Ja ihminen loi älyn -podcast

Jakso 1: Tekoäly muuttaa ihmisyyden ja yhteiskunnan − Harri Valpola

Arno Solin: Mikä se visio silloin oli?

Harri Valpola: Rakennetaan yleinen tekoäly – six easy steps to AGI. Tämä muutos tulee tapahtumaan. Sitä ei voi pysäyttää. Niiden hyötyjen pitää jakautua kaikille ihmisille, tai muuten tämä menee todella huolella vihkoon. Tekoälyt saavat sitten määritellä mitä se tekoälykkyys on. Mutta minä kuitenkin vielä luotan siihen, että ihmiset saavat määritellä tämän ihmisyytensä.

Eric Malmi: Tervetuloa Aalto-yliopiston "Ja ihminen loi älyn" -podcastin ensimmäisen jakson pariin. Tekoälystä puhutaan nykyään kaikkialla. Mutta miten tekoäly oikeastaan toimii, ja mikä on ollut Suomen ja suomalaisten rooli tekoälyn kehityksessä? Näihin kysymyksiin haetaan vastauksia tässä podcastissa. Lisäksi päästään tutustumaan suomalaisiin tekoälyalan pioneereihin ja keskustellaan ajankohtaisista tekoälyyn liittyvistä aiheista.

Eric Malmi: Minun nimeni on Eric Malmi. Toimin vierailevana professorina Aalto-yliopistossa ja Suomen ELLIS-instituutissa, jossa tutkimukseni keskittyy suuriin kielimalleihin. Kanssani tätä podcastia juontaa..

Arno Solin: Arno Solin. Toimin niin ikään Aalto-yliopistossa ja ELLIS-instituutissa. Minun tittelini on koneoppimisen professori. Oma tutkimukseni liittyy yleisemmin generatiivisiin malleihin, epävarmuuksien mallintamiseen ja tekoälyn rakennuspalikoiden luomiseen.

Arno Solin: Tänään meillä on ilo saada vieraaksi tekniikan tohtori Harri Valpola. Harri on toiminut koneoppimisen ja neurotieteen risteyskohdassa useiden vuosikymmenten ajan, ja hänet tunnetaan yhtenä Suomen merkittävimmistä tekoälyvisionääreistä. Harri tuntuu aina olevan jotenkin askeleen edellä meitä muita ja hän on toiminut perustajana useissa mielenkiintoisissa, mullistavissakin tekoälyalan yrityksissä, kuten ZenRobotics ja Curious AI. Sen lisäksi hän on työskennellyt Applella Sveitsissä. Harri, lämpimästi tervetuloa!

Harri Valpola: Kiitos.

Arno Solin: Tässä jaksossa puhutaan Harrin kanssa sellaisista käsitteistä kuin yleinen tekoäly ja singulariteetti. Ennen kuin mennään kysymyksiin,
niin Eric, haluaisitko vähän avata, mitä näillä käsitteillä tarkoitetaan?

Eric Malmi: Joo. No, yleinen tekoäly, tai AGI, eli Artificial General Intelligence, viittaa sellaiseen tekoälyyn, joka suoriutuu ihmisen tasoisesti kaikista älykkyyttä vaativista tehtävistä, vastakohtana perinteisiin niin sanottuihin kapeisiin tekoälyihin, jotka on kehitetty jotain tiettyä sovelluskohdetta varten, kuten vaikka shakkitekoäly. Siinä mielessä voidaankin sanoa, että näissä nykyisissä kielimalleissa, kuten vaikka ChatGPT:ssä, on jotain viitteitä yleisestä tekoälystä. Niiltä voi kysyä periaatteessa mitä tahansa ja saada jonkinlaisen vastauksen. Toisaalta ne eivät vielä anna ihmisen tasoisia vastauksia tai varsinkaan asiantuntijatasoisia vastauksia ihan kaikkiin kysymyksiin.

