Linkki tulevaisuusvaliokunnan RTI-sivulle: Radical Technology Inquirer (RTI) | Eduskunta

RTI-GPT:n kanssa käyty keskustelu Ennakointifoorumissa 19.8.2026

Linkki keskusteluun: https://chatgpt.com/share/6a859e0a-f0c4-83ed-9ecc-94921a339dab

Kopio keskustelusta:

Kysymys RTI-GPT:lle: Yleisöä kiinnostaa AI agentit ja uusi työnjako. Tunnista RTI-aineistosta siihen parhaiten liittyvät ARTit. Mitä konkreettisia heikkoja signaaleita kannattaisi seurata, jos haluamme tunnistaa ajoissa, onko tämä murros todella etenemässä? Erottele ainakin teknologian kypsyminen, käyttöönotto sekä työn ja osaamisen muutoksesta kertovat signaalit.

Johtopäätös

RTI-aineistossa **AI-agenttien ja uuden työnjaon selvästi tärkein ART-ankkuri on ART 03.019 Collaborative Multi-Agent AI**. Sen ydin ei ole yksittäinen chatbot tai assistentti, vaan useiden ohjelmisto- tai robottiagenttien kyky jakaa tehtäviä, neuvotella rooleista, ylläpitää yhteisiä tilannekuvia ja toimia ihmisen asettamien tavoitteiden puitteissa. Group3 arvioi teknologian kypsyydeksi **6/7**, eli ensimmäisiä tuotteita ja todellisia käyttöönottoja on jo nähtävissä, vaikka laajamittainen, luotettava agenttien koordinointi ei vielä ole kaikkialla vakiintunutta.

Työn murroksen kannalta sitä kannattaa tarkastella yhdessä kolmen muun ARTin kanssa:

- **ART 03.019 Collaborative Multi-Agent AI** – varsinainen *agenttityönjako*: koneet jakavat, delegoivat ja koordinoivat tehtäviä keskenään.
- **ART 03.018 Automated Machine Learning (AutoML)** – *agenttien tuotantokoneisto*: aineisto kuvaa kypsää versiota suorastaan “model and agent factoryksi”, joka voi tuottaa tehtäväkohtaisia malleja ja agentteja huomattavasti nykyistä pienemmällä asiantuntijapanoksella.
- **ART 03.017 Trustworthy AI Systems** – *delegoinnin mahdollistava luottamuskerros*: valvonta, audit trailit, fallbackit ja ihmisen eskalointi ratkaisevat, kuinka pitkälle organisaatiot uskaltavat antaa agenteille oikeuksia. GVN 10:ssä Automation of Work tämän vaikutus on matriisissa maksimitasolla 20.
- **ART 04.026 AI Understanding of Human Intent & Context** – *ihmisen ja agentin työnjaon käyttöliittymä*: tekoäly päättelee käyttäjän tavoitteita ja kontekstia sen sijaan, että ihminen joutuisi määrittelemään jokaisen vaiheen eksplisiittisesti. Tämän ARTin kypsyys on aineistossa vielä alempi, **4/7, proof-of-concept demonstrated**.

Matriisi vahvistaa priorisoinnin. **ART 03.019 saa GVN 10 Automation of Work -vaikutukseksi 20 ja GVN 11 Work and Income -vaikutukseksi 10**. ART 03.018 saa vastaavasti 10 ja 20; ART 03.017 20 ja 10. Kyse ei siis aineiston logiikassa ole vain uuden AI-tuotteen tulosta vaan potentiaalisesti sekä työn automatisoinnin että työmarkkinoiden organisointitavan muutoksesta.

Mitä heikkoja signaaleja kannattaa seurata?

Tärkeää olisi olla mittaamatta ensisijaisesti sitä, **kuinka moni käyttää AI:ta**, vaan sitä, **muuttuuko delegoinnin yksikkö ihmiseltä koneelle**. Varhainen murros näkyisi esimerkiksi siinä, että työntekijä ei enää itse suorita kymmentä digitaalista vaihetta, vaan määrittelee tavoitteen ja valvoo agenttijoukkoa, joka jakaa välivaiheet keskenään.

1. Teknologian kypsyminen

Näistä saadaan aikaisin selville, onko agenttisuus muuttumassa demosta käyttökelpoiseksi infrastruktuuriksi.

