
Tekoälyn kyvyt ja roolit tekstin- tuotannossa:

**Käyttätymistieteellinen tarkastelu
kirjoittajan ja vastaanottajan näkökulmasta**

Janne Kauttonen ja Jyrki Suomala

ProCom – Viestinnän ammattilaiset ry
ISBN 978-952-65488-4-5

Generatiivinen tekoäly on jo laajassa käytössä tekstin tuottamisessa, mutta miten tämä vaikuttaa tekstin kirjoittamisprosessiin, tekstin ominaisuuksiin ja lukijan reaktioihin? Pirstaleisessa informaatioympäristössä ihmisiä puhutteleva viestintä on vaikeaa. Jotta ymmärrämme paremmin generoivan tekoälyn vaikutukset viestintäympäristöön, on tärkeää ymmärtää, millaiset tekijät ohjaavat ihmisen informaation prosessointia ja mikä tekee tietyistä sisällöistä vaikuttavampia kuin toisista. Tarkastelemme tekoälyavusteista sisällöntuotantoa käyttätymistieteen eri näkökulmista. Generoiva tekoäly auttaa viestinnän ammattilaista, mutta voi toisaalta ohjata viestien sisältöä ja tuottaa lukijassa negatiivisia reaktioita. Tekoälyn rooli kannattaa sovittaa sellaiseksi, että se ei korvaa kriittistä ajatustyötä.

*Generatiivinen tekoäly, kirjoitettu viestintä, kausaaliset mallit,
ihminen –tietokone-vuorovaikutus*

Johdanto

Generoivat tekoälymallit ovat yksi tekoälyn ja syväoppimisen osa-alue, joka keskittyy kerätyn datan tulkittamiseen lisäksi uuden datan tuottamiseen. Ne muuttavat viestintää useilla tavoilla. Nykyiset tekoälymallit tuottavat riittävän korkeatasoisia tekstejä myös ammattilaisille ja syötesuunnittelun avulla tekstituotantoa voidaan kontrolloida tarkasti. Tekoälyllä generoitu tekstisisältö (tekoälyteksti) on monilta osin vertailukelpoista ihmisen tuottaman tekstin (ihmisteksti) kanssa. Edes kirjoittamisen ammattilaiset eivät pysty luotettavasti erottamaan kielimallilla tuotettua tekoälytekstiä ihmistekstistä (Casal & Kessler 2023). Toisaalta tekoälyteksteistä löytyy toistuvia maneeereja esimerkiksi tyylin, sanaston, tunneilmaisun ja tekstin vaihtelevuuden osalta (Herbold ym. 2023; Mirowski ym. 2023; Schaaff ym. 2023). Lukijat saattavat myös kokea tekoälytekstit tai niiksi luulemansa viestit erilaisina (Zhang & Gosline 2023).

Generatiivisen tekoäly nopeuttaa kirjoittamista ja ammattilaisten tuottavuutta (Noy & Zhang 2023). Se voi vaikuttaa myös käyttäjän kirjoittamiseen ohjaten tekstinmuodostumista tiettyyn suuntaan ja ulkoistamalla omaa ajatustyötä (Jakesch ym. 2023; Padmakumar & He 2023). Generatiiviset kielimallit heijastavat opetusdatassaan olevia vinoumia, jolloin sisällöntuotannon eettiset näkökulmat on huomioitava. Luova kirjoittaminen on yhteydessä loogiseen päättelyyn, kykyyn yleistää asioita ja uusien yhteyksien löytymiseen asioiden välille. Korkean tason kausaalinen päättely on kuitenkin edelleen tekoälylle haastavaa.

Keskitymme tässä artikkelissa tekstimuotoiseen viestintään ja generatiivisiin isoihin kielimalleihin (esim, GPT-4 ja Gemini), joiden käyttö sisällöntuotannossa on kasvanut räjähdysmäisesti alkaen vuodesta 2023. Koska generoivat kielimallit ovat olleet käytössä vasta vähän aikaa, vertaisarvioitua tutkimustietoa niiden vaikutuksista on vielä verrattain vähän saatavilla. Tiedämme hyvin vähän siitä, miten kielimallien käyttö vaikuttaa ammattikirjoittajan luomistyöhön ja ajatusprosesseihin. Vaikka useimmat tutkimukset on toteutettu pääosin englannin kielellä, oletamme niiden päätulosten olevan pitkälti kieliriippumattomia.

Käymme ensin läpi ihmisen informaation prosessointia ja päättelyä sekä kuvaamme vaikuttavaa viestintää. Sitten pohdimme, miten generatiivisen tekoälyn käyttö vaikuttaa kirjoittajaan ja tekstin vastaanottajaan

ja mitä eroja ihmistekstin ja tekoälytekstin välillä on havaittu. Käymme läpi syötesuunnittelua ja sitä, miten siinä voidaan korostaa aktivoivia tekijöitä ja päättelyä. Lopuksi pohdimme, miten tekoälyn negatiivisia vaikutuksia kirjoittamiseen voidaan minimoida. Tavoitteenamme on esittää tuoreiden tutkimusten valossa kooste ihmisen informaation prosessoinnista ja sen yhteydestä tekoälyn hyödyntämiseen tekstintuotannossa.

Tätä tekstiä kirjoitettaessa hyödynsimme ChatGPT (GPT-4) ja Perplexity.ai-tekoälypalveluita artikkelien etsintään, kääntämiseen ja oikolukuun.

Ihminen sopeutuu ympäristöihin informaation prosessoinnin avulla

Elinympäristömme sisältää lähes loputtomasti dynaamista informaatiota (Suomala & Kauttonen 2023). Tutkimusten mukaan ihmisäivot ratkaisevat tämän informaation runsauden ongelman hyödyntämällä aiempia kokemuksia ja kontekstuaalista tietoa toiminnassaan. Kehittyessään ihmisäivot muodostavat intuitiivisia representaatioita eli malleja ilman tietoista ponnistelua toimimalla monipuolisesti ympäristöissään (Emt.).