Eric Malmi: Sitten toisaalta singulariteetti liittyy läheisesti tähän. Se viittaa sellaiseen hypoteettiseen käännekohtaan, jossa tekoälyn älykkyys ylittää ihmisen kyvyt, erityisesti niin, että se pystyy kehittämään itse itseään entistä älykkäämmäksi. Tämän on arveltu johtavan mahdollisesti sellaiseen älykkyyden räjähdysmäiseen kasvuun eli intelligence explosion -ilmiöön, jonka seurauksia voi olla hankalaa ihmisen ennakoida. Mutta näihin liittyen, Harri, muistan yhden seminaariesitelmän sinulta noin 10 vuoden takaa Otaniemestä. Puhuit jostain teidän konenäkömenetelmistä, mutta puhuit myös yleisestä tekoälystä ja muistaakseni myös singulariteetista. Tämä oli aikana, jolloin näistä ei puhuttu läheskään niin paljon kuin nykyisin. Ja minusta tuntuu, että nämä eivät olleet mitenkään erityisen suosittuja, varsinkaan yliopistopiireissä. Ne eivät ole ehkä vieläkään kauhean suosittuja, mutta minusta tuntuu, että varsinkin silloin jotkut meidänkin kollegat saattoivat vähän mielessään asetella sinun päähäsi pientä foliohattua, kun otit nämä teemat esille. Haluaisin kysyä, että miten sinun oma ajattelusi yleisen tekoälyn ja singulariteetin suhteen on kehittynyt tai muuttunut mahdollisesti viimeisen 10 vuoden aikana? Ja olemmeko me lähellä näitä nyt?

Harri Valpola: Olen varmaan oikeastaan ihan urani alkuajoista, silloin vuodesta 1992, tai jotain, ollut sitä mieltä, että kyllähän me tämmöisen yleisen tekoälyn rakennamme tässä minun elinaikanani. Toisaalta sitten taas on ollut sellaisia ihmisiä, jotka ovat ajatelleet, että yksi päivä yhtäkkiä herätään ja huomataan, että jaaha, täällä nämä robotit lentelevät ympäriinsä, ja kaikki on muuttunut. En ole siihen oikein uskonut,

koska vaikka älykkyys olisi täysin ratkaistu ongelma, tekoäly olisi täysin ratkaistu ongelma, niin on paljon muita pullonkauloja, joita tulee esiin. Eli joo, minä olen koko aikuisikäni varmasti niin kauan kuin muistan uskonut, että kyllä pystytään rakentamaan sellainen yleinen tekoäly, joka sitten johtaa niin kutsuttuun singulariteettiin. Enhän minä sitä termiä tuntenut, mutta tämä rekursiivinen itsensä kehittäminen oli kyllä ihan tuttu ajatus jo 1990-luvun alkupuolelta asti,

Arno Solin: Palataan näihin teemoihin myöhemmin, mutta mennään ensin sinne sinun urasi alkuvaiheisiin. Aloitit urasi 1990-luvun alussa Teknillisessä korkeakoulussa. Työskentelit tekoälyn suomalaisten pioneerien, Teuvo Kohosen ja Erkki Ojan, alaisuudessa. Mikä sai sinut alun perin tekoälyn pariin? Mikä veti puoleensa?

