Merkkijonot, merkitykset ja tilastolaskenta: suurten kielimallien ja luonnollisen kielen erikoinen liitto

Marko Latvanen
Digi- ja väestötietovirasto

DIGI- JA VÄESTÖTIETO-VIRASTO

PLAYING THE IMITATION GAME

Luonnollista kieltä koneella jo 75 vuotta

- 1950: **Alan Turing** julkaisee artikkelin ”Computing Machinery and Intelligence”, jossa hän ehdottaa koneen älykkyyden kriteeriksi **tietokoneohjelman kykyä esittää ihmistä reaaliaikaisessa kirjallisessa keskustelussa** niin hyvin, että ihminen ei pysty luotettavasti erottamaan - pelkästään keskustelun sisällön perusteella - tietokoneohjelmaa ja oikeaa ihmistä toisistaan.
- 1954: ”Georgetownin kokeessa” käännetään täysin automaattisesti yli kuusikymmentä venäjänkielistä lausetta englanniksi.
- 1966: ALPAC-raportissa todetaan, että kymmenen vuotta kestänyt tutkimus ei ollut täyttänyt odotuksia → ”tekoälyn ensimmäinen talvi” liittyy tähän
- Konekääntämistä tutkitaan vain vähän ennen 1980-luvun loppua, jolloin kehitetään ensimmäiset **tilastolliset** konekäännösjärjestelmät. Tilastollisten lähestymistapojen kehittymistä edistävät **laskentatehon lisääntyminen** ja **suurten tietokokonaisuuksien saatavuus → uusia menetelmiä → ChatGPT 2022 → suurten kielimallien räjähdys**

Turing: tavoite (melkein) saavutettu

- Alan Turing asetti koneen **älykkyyden** kriteeriksi, että sillä on ”*the ability … to impersonate a human in a real-time written conversation*”.
- Nyt tähän on toiminnallisesti päästy, ja vieläpä useilla kielillä. Lisäksi ne eivät rajoitu kirjoitettuun keskusteluun, vaan niiden kanssa voi keskustella puhuen – ainakin melko sujuvasti ja tämän hetken suuria kieliä käyttäen.
- Tavoitteliko Turing varsinaista **älykästä** tietokonetta tai tietokoneohjelmaa? On ohjelmia ja sovelluksia, jotka **imitoivat inhimillisen älykkyyden kielillisiä ilmenemismuotoja**, mutta ne eivät ole millään tavalla älykkäitä ihmisen tavalla.
- Tarkoittiko Turing sen sijaan “koneälyä”, tietoisena erotuksena ihmisen älykkyydestä?

Rebooting the problem

- Turing kiersi, tai käynnisti uudelleen / rebuuttasi, koko ongelman tavalla, jonka vaikutukset tunnemme edelleen:
  - **“Since the words "think" and "machine" *cannot be clearly defined*, we should replace the question (‘can machines think?’) by another, which is closely related to it and is expressed in relatively unambiguous words.“**
- Käytännössä Turing pyrki siis nollaamaan keskeisimmät käsitteet, korvaamaan ne uusilla ja luomaan uuden paradigman, jossa kysymys koneiden älykkyyden suhde ihmisen älykkyyteen olisi tietyssä mielessä yhdentekevä.

”Emmehän me ymmärrä koirankaan aivoja”

- Turingin tekemä keskeisten käsitteiden nollaus suhteessa koneälyyn, ajatteluun ja näiden kautta kieleen, vaikuttaa edelleen tekoälykehityksessä.
- Emme useimmiten pysty selvittämään ainakaan täsmällisesti, miten jokin AI-järjestelmä tuottaa lauseen, ennusteen, yhteenvedon tai kuvan, jos kone toimii → tekoälyn musta laatikko –dilemma.
- Kun olen nostanut tästä hallinta-, luotettavuus- ja läpinäkyvyyshuolia, olen saanut mm. seuraavan vastauksen kehittäjäpuolelta: ”emmehän me ymmärrä kunnolla koirien aivojen toimintaakaan mutta hyvin me niiden kanssa toimimme.”
- Uskallan sanoa, että Turing nerokkuudestaan huolimatta teki yhden pitkäkantoisen virheen siinä, että hänen käsitekiepautuksensa avasi oven keskeisten dilemmojen ohittamiselle ja vähättelylle.