Ihmisaivot mallintavat fysikaalista, psykologista ja kulttuurista todellisuutta. Tällöin fysiikan mallintamisessa keskeistä on informaatio esimerkiksi esineiden massasta ja niiden liikkeestä (Lake ym. 2017), kun taas psykologisessa mallintamisessa kohteena ovat toisten ihmisten käsitukset ja aikomukset (Tomasello 2014). Kulttuurisessa mallintamisessa puolestaan olennaisia ovat yhteisesti omaksuttavat arvot ja toimintatavat (Geary 2005). Intuitiiviset mallit eivät ole kopioita maailmasta, vaan rajattuja subjektiivisia mallinnuksia maailmasta. Ne auttavat meitä olemaan vuorovaikutuksessa tarkoituksenmukaisesti kontekstien esineiden, ihmisten ja kulttuuristen käytäntöjen kanssa. Lisäksi ne auttavat ihmisiä valitsemaan informaatiota ja luomaan odotuksia tulevista tapahtumista (Lake ym. 2017). Kaikki tämä on oleellista sekä arkielämässä että luovien sisältöjen, kuten tekstien, kuvien ja musiikin, tuottamisessa.

Aivot mallintavat informaatiota hierarkkisesti. Kullekin tasolle on muodostunut edellä kuvattuja intuitiivisia malleja, jotka luovat tavoitteita tiettyihin tilanteisiin liittyen. Seuraavaksi kuvaamme näitä tasoja tarkemmin.

Informaation prosessoinnin kolme tasoa

Ihmisaivot mallintavat informaatiota **assosiaatioiden, interventioiden ja mielikuvituksen** tasoilla (Pearl 2019; Pearl & MacKenzie 2020). Nämä kolme tasoa ovat kietoutuneet toisiinsa. **Assosiaatioiden** avulla muodostetaan säännönmukaisuuksia havainnoimalla ympäristöä. Kuvitellaan mediafirman markkinointijohtajaa, joka käyttää assosiativista päättelyä. Johtaja esittää seuraavan kysymyksen: *Kuinka todennäköisesti henkilö, joka seuraa heidän maksullista urheilukanavaansa, tilaa heidän urheiluun liittyvän maksullisen e-uutiskirjeensä?* Tällaiseen kysymykseen voi vastata keräämällä dataa niistä asiakkaista, jotka ovat ostaneet urheilukanavan palvelut ja e-uutiskirjeen. Kun data on kerätty, voidaan laskea korrelaatiot ja käyttää regressioanalyysia ja nähdä, millainen osuus urheilukanavaa katsovista ostaa myös e-uutiskirjeen. Nämä tilastolliset analyysit assosiaatioihin pohjautuen eivät kuitenkaan kerro *kausallisista suhteista*, eli emme tiedä, *miksi* asiakas ostaa urheilukanavan palvelut ja e-uutiskirjeen. Pearl ja Mckenzie (2020) esittävät, että syväoppivat neuroverkot, joihin nykyiset generoivat mallit perustuvat, operoivat enimmäkseen tällä assosiativisella tasolla. Ne voivat etsiä assosiaatioita ja laskea todennäköisyyksiä siitä, miten eri asiat ovat yhteydessä keskenään. Varsinaisena älykkyytenä tätä prosessia ei kuitenkaan voida pitää. Kun sisällöntuottaja hyödyntää generatiivista tekoälyä omassa työssään, hänen on tarpeen tietää, että valtaosa tekoälyn sovelluksista toimii assosiaatioiden tasolla.

Interventiossa sovelletaan keinoja, joilla voidaan vaikuttaa ympäristöön. Sen avulla päätellään kausaalisesti. Tällöin markkinointijohtaja saattaa esittää kysymyksen: *Millä tavoin e-uutiskirjeen hinnan kaksinkertaistaminen vaikuttaa urheilukanavan myyntiin?* Nyt ei ainoastaan tehdä havaintoja ja kerätä dataa, vaan pyritään aktiivisesti vaikuttamaan tapahtumiin hintavaihtelulla. Vaikka asiakasdatan määrä voi olla valtava, kuten tekoälyn tapauksessa usein on, ja vaikka voimme hyödyntää sitä assosioimalla, emme silti kykene sen perusteella vastaamaan interventiotason kysymyksiin. Emme voi vastata hintakysymykseen vain havainnoimalla ja keräämällä dataa aikaisemmasta myyntihistoriasta. Tämä ei onnistu, koska aiempi hintavaihtelu on saattanut olla seurausta lukuisista markkinoilla olevista tekijöistä (Emt.). Interventiossa pyritään toimimaan kontrolloidusti esimerkiksi seuraamalla, miten hinnan kaksinkertaista-

minen vaikuttaa ostopäätöksiin. Tämä on nykyisin suhteellisen helppoa sähköisten asiakasjärjestelmien ansiosta. Voidaan järjestää myös kontrolloitu koe, jossa hinnannoston vaikutusta simuloidaan.

Pearl ja Mckenzie (2020) esittävät, että intervention vaikutusta voidaan tutkia myös ilman koetta. Tällöin markkinointijohtaja tiimeineen voisi kehittää kausaalisen mallin asiakkaan käyttäytymisestä tietyssä markkinatilanteessa. Tällaisen kausaalisen mallin rakentaminen edellyttää älykkyyttä, jota ihmisillä tyypillisesti on. Sen sijaan tekoälymallit, jotka datahavainnoinnin avulla assosioivat, eivät osaa vastata interventio-tyyppeihin kysymyksiin (Emt.). Sisällöntuottajan kannattaa tällä tasolla toimiessaan tukeutua monipuolisesti todistusaineistoihin ja omaan päätelyyn, sillä data itsessään ei kerro, millaisia syy-seuraussuhteet ovat.

Vaikka interventio on laadukkaampi päätelyn muoto kuin havainnointi, se ei silti vielä ole kaikkein vaativinta päätelyä. **Mielikuvituksen** tasolla pohditaan aikaisempien kokemusten perusteella, miten asiat olisivat voineen olla tai voisivat olla (Emt.). Se kohdistuu vaihtoehtoi- siin menneisyyden ja tulevaisuuden tapahtumiin ja tekoihin. Tällöin verrataan tosiasiallista tapahtumaa vaihtoehtoihin mahdollisuuksiin. Pearl ja Mackenzie (2020) painottavat, että kyky mielikuvituksen varassa tapahtuvaan päätelyyn erottaa ihmisen älykkyyden eläimistä ja kone- oppivista algoritmeista. Mediafirman markkinointijohtaja voisi käyttää mielikuvitusta kysymällä: *Mikä on todennäköisyys sille, että asiakas, joka ostaa urheilukanavapalvelun, ostaa sen myös kaksinkertaisella hinnalla?* Tällöin markkinointijohtaja vertaa realistista tilannetta, jossa asiakas ostaa palvelun nykyisellä hinnalla, kuviteltuun tilanteeseen, jossa asiakas ostaa paidan kaksinkertaisella hinnalla. Vastaamalla tällaisiin ”*entä jos asiat tapahtuisivat eri tavoin*”-kysymyksiin, voidaan oppia historiasta ja toisilta ihmisiltä. Tälle kolmannelle tasolle kuuluvat myös miksi-kysymykset.