Agentit alkavat ratkaista pitkiä työnkulkuja ilman jatkuvaa uudelleenohjausta. Seuratkaa erityisesti tapauksia, joissa agenttijärjestelmä ei vain vastaa kysymykseen vaan pitää tehtävän tilaa yllä, jakaa sen osatehtäviksi, käyttää useita järjestelmiä ja korjaa suunnitelmaa epäonnistumisen jälkeen. RTI:n ART 03.019:n ydinkyvykkyyksiä ovat juuri dynaaminen tehtävien jako, roolien neuvottelu ja yhteinen tilannekuva.

Monen agentin järjestelmille syntyy standardoituja yhteisiä rajapintoja. Erityisen vahva signaali olisi se, että agentit eri toimittajien ohjelmistoissa alkavat vaihtaa tehtäviä, tilatietoa ja oikeuksia ilman räätälöityä integraatiota. RTI nostaa eri GVN-yhteyksissä esiin yhteiset task boardit, agenttien väliset viestistandardit ja orchestration-platformit.

Poikkeustilanteiden käsittely paranee nopeasti. Agentin tavallinen onnistumisprosentti ei yksin riitä. Seuratkaa, kykeneekö järjestelmä tunnistamaan epävarmuuden, palaamaan turvalliseen tilaan, siirtämään työn toiselle agentille tai pyytämään ihmistä mukaan. Trustworthy AI -aineistossa juuri fallbackit, incident replay, provenance ja ihmisen eskalointipolut ovat käyttöönottoa avaavia tekijöitä.

Agenttien tekeminen automatisoituu. Erittäin kiinnostava signaali olisi, että yritys pystyy tuottamaan uuden tehtäväkohtaisen agentin omista prosessilokeistaan ilman varsinaista ML-projektia. RTI:n heikko signaali ART 03.018:ssa on suoraan: *firms generate task-specific decision agents from logs*.

Agenttien välisten virheiden käsittelystä tulee oma insinööriala. Tämä on hieman paradoksaalinen mutta vahva kypsymissignaali. Kun incident-raporteissa puhutaan agenttien välisestä virheketjusta eikä vain “malli hallusinoi”, teknologia on jo riittävän syvällä prosesseissa aiheuttaakseen järjestelmätason ongelmia. RTI nostaa juuri agent-to-agent failure cascades esiin.

2. Käyttöönotto organisaatioissa

Tässä pitäisi etsiä siirtymää **AI-avusteisesta ihmisestä agenttivetoiseen prosessiin**.

Yritykset nimeävät ensimmäisiä “agent fleet supervisor” -rooleja. Tämä on RTI-aineiston ehkä kaikkein kiinnostavin yksittäinen heikko signaali: yritys palkkaa yhden ihmisen valvomaan agenttijoukkoa työssä, jonka aiemmin teki ihmistiimi.

Yhden ihmisen valvontajänne kasvaa. GVN 10:n signaali on, että *one supervisor covers more machines under assurance dashboards*. Digitaalisessa työssä vastaava olisi esimerkiksi yksi palvelupäällikkö, joka valvoo kymmeniä myynti-, laskutus-, analyysi- tai asiakaspalveluagenteja.

Pienet yritykset alkavat toimia aikaisemmin vain suurelle yritykselle mahdollisella kapasiteetilla. RTI nostaa suoraan signaaliksi sen, että erittäin pieni yritys operoi multi-agent-palvelua. Tämä on erityisen kiinnostava signaali, koska silloin agentit eivät vain säästä kustannuksia vaan **muuttavat yrityksen optimaalista kokoa ja organisaatiorakennetta**.

Ohjelmistojen hinnoittelu siirtyy käyttäjäpaikoista tehtäviin tai tuloksiin. RTI:n signaali on siirtyminen *seat pricingistä task pricingiin*. Tämä olisi vahva ulkoinen markkinasignaali: ohjelmistotoimittaja ei enää oleta, että jokaisen lisenssin takana istuu ihminen.

Agenttien välisistä API-rajapinnoista tulee palvelumarkkinan perusrakenne. Palvelualustat alkavat tarjota agent-to-agent-rajapintoja tehtävien vaihtamiseen. Tämä kertoo, että agentit eivät ole enää vain käyttöliittymäominaisuus vaan taloudellisia toimijoita prosessien sisällä.

Pilottien aika lyhenee kuukausista viikkoihin. AutoML:n yhteydessä RTI pitää tätä signaalina siitä, että uuden automaation rakentamisen oma työpanos romahtaa. Silloin myös agenttien diffuusio yritysten välillä voi muuttua olennaisesti nopeammaksi.

Autonomialle syntyy pysyviä hyväksyntöjä pilottilupien sijasta. RTI:n Trustworthy AI -signaaleissa erityisen tärkeä institutionaalinen käännekohta on siirtymä yksittäisistä piloteista jatkuviin lupiin ja vakiintuneisiin auditointikäytäntöihin.

