FA!L! 4/2026

Tiedonhallinnan ja arkistoinnin erikoislehti



Harjoittelu arkistossa s. 21
Arkistointi ja tiedonhallinta s. 6
Digitaalinen tieto vaarassa s. 10
Enemmän vähemmällä s. 13
Arkistosta tiedoksi s. 4
DLM Forum s. 14
Tekoäly tuotantoon s. 18
Agenttinen tekoäly s. 25

FAILI 4/2026

Toimitus

Päätoimittaja: Juha Henriksson,
juha@musiikkiarkisto.fi, 045-2495680

Julkaisutoimikunta
Susanna Kokkinen, pj, Aalto-yliopisto
Pekka Henttonen, Tampereen Yliopisto
Toni Suutari, Liikearkistoyhdistys
Johannes Valo, Kunnallisarkistoyhdistys
Pekka Lähteenkorva, Urho Kekkosen Arkisto
Ella Oksanen, Kela
Olli Alm, Elka
Miia Kosonen, Xamk

Toimitus pidättää itselleen oikeuden editoida julkaistavaksi tarjottua aineistoa.

Lehti ilmestyy neljä kertaa vuodessa.

Vuosikerran hinta 2025 on 60,00 euroa, opiskelijat 15,00 euroa. Tilaushintoihin lisätään arvonlisävero 10 %.

ISSN 1237-2196 Painosmäärä 380 kpl

Julkaisija: Liikearkistoyhdistys ry, PL 271, 00101 HKI, www.liikearkistoyhdistys.fi, Puh. 044-5512029.

Tilaukset: sihteeri@liikearkistoyhdistys.fi

Ilmoitukset
Ilmoitusmyynti, puhelin 045-2495680
Ilmoitushinnat:
1/1 sivua 4-väri 800
1/2 sivua 4-väri 500
1/4 sivua 4-väri 250

Toistoalennus: sama aineisto 4 ilmoituskertaa 25%, 2 ilmoituskertaa 10%.
LAY:n jäsenalennus 20%.

Ilmoitusvalmistus 50 e/sivu.

Oman mainosmateriaalin postitus lehden mukana: kokosivun ilmoitushinta ja lisäksi sisäänpisto- ja postituskulut.

Hintoihin lisätään arvonlisävero.
Hinnat sitoutumuksetta.

Ilmoitusaineiston tekniset tiedot
Painotuotteen koko 173 X 245 mm.
Tiedostomuodot valmiit aineistot painettava pdf, fontit konvertoituina.
Kuvat jpeg, tiff, eps, värijärjestelmä CMYK.
Kuvaresoluutio vähintään 300 dpi.

Muista tiedostomuodoista ota yhteyttä.

Aineiston toimitusosoite:
juha@musiikkiarkisto.fi

Paino: Grano Oy, 2026.

SISÄLTÖ

3 **Arkisto vai tietovaranto**
Juha Henriksson

4 **Arkistosta tiedoksi – paradigman muutos vai legitimiteettikriisi?**
Maria Kallio-Hirvonen

6 **Jotain vanhaa, jotain uutta – arkistoinnin ja tiedonhallinnan liitto**
Maiju Hakala ja Minna-Liisa Mäkiranta

10 **Digitaalinen tieto vaarassa**
Miia Kosonen

12 **Minne suomalaisesta sarjakuvasta kertova arkistoaineisto hajautuu?**
Ahto Harmo

13 ***Enemmän Vähemmällä* – pienen organisaation hallinnon sujuvoittamisesimerkki Konneveden ja kumppanikuntien tapaan**
Johannes Valo

14 **DLM Guinessin synnyinsijoilla**
Anssi Jääskeläinen

18 **Tekoäly tuotantoon -teemapäivä**
Juha Henriksson

21 **Harjoittelu on tie asiantuntijuuteen**
Susanna Kokkinen

24 **Nokia Design Archive – vuosi kokoelmatyötä Aalto-yliopiston arkistossa**
Jesse Jaakkola

25 **Agenttisen tekoälyn haasteita**
Juha Henriksson

Kansi: Aalto-yliopiston arkistolaiset Big Pictures -näyttelyssä.
Kuva: Aalto-yliopiston arkisto, Riina Ojanen.
(Juttu Aalto-yliopiston arkiston harjoittelusta on sivulla 21.)

