroon. Kun LLM viittaa tutkimuksiin, voivat tutkimukset näyttää oikeilta, vaikka ne ovatkin täysin tekoälyn tekaisemia. (Alkaissi & McFarlane, 2023.)

Vaikka jo avoimesti jaossa olevien kielimallien konepellin alle katsominen on liki mahdotonta jopa mallit rakentaneille ihmistutkijoille, kaupalliset sovellukset ovat täydellisiä mustia laatikkoja – eritoten käyttäjän näkökulmasta. Samaa tulosta ei voida taata, vaikka kehote pysyisi samana. (Ollion ym., 2023; Reiss, 2023.)

Hiljattain havaittu trendi käyttää LLM-chattibotteja tiedonlähteinä perinteisen googlettamisen taikka Wikipediasta tarkistamisen sijaan kuulostaa tästä näkökulmasta huolestuttavalta. Tekoälyn generoiman verkkosisällön suhteellisen määrän kasvaessa alituiseen luotettavan tiedon äärelle löytäminen muuttuu hetki hetkeltä haastavammaksi. (Donath & Schneier, 2024.)

Google ei ole sekään millään tavalla syytön tähän trendiin. Vaikka se luopui käyttäjän kannalta parhaiden hakutulosten näyttämisestä mainostajiensa eduksi aikaa sitten, vasta tekoälysisältöjen työntäminen oletusarvoisesti joka haun yhteydessä teki siitä hakukoneena monelle käyttökelvottoman. Myös Googlen omat tutkijat ovat todenneet, että generatiivinen tekoäly voi vääristää sekä yhteiskuntapoliittisen todellisuuden että tieteellisen konsensuksen kollektiivista ymmärrystä. (Marchal ym., 2024.)

Olen hiljattain tutkinut LLM-avusteista tekstianalyysia toistettavuuden näkökulmasta, mikä on yleisesti ottaen osoittautunut erittäin ongelmalliseksi (Mervaala & Kousa, 2024; Ollion ym., 2023). Lisäksi suuret kielimallit ovat muun muassa vahvistaneet misinfirmaatiota, stereotyyppejä ja salaliittoteorioita (Khatun & Brown, 2023; Makhortykh ym., 2024) ja kehottaneet ihmisiä syömään kiviä ja laittamaan pizzaan liimaa (Hart, 2024). Positiivisena käyttötapauksena mainittakoon, että yhdessä tutkimusartikkelissa todettiin tekoälykkäiden keskustelujen voivan vähentää alttiutta uskoa salaliittoteorioihin (Costello ym., 2024).

Toistettavuuden ongelma koskee tietysti myös ihmisenkin tekemää tutkimusta, ja kyllähän tekoäly tekee tietyt asiat huomattavasti ihmistä nopeammin. Generatiivisen tekoälyn perustavia piirteitä on kuitenkin se, ettei sen ole mahdollista toistaa itseään. Ihmistä voi pyytää olemaan läpinäkyvä tutkimuksensa kulusta, mutta LLM:n sekunneissa generoima kirjallisuuskatsaus ei välttämättä ole millään tavalla kurantti, vaikka se aidolta näyttäisikin.

Edellä mainitsemistani seikoista huolimatta en anna kuitenkaan missään nimessä generatiivisten tekoälymallien avustamalle tutkimukselle täystyrmäystä. Itse olen pyytänyt sitä monesti luomaan kätevän pätkän koodia taikka jäsentelemään jotain pientä osaa tekstistäni. En ole intoutunut kuitenkaan luomaan definitiivisiä tiivistelmiä tutkimusartikkeleista, sillä siihen se ei ole luotettava työkalu. Mutta jos

tietää, mitä mallilla on mahdollista tehdä, se voi kuitenkin olla juuri sitä – omaa työtä tehostava työkalu. Tutkimukseen liittyvien analyysien näkökulmasta kynnys on mielestäni paljon suurempi, mutta mikäli tietää, mitä tekee ja selittää sen tutkimusartikkelissa avoimen läpinäkyvästi – päivämäärineen, versionumeroineen ja kehotteineen päivineen – niin en koe senkään olevan täysin poissuljettua.

Tekoälyn planetaariset rajat

Lopuksi on vielä nostettava esiin ajan mittaan kenties tärkein generatiivisten tekoälysovellusten käyttöön liittyvä seikka: niiden valtava hiilijalanjälki. Niin mallien kouluttaminen kuin niiden ylläpitäminen ja käyttäminen ovat niin kutsuttuun perinteiseen internetin käyttämiseen verrattuna energiankulutukseltaan varsin suurta. Generatiivisen tekoälyn omaksumisen jälkeen Googlen hiilidoksidipäästöt ovat kohonneet 48 prosenttia vuodesta 2019 ja Microsoftin 31 prosenttia vuodesta 2020 (Nguyen, 2024). Niin ikään näytönohjaimiin ja muihin malleille kelpaaviin teknologiatuotteisiin tulee tällä kehitysellä kulumaan yhä enemmän luonnonvaroja.

Tilanteessa, jossa nykyisenkään kaltainen kulutusyhteiskuntamme ei ole kestävällä pohjalla, kuulostaa edesvastuuttomalta puskea tuhottomasti tehoa vieviä mutta vain vaatimattomasti sitä tuovia työkaluja osaksi sitä tietoyhteiskuntaa, jona olemme tottuneet Suomea pitämään. Tokihan myös ihmistyö on hiili-intensiivistä. On kuitenkin hyvä arvioida, mihin ja milloin moisia suurikuormaisia palveluita käyttää ja milloin riittää vain Excelin käyttö, vanhanaikainen googlaaminen taikka ihan vain ajattelu – ilman tekoälyhöysteitä. ■

Puheenvuoro perustuu osin Puistokatu 4:ssä 14.3.2024 pidettyyn alustukseen tekoälyn tutkimuskäytöstä.