”STOKASTISIA PAPUKAIJOJA”: Luonnollisen kielen imitaatio varastettuna datana ja koodeina vailla merkityksiä

”Stokastinen papukaija” -termiä käytettiin ensimmäisen kerran artikkelissa **”On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?”** *(Bender, Gebru, McMillan-Major, 2021).*

Tekijät huolestuivat, että suuret kielimallit aiheuttavat riskejä: ympäristöhaitat ja rahoituskustannukset; läpinäkymättömyydestä johtuvat, hallitsemattomat vinoumat; kyky tuottaa harhaanjohtavaa aineistoa, ja se, että kielimallit **eivät pysty ymmärtämään käsitteiden merkityksiä, jotka ovat niiden ”oppimien” aineistojen taustalla ja perustana.**

Ei merkityksiä vaan todennäköisyyksiä

- Tekoälysovellukset ovat tilastolaskimia. Aineistot muutetaan neutraaleiksi koodeiksi (token) matemaattista käsittelyä varten.

cat = 34567
did = 12345
mat = 67890
on = 56789
sit = 45678
the = 23456

"Did the cat sit on the mat?“ → [12345] [23456] [34567] [45678] [56789] [23456] [67890]

esimerkki: https://www.forbes.com/sites/lanceeliot/2024/02/25/exposing-the-brittleness-of-generative-ai-as-exemplified-by-the-recent-gibberish-meltdown-of-chatgpt/

- Tekoälyjärjestelmät tuottavat ”mielestään” (koneella ei ole mieltä) parhaan tilastollisen arvion käsittelemästään asiasta
  - ”tässä kuvassa on kissa 87 % todennäköisyydellä”; ”lauseen ’kissat syövät mielellään ___’ seuraava sana on ’kalaa’ 63 % todennäköisyydellä”: todennäköisyydet lasketaan tokeneiden ”vektorietäisyyksistä” toisistaan siinä aineistossa, jolla malli on opetettu
- Tekoälysovellukset tuottavat siis ”tilastollisesti hyviä arvauksia”.
- **GPT-mallit ovat vain ”auto-correct on steroids”** (Linus Torvalds)

Papukaija, vai suorastaan bullshittiä?

“Harry Frankfurt, among others, analyzed the concept of **bullshit** as related to, but distinct from, lying.

The liar tells untruth, but the bullshitter aims to convey a **certain impression of themselves without being concerned about whether anything at all is true; it may be, or not”.**

Kaksi eettistä pähkinää

1. OpenAI:n, Googlen, Metan etc suuret kielimallit on koulutettu käytännössä kaikella, mitä julkisessa internetissä on, piittaamatta mistään oikeuksista. Toiminta on itsessään ollut epäeettistä ja laitonta.

   Miksi olemme hyväksyneet tämän olkia kohauttamalla (senkin jälkeen kun tästä tuli yleistä tietoa)? Mukavuus, helppous, uutuus, kiire, laiskuus…?

2. Antropomorfismi

”Sano hei uudelle työkaverillesi”