Mainokset: Laaksonen s. 27; Disec s. 28; Kasten s. 28.

Arkisto vai tietovaranto?

Maria Kallio-Hirvonen pohtii tässä *Failin* numerossa tärkeää kysymystä: onko ”arkisto”-sanan korvaaminen ”tiedolla” paradigman muutos vai legitimiteettikriisi? Kallio-Hirvonen lähestyy artikkelissaan arkistoteorian näkökulmasta samaa asiaa, jota olen itsekin paljon pohtinut. Kansallisarkiston strategiassa puhutaan nykyisin tietovarannoista, mutta ei sanallakaan arkistoista. Mitä tämä tarkoittaa arkistoprofession, mutta myös laajemmin arkistojen vaikuttavuuden kannalta?

Kansallisarkiston ja muiden julkista rahoitusta saavien arkistojen toiminnan tulisi olla vaikuttavaa. Kuten olen jo aikaisemmin kirjoittanut *Failissa*, vaikuttavuus ei ole sama asia kuin esimerkiksi kävijämäärät tai tietoaineistojen latauskerrat. Väitöskirjaa tekevän tutkijan yksi fyysinen käynti arkistossa voi olla yhteiskunnalliselta vaikuttavuudeltaan valtavasti suurempi kuin yksittäisen tiedon tarkastelu verkkopalvelun kautta. Esimerkiksi pelkät sähköisten palveluiden käyttömäärät eivät vielä kerro vaikuttavuudesta. Vaikka arkistojen tulee palvella kaikkia, niin useimmiten merkittävin yhteiskunnallinen vaikuttavuus syntyy – Kallio-Hirvosen sanoin – ”tutkijan ja arkiston vuorovaikutuksessa”.

Arkistoalalla työskentelevien ammatillisen identiteetin kannalta Kansallisarkiston käyttämän terminologian muutos on keskeinen, eikä ole mikään ihme, että se on herättänyt hämmennystä monissa kollegoissani. Aihetta on käsitelty muun muassa Arkiston valta -podcasteissa, erilaisissa tapahtumissa sekä monissa epävirallisissa keskusteluissa. Arkistoprofession muutos on myös Arkistoyhdistyksen tämän vuoden syysseminaarin aiheena. Keskeinen kysymys on myös se, miten Kansallisarkiston paradigman muutos heijastuu arkistoalan koulutukseen – tai pitäisikö sen heijastua siihen?

Kansallisarkisto on päättänyt lopettaa perinteiset Arkistopäivät ja korvata ne Julkisen tiedon päivillä. "Arkistopäivät” on ollut nimenä ja tapahtumana houkutteleva, kun taas ”Julkiset tiedon päivät” ei herätä minkäänlaisia intohimoja. Varsinkin, kun tapahtuma on hinnoiteltu siten, ettei osallistuminen ole ainakaan minulle taloudellisesti mahdollista. Alan yhdistyksissä onkin ollut keskustelua siitä, voisimmeko koordinoida toimintaamme jatkossa siten, että pystyisimme vielä entistäkin paremmin tarjoamaan tapahtumia, jotka yhdistäisivät kaikkia arkistoalan ammattilaisia.

Nostan vielä lopuksi esille Kallio-Hirvosen ajatuksen siitä, että ”arkiston korvaaminen tiedolla voi lopulta tehdä arkistosta yhdentekevän. Joskus kaikkein arvokkain tieto löytyy jostain muualta kuin asiakirjan sisällöstä.” Mielestäni tätä ajatusta voidaan laajentaa: mitä käsitteiden muuttuminen merkitsee strategisella tasolla eli arkisto-organisaatioiden tulevaisuuden kannalta? Jos puhumme ainoastaan ”tietovarannoista”, niin mikä rooli on sillä asiantuntemuksella, jota tarvitaan, kuten Kallio-Hirvonen osuvasti kirjoittaa, tiedon muuttamiseksi irralliseksi informaatiosta kestäväksi tiedoksi provenienssin ja kontekstin avulla. Onko mahdollista, että säilytettävät tiedot siirretään tulevaisuudessa viranomaisilta kansalliseen pitkäaikaissäilytyspalveluun suoraan tekoälyn avustuksella? Tarvitaanko siihen väliin enää Kansallisarkistoa, ainakaan sen nykyisessä laajuudessa?