Ovatko suuret generoivat kielimallit saavuttaneet edes tason kaksi? Luultavasti eivät, mutta kysymykseen on hankala vastata täsmällisesti, sillä kausaalisten tehtävien muotoilu ja tarkistus pelkkää tekstidataa käyttäen on haastavaa (Pearl & Mackenzie 2023). Keskeinen kysymys on, miten tekoälykäs viestintä tukee informaationkäsitteilyn kolmea tasoa. Erityisesti kolmas taso on kiinnostava, koska se on ihmisen luovuuden keskeisin elementti. Ihminen ei vain toimi suoraviivaisesti, vaan hän etsii vaihtoehtoisia ratkaisuja.

Ihmisaivojen mielihaluverkoston aktivoituminen viestinnän avulla

Ihminen lukee ja jakaa sellaisia sisältöjä, jotka hän kokee mielekkäinä (Chan ym. 2023; Cosme ym. 2023; Genevsky & Knutson 2015). Ymmärrettävät, uudet, myönteiset, emotionaaliset, hyödylliset, luotettavat ja helposti samaistuttavat sisällöt koetaan yleensä mielekkäiksi. Vaikka runsaassa informaatioympäristössä viestintä on vaikeaa, tutkimukset osoittavat, että aivojen mielihaluverkosto säätelee ihmisen käyttäytymistä toimimalla portinvartijana hänen mallintaessaan informaatiota (Genevsky & Knutson 2015; Suomala 2023; Suomala & Kauttonen 2023).

Mielihaluverkoston ytimen muodostavat aivopiirit etuotsalohkon keskialueilla ja aivojuoviossa. Nämä aivopiirit ovat kiinteässä yhteydessä toisiinsa ja muihin verkostoihin. Muista verkostoista keskeisimpiä ovat aivojen sosiaalista informaatiota mallintava verkosto (Suomala & Kauttonen 2023) ja mielikuvituksen pohjautuvaa päättelyä tukeva verkosto (Heinonen ym. 2016).

Mielihaluverkoston aktivoituminen on yhteydessä seuraaviin aktivaattoreihin: minä, me, tunne, tieto, tavoite, aistit ja erottautuminen (Suomala 2023; Suomala & Berg 2024). Näitä aktivaattoreita kannattaa hyödyntää viestinnässä ja tekoälyä käytettäessä. Taulukko 1 kuvaa aktivaattorit ja selittää, mitä ne tarkoittavat.

Aktivoivia tekijöitä kannattaa soveltaa oman asiantuntemuksen perusteella. Generoivaa tekoälyä käyttäessä syötesuunnittelussa voi etsiä asioita, jotka vetoavat mielihaluverkostoon ja sitä kautta aktivoivat vastaanottajaa. Esimerkiksi ”minä”-aktivaattorin yhteydessä voi pohtia, millainen viesti aktivoi vastaanottajan minuutta. Tähän voi saada vihjeitä terveysviestien tehokkuutta käsittelevistä tutkimuksista (Falk ym. 2016). Tällöin faktoihin pohjautuva viesti, kuten *tupakoinnin lopettaminen lisää keuhkojen hapenottokykyä*, ei nostanut mielihaluverkoston aktiivisuutta. Sen sijaan minuutta aktivoiva viesti, kuten *sinun keuhkosi hapenottokyky paranee, kun lopetat tupakoinnin*, nosti mielihaluverkoston aktiivisuutta ja sai ihmiset myös toimimaan eli ilmoittautumaan tupakasta vieroittamiskursseille.

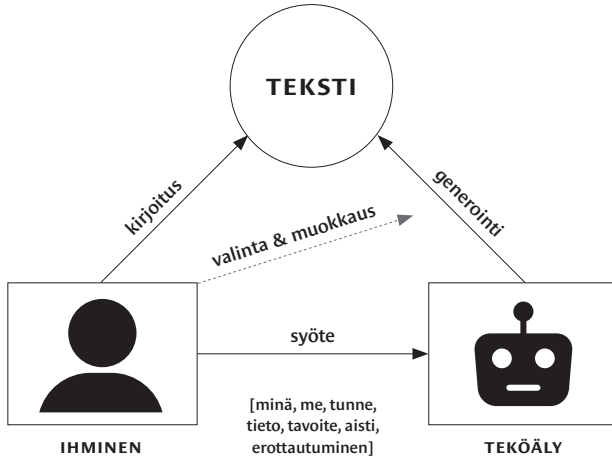
Luovaan kirjoittamiseen sisältyvät (1) ideoiden ja tavoitteiden *suunnittelu*, (2) sanallistaminen eli *kääntäminen* ja (3) tekstin arviointi ja muokkaaminen eli *tarkistaminen* (Flower & Hayes 1981). Chakrabar-

Taulukko 1.

Mielihaluverkoston aktivaattorit, aktivaattorien yhteys aivoverkostoihin ja niiden vaikutus. Muokattu lähteiden Suomala & Berg 2024 ja Suomala & Kauttonen 2022 pohjalta.

AKTIVOIVA TEKIJÄ	PÄÄASIALLINEN AIVOVERKOSTO	TEKIJÄN VAIKUTUS
Minä	Mielihaluverkosto	Minuutta vahvistavat viestit. Miten viestisi vetoaa vastaanottajan minuuteen?
Me	Sosiaalisuusverkosto	Yhteyttä toisiin ihmisiin vahvistavat viestit. Miten viestisi vahvistaa vastaanottajan yhteyttä toisiin ihmisiin ja ryhmiin?
Tunne	Aivojuovion aktivoituminen (osa mielihaluverkostoa)	Myönteisiä tunteita herättävät viestit. Miten viestisi aktivoi vastaanottajan myönteisiä tunteita?
Tieto	Etuoslohkon aktivoituminen (osa mielihaluverkostoa)	Vastaanottajan kannalta tärkeää tietoa sisältävät viestit. Miten viestisi aktivoi vastaanottajan tiedontarpeita?
Tavoite	Etuoslohkon aktivoituminen	Vastaanottajan henkilökohtaisia tavoitteita aktivoivat viestit. Miten viestisi aktivoi vastaanottajan tavoitteita?
Aisti	Mielihaluverkoston aktivoituminen	Aistikanavia aktivoivat viestit. Millaisia aistimuksia viestisi aktivoi vastaanottajassa?
Erottautuminen	Etuoslohkon aktivoituminen	Erottautumista vahvistavat viestit. Mikä erottaa sinun viestisi muista viesteistä?