Harri Valpola: Tämä on hauska juttu sillä tavalla, että minä voin antaa ihan päivämäärän milloin tämä tapahtui. Minä olin koko lapsuuteni ja kouluikäni ollut valtavan kiinnostunut kaikesta tieteestä ja teknologiasta ja muun muassa tekoälystä. Silloinhan oli kuitenkin jo Star Wars, Ritari Ässän ihmeauto KITT ja joitain muita tämmöisiä medioissa. Ihmiset ovat näitä tarinoita kertoneet pitkän aikaa. Mutta se oli silloin jo tavallaan konkreettinen mahdollisuus. Oli tietokoneet, ja kävin kirjastosta etsimässä jotain tekoälykirjaa. Ei löytynyt mitään hyvää silloin. Olin hyvin pettynyt Outokummun kirjastoon joskus 10-vuotiaana. No, ei silloin 1980-luvun alussa ollut vielä sellaista tekoälyä. Minä en tiennyt tekoälystä hirveästi. Olin kiinnostunut matematiikasta ja biokemiasta. Minä olin ajatellut, että lähden ohjelmoimaan soluja. Geeniteknologiaa ja tämmöistä. Se oli minun ykkösuravaihtoehtoni lukiossa. Mutta sitten lukion kakkos- ja kolmosvuoden välissä Hesarin tiedesivuilla muistaakseni 29.6.1990 oli artikkeli, minulla on vieläkin tallessa se artikkeli, siellä tiedesivuilla Teuvo Kohosen hermoverkkotutkimuksista. Olin, että ei jummi jammi, tässä se on. Tätä minä rupean tekemään. Pistin sen artikkelin talteen ja kirjoitin ylioppilasaineenikin hermoverkoista, jotka mullistavat maailmaa. Sitten kolme vuotta myöhemmin minä olin Teuvon tutkimusapulainen. Tämmöinen tarina.

Arno Solin: Tosi siistiä. Koen tämän tosi inspiroivana.

Eric Malmi: Jotkut voivat antaa sellaisen uskoontulopäivämäärän, mutta sinä voit antaa tekoälyuskoontulon päivämäärän.

Harri Valpola: Kyllä minä olin tekoälyuskossa tavallaan ollut jo, mutta silloin minä löysin sen, että täällä oikeasti tehdään tätä hommaa. Vielä sillä tavalla, joka tuntui minun mielestäni järkeenkäyvältä. Meillä ihmisillä on päässämme hermoverkkoja. Joku tekee niistä keinotekoisia hermoverkkoja ja yrittää ymmärtää niitä matemaattisesti. Tämä on minun juttuni.

Arno Solin: Me suomalaiset tekoälytutkijat tykkäämme sanoa, että Suomi on kokoaan isompi tekoälymaa. Täällä on pitkät perinteet tekoälyn parissa. Tässä podcast-sarjassakin yritetään tuoda esiin niitä semmoisia käänteentekeviä tekijöitä ja asioita. Sinä tosiaan olet työskennellyt Teuvo Kohosen ja Erkki Ojan kanssa. Haluaisitko vähän avata, että millaista oli työskennellä heidän kanssaan ja mikä vaikutus sillä oli sinun uraasi?

Harri Valpola: Teuvohan on varmaan se grand old man, joka Suomessa tämän hermoverkkohomman on pistänyt liikkeelle. Hän innostui siitä asiasta jo olisikohan se ollut 60-luvun lopussa tai 70-luvun alussa. Erkki tietää paremmin, koska Erkki oli Teuvon ensimmäinen oppilas. Minä sattumalta olen Teuvon viimeinen oppilas. Minä olin 20-vuotias, valtavan innostunut tästä asiasta. Tulee vähän mieleen nämä minun nykyiset tekoälyagenttini. Intoa on ehkä vähän enemmän, kuin sellaista elämänkokemusta. Tein Teuvon kanssa hommia.

Harri Valpola: Se ei ollut helppoa. Teuvo ei ehkä ollut maailman helpoin ihminen monella tavalla, mutta olin myös itse nuori ja innokas. Se ei ollut täydellinen mätsi, mutta monella tavalla tunnen kyllä sielujen sympatiaa Teuvon kanssa. Meitä molempia on kiehtonut ihmisen aivojen toiminta ja se, miten me saamme rakennettua jotain sellaista teknologiaa, joka sitä kopioi tai vähintäänkin ottaa inspiraatiota siitä. Kolmen ensimmäisen vuoden jälkeen alkoi olla jo Teuvon kanssa sen verran sukset ristissä, että olisin lähtenyt jonnekin ihan muualle, mutta sitten Erkki otti koppia. Pysyinkin sitten siellä labrassa väitöskirjatutkijana vielä. Ehkä Teuvon ja minun vahvimpia puolia ei ole tämä human relations. Vähän kulmikkaita taidamme olla molemmat. Sanotaan asiat niin kuin ajatellaan.