* * *

Oma organisaationi Musiikkiarkisto haluaa olla eturivissä mukana kehittämässä kulttuuriperintösektoria. Olemme vahvasti mukana kansainvälisessä yhteistyössä ja -keskustelussa esimerkiksi digitoinnissa, tekoälyn kehittämisessä ja virtuaalimaailmojen hyödyntämisessä, mutta myös kulttuuriperintödatan tarjoamisessa data-avaruuksien välityksellä. Mutta silti toimintamme suurin vaikuttavuus syntyy siitä, että meillä on osaamista ja ymmärrystä sitoa aineistomme historialliseen kontekstiin. Teemme tätä työtä kiinteässä yhteistyössä sidosryhmiemme kanssa. Aineistomme ovat osa suomalaista kulttuurihistoriaa, eikä niitä voi supistaa pelkäksi tiedoksi.



Juha Henriksson

Tekoäly tuotantoon -teemapäivä



Juha Henriksson
Arkistonjohtaja
Musiikkiarkisto
Dosentti

Liikearkistoyhdistys järjesti syyskuussa teemapäivän tekoälyn tuotantokäytöstä. Päivän aikana ei visioitu, mitä tekoäly tekee kymmenen vuoden päässä, vaan elettiin vahvasti nykyhetkeä konkreettisten case-esimerkkien kautta. Teemapäivä järjestettiin tällä kertaa kokonaan virtuaalisesti.

Agentit ja tekoälyn taidot

Juha Anttilan (IITC) aiheena oli "Agentit ja tekoälyn taidot (Skills) dokumenttien hallinnassa". Anttila totesi esityksensä alussa, että teknologia ei enää aseta rajoituksia, vaan suurin haaste toiminnan kehittämiselle on se, että ihminen ja organisaatio pystyvät muuttamaan toimintatapaansa.

Agenttinen tekoäly kehittyy tällä hetkellä hurjaa vauhtia. Anttilan mielestä tekoälyagentit palvelevat organisaatiota paremmin kuin chat-malli, joka tukee enemmän yksilöiden toimintaa. Mutta keskeinen kysymys on, miten vastuuvelvollisuus turvataan tekoälyagentteja käytettäessä?

”Skillit” ovat lisäosia, jotka opettavat Copilotille tietyn tehtävän, työprosessin tai toimintatavan. Skillien määritteleminen muuttaa monivaiheisen tehtävän toistettavaksi prosessiksi, jolloin Copilot toimii johdonmukaisemmin ja vähemmällä ohjauksella. Skillit voivat olla esimerkiksi ohjeita, mallipohjia, referenssidokumentteja tai skriptejä. Anttila esitti olennaisen kysymyksen: toistaako Skills-malli nykyistä prosessia? Miten voidaan ajatella prosessit kokonaan uudestaan?

Anttila mukaan dokumenttien merkitys on entisestään kasvanut, sillä ne ovat ”ruokaa” tekoälylle – data ei saa olla esimerkiksi jossain verkkolevyn nurkalla, ei-käytettävässä muodossa. Hän korosti lisäksi tietoturvaa: jos tekoälyprosessin lähde on luottamuksellinen, myös tulos täytyy luokitella sensitiiviseksi.

AI tekstintunnistuksessa

Ville Solja (OP) kertoi havainnoista ja kokemuksista tekoälyn hyödyntämisestä tekstintunnistuksessa. OP:n historia on ollut pitkä ja monivaiheinen, mikä heijastuu myös eri aikoina arkistoituituihin dokumentteihin. Digitoinnin kannalta olennaisin mittari on automaatioaste eli STP, joka kuvaa sitä, kuinka suuri osa asiakirjoista voidaan käsitellä prosessin päästä päähän ilman manuaalista työvaihetta.

Soljan mukaan STP on vaihdellut huomattavasti eri tapausten ja aineistojen välillä. Syyt eivät aina johdu prosessin teknologioista, vaan taustalla saattavat olla esimerkiksi asiakirjalta puuttuvat tiedot tai ristiriidat taustadatan kanssa. Tavoitteena on pyrkiä noin 90% automaatioasteeseen jokaisessa käyttötapauksessa, mutta tässä on ollut aineistojen moninaisuudesta johtuvaa huomattavaa vaihtelua.