3. Työn ja osaamisen muutos

Juuri tästä näkisin kaikkein aikaisimmat yhteiskunnallisesti kiinnostavat merkit.

Työtehtävien nimet alkavat muuttua tekemisestä orkestrointiin. Esimerkiksi “analyytikon” rinnalle tulee agent supervisor, AI operations, workflow orchestrator tai vastaava rooli. RTI:n aineistossa työn painopiste siirtyy toistuvasti suorittamisesta valvontaan, poikkeusten käsittelyyn ja agenttien koordinointiin.

Koulutuksessa painotus siirtyy mallien rakentamisesta niiden valvontaan. ART 03.018:n GVN-vaikutuksissa RTI antaa signaaliksi, että työvoimaohjelmat siirtyvät ohjelmointitaidoista

model oversight skills -osaamiseen. Se olisi käytännössä erittäin hyvä seurattava indikaattori ammattikorkeakoulujen, yrityskoulutusten ja täydennyskoulutuksen sisällöissä.

Poikkeustilanteiden ratkaisemisesta tulee arvokkaampi taito kuin normaaliprosessin suorittamisesta. Tätä aineisto ei ilmaise yhtenä indeksinä, mutta se seuraa varsin suoraan sen kuvaamasta työnjaosta: agentit hoitavat rutiinin, ihminen ottaa eskalaatiot, epäselvät tavoitteet, vastuun ja neuvottelun. Tämä on siis **aineistosta tehtävä varovainen päätelmä**, ei RTI:n valmis pistemittari.

Junioritehtävien rakenne alkaa muuttua. Tätä kannattaa seurata erityisen herkästi: vähenevätkö työpaikat, joiden perinteinen tehtävä on ollut tiedonhaku, ensimmäisen luonnoksen tekeminen, dokumentointi, luokittelu tai prosessin läpivienti? RTI ei anna tästä yhtä valmista “junior jobs” -signaalia, joten tätä on syytä käsitellä johdettuna seurantahypoteesina, ei aineiston suorana väitteenä.

Työehtosopimuksiin tai yhteistoimintaan ilmestyvät agenttien delegointisäännöt. RTI:ssa jo nyt näkyvät signaalit, joissa työntekijäjärjestöt neuvottelevat läpinäkyvistä eskalointisäännöistä ja joissa automaation käyttöönotosta sekä hyötyjen jakamisesta tulee neuvottelukysymys.

Työn jakamisesta tulee vaikeammin näkyvää työntekijälle. RTI nostaa negatiiviseksi heikoksi signaaliksi sen, että tehtävien allokaatio muuttuu dynaamisemmaksi ja läpinäkymättömämmäksi. Se olisi merkki siitä, että työnjohtofunktio on siirtymässä ihmiseltä orkestrointialgoritmille.

Viiden signaalin “varhainen varoitustaulu”

Jos yleisölle pitäisi rakentaa erittäin kompakti seuranta, ottaisin nämä viisi:

1. **Kuinka monta ihmistä yksi työntekijä korvaa valvomalla agenttijoukkoa?** Seuratkaa todellista supervisor-to-agent-suhdetta.
2. **Syntyykö yrityksiin pysyviä “agent fleet / AI operations” -rooleja?**
3. **Siirtyvätkö ohjelmistot seat-hinnoittelusta tehtävä-, transaktio- tai tuloshinnoitteluun?**
4. **Pystyvätkö hyvin pienet yritykset pyörittämään aiemmin kymmenien ihmisten vaatimia palveluprosesseja?**
5. **Muuttuuko koulutus “AI:n käyttämisestä” agenttien tavoitteiden asettamiseen, laadunvarmistukseen, auditointiin ja poikkeusten ratkaisuun?**

Yksittäinen näyttävä agenttidemo ei vielä olisi minusta RTI-aineiston perusteella ratkaiseva merkki murroksesta. **Paljon vahvempi näyttö syntyy, jos tekninen autonomia, organisaatioiden todellinen delegointi ja työn rakenteen muutos alkavat näkyä**

samanaikaisesti. Silloin kyse ei enää olisi vain paremmasta käyttöliittymästä tekoälyyn, vaan uuden työnjaon muodostumisesta ihmisten, yksittäisten agenttien ja agenttijoukkojen välille.