Eric Malmi: (Eric) Haluatko avata jonkun esimerkin kautta tätä?

Harri Valpola: (Harri) No, muistan semmoisen, kun olin taas ollut nuori ja innokas ja tehnyt kaikkea mitä Teuvo ei ollut nimenomaan pyytänyt. Teuvo saattoi olla hyvin

tarkkakin siitä, että tehdään niin kuin hän sanoi. Minun opiskelukaverini sattui sitten tulemaan sinne labraan ja kertoi minulle jälkikäteen, että hän kuuli sieltä oven takaa semmoista huutoa, että "sinun työlläsi ei ole mitään arvoa, mitä sinä olet tehnyt". Teuvon palaute oli joskus hyvinkin terävää, mutta onneksi olen sille aika kovanahkainen, tai sitten vain puuttuu päästä semmoinen palikka, joka arvioisi näitä sosiaalisia vivahteita. Että en nyt siitä sitten liikaa ottanut kuitenkaan itseeni.

Arno Solin: (Arno) Ja ehkä kuulijoille...

Harri Valpola: (Harri) Tai sitten riittävästi, ihan näkökulmasta riippuen.

Arno Solin: (Arno) Ehkä kuulijoille voisi sanoa, että akateeminen maailma on täynnä persoonallisuuksia monella tavalla. Ehkä tämä tekoälypuoli ei ole mitenkään poikkeus tässä.

Harri Valpola: (Harri) Teuvoa arvostan suuresti siinä, että hän oli tutkija, tiedemies ja mikähän olisi hyvä suomennos tälle sanalle "scholar". Oli integriteettiä. Hän ei missään tapauksessa olisi ikinä mennyt maailmalle väittämään jotain, mitä hän ei ollut kunnolla huolellisesti tutkinut, testannut, selvittänyt. Teuvo oli hyvä markkinoimaan sitä työtä. Teuvohan kävi paljon maailmalla puhumassa näistä asioista, mutta koskaan ei minun uskoakseni lipsauttanut asioita, jotka eivät pitäneet paikkaansa. Siitä ehkä jotkut nykyaikaiset markkinoijahenkilöt voisivat ottaa onkeensa tuolla maailmalla.

Arno Solin: (Arno) Palataan varmaan siihenkin myöhemmin. Sinun urasi alkuvaiheissa teit monien erilaisten tekoälyteknologioiden kanssa töitä. Tietysti Teuvo Kohonen on tunnettu itseorganisoituvista kartoista. Self-organizing maps. Sinä teit sen parissa hommia, sitten harvan koodauksen ja itsenäisten komponenttien analyysin, ICA:n, parissa. Miten sanoisit nyt, kun on aikaa kulunut, että miten nämä tutkimusaiheet yhdistyvät niihin juttuihin, joita sinä teet nykyisin? Mikä niiden rooli on ollut siinä isossa kuvassa?

Harri Valpola: (Harri) Siihen maailman aikaan minua kiinnosti eniten se, miten näitä sisäisiä representaatioita syntyy. Ohjaamaton oppiminenhan oli se juttu, jota nimenomaan siinä Teuvon labrassa tehtiin. Teuvo oli itse kehittänyt tämän itseorganisoivan kartan, josta hänet tunnettiin maailmalla ja joka oli se juttu, jota piti tutkia. Tämä olikin sitten sitä ristiriitaa aiheuttavaa hommaa. Koska minä katsoin, että ei

tämä riitä, ei tämä toimi. Minä halusin tehdä joitain muita juttuja, ja niin kauan kuin Teuvon leipää syödään, niin Teuvon lauluja pitäisi laulaa. Se ei oikein minun pirtaani istunut. Jos en usko johonkin, niin en minä sitä voi tehdä sillä lailla virkamiesmäisesti.