tyn ja kumppanien (2023) tutkimuksessa havaittiin tekoälyn (GPT-3.5) hyödyllisyys kaikissa kolmessa vaiheessa, mutta erityisesti kahdessa viimeisessä. Tällöin tekoälyllä olisi kirjoittamista tukeva tehtävä. Kuvassa 1 on havainnollistettu miten tekoäly voi tukea kirjoitustyötä. Tekoälyn vaikutus lopulliseen tekstiin riippuu siitä, kuinka paljon sisällöntuottaja kirjoittaa tekstiä itsenäisesti, kuinka hän suunnittelee syötteet ja kuinka paljon ja miten hän valitsee ja muokkaa generoituja tekstejä.



Kuva 1.

Tekoälyavusteinen kirjoitusprosessi (mukailtu lähteestä Woo & Guo 2023).

Generoivien isojen kielimallien kyvyillä on rajansa

Generoivat isot kielimallit tietävät lähes kaikista asioista jotakin, mutta ne eivät ole varsinaisia asiantuntijoita millään alalla. Vaikka uusimmat kielimallit ovat hyviä yleisessä (eng. common-sense) päättelyssä ja osoittavat jonkinasteista älykkyyttä (Bubeck ym. 2023), niissä on myös huomattavia puutteita.

Ihmisten älykkyyden vahvuus on sääntöjen yleistäminen, mielikuvitukseen pohjautuva päättely ja älykkyyden soveltaminen joustavasti eri konteksteissa, mitkä ovat tekoälylle vaikeita (Pearl & MacKenzie 2020; Sun ym. 2023; Webb ym. 2023). Esimerkiksi GPT-4-malli epäonnistuu yksinkertaisissa laskutoimituksissa, mikäli laskenta tehdään muussa kuin tutussa kymmenkantaisessa numerojärjestelmässä, vaikka laskenta noudattaa samaa logiikkaa. Tämä johtuu siitä, että sen opetusdatassa esiintyy hyvin vähän ei-kymmenkantaista numerojärjestelmää eikä malli pysty oppimaan matematiikkaa abstraktilla tasolla. Sama ongelma pätee

muillakin osa-alueilla, kuten shakissa ja muissa peleissä, piirtämisessä ja avaruudellisessa hahmottamisessa (Wu ym. 2023). Myös käänteinen päättely on hankalaa. Tutkimuksessa GPT-4-malli osasi ratkaista tehtävät, jotka noudattivat kaavaa ”A on B”, mutta tarkkuus putosi dramaattisesti tehtävän käänteisessä versiossa, eli ”B on A” (Berglund ym. 2024). Tämä on seurausta siitä, että kielimallit on koulutettu ennustamaan sitä *mitä* ja *miten* ihmiset kirjoittavat, ei sitä mikä on totta.

Tekoäly ei ole vielä saavuttanut informaation prosessoinnin kolmatta ja vaativinta tasoa. Tätä selvittivät Jin ja kumppanit (2023) erityisellä kausaalista päättelyä vaativalla tehtävistöllä. GPT-4-malli saavutti tasoilla 1–3 tarkkuudet 61...66 %, mikä on vain hieman sattu-maa (50 %) parempi. Kim ja kumppanit (2023) puolestaan tutkivat tekoälyn kykyä ns. *mielen teoriassa* käyttäen simuloitua kolmen ihmisten välistä dialogia, jossa tekoälyn piti asettua eri keskustelijoiden asemaan. Testissä tekoäly (GPT-4) epäonnistui jääden kauas vastaavista ihmisten tuloksista.

Tekoälyn älykkyyden ja kognitiivisten kykyjen testaaminen on tärkeää sekä mallien kehittäjille, että tekoälyä työssään käyttäville. Jos viestinnän ammattilainen – esimerkiksi journalisti tai yritysviestijä – käyttää tekoälyä tiedonhakuun ja tekstin tuottamiseen, hänen on ymmärrettävä, millaisissa tilanteissa tekoäly mahdollisesti epäonnistuu ja tekee virheitä.

Tekoälyn tuottamassa tekstissä on erityispiirteitä

Ovatko kielimallien tuottamat tekstit samanlaisia kuin ihmisten kirjoittamat ja voidaanko tekoälyn tuotoksia tunnistaa automaattisesti? Koneen tuottamassa tekstissä on usein tiettyjä erityispiirteitä, kuten tilastollisia poikkeamia ja kielellisiä vivahteita, jotka erottavat ne ihmisten kirjoittamista (Herbold ym. 2023). Esimerkiksi Schaafin ja muiden (2023) tutkimuksissa selvitettiin, voiko GPT-3.5:n kirjoittamia englannin-, ranskan-, saksan- ja espanjankielisiä tekstejä tunnistaa koneoppimismalleja käyttämällä. Tunnistuksen osumatarkkuus oli yleisesti ottaen heikko (70–89 % kielestä riippuen) ja uudempien mallien myötä oletettavasti vielä huonompi. Tunnistuksen teho riippui kuitenkin myös siitä, miten ja millaisilla syötteillä tekoälyn tekstit oli generoitu (Emt.).

Erityisesti luovassa kirjoittamisessa tekoälyn tekstejä on kritisoitu subtekstin heikkouksista ja tekstejä on kuvailtu pinnallisiksi, kliseisiksi, yllätyksettömiksi ja niissä on tunnistettu monimuotoisuuden puutteita ja taipumusta moralisointiin ja yllätyksettömyyteen (Casal & Kessler 2023; Chakrabarty ym. 2024; Mirowski ym. 2023).