Aineistolähteet: KNOWLEDGE_MAP.txt; 100ART_Descriptions.docx; arts_signal_scores_matrix.xlsx; Group3.docx; Group4.docx; GVN_ART_Impact.docx. ART 03.017 *Trustworthy AI Systems*, ART 03.018 *Automated Machine Learning (AutoML)*, ART 03.019 *Collaborative Multi-Agent AI*, ART 04.026 *AI Understanding of Human Intent & Context*; erityisesti GVN 10 *Automation of Work* ja GVN 11 *Work and Income*.

Milloin AI alkaa korjata ihmisen ongelmanmäärittelyitä pikemminkin kuin ottaa ihmiseltä valmiita tehtävänantoja? Mitä tulisi seurata sen hahmottamiseksi onko näin tapahtumassa?

Johtopäätös

RTI-aineiston perusteella tätä rajaa ei ole vielä ylitetty yleisesti. Lähin ART-ankkuri on **ART 04.026 AI Understanding of Human Intent & Context**, jonka maturiteetti on vasta **4/7: Proof-of-Concept demonstrated**. Sen visio on kuitenkin juuri siirtymä pois kirjaimellisista komennoista kohti käyttäjän tavoitteiden, tilanteen ja tarkoituksen tulkintaa. Aineisto kuvaa myös järjestelmiä, jotka ennakoivat käyttäjän tarpeita ja toimivat proaktiivisesti ilman eksplisiittistä pyyntöä.

Mutta **ihmisen ongelmanmäärittelyn korjaaminen on vielä tätä astetta vaativampi kyvykkyys**. Siinä AI:n pitäisi pystyä sanomaan käytännössä:

“Pyydät minua ratkaisemaan X:n, mutta havaintojeni perusteella X ei ehkä ole varsinainen ongelma. Todennäköisempi tavoite on Y, ja meidän pitäisi ensin selvittää Z.”

RTI ei nimeä tällaista kyvykkyyttä omaksi ARTikseen. Se syntyy aineiston perusteella lähinnä **ART 04.026:n intentin ymmärtämisen ja ART 03.015 Causal Learning & Reasoningin yhdistelmästä**. Jälkimmäinen pyrkii siirtymään korrelaatioista eksplisiittiseen syy-seurauspäättelyyn ja vastaamaan kysymyksiin “miksi?”, “mitä jos?” ja “mitä tapahtuu, jos muutamme X:ää?”.

Milloin sanoisin, että muutos on todella tapahtumassa?

En käyttäisi rajana sitä, että AI osaa kysyä hyvän tarkentavan kysymyksen. Se voidaan tehdä melko pinnallisestikin. Vahvempi kriteeri olisi, että AI tekee **toistuvasti neljä asiaa oikein**:

1. se erottaa käyttäjän **lausutun tehtävän** tämän **todennäköisestä tavoitteesta**;
2. se tunnistaa, että annettu ongelmanrajaus voi olla väärä;
3. se tuottaa vaihtoehtoisen ongelmahypoteesin ja perustelee sen havaittavilla syy-seuraussuhteilla;
4. käyttäjä tai organisaatio hyväksyy uuden ongelmanmäärittelyn riittävän usein, koska se tuottaa paremman lopputuloksen.

Tuo neljäs kohta on ratkaiseva. **Teknologinen kyvykkyys muuttuu työnjaoksi vasta silloin, kun ihmiset alkavat antaa AI:lle myös oikeuden haastaa kysymys, eivät vain vastata siihen.**

Mitä pitäisi seurata?

Seuraisin ennen kaikkea seuraavia signaaleja.

1. AI alkaa ehdottaa "älä tee tätä" -ratkaisuja. Varhainen vahva signaali on se, että liiketoiminta-, suunnittelu-, terveys-, tutkimus- tai operatiivisissa järjestelmissä AI ei vain optimoi annettua vaihtoehtoa vaan ehdottaa toisen intervention. ART 03.015:n aineistossa tätä kohti menevät jo esimerkiksi järjestelmät, jotka näyttävät "best intervention" pelkän poikkeamahälytyksen sijasta ja testaavat automaattisesti what-if-toimenpiteitä.

2. Tarkentavat kysymykset muuttuvat oletusten haastamiseksi. Heikko signaali ei ole "mitä tarkoitit?", vaan esimerkiksi "miksi oletat, että toimitusviive johtuu kapasiteetista?" tai "onko tavoite todella vähentää käsittelyaikaa vai vähentää asiakkaan odotettua epävarmuutta?". Tämä olisi merkki siitä, että intentin ymmärtäminen alkaa yhdistyä kausaaliseen malliin eikä jää kielelliseksi disambiguoinniksi.