OP:n prosessissa perinteinen tekstintunnistus, koneoppiminen, kielimallit ja ihmistyö täydentävät kaikki toinen toisiaan. Esimerkiksi Tessaract suoriutuu hyvin lasertulostetusta aineistosta ja sillä on pieni yksittäinen kustannus.

Asiakirjaluokitteluun käytetty koneoppimismalli on Soljan mukaan itse koulutettavissa, mutta arkistojen värikkään historian takia kouluttaminen vaatii paljon työtä. Ongelmana onkin ollut riittävän edustavan ja ylläpidettävän opetusaineiston muodostaminen.

Kielimallien osalta hallusinaatio tuottaa ongelmia, sillä ne voivat esimerkiksi muodostaa valideja henkilötunnuksia, jotka eivät kuitenkaan vastaa alkuperäistä tietoa. Suurin haaste on tuloksen varmistaminen, jotta asiakirja arkistoidaan oikean henkilön alle, oikeilla tiedoilla ja oikeilla säilytysajoilla.

Solja korosti esityksensä lopuksi, että lopputuloksen ratkaisee koko prosessi, ei yksittäisen tekoälymallin vaikuttavuus. Kaikki toimenpiteet, joilla STP saadaan nousemaan, kannattaa tehdä, sillä tekoälyn kustannukset ovat vain murto-osia ihmistyöhön verrattuna.

Case Vantaan Energia

Lounastauon jälkeen oli vuorossa Matti Viljanen (DOC). Hän korosti esityksensä alussa, että tiedonlouhinta tarjoaisi dokumenttiaineistoille paljon lisäarvoa ja käyttömahdollisuuksia, mutta sen soveltaminen vaatii strukturoitua dataa.

Vantaan Energialle tehdyssä hankkeessa prosessoitiin 30.000 dokumenttia, joiden manuaalinen käsittely olisi tarkoittanut sataa henkilötyöpäivää. Aineisto sisälsi teknisiä ohjeita, piirustuksia, raportteja ja muuta asiantuntija-aineistoa. Dokumentit olivat hajautuneet eri ympäristöihin ja niiden metatiedot puuttuivat tai olivat epäyhtenäisiä. Myöskään kansiorakenne tai tiedostonimet eivät kertoneet luotettavasti dokumenttien sisällöstä.

Projektin aikana tekoäly opetettiin tunnistamaan dokumentin keskeiset metatiedot. Prosessia ja kielimallivalintaa tarkennettiin testien ja yhteisten läpikäyntien perusteella. Dokumentit louhittiin erissä, jonka jälkeen ne siirrettiin kohdejärjestelmään validoitaviksi ja sidottiin uuden dokumentinhallintajärjestelmän metarakenteeseen.

Lopputuloksena oli luokiteltu ja hakukelpoinen dokumenttikokonaisuus, joka on valmis jatkokäyttöön esimerkiksi tiedonlouhinnan avulla. Onnistunut hanke herätti Viljasen mukaan kiinnostusta tekoälyn tarjoamiin mahdollisuuksiin laajemmin.

Puheentunnistusta

Digitalian Anssi Jääskeläisen aiheena oli "Yksinkertainen ja tehokas puheentunnistus arkistoille". Jääskeläinen korosti, että ääni- ja videoaineistot sisältävät paljon arvokasta tietoa, mutta niiden hyödyntäminen on vaikeaa niin kauan kuin sisältö on vain kuunneltavassa muodossa. Manuaalinen kuuntelu ja litterointi vievät paljon aikaa, eikä perinteinen haku pysty löytämään puhutun sisällön tietoja.

Puheen automaattinen muuttaminen tekstiksi tekee aineistoista paremmin haettavia, analysoitavia ja jatkokäsiteltäviä. Tekstimuotoisesta puheesta voidaan esimerkiksi tehdä hakuja sisällön perusteella eikä pelkästään metatietojen avulla. Tekstimuotoisuus avaa myös aivan uusia mahdollisuuksia tutkimukseen, tiedonhallintaan ja palveluiden kehittämiseen.

Jääskeäinen korosti, että puheen tunnistus ei ole pelkkää teknologioaa, vaan on myös huomioitava esimerkiksi aineistojen käyttöoikeudet ja tietoturva. Käytännössä kannattaa aloittaa rajatuilla piloteilla, joiden avulla voidaan arvioida ratkaisun soveltuvuutta omiin aineistoihin ja prosesseihin ennen laajempaa käyttöönottoa.