Mutta joka tapauksessa tämä ohjaamaton oppiminen oli hirvittävän tärkeää minun oman ymmärrykseni kehittymisen kannalta. Minähän itse tein sitä riippumattomien komponenttien analyysiä osittain senkin takia, että se oli sitten taas se Erkin juttu. Mutta minua itseäni kiinnosti yleisemmin tämmöinen aivojen toiminnasta inspiraatiota ottava harvakoodaus. Sitten tutustuin siinä 90-luvun loppupuolella bayesiläiseen todennäköisyyslaskentaan ja siihen, miten se on oikeastaan ajattelun, oppimisen ja päättelyn semmoinen matemaattinen perusta. Minua itseäni kiinnosti se erityisesti.

Minähän sitten tein tällaista bayesiläistä neuroverkkoja yhdistelevää tutkimusta silloin. Se oli ehkä vähän liikaakin aikaansa edellä sillä tavalla, että tuli tehtyä joitain sellaisia juttuja, joita sitten vuosikymmen myöhemmin jotkut ekivät paljon paremmin. Ei ole aina hyvä olla liikaa pioneerina, koska kun yksinään tarpoo siellä suossa, niin ei välttämättä tajua, että nyt on iskenyt pään puuhun. Jos on isompi porukka, niin se homma menee paremmin eteenpäin. Mutta se oli tärkeää oppia, "Ollin oppivuosia" siinä mielessä, että sain hyvät pohjat jatkaa siitä eteenpäin.

Arno Solin: (Arno) Sinä siirryit vuosituhannen vaihteessa vetämään omaa ryhmää, omaa tiimiä, laskennallisen neurotieteen ryhmää, ja se keskittyi nimenomaan kognitiivisiin arkkitehtuureihin ja muuhun. Kuinka tärkeää on tekoälyä kehitettäessä hakea inspiraatiota nimenomaan aivoista ja aivojen toiminnasta?

Harri Valpola: (Harri) Minulle itselle se on ollut hirveän tärkeä juttu. Jotenkin en pysty miettimään tällaisen tekoälyn neuroverkon oppimista, jos en jotenkin itse samaistu siihen. Minun pitää pystyä kuvittelemaan itseni pieneksi neuroniksi siellä hermoverkossa. Eli miten minä näen maailman? Minkälaisia signaaleja minä saan ja mitä minä sitten teen niiden kanssa? Miten minä opin? Inspiraatio aivoista on aina ollut minulle hirveän tärkeä. Minä olen ihan tarkoituksella opiskellut neurotieteitä. Oli hirveän onnellinen tilanne, että samassa yliopistossa oli paitsi uraauurtava hermoverkkotutkija, niin myös uraauurtavia aivotutkijoita. Eli minulla on ollut ihan alusta asti onni päästä heidän luentojansa kuuntelemaan ja heidän oppejansa seuraamaan. Olin 10 vuotta ehtinyt tehdä koneoppimisen parissa tekniikkaa, matematiikkaa, sen parissa tutkimusta,

ja nimenomaan ohjaamatonta oppimista. Siinä vaiheessa päätin, että minun on pakko lähteä eteenpäin. Minä haluan soveltaa tätä johonkin. Haluan ymmärtää, että mihin aivot ovat ratkaisu, koska ohjaamaton oppiminen on semmoinen väline jonkun asian ratkaisemiseksi. Mutta mikä se kysymys on? Vähän kuin Douglas Adamsin Hitchhiker's Guide. Että okei, tämä on se ratkaisu, mutta mikä on se kysymys? Sen takia lähdin sitten robotiikan pariin vuonna 2003. Eli tein tavallaan alanvaihdoksen, mutta otin sinne mukaan nämä neuroverkot, koneoppimisen ja muut. Omassa mielessäni se on ollut jatkumo siihen, että halusin edelleenkin ymmärtää, miten aivot toimivat. Aivot ovat kehittyneet ohjaamaan jotain fyysistä ruumista fyysisessä maailmassa. Minun mielestäni nimenomaan bioinspiroitunut robotiikka tuntui järkevältä suunnalta, että lähden sinne suuntaan tutkimaan sitä, mitä ne aivot oikeasti tekevät. Mikä on se ongelma, mitä ne ratkaisevat?