3. Agentti käyttää muuta kontekstia kuin käyttäjän promptia ongelman uudelleenrajaamiseen. ART 04.026 perustuu jatkuvaan kontekstimalliin: kuka tekee mitä, mikä käyttäjän huomion kohde on ja mikä hänen todennäköinen tavoitteensa on. Aineiston nykyiset esimerkit ovat vielä suhteellisen rajattuja, kuten katseen ja eleiden käyttö epäselvän pyynnön tulkitsemiseen. Vahvempi signaali olisi, että työagentti yhdistää prosessihistorian, organisaation tavoitteet, käyttäjän aiemmat päätökset ja tilanteen ja sanoo niiden perusteella, että annettu tehtävä on väärin muotoiltu.

4. AI:lta aletaan pyytää tavoitteita eikä tehtäviä. Käyttöliittymässä tämä olisi hyvin näkyvä muutos. Nykyisen "laadi raportti / analysoi nämä luvut / tee kampanja" -mallin sijasta käyttäjä antaisi esimerkiksi "pienennä asiakkaiden poistumaa" tai "selvitä miksi tuotanto ei saavuta tavoitetta". AI päättäisi itse, mitä kysymyksiä kannattaa ratkaista. Tämä olisi yksi parhaista työnjaon muutoksen indikaattoreista.

5. Organisaatiot mittaavat AI:n kykyä löytää väärä ongelma. Kun benchmarkit ja hankintakriteerit alkavat mitata "problem reframing", "goal inference", "root-cause identification" tai vastaavaa eikä vain tehtävän suorittamista, olemme lähellä merkittävää käännettä. ART 03.015:n nykyiset maturiteettikriteerit painottavat jo tunnetun ground truthin kausaalista rakennetta, intervention arviointia ja root-cause-analyysiä.

6. Ihmisten rooli muuttuu ongelmanmäärittelijästä ongelmanmäärittelyn hyväksyjäksi. Tämä olisi ehkä tärkein työelämän signaali. Jos esimerkiksi konsultti, analyytikko, lääkäri, insinööri tai johtaja alkaa saada AI:lta 2–3 vaihtoehtoista tulkintaa "mikä tässä oikeastaan on ongelma" ja hänen tehtäväkseen jää valita, rajata ja ottaa vastuu, työnjako on muuttunut olennaisesti.

7. Virheitä aletaan käsitellä “vääränä tavoitteena”, ei vain vääränä vastauksena. Kypsissä agenttijärjestelmissä kiinnostava incident-luokka olisi esimerkiksi: AI optimoi väärää tavoitetta, tulkitsi käyttäjän intentin väärin tai rakensi väärän kausaalimallin. Se kertoo samalla, että ongelmanmäärittelyä on jo delegoitu koneelle. RTI korostaa muutenkin intenttijärjestelmissä käyttäjän mahdollisuutta tarkistaa ja ohittaa AI:n päätelmät sekä kausaalisissa järjestelmissä oletusten ja malliversioiden auditoitavuutta.

Kolme porrasta, joita seuraisin

Tämän voisi tiivistää näin:

Taso 1 – AI suorittaa:
“Ihminen sanoo mitä tehdään → AI tekee.”

Taso 2 – AI tulkitsee:
“Ihminen sanoo jotakin → AI päättelee, mitä hän oikeastaan tarkoittaa.”

RTI:n **ART 04.026** on selvästi menossa tähän suuntaan, mutta yleinen versio on vielä PoC-vaiheessa.

Taso 3 – AI problematisoi:
“Ihminen määrittelee ongelman → AI arvioi, onko ongelma oikein määritelty, ja ehdottaa parempaa määrittelyä.”

Tämä kolmas taso **ei ole RTI-aineistossa vielä valmiiksi kypsänä ARTina**, vaan se on varovainen päätelmä ART 04.026:n ja ART 03.015:n konvergenssista.

Ehkä kaikkein käyttökelpoisin yksittäinen indikaattori olisi siksi:

Kuinka usein organisaatioissa toteutetaan AI:n ehdottama ongelman uudelleenmäärittely sen sijaan, että AI vain toteuttaisi ihmisen alkuperäisen tehtävänannon?

Kun tästä tulee mitattava ja toistuva käytäntö esimerkiksi konsultoinnissa, tutkimuksessa, johtamisessa, tuotekehityksessä tai operatiivisessa päätöksenteossa, sanoisin että murros on todella käynnissä.

Aineistolähteet: KNOWLEDGE_MAP.txt; 100ART_Descriptions.docx; Group3.docx; Group4.docx; GVN_ART_Impact.docx; arts_signal_scores_matrix.xlsx. ART 04.026 *AI Understanding of Human Intent & Context*; ART 03.015 *Causal Learning & Reasoning*.