Koska kielimallit heijastavat sitä dataa, jolla ne on koulutettu, ne sisältävät erilaisia datan vääristymiä. Kielimallit eivät esimerkiksi edusta kaikkia ihmisryhmiä tasapuolisesti (Santurkar ym. 2023). Tekoälymallin käytön kirjoittamisen apuna on huomattu vähentävän kirjoitusten diversiteettiä ja häivyttävän kirjoittajan omaa persoonallista kirjoitustapaa (Padmakumar & He 2023). Kielimalleissa olevat vääristymät voivat vaikuttaa suoraan luovaan kirjoittamiseen. Jakeschin ja kumppanien (2023) tutkimuksessa koehenkilöt kirjoittivat mielipidekirjoituksia käyttäen GPT-3-pohjaista kirjoitusavustajaa, jotka oli esiohjelmoitu puoltamaan tai vastustamaan tiettyä mielipidettä. Nämä vääristyneet mallit saivat koehenkilöt todennäköisemmin tukemaan mallin suosimaa näkökulmaa. Vuorovaikutus mielipiteellisen tekoälyn kanssa voi siis vaikuttaa käyttäjän mielipiteisiin huomaamatta. Tutkijat kutsuivat tätä ilmiötä *latentiksi taivutteluksi*.

Lukija voi kokea tekoälyn tuottaman tekstin erilaisena

Pelkkä tieto siitä, että tekoäly on tuottanut sisältöä, voi jo itsessään vaikuttaa lukijoiden asenteeseen negatiivisesti (Sundar ym. 2019). Psykologiassa kyseessä on yleisesti tuttu *pohjustusilmiö*, missä aiemmat tiedot vaikuttavat tiedostamattomasti siihen, mihin ihminen kiinnittää huomiota. Tekoälyn yhteydessä puhutaan jopa *algoritmien vastenmielisyydestä* (eng. algorithm aversion; Dietvorst ym. 2015). Ihmiset voivat suosia koneita erityisesti tilanteissa, joissa vaaditaan objektiivisuutta, tarkkuutta ja neutraalia näkemystä (Sundar ym. 2019). Toisaalta ihmisten välisessä tekstipohjaisessa, päivittäisessä ja henkilökohtaisessa viestinnässä tekoälyn käyttö koettiin negatiivisena (Hohenstein ym. 2023), mutta vain jos tekoälyn osallisuus oli tiedossa. Kun viestinnän vastapuolena on kone, ihmisille syntyy helposti stereotyyppisiä mielikuvia tietokoneesta mekaanisena, objektiivisena, tunteettomana ja kylmänä osapuolena (Meng & Dai 2021).

Köbisin ja Mossinkin (2021) tutkimuksen mukaan ihmiset kokivat tekoälyn (GPT-2) tuottamat runot yhtä laadukkaina kuin ihmisten kirjoittamat, vaikka suhtautuivatkin niihin negatiivisemmin. Negatiivisuus ilmeni sekä ennen tekijän paljastamista että sen jälkeen. Toisaalta myöhemmissä tutkimuksissa ei löydetty eroa lukijoiden reaktioissa ihmisen ja tekoälyn (GPT-3.5) kirjoittamiin runoihin luovuuden, omaperäisyyden tai nautinnon osalta (Landa-Blanco ym. 2023). Zhangin ja Goslinen (2023) tutkimuksessa ihmisiä pyydettiin arvioimaan markkinointitekstien ja mainoskampanjoiden laatua, joiden kirjoittajana oli ihminen, tekoäly tai molemmat. Havaittiin, että lukijoiden tietäessä tekstin lähteen he suhtautuivat ihmisten luomaan sisältöön myönteisesti, mutta muutoin suosivat tekoälyllä luotua sisältöä. Kirkby ja kumppanit (2023) tutkivat kuluttajien suhtautumista brändin tuotekuvauksiin ja chat-keskusteluihin, joiden kirjoittajia ei paljastettu tai paljastettiin osittain. Tutkijat eivät löytäneet eroa ihmisten suhtautumisessa tekoälyyn, ja kaikki tekstit arvioitiin yhtä autenttisiksi.

Tutkimustulosten perusteella suhtautuminen riippuu siis vahvasti kontekstista ja tekstin lajista. Esimerkiksi runot ja kaunokirjallisuus vaikuttavat voimakkaasti emotionaalisesti, kun taas markkinointimateriaalit voidaan kokea kevyempänä ja heikosti lukijaa sitouttavana sisältönä. Epäilemättä myös valitun tekoälyn kyvykkyys ja syötesuunnittelu vaikuttavat huomattavasti kokonaisuuteen.

Hyvällä syötesuunnittelulla kohti vaikuttavaa viestintää

Syötesuunnittelu on toistaiseksi ratkaisevassa osassa generatiivisen tekoälyn käytössä. Se on yhdistelmä taidetta ja tiedettä ja vaatii teknistä ja strategista ajattelua sekä luovuutta (Stechly ym. 2023). Viestinnän ammattilaisen on tarpeen tuntea keskeiset syötesuunnittelutekniikat, jotka riippuvat paitsi tekoälymallista ja mediasta (kuvat vai teksti). Seuraavat huomiot koskevat lähinnä tekstin tuottamista GPT-4-tasoisilla malleilla.

Hyvä syöte on eksakti, ja se sisältää *kontekstin, tavoitteen, tyylin, sävyn, yleisön ja vastausformaatin* (ks. COSTAR-menetelmä, Teo 2023). Kontekstilla kerrotaan, mitä tehdään ja missä tilanteessa. Tavoitteella kerrotaan, mitä tekoälyn halutaan tekevän. Tyyli määrittää, miten teksti

kirjoitetaan. Halutaanko tekstin simuloivan tietyssä roolissa olevan henkilön tekstiä. Sävy kertoo esimerkiksi sen, onko kyseessä muodollinen, vakava, iloinen tai hauska kirjoitus. Kohderyhmä kertoo, kenelle teksti suunnataan, ja vastausformaattilla määritellään, missä muodossa vastaus halutaan (esim. ranskalaisilla viivoilla).

Monimutkaisissa päättelyä vaativissa kirjoitustehtävissä kannattaa suosia esimerkkeihin pohjautuvia syötteitä. Tällöin tekoälymallia voi ohjata ja auttaa ratkaisussa. Kielimallin loogisen päättelyketjun parantamisessa voi hyödyntää syötteitä, kuten ”pohditaan asiaa askel askeleelta” tai ”lue ensin ongelma ja laadi suunnitelma sen ratkaisemiseksi. Jatka sitten suunnitelman toteuttamista ja ongelman ratkaisemista askel askeleelta.” (Sun ym. 2023). Täsmällisyyden nimissä kannattaa syötteen ”kirjoita muutama lause” sijasta pyytää ”kirjoita 5 lausetta”. Tärkeitä asioita voi korostaa isoilla kirjaimilla.

Otetaan esimerkki siitä, millainen voisi olla hyvä mallipohjaan perustuva syöte lyhyen mainostekstin kirjoitukseen. Korostetaan syötteessä Taulukon 1 tekijöitä ”Me”, ”Tunne” ja ”Aisti” (merkitty lihavoituna).

