U 18/2024
Yhteiskuntasuhdejohtaja Patrik Gayer
Silo AI
TaV 23.4.2024

Asia: U 18/2024 vp
Valtioneuvoston kirjelmä eduskunnalle komission ehdotuksesta neuvoston EuroHPC-asetuksen muuttamisesta

Johdanto

Silo AI on Euroopan suurin yksityinen tekoälylaboratorio, jonka missiona on rakentaa eurooppalaista tekoälyn lippulaivaa. Uskomme, että tekoäly synnyttää arvoa digitaalisten tuotteiden ja palveluiden kautta. Tekoälylaboratoriona olemme kumppani, joka tuo tekoälyn avulla kilpailuetua tuotteisiin. Tarjoamme tekoälymalleja, työkaluja ja palveluita, joiden avulla autamme asiakkaitamme luomaan kilpailuetua. Osana tätä missiota ja tarjoamaa, Silo AI rakentaa globaalissa vertailussa suorituskykyisiä avoimeen lähdekoodiin perustuvia monikielisiä suuria kielimalleja (LLM), ja laajemmin pohjamalleja (*engl.* foundation models), edistääksemme Euroopan digitaalista suvereniteettiä ja demokratisoidaksemme kielimallien saatavuutta.

Tekoäly, ja erityisesti generatiivinen tekoäly, on jo nyt laajasti taloutta ja yhteiskuntaa mullistava voima, jonka strateginen vaikutus ulottuu merkittävästi lyhyen aikavälin tehokkuushyötyjä pidemmälle. Suomen ja Euroopan kilpailukyky tekoälyn osalta ei ole vain taloudellinen kysymys, vaan sillä on myös laajoja geopoliittisia ja turvallisuuspoliittisia ulottuvuuksia ja seurauksia. Generatiiviset tekoälymallit, ja etenkin pohjamallit, voidaan mieltää taloudellisen innovaation ja tuotekehityksen ja sitä kautta arvonluonnin infrastruktuuriksi. Generatiivinen tekoäly on siis keskeisessä roolissa pyrkimyksessä vahvistaa Euroopan strategista kilpailukykyä.

Silo AI on kouluttanut laajoja kielimalleja (LLM) EuroHPC JU:hun kuuluvalla LUMI-supertietokoneella kesästä 2023 alkaen ja ollut mukana Turun Yliopiston kanssa pilotoimassa LUMI:a alusta alkaen.

Silo AI:n näkemykset U-kirjelmästä

Kokemuksemme mukaan EuroHPC-infrastruktuurilla on potentiaalia luoda merkittävää arvia, mutta se ei sellaisenaan sovellu optimaalisesti laajojen generatiivisen tekoälymallien kouluttamiseen. Osittain siksikin kannatamme ehdotuksen tavoitetta siitä, että EuroHPC -yhteisyritys voi nykyistä paremmin tarjota eurooppalaisten supertietokoneiden laskentakapasiteettia ja palveluita tekoälytutkimusta varten niin julkisille tutkimusorganisaatioille kuin yrityksille.

Tekoälytehtaiden todellinen toiminta on tässä vaiheessa vielä epäselvää. Tekoälytehtaiden rajaaminen pelkästään startup- ja PK-yrityksiin olisi kuitenkin haitallista Euroopan strategisen

kilpailukyvyn näkökulmasta, koska silloin suhteellisen matalan PK-yritysmääritelmien kynnysarvojen ylittävä yritys ei voisi hyötyä järjestelystä.

LLM:t ja muut generatiiviset tekoälymallit koostuvat pohjimmiltaan kolmesta pääkomponentista: datasta, joka toimii mallin koulutuksen perustana, laskennasta, joka tarjoaa mahdollistaa mallien kehityksen ja koulutuksen, ja algoritmeista, jotka määrittävät, miten dataa käsitellään ja tulkitaan. Laskentakapasiteetti on näistä kustannuksiltaan ylivoimaisesti merkittävin komponentti. Esimerkiksi OpenAI:n ChatGPT-palvelun taustalla toimivan GPT-4 mallin yhden sukupolven koulutus on arvioitu maksaneen noin 100 miljoonaa dollaria. Kokonaiskustannus lienee tällä hetkellä jo monta suuruusluokkaa korkeampi.

Käytännössä jokainen yritys joka kouluttaa kilpailukykyisiä malleja on ylittänyt tai tulee pian ylittämään startup- ja PK-yritysmääritelmien kynnysarvot. Mikäli EuroHPC-suurteholaskentakapasiteetti ohjataan vain startup- ja PK-yrityksille, ei poisteta pullonkaulaa scale-up vaiheen, laajoja generatiivisen tekoälyn malleja kouluttavien yritysten tieltä eikä heidän asemaa kansainvälisessä kilpailussa tueta.

LUMI 2-investointi

On kuitenkin tärkeää mainita, että EuroHPC-supertietokoneiden laskentakapasiteetin parempi soveltuvuus yrityskäyttöön ei tapahdu pelkästään tämän lausunnon kohteena olevan asetuksen muutoksella. Suomessa toimivaa LUMI-supertietokonetta hallinnoi CSC.

Suurteholaskennan työkuorman kysyntä on muuttumassa siten, että se sisältää yhä enemmän tehtäviä, joilla on korkeat käyttöaikavaatimukset. Toisin kuin perinteisen laskennan työmäärät, jotka eivät ole yhtä herkkiä suoritusviiveille, suurten kielimallien kouluttamisen kaltaiset työt ovat haastavia, koska ne voivat skaalautua vertikaalisesti vain rajoitetusti tehokkuuden kärsimättä ja koska niissä on käytettävä laskentatehoa tasaisesti ja tehokkaasti kuukausia kerrallaan projektin tavoitteiden saavuttamiseksi.

Suomen ja Euroopan kilpailukyvyn ja strategisen kilpailukyvyn näkökulmasta on elintärkeää, että Talouspoliittisessa ministerivaliokunnassa linjatun, uuden supertietokonehankinnan suunnittelun lähtökohdaksi otetaan generatiivinen tekoäly ja sen kehittäminen globaalin kilpailun edellyttämin ehdoin. Käytännössä tämä tarkoittaa esimerkiksi jonotuspolitiikan kehittämistä, jotta koneella voi suorittaa tehtäviä, joissa hyödynnetään monia solmuja ja pitkään, sekä supertietokoneen käyttösääntöjen, asiakaspalvelun ja muun hallinnon optimointia, jotta LUMI:lla voi kouluttaa kansainvälisen huipputason malleja. Näin toimiva LUMI voi poistaa pullonkaulan resurssi-intensiivisestä tekoälyn kehitystyöstä ja tarjota merkittävää vipuvartta tekoälyn kehitystyölle ja kaupallistamiselle Suomessa.

Patrik Gayer
Yhteiskuntasuhdejohtaja, Silo AI
patrik.gayer@silo.ai
+358 40 722 6644