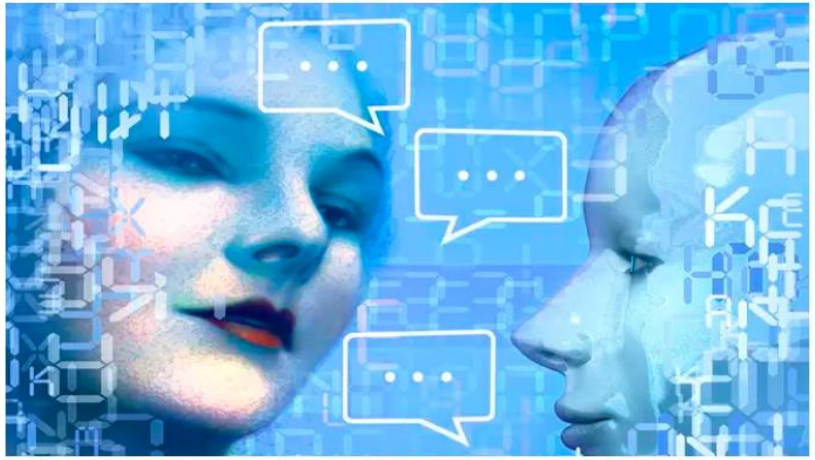
- Se, että generatiivisen AI:n NLP-ohjelmistot kykenevät tilastolaskemaan todennäköisyyksiä lauseessa olevaa sanaa seuraavalle sanalle, ei ole itsessään mikään ongelma.
- Tieto/asiantuntijatyössä genAI-työkalut voivat olla ihan näppäriä, kunhan a) ymmärtää niiden rajoitteet ja perusteet, ja b) asiantuntija itse on sen verran ekspertti, että osaa havaita koneen mokat ja korjata ne.
- Heikolle jäälle mennään, kun alamme (työ)kulttuureina puhua tietojenkäsittelyohjelmistoista käsitteillä, jotka eivät niihin kuuluisi
- Copilot, ChatGPT-4o tai Claude ei ole kaverisi, työkaverisi, assistenttisi, sihteerisi, konsulttisi eikä sielunkumppanisi.
- Se on ohjelmisto. Se ei ymmärrä käsitteiden merkityksiä, vaikka sillä on **kyky esittää ihmistä reaaliaikaisessa kirjallisessa keskustelussa.**

Vain muutama vuosi sitten…

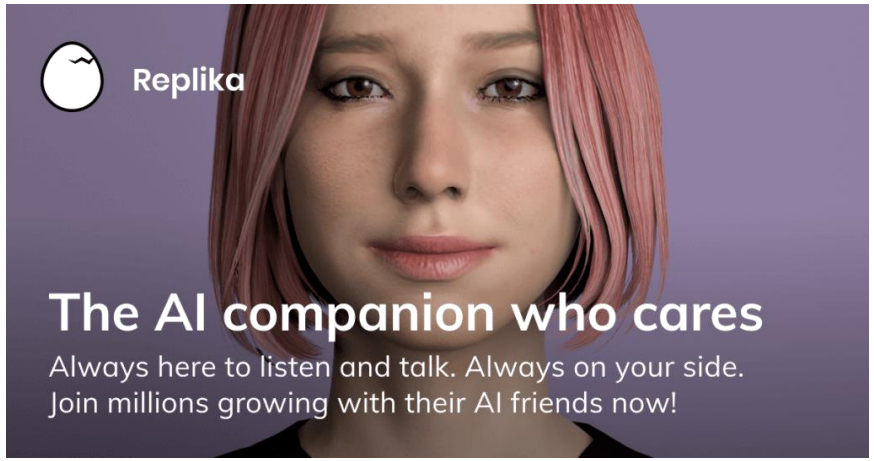
- Vielä ennen pandemiaa niin yritysmaailma kuin julkinen sektori painottivat, että
  - asiakaspalvelubotti ei koskaan saa olla sellainen, että ihminen luulee sitä toiseksi ihmiseksi; sillä ei saa olla liian ihmismäisiä ominaisuuksia
  - tekoäly ei koskaan tule astumaan luovan työn ja ihmissuhteiden alueelle.

Tekoälylle kannattaa olla kohtelias

Tekoälyt | Kohtelias kieli saa tekoälyn hakemaan tietoa paikoista, joissa on tarkkaa tietoa ja dataa.



Äksyily vie tekoälyn harhapoluille. Kuva: Kimmo Taskinen / HS



In Hollywood writers' battle against AI, humans win (for now)

1 of 4 | Actor and SAG-AFTRA negotiator Frances Fisher, middle, raises her sign on a picket line outside Netflix studios on Tuesday, Sep. 26, 2023, in Los Angeles. (AP Photo/Damian Dovarganes)

Luonnollinen kieli, luonnollinen mieli?