KONTEKSTI

Haluan kirjoittaa mainostekstin, joka mainostaa yritystäni, joka on pieni kukkakauppa Helsingissä.

TAVOITE

Tee minulle NELJÄN LAUSEEN mainosteksti, joka saa ihmiset ostamaan kukkia **omille läheisilleen**.

TYYLI

Mainoksen tyyli on epämuodollinen ja **tunteita herättävä**.

SÄVY

Lämmin ja tunteisiin vetoava. Korosta kukkien **kauneutta ja läheisten ilahduttamista**.

YLEISÖ

Yritykseni asiakaskuntaa ovat tyypillisesti keski-ikäiset ja sitä iäkkäämpi väestö. Räätelöi viestisi sen mukaan.

GPT-4:n vastaus on seuraava ja täyttää hyvin vaatimukset:

”Löydä Helsingin sydäimestä piilotettu aarre – meidän kukkakaupamme, joka täyttää päiväsi kauneudella ja ilolla. Anna kukkien puhua puolestasi ja tuo hymy rakkaidesi huulille; jokainen kimpumme on suunniteltu erityisesti heitä ajatellen. Kukkien lumoava tuoksu ja värien loisto kutsuvat sinut jakamaan ainutlaatuisia hetkiä. Tule, anna meidän auttaa sinua ilahduttamaan niitä, jotka merkitsevät sinulle eniten – kauneudella, joka kertoo enemmän kuin tuhat sanaa.”

Heikentäkö tekoälyn käyttö kirjoittamisessa kriittistä ajattelua?

Tekstiä generoivat tekoälymallit ovat olleet laajassa käytössä vasta vähän aikaa. Vaikka niiden vaikutus sisällöntuotannon tuottavuuden lisääntymiseen on selvä (Noy & Zhang 2023), tutkimuksellinen näyttö tekoälymallien pitkäaikaisen käytön vaikutuksesta sisällöntuottajien kriittiseen ajatteluun, luovuuteen ja jatkuvaan oppimiseen on vasta alkutekijöissään.

Ei-rutiininomaisessa kirjoittamisessa kirjoittaja liikkuu ajattelun ja tekstin tuottamisen välillä. Prosessi on vaativa sisältäen tekstin rakenteen, kielen ja tyylin valintaa sekä argumenttien muotoilua. Lisäksi kirjoittajat voivat visualisoida ajatuksiaan paperille tai näytölle etsien ideoita, joiden avulla muokata tekstiä. Tällainen luova kirjoittaminen voi muuttaa kirjoittajaa, avata uusia näkökulmia tai paljastaa uuden oppimisen tarpeen. Jos aihe on uusi ja tekoäly tuottaa tekstin automaattisesti ohittaen edellä kuvatun prosessin, se voi vähentää luovuutta ja uuden oppimista (Anson 2023). Olemme tästä samaa mieltä. Generoiva tekoäly voi toimia erinomaisena työkaluna, mutta sisällöntuottajien on pohdittava, missä ja miten sitä kannattaa käyttää. Parhaimmillaan luova kirjoittaminen parantaa omaa ajattelua ja kirjoitustaitoja, ja siksi sitä ei kannata ulkoistaa koneelle. Ihminen voi tuottaa sisältöä havaitsemisen, intervention ja kontrafaktuaalisen päättelyn avulla. Tekoälymallilla ei kannata luoda sellaista tekstiä, joka nojaa vahvasti kausaaliseen päättelyyn tai tiedon täsmällisyyteen (Jin ym. 2023).

Lopuksi

Vaikka nykyinen generatiivinen tekoäly ei vielä kykene aidosti mielikuvitukseen ja toimimiseen korkeimmalla päättelyn tasolla, tilanne voi olla toinen jo muutaman vuoden kuluttua. Tämän vuoksi näemme, että taitavan sisällöntuottajan rooli kasvaa lähitulevaisuudessa, mikäli tekoälyllä hoidetaan lähinnä vähemmän vaativia rutiinitöitä. Kuluttajat tottuvat tekoälyyn sekä sen tuotoksia kuluttamalla että myös käyttämällä sitä itse. Vaikka tekoälyn tuotoksia aikaisemmin saatettiin jopa karsastaa, tilanne voi jopa kääntyä tekoälyn sisältöjen suosimiseen mallien parantuessa. Tiivistämme neuvomme viestinnän ammattilaisille seuraavasti:

1. Panosta syötesuunnitteluun: Käytä mallipohjaa ja päättelyketjuja ja sisällytä syötteisiin sisällön vastaanottajaa aktivoivia tekijöitä (ks. Taulukko 1).
2. Älä ulkoista kriittistä ajattelutyötä tekoälylle: Luova kirjoittaminen kehittää ajatteluasi ja tehostaa uusien asioiden oppimista.
3. Pysy kartalla tekoälymallien kehityksestä: Eri tekoälymalleilla on erilaisia rajoituksia, esimerkiksi kaikki mallit eivät toimi hyvin suomeksi tai ovat heikompia päättelyssä.
4. Huomioi kohderyhmäsi ja ole avoin tekoälyn käytöstä: Jotkut lukijat saattavat karsastaa tekoälyllä tuotettua sisältöä.

Näistä voit aloittaa

- 1) DAIR.AI:n ylläpitämä ajantasainen syötesuunnitteluohje: <https://www.promptingguide.ai/>
 - 2) ATLAS: An LLM Inquiry Principle Benchmark.
26 tutkimukseen perustuvaa syötesuunnittelun periaatetta:
ATLAS: An LLM Inquiry Principle Benchmark
 - 3) Avointen tekoälymallien ranking: https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard ja äänestykseen perustuva ranking: <https://huggingface.co/spaces/lmsys/chatbot-arena-leaderboard> joilla voit ajantasaisesti seurata esim. kielimallien päättelykykyjä.
-

Kuva: Sanni Laine



FT Janne Kauttonen on vanhempi tutkija Haaga-Helia ammattikorkeakoulussa. Hän väitteli teoreettisen fysiikan alalta 2012 ja on sen jälkeen tehnyt poikkitieteellistä tutkimustyötä neurotieteen, kognitiotieteen ja datatieteen parissa. Hän keskittyy erityisesti data-analyysiin, laskennallisiin menetelmiin sekä tekoälyn soveltamiseen ja käyttöönottoon organisaatioissa. Erityisenä kiinnostuksen kohteenaan ovat luonnollisen kielen analyysi ja generatiiviset tekoälymenetelmät.