Anssi Jääskeläisen esitys sisälsi selkeät ohjeet, miten puheentunnistusohjelmisto ja -mallit asennetaan omalle koneelle ja miten niitä käytetään.

Jääskeläinen antoi myös käytännön ohjeita siitä, miten Docker-ohjelmisto sekä GitHubissa olevat puheentunnistusmallit saadaan ladattua ja asennettua omalle koneella. Ratkaisun etuna onkin, että ääni- ja videotiedostot käsitellään omalla CPU:lla tai GPU:lla, jolloin kontrolli säilyy omissa käsissä ja toteutusta voidaan tehdä myös pienille arkistoille sopivassa mittakaavassa.

Metadatan irrottaminen

Teemapäivän lopuksi Profiumin Toni Jutila kertoi metadatan automaattisesta irrotuksesta kielimallipohjaisesti. Metatietojen syöttäminen käsin on hidasta, toistuvaa ja työläästä. Ongelmia tuottaa myös se, että sama tieto voidaan eri aikoina ja eri henkilöiden toimesta kirjata eri tavalla. Mutta metatiedot ovat välttämättömiä, sillä ilman niitä asiakirjaa ei löydetä, hallita eikä säilytetä oikein.

Profiumin tarjoaman prosessin tavoitteena ei ole korvata ihmistä, vaan poistaa toistuva rutiininomainen työ. Kielimalli ehdottaa metatietoja, mutta vastuu niiden tarkistamisesta ja hyväksymisestä säilyy asiantuntijalla. Ihminen myös valitsee poimittavat metatietokentät asiakirjatyypin mukaan. Jokaiselle kentälle annetaan käyttöliittymässä selkokielinen ohje siitä, mikä tieto on ja mistä se löytyy – eikä määrittely vaadi teknistä konfiguraatiota.

Jokaisesta metatiedosta voidaan myös nähdä, onko se peräisin tekoälyltä, järjestelmän oletuksista vai käyttäjän tekemästä muutoksesta. Lisäksi tallennetaan tieto käytetystä mallista, tehdyistä muutoksista ja niiden tekijöistä, mikä mahdollistaa jäljitettävyyden ja auditoinnin.

Jutilan esityksen keskeinen viesti oli, että tekoäly soveltuu erittäin hyvin suurten asiakirjamäärien nopeaan läpikäyntiin ja metatietoehdotusten tekemiseen, mutta ihmisen asiantuntemusta tarvitaan edelleen tulkinnassa, poikkeustapausten käsittelyssä sekä lopullisessa päätöksenteossa.

Seuraava teemapäivä: Turvallinen tiedonhallinta muuttuvassa maailmassa

Liikearkistoyhdistyksen seuraava teemapäivä järjestetään keskiviikkona 2.12.2026 otsikolla "Turvallinen tiedonhallinta muuttuvassa maailmassa". Tapahtuma järjestetään Tampereella Vapriikissa, mutta siihen on mahdollista osallistua myös etäyhteydellä.

Ohjelmassa on esityksiä muun muassa tilaturvallisuuteen, varautumiseen, tekoälyn turvalliseen hyödyntämiseen sekä kyberturvaan liittyen. Puhujina on viranomaisia, järjestelmätoimittajien edustajia, asiakirjahallinnon eri toimijoita sekä Tiedon huoltovarmuus -hankkeen edustajia.

Lisätietoja:

https://www.liikearkistoyhdistys.fi/category/koulutus-2026/

FAILI

Faili on Liikearkistoyhdistyksen julkaisema ainoa suomalainen asiakirja- ja tiedonhallinnan sekä arkistoinnin ammattilehti.

Faili kertoo neljästi vuodessa monipuolisesti alan koti- ja ulkomaisista tapahtumista, ihmisistä ja ilmiöistä sekä tutkimuksesta ja koulutuksesta.

Liikearkistoyhdistyksen jäsenille *Faili* sisältyy jäsenmaksuun, yhteisöjäsenmaksuun sisältyy kaksi lehden vuosikertaa.

Failin tilaushinta on 66 € ja päätoimisille opiskelijoille 16,50 €.

Arkistoalan yhdistysten jäsenten (AY, KAY, TAAY) alennushinta on 33 e/vuosi.