- Mitä pitemmälle generatiivisen AI:n sovellukset kehittyvät ihmismäiseen suuntaan, sitä vaikeampi meidän on suhtautua niihin algoritmeja sisältävinä tilastolaskimina, tai edes työvälineinä.
- Homo sapiens ei muutenkaan tee jatkuvasti objektiivista analyysiä ympäristöstään, vaan toimii kognitiivisten oikopolkujen avulla, mahdollisimman paljon energiaa säästäen.
- Kuuluu ihmislajin valuvikoihin, että alamme pitää koneita inhimillisinä. Ei siksi, että ne generoivat näppärästi kuvia, puhetta ja musiikkia, vaan siksi, että ne käyttävät **käsitteellistä kieltä**, jota me ihmiset olemme tottuneet pitämään vain meidän ominaisuutenamme.

Informaatio, merkitys ja tieto

- Generatiivisen AI:n sovellukset kykenevät kokoamaan ja toisintamaan kertaalleen tuotettua aineistoa; voidaan sanoa, että ne tuottavat informaatiota (mutta eivät luo sitä).
- Merkityksistä ja käsitteiden kontekstuaalisuudesta koneella ei ole mitään todellista ymmärrystä, koska kone ei ajattele. Turingin käsitekiepautus on kuitenkin mahdollistanut sen, että helposti ohitamme tämän olennaisen faktan.
- En pidä suurena riskinä sitä, että käytämme joitain generatiivisen AI:n sovelluksia tietotyön tehostamiseen ja nopeuttamiseen; itse käytän toisinaan konekääntäjiä kielissä, joita itse osaan riittävästi tuotoksen tarkistamiseen.
- Inhimillinen riski syntyy, jos erehdymme pitämään koneen tuotosta **tietona.**
- Riskin realisoitumisesta esimerkkinä on yleistyvä tapa käyttää ChatGPT:n kaltaista sovellusta hakukoneena → yksi black box –vastaus sen sijaan että kävisimme läpi hakutuloksia ja niiden lähteitä; tästä mm. toimittaja Johanna Vehkoon ilmaisema huoli 6.9. Kirkon viestintäpäivillä.

Mediakriittisyydestä AI-kriittisyyteen

- Perinteinen mediakritiikki ja –lukutaito perustuvat siihen, että voimme tarkastella lähteitä ja arvioida niiden luotettavuutta.
- Generatiivisen AI:n sovellukset eivät anna tähän mahdollisuutta. Lisäksi emme voi olla varmoja, missä määrin järjestelmät ”ymmärtävät” käsitteiden merkityksiä ja **niiden kontekseja ja keskinäisiä suhteita.**
- Tästä syystä on toivottava, että kaikilla koulutustasoilla käsiteltäisiin AI-järjestelmien luonnetta ja tekoälyn periaatteita; perusasiat eivät ole rakettitiedettä eivätkä edes matematiikkaa, vaan varsin yleistajuisesti välitettäviä asioita.
- Jos emme tee näin, otamme todellisen riskin, että suurta osaa väestöstä voidaan viedä kuin sitä kuuluisaa litran mittaa.

Tekoäly on tietokoneohjelmia, tekoäly on tutkimusala, tekoäly on haave ja painajainen, tekoäly on markkinointitermi. Tekoäly on metafora ja tarina.

"Mitä tekoäly on?" on professori Hannu Toivosen tietokirja tekoälystä. Hannu Toivonen on myös yleistajuinen puhuja sekä tekoälytaiteilija.

Hannu Toivosen ***Mitä tekoäly on? 100 kysymystä ja vastausta*** (Teos 2023) on helppotajuinen tietokirja, jonka voi lukea missä järjestyksessä haluaa!

Osta kirja | Lue lisää

Hannu Toivonen on tietojenkäsittelytieteen professori ja tekoälytutkija. Hän on myös pidetty puhuja ja osaa kertoa yleistajuisesti, mitä tekoäly on – ja mitä se ei ole. (Kuva: Teos Oy)

Lue lisää

Hannu Toivonen on myös tekoälytaiteilija. Hän sekä tutkii että soveltaa luovaa tekoälyä. Esimerkiksi tämä syntymäpäiväkuva 1970-luvulta on tehty kokonaan tekoälyn avulla.

Katso lisää

Kiitos!

marko.latvanen@dvv.fi