KT, dosentti Jyrki Suomala työskentelee Laurea-ammattikorkeakoulun yliopettajana. Hän tutkii ihmisen päätöksentekoa moniulotteisissa luonnollisissa tilanteissa. Hän soveltaa tutkimuksissaan moderneja neurotieteen ja tekoälyn menetelmiä. Suomalalla on laaja kokemus tutkimuksesta ja opettamisesta sekä suomalaisissa että ulkomaisissa korkeakouluissa.

Kirjallisuus

- Anson, Chris (2023). What are the cognitive consequences of machine-generated writing? *NC State University News* 27.3.2023. Saatavilla: <https://news.ncsu.edu/2023/03/is-ai-changing-the-way-we-write> (luettu: 19.4.2024).
- Berglund, Lukas; Tong, Meg; Kaufmann, Max; Balesni, Mikita; Stickland, Asa Cooper; Korbak, Tomasz & Evans, Owain (2024). The reversal curse: LLMs trained on "A is B" fail to learn "B is A". *Proceedings of International Conference on Learning Representations (ICLR) 2024*. Saatavilla: <http://arxiv.org/abs/2309.12288>
- Bubeck, Sébastien; Chandrasekaran, Varun; Eldan, Ronen; Gehrke, Johannes; Horvitz, Eric; Kamar, Ece; Lee, Peter; Lee, Yin Tat; Li, Yuanzhi; Lundberg, Scott; Nori, Harsha; Palangi, Hamid; Ribeiro, Marco Tulio & Zhang, Yi (2023). Sparks of artificial general intelligence: early experiments with GPT-4. Saatavilla: <http://arxiv.org/abs/2303.12712>
- Casal, J. Elliott & Kessler, Matthew (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, 2(3), 100068. doi: 10.1016/j.rmal.2023.100068
- Chakrabarty, Tanmoy; Laban, Philip; Agarwal, Divyanshu; Muresan, Smaranda & Wu, Chien-Shiung (2024). Art or artifice? Large language models and the false promise of creativity. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Saatavilla: <http://arxiv.org/abs/2309.14556>
- Chakrabarty, Tuhin; Padmakumar, Vishakh; Brahman, Faeze & Muresan, Smaranda (2023). Creativity support in the age of large language models: an empirical study involving emerging writers. Saatavilla: <https://arxiv.org/abs/2309.12570>
- Chan, Hang-Yee; Scholz, Christin; Cosme, Danielle & Falk, Emily B. (2023). Neural signals predict information sharing across cultures. *Proceedings of the National Academy of Sciences*, 120(44), e2313175120. doi: 10.1073/pnas.2313175120
- Cosme, Danielle; Scholz, Christin; Chan, Hang-Yee; Doré, Bruce; Pandey, Prateekshit; Carreras-Tartak, José & Cooper, Nicole (2023). Message self and social relevance increases intentions to share content: correlational and causal evidence from six studies. *Journal of Experimental Psychology: General* 152(1): 253–67. doi: 10.1037/xge0001270.
- Dietvorst, Berkeley J.; Simmons, Joseph P. & Massey, Cade (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. doi: 10.1037/xge0000033
- Falk, Emily; O'Donnell, Matthew; Tompson, Steven; Gonzalez, Richard; Sonya, Cin; Strecher, Victor; Cummings, Kenneth & An, Lawrence (2016). Functional brain imaging predicts public health campaign success. *Social Cognitive and Affective Neuroscience*, 11(2): 204–14. doi: 10.1093/scan/hsv108.

-
- Flower, Linda & Hayes, John R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365. doi: 10.2307/356600
- Geary, David (2005). *The origin of mind: evolution of brain, cognition, and general Intelligence*. 1st ed. Washington, DC: American Psychological Association.
- Genevsky, Alexander & Knutson, Brian (2015). Neural affective mechanisms predict market-Level microlending. *Psychological Science* 26(9): 1411–22. doi: 10.1177/0956797615588467
- Heinonen, Jarmo; Numminen, Jussi; Hlushchuk, Yevhen; Antell, Henrik; Taatila, Vesa & Suomala, Jyrki (2016). Default mode and executive networks areas: association with the serial order in divergent thinking. *PLOS ONE*, 11(9): e0162234. doi: 10.1371/journal.pone.0162234.
- Herbold, Steffen; Hautli-Janisz, Annette; Heuer, Ute; Kikteva, Zlata & Trautsch, Alexander (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, 13(1), 1–27. doi: 10.1038/s41598-023-45644-9
- Hohenstein, Jess; Kizilcec, René F.; DiFranzo, Dominic; Aghajari, Zhila; Mieczkowski, Hannah; Levy, Karen; Naaman, Mor; Hancock, Jeffrey T. & Jung, Malte F. (2023). Artificial intelligence in communication impacts language and social relationships. *Scientific Reports*, 13, 5487. doi: 10.1038/s41598-023-30938-9
- Jakesch, Maurice; Bhat, Advait; Buschek, Daniel; Zalmanson, Lior & Naaman, Mor (2023). Co-Writing with opinionated language models affects users' views. *Conference on Human Factors in Computing Systems - Proceedings*, 1(1), 1–22. doi: 10.1145/3544548.3581196
- Jin, Zhijing; Chen, Yuen; Leeb, Felix; Gresele, Luigi; Kamal, Ojasv; Lyu, Zhiheng; Blin, Kevin; Gonzalez Adatao, Fernando; Kleiman-Weiner, Max; Sachan, Mrinmaya & Schölkopf, Bernhard. (2023). A benchmark to assess causal reasoning capabilities of language models. Saatavilla: <https://arxiv.org/abs/2312.04350>
- Kim, Hyunwoo; Sclar, Melanie; Zhou, Xuhui; Bras, Ronan; Kim, Gunhee; Choi, Yejin & Sap, Maarten (2023). FANToM: A benchmark for stress-testing machine theory of mind in interactions. *Conference on Empirical Methods in Natural Language Processing*, 14397–14413. doi: 10.18653/v1/2023.emnlp-main.890
- Kirkby, Alexandra; Baumgarth, Carsten & Henseler, Jörg (2023). To disclose or not disclose, is no longer the question – effect of AI-disclosed brand voice on brand authenticity and attitude. *Journal of Product and Brand Management*, 32(7), 1108–1122. doi: 10.1108/JPBM-02-2022-3864
- Köbis, Nils & Mossink, Luca D. (2021). Artificial intelligence versus Maya Angelou: experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, 114, 13. doi: 10.1016/j.chb.2020.106553