Eric Malmi: (Eric) Tässä on noussut useamman kerran esiin ohjaamaton oppiminen. Jos jotkut meidän kuulijoista eivät tiedä mitä sillä tarkoitetaan, niin voitko avata vähän tätä?

Harri Valpola: (Harri) Ohjaamaton oppiminen tarkoittaa sitä, että on paljon jotain dataa. Vaikkapa me saamme silmien, korvien ja tuntoaistin kautta hirveän määrän dataa maailmasta. Mutta siinä ei ole mitään rakennetta. Jos katsoo niitä signaaleja, niin siellä ei ole kauheasti mitään. Tai siis ei siitä saa mitään tolkkua. Ohjaamaton oppiminen tarkoittaa sitä, että yritetään tämmöisestä valtavasta datamassasta löytää rakennetta. Vähän samalla tavalla kuin miten pieni ihmisvauva syntyy maailmaan ja alkaa nähdä kaikkia juttuja. Se alkaa hahmottaa ympäriltään niitä. Että ahaa okei, täällä on tämmöisiä hahmoja, tommoisia juttuja, ja nämä liittyvät yhteen. Sitten mitä se ei ole, niin on ohjattu oppiminen, joka on sitä, että on annettu valmis vastaus.

Harri Valpola: Neuroverkko saa hirveän kasan jotain kuvaa tai jotain ja sitten annetaan vastaus, että tämä on koira tai tämä on kissa, tai tässä on A-kirjain. Se on ohjattua oppimista. Sitten on vielä tämmöinen vahvistusoppiminen, re-enforcement learning englanniksi, jossa annetaan jotain palautetta, mutta se on enemmän sellaista "hyvä", "nyt menee hyvin", "nyt ei mene niin hyvin". Eli ei anneta valmista vastausta, vaan annetaan vain tämmöistä plussaa tai miinusta -tyyppistä palautetta. Näitä kaikkia käytetään nykyään toki näiden suurten kielimallienkin opettamisessa esimerkiksi. Mutta isoimman osan siitä kuormasta hoitaa, voisi sanoa, se ohjaamaton oppiminen.

Annetaan vain internetin verran dataa, ja sieltä sitten pitää löytää sitä rakennetta. Se tehdään nykyään sillä tavalla, että yritetään ennustaa vain sitä seuraavaa sanaa esimerkiksi. Mutta se on yksi tapa tehdä ohjaamatonta oppimista tietyllä tavalla.

Arno Solin: (Arno) Se on tietty taloudellista, kun ei tarvitse olla niitä oikeita vastauksia, vaan pelkkä data riittää.

Harri Valpola: (Harri) Kyllä.

Arno Solin: (Arno) Puhuit tuosta aivojen rakenteesta ja siitä inspiraatiosta. Jos nyt katsoo näitä nykymalleja, kielimalleja ja kuvamalleja, niin ne perustuvat melkein poikkeuksetta transformer-arkkitehtuuriin. Niin onko siinä jotain syvempää linkkiä, että onko se transformer jotenkin enemmän aivojen kaltainen vai mikä siinä on ratkaissut jotenkin tätä älykkyyttä?

Harri Valpola: (Harri) Sanotaan, että transformer on semmoinen ratkaisu, ehkä Marvin Minsky, vai kukahan se puhui näistä cognitive wheel -ajatuksista, että teknologiassa ei kaikkea tarvitse ratkaista ihan samalla tavalla miten biologia on sen ratkaissut. Vaikka kaikilla eläimillä on jalat, niin ei meidän tarvitse rakentaa autoille jalkoja. Pyörä on itse asiassa ihan tavattoman tehokas ratkaisu, johon biologia ei ole pystynyt. Jalat ovat jossain kohtaa parempia kuin pyörät. Mutta pyörät ovat monella tapaa erinomainen ratkaisu meidän teknologialle. Sanoisin, että transformer on ehkä enemmän tämän tyyppinen ratkaisu. Se ratkaisee jonkun todellisen ongelman. Minä itse asiassa tuossa 10 