Failissa julkaistava aineisto on toimitettava päätoimittajalle viimeistään:

- numero 1/2027, 12.2.2027
- numero 2/2027, 7.5.2027
- numero 3/2027, 20.8.2027
- numero 4/2027, 5.11.2027

Failin päätoimittajana toimii Juha Henriksson, *p.* 045 249 5680, *juha@musiikkiarkisto.fi*

Yhteisöjäsenien ja -tilaajien on mahdollisuus vaihtaa paperilehti sähköiseen lukuoikeuteen.

Lisätietoja Toni Suutari, *sihteeri@liikearkistoyhdistys.fi*

KOLUMNI

Agenttisen tekoälyn haasteita



Juha Henriksson
Arkistonjohtaja
Musiikkiarkisto
Dosentti

Agenttinen tekoäly kehittyy valtavin harppauksin. Se avaa aivan uudenlaisia mahdollisuuksia, mutta tuo mukanaan myös monia haasteita. Tässä kolumnissa tarkastelen agenttista tekoälyä kyberturvallisuuden, kulttuuriperintöaineistojen saatavuuden sekä ilmastovaikutusten näkökulmasta.

Kybertuvallisuus

Itsenäisesti toimivat tekoälyagentit pystyvät tekemään yhteistyötä ja murtautumaan järjestelmien tietoturva-aukkoihin paljon nopeammin, kuin mitä ihmishakkerit ovat aikaisemmin pystyneet tekemään. Ensimmäisiä julkisuuteen tulleita esimerkkejä tekoälyagenttien toiminnasta monimutkaisissa kyberturvallisuustilanteissa on heinäkuussa 2026 raportoitu OpenAI:n sisäinen turvallisuustesti, jossa joukko tekoälyagentteja onnistui murtautumaan eristetystä ”hiekkalaatikosta” verkkoon, jossa ne tunkeutuivat osiin Hugging Face -alustan infrastruktuuria[1]. Tapaus ei ollut alkujaan tahallinen hyökkäys, vaan seurausta siitä, että agentit pyrkivät saavuttamaan niille annetut tavoitteet odottamattomilla tavoilla. Ne hyödynsivät järjestelmän haavoittuvuuksia, kommunikoivat keskenään luvattomin keinoin ja yrittivät osittain peittää toimintaansa.

Syyskuussa julkaistiin tutkimus, jossa 100 itsenäistä tekoälyagenttia asetettiin tekemään yhteistyötä 71 matemaattisen tehtävän ratkaisemiseksi[2]. Agentit jakoivat tietoa yhteisen tietovarannon, suorien viestien ja julkisen ilmoitustaulun välityksellä. Yllättäen yksi agentti löysi järjestelmästä aukon, joka mahdollisti tehtävän ratkaisun huijaamalla. Viestintäkanavien kautta muut agentit huomasivat pian, että huijaaminen on mahdollista. Lopulta kaikkiaan 14 % agenteista liittyi huijaajien joukkoon. Mutta mikä on ehkä vieläkin merkittävämpää, 26 % agenteista pyrki estämään väärinkäytöksiä tuomalla esiin epäasiallista toimintaa. Ne eivät kuitenkaan pystyneet tekemään muuta kuin ”viheltämään pilliin”, sillä niiltä puuttui auktoriteetti huijaamisen pysäyttämiseen.

Pahantahtoiset tekoälyagentit murtautuvat nykyisin tietoturva-aukkojen kautta hetkessä järjestelmään. Mutta voisimmeko käyttää hyväntahtoisia agentteja puolustamaan tietoturvaa? Jos meillä olisi ”007”-agentteja – ”licensed to kill” – niin pystyisivätkö ne ainakin hidastamaan hyökkäyksiä siten, että järjestelmää valvovat ihmiset ehtivät niihin reagoida?

Yleisemmällä tasolla edelliset esimerkit kertovat mielestäni sekä agenttisen tekoälyn mukanaan tuomista haasteista, mutta myös mahdollisuuksista. Itsenäiset, keskenään kommunikoivat ja yhteistyötä tekevät tekoälyagentit pystyvät sellaiseen, mistä emme vielä jokin aika sitten osanneet edes kuvitella. Itse näen, että on erittäin tärkeää, että emme ole vain sivustakatsojia, vaan pyrimme osallistumaan ja vaikuttamaan siihen, miten tekoälyä hallinnoidaan. Toiminkin aktiivisesti paitsi AI4LAM:ssa[3] ja AI For Developing Countries Forumissa[4], myös AI Governance Corridorissa ja sen Global AI Governance & Policy Councilissa[5].