-
- Lake, Brenden M.; Ullman, Tomer D.; Tenenbaum, Joshua B. & Gershman, Samuel J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, e253. doi: 10.1017/S0140525X16001837
- Landa-Blanco, Miguel; Agüero-Flores, Maitée & Mercado, Miguel (2023). Human vs. AI authorship: does it matter in evaluating creative writing? A pilot study using ChatGPT. Saatavilla: <https://osf.io/preprints/psyarxiv/wjwsm3>
- Meng, Jiaying & Dai, Yilin Nancy (2021). Emotional support from AI chatbots: should a supportive partner self-disclose or not? *Journal of Computer-Mediated Communication*, 26(4), 207–222. doi: 10.1093/jcmc/zmab005
- Mirowski, Piotr; Mathewson, Kory W.; Pittman, Jaylen & Evans, Richard (2023). Co-writing screenplays and theatre scripts with language models: Evaluation by industry professionals. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 355, 1–34. doi: 10.1145/3544548.3581225
- Noy, Shakked & Zhang, Whitney (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192. doi: 10.1126/science.adh2586
- Padmakumar, Vishakh; & He, He. (2023). Does writing with language models reduce content diversity? *Proceedings of International Conference on Learning Representations (ICLR) 2024*. Saatavilla: <http://arxiv.org/abs/2309.05196>
- Pearl, Judea (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3), 54–60. doi: 10.1145/3241036
- Pearl, Judea & Mackenzie, David (2023). *Judea Pearl, AI, and causality: what role do statisticians play?* AMSTATNEWS 1.9.2023. Saatavilla: <https://magazine.amstat.org/blog/2023/09/01/judeapearl> (luettu: 19.4.2024).
- Pearl, Judea & Mackenzie, David (2020). *The book of why: The New science of cause and effect*. First trade paperback edition. New York: Basic Books.
- Santurkar, Shibani; Durmus, Esin; Ladhak, Faisal; Lee, Cinoo; Liang, Percy & Hashimoto, Tatsunori. (2023). Whose opinions do language models reflect? *Proceedings of Machine Learning Research*, 202, 29971–30004. doi: 10.5555/3618408.3619652
- Schaaff, Kristina; Schlippe, Tim & Mindner, Lorenz (2023). Classification of human- and AI-generated texts for English, French, German, and Spanish. *Proceedings of the 6th International Conference on Natural Language and Speech Processing (ICNLSP 2023)*. Saatavilla: <https://aclanthology.org/2023.icnls-1.1>
- Stechly, Kaya; Marquez, Matthew & Kambhampati, Subbarao (2023). GPT-4 doesn't know it's wrong: an analysis of iterative prompting for reasoning problems. *Proceedings of Foundation Models for Decision Making @NeurIPS 2023*. Saatavilla: <https://openreview.net/pdf?id=PmtZjDYB68>
-

-
- Sun, Jiankai; Zheng, Chuanyang; Xie, Enze; Liu, Zhengying; Chu, Ruihang; Qiu, Jianing; Xu, Jiaqi; Ding, Mingyu; Li, Hongyang; Geng, Mengzhe; Wu, Yue; Wang, Wenhai; Chen, Junsong; Yin, Zhangyue; Ren, Xiaozhe; Fu, Jie; He, Junxian; Yuan, Wu; Liu, Qi; Liu, Xihui; Li, Yu; Dong, Hao; Cheng, Yu; Zhang, Ming; Heng, Pheng Ann; Dai, Jifeng; Luo, Ping; Wang, Jingdong; Wen, Ji-Rong; Qiu, Xipeng; Guo, Yike; Xiong, Hui; Liu, Qun; Li, Zhenguo. (2023). A survey of reasoning with foundation models. Saatavilla: <http://arxiv.org/abs/2312.11562>
- Sundar, S. Shyam & Kim, Jinyoung (2019). Machine heuristic: when we trust computers more than humans with our personal information. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*, 538, 1–9. doi: 10.1145/3290605.3300768
- Suomala, Jyrki (2023). *Ostavat aivot (in English The Buying Brain)*. BasamBooks.
- Suomala, Jyrki & Berg, Pekka (2024). Brain's desire network as the knowledge sources for the innovation process concept development stage. Teoksessa: Luiz. Moutinho & Cerf, Moran (toim.). *Biometrics and Neuroscience Research in Business and Management: Advances and applications*. De Gruyter.
- Suomala, Jyrki & Kauttonen, Janne (2022). Human's intuitive mental models as a source of realistic artificial intelligence and engineering. *Frontiers in Psychology*, 13, 873289. doi: 10.3389/fpsyg.2022.873289
- Suomala, Jyrki & Kauttonen, Janne (2023). Computational meaningfulness as the source of beneficial cognitive biases. *Frontiers in Psychology*, 14, 1189704. doi: 10.3389/fpsyg.2023.1189704
- Teo, Sheila (2023). *How I Won Singapore's GPT-4 Prompt Engineering Competition*. Saatavilla: <https://towardsdatascience.com/how-i-won-singapores-gpt-4-prompt-engineering-competition-34c195a93d41> (luettu: 19.4.2024).
- Tomasello, Michael (2014). The ultra social animal. *European Journal of Social Psychology*, 44(3): 187–94. doi: 10.1002/ejsp.2015.
- Webb, Taylor; Holyoak, Keith J. & Lu, Hongjing (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour*, 7(9), 1526–1541. doi: 10.1038/s41562-023-01659-w
- Woo, David James & Guo, Kai (2023). Exploring an AI-supported approach to creative writing: Effects on secondary school students' creativity. Saatavilla: <https://doi.org/10.13140/RG.2.2.24489.06247>
- Wu, Zhaofeng; Qiu, Linlu; Ross, Alexis; Akyürek, Ekin; Chen, Boyuan; Wang, Bailin; Kim, Najoung; Andreas, Jacob & Kim, Yoon (2023). Reasoning or reciting? Exploring the capabilities and limitations of language models through counterfactual tasks. Saatavilla: <http://arxiv.org/abs/2307.02477>
- Zhang, Yunhao & Gosline, Renée (2023). Human favoritism, not AI aversion: People's perceptions (and bias) toward generative AI, human experts, and human-GAI collaboration in persuasive content generation. *SSRN Electronic Journal*, 18, e41. doi: 10.2139/ssrn.4453958