Kiitokset

Kiitokset Puistokatu 4:ssä käydyille keskusteluille sekä Strategisen tutkimuksen neuvoston rahoittamalle ORSI-hankkeelle (327768, www.ecowelfare.fi).

Lähteet

Alkaissi, H. & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: implications inscientific writing. *Cureus*, *15*(2), e35179. https://doi.org/10.7759/cureus.35179

Anguiano, D. & Beckett, L. (1.10.2023). How Hollywood writers triumphed over AI – and why it matters. *The Guardian*. https://www.theguardian.com/culture/2023/oct/01/hollywood-writers-strike-artificial-intelligence

Costello, T. H., Pennycook, G. & Rand, D. G. (2024). *Durably reducing conspiracy beliefs through dialogues with AI.* https://doi.org/10.31234/osf.io/xcwdn

Donath, J. & Schneier, B. (22.4.2024). It's the end of the web as we know it. *The Atlantic*. https://www.theatlantic.com/technology/archive/2024/04/generative-ai-search-llmo/678154/

Goldman Sachs. (25.6.2024). Gen AI: Too much spend, too little benefit. *Goldman Sachs Global Macro Review*, 129. https://www.goldmansachs.com/insights/top-of-mind/gen-ai-too-much-spend-too-little-benefit

Hart, R. (31.5.2024). Google restricts AI search tool after 'nonsensical' answers told people to eat rocks and put glue on pizza. *Forbes*. https://www.forbes.com/sites/roberthart/2024/05/31/google-restricts-ai-search-tool-after-nonsensical-answers-told-people-to-eat-rocks-and-put-glue-on-pizza/

Khatun, A. & Brown, D. (2023). Reliability check: n analysis of gpt-3's response to sensitive topics and prompt wording. Teoksessa *Proceedings of the 3rd workshop on trustworthy natural language processing (TrustNLP 2023)* (s. 73–95). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.trustnlp-1.8

Luukkonen, R., Burdge, J., Zosa, E., Aarne Talman, Komulainen, V., Hatanpää, V., Sarlin, P. & Pyysalo, S. (2024). Poro 34B and the blessing of multilinguality. *arXiv*. https://arxiv.org/pdf/2404.01856

Makhortykh, M., Sydorova, M., Baghumyan, A., Vziatysheva, V. & Kuznetsova, E. (2024). Stochastic lies: how LLM-powered chatbots deal with Russian disinformation about the war in Ukraine. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-154

Marchal, N., Xu, R., Elasmar, R., Gabriel, I., Goldberg, B. & Isaac, W. (2024). Generative AI misuse: A taxonomy of tactics and insights from real-world data (Version 2). *arXiv*. https://doi.org/10.48550/ARXIV.2406.13843

Mervaala, E. & Kousa, I. (2024). Order up! Micromanaging inconsistencies in ChatGPT-4o text analyses. Teoksessa M. Hämäläinen, E. Öhman, S. Miyagawa, K. Alnajjar & Y. Bizzoni (toim.), P*roceedings of the 4th*

International Conference on Natural Language Processing for Digital Humanities (s. 521–535). Association for Computational Linguistics. https://aclanthology.org/2024.nlp4dh-1.51

Monahan, K. & Burclau, G. (23.7.2024). *From burnout to balance: AI-enhanced work models.* Upwork. https://www.upwork.com/research/ai-enhanced-work-models

Morehouse College. (10.7.2024). *Morehouse to use AI teaching assistants this fall.* https://news.morehouse.edu/morehouse-college-to-use-ai-teaching-assistants-this-fall

Nguyen, B. (24.7.2024). *AI is making Google and Microsoft big contributors to climate change.* Quartz. https://qz.com/ai-google-microsoft-climate-change-data-center-energy-1851589453

Nicoud, A. (30.7.2024). *Will generative AI live up to its hype?* Ibm. Com. https://www.ibm.com/blog/gen-ai-live-up-to-hype/

Ollion, E., Shen, R., Macanovic, A. & Chatelain, A. (2023). ChatGPT for text annotation? Mind the hype! *SocArXiv*. https://doi.org/10.31235/osf.io/x58kn

Palmer, K. (29.7.2024). Taylor & Francis AI deal sets 'worrying precedent' for academic publishing. *Inside Higher Ed.* https://www.insidehighered.com/news/faculty-issues/research/2024/07/29/taylor-francis-ai-deal-sets-worrying-precedent

Reiss, M. V. (2023). Testing the reliability of ChatGPT for text annotation and classification: A cautionary remark. *arXiv*:2304.11085. arXiv. https://doi.org/10.48550/arXiv.2304.11085

Stack Overflow. (6.5.2024). *Stack Overflow and OpenAI partner to Strengthen the world's most popular large language models.* https://stackoverflow.co/company/press/archive/openai-partnership

Stokel-Walker, C. (2024). Chatbot Invasion. *Scientific American*, *331*(1), 16. https://doi.org/10.1038/scientificamerican072024-7bVpuVjnKZbJ8IQrkgMTlR

Tangalakis-Lippert, K. (26.6.2024). OpenAI's CTO treats creativity like a problem to be solved—And that itself is the problem. *Business Insider*. https://www.businessinsider.com/openai-cto-mira-murati-creative-jobs-eliminated-ai-2024-6

Yegge, S. (24.6.2024). The death of the junior developer. *Sourcegraph.Org.* https://sourcegraph.com/blog/the-death-of-the-junior-developer