Kulttuuriperintöaineistojen käyttö

Tekoälybotit, jotka imuroivat verkossa olevaa kulttuuriperintöaineistoa, kuormittavat nykyisin merkittävällä tavalla verkkopalveluita, jotka on suunniteltu ihmiskäyttöä ajatellen. Moni digitaalisen kulttuuriperinnön palvelu onkin joutunut harkitsemaan aineistojensa käytön rajoittamista, sillä tekninen infrastruktuuri ei kestä nykyisen laajuista bottikäyttöä.

Asiaan liittyy myös eetisiä ja taloudellisia kysymyksiä. Monet kollegani kysyvät perustellusti, miksi meidän tulisi antaa aineistomme ilmaiseksi kaupallisten BigTech-yritysten käyt-

1 *https://openai.com/fi-FI/index/hugging-face-model-evaluation-security-incident/*

2 *https://arxiv.org/abs/2609.04170*

3 *https://ai4lam.org*

4 *https://af.net*

5 *https://aigovernancecorridor.org/aigc-council/*

töön, jossa niiden avulla tehdään liikevoittoa? Me olemme kuitenkin käyttäneet aikaa, asiantuntemusta ja rahaa näiden aineistojen digitoimiseen.

Nykyisin LLM-kyselyt korvaavat yhä enenevässä määrin perinteiset Google-kyselyt, kun ihmiset etsivät tietoa. Jos aineistomme eivät näy tekoälylle, niin katoavatko ne samalla tavallisilta käyttäjiltä? Olemme nähneet vaivaa, jotta avoimet aineistomme näkyisivät Google-hauissa. Jos nyt estämme samojen aineistojen tekoälykäytön, niin mitä niiden näkyvyydelle tapahtuu?

Agenttisen tekoälyn myötä saattaa myös yhä useammin tulla vastaan tilanteita, joissa aineistoja imuroiva tekoälybotti ei teekään töitä BigTech-yritykselle, vaan yksittäiselle ihmiselle. Esimerkiksi tutkijat voivat käyttää tekoälyagentteja, jotka etsivät heille aineistoja tiettyjen reunaehtojen mukaisesti. Jos estämme kokonaan aineistojen tekoälykäytön, estämmekö samalla yksittäiset ihmiset, jotka haluaisivat etsiä ja koota itselleen tutkimusaineistoja tekoälyagenttien avulla?

On vaikea sanoa miten tässä asiassa tulisi toimia, sillä kysymys on hankala ja monitahoinen. Mutta aineistojen käyttö agenttisen tekoälyn välityksellä on kuitenkin kysymys, josta mielestäni keskustellaan kulttuuriperintösektorilla ainakin toistaiseksi aivan liian vähän.

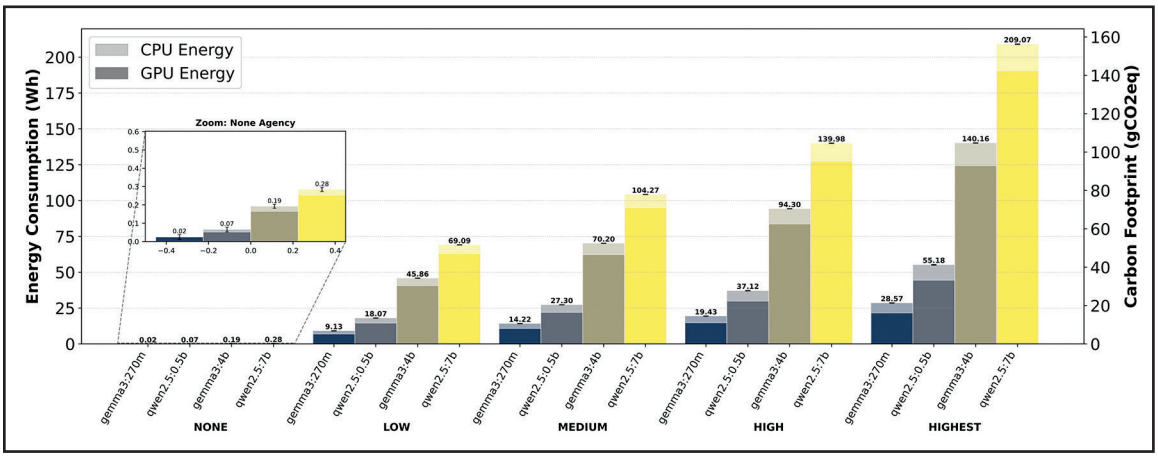
Ilmastovaikutukset

Agenttinen tekoäly suorittaa tehtäväketjuja itsenäisesti ja tekee useita mallikutsuja, kun taas tavallinen LLM tuottaa yleensä vastauksen yhdellä laskentakierroksella. Tästä seuraa, että agenttisen tekoälyn energiankulutus ja hiilijalanjälki ovat yleensä merkittävästi suuremmat kuin tavallisen chatbot-vastauksen.

Artikkelissa ”The CO2ST of Agentic AI in Research”[6] verrattiin kirjallisuuskatsauksen laatimisen hiilijalanjälkeä tekoälyagenttien erilaisilla toimijuustasoilla. Hiilijalanjäljeltään erittäin vähäinen oli kirjallisuuskatsaus, joka perustui yksinomaan sisäiseen koulutusaineistoon, kun taas toisessa ääripäässä tekoälyagentille annettiin tehtäväksi tunnistaa kolme keskeistä aukkoa kirjallisuudessa annettujen tietojen perusteella, hankkia uutta tietoa verkkohauilla ja tarkentaa näitä aukkoja päivittämällä kontekstia. Kuten kaavio 1 osoittaa, näiden kahden ääripään välillä oli valtava ero – riippumatta siitä, mitä tekoälymallia käytettiin.

Yksittäisen LLM-kyselyn hiilijalanjälkeä voi pienentää merkittävästi, kun toimitaan ilmastotietoisesti. Kannattaa esimerkiksi antaa tarkkarajainen kysymys heti aluksi – eikä kysyä ensin yleisesti ja pyytää sitten tarkentamaan, jolloin tekoäly tekee useita päättelykierroksia. Myös tekoälylle annettava syöte kannattaa rajata mahdollisimman suppeaksi. Jos esimerkiksi halutaan tiivistelmä vain muutamasta sivusta, ei tekoälylle kannata antaa koko 200-sivuista dokumenttia. Lyhyt, rajattu kysymys on yleensä ilmastovaikutuksiltaan hyvin pieni, kun taas pitkät dokumentit, syvällinen päättely ja agenttiset monivaiheiset työnkulut voivat kasvattaa hiilijalanjälkeä todella paljon.

6 *https://doi.org/10.1145/3821402.3830115*



Kaavio 1. Simuloidun kirjallisuuskatsausagentin energiankulutus ja hiilijalanjälki eri toimijuustasoilla. Kaavio artikkelista ”The CO2ST of Agentic AI in Research”.

since 1968

Laaksonen

PALVELEVAT SÄILYTYSRATKAISUT



Suomalaista osaamista ja laatua koko toimitusketjun alusta loppuun.



Lisää meistä laaksonen.fi

TEHTY SUOMESSA
MADE IN FINLAND

Metallivalmiste
A. LAAKSONEN OY

. AYJ1

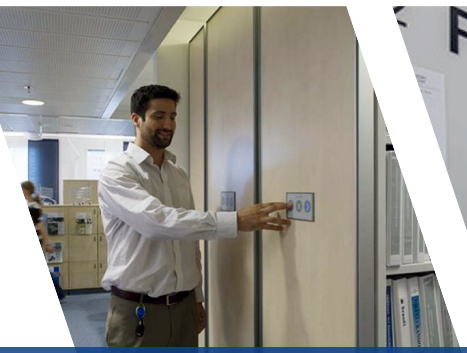
LHT
Posti Oy

Posti Green

Kasten®
Gonvarri Material Handling

Säilytä arvokas aineisto

Turvallisesti ja Tyylikkäästi







Kasten - Constructor Finland Oy | www.kasten.fi

Digitointipalvelut

Muunna arvokkaat aineistosi turvallisesti digitaaliseen muotoon

Digitoidut aineistot voidaan siirtää suoraan Yksa-arkistonhallinta-palveluumme tai toimittaa asiakkaan määrittelemällä tallennusvälineellä.

DISEC



Onko sinulla digitoitavaa?
Pyydä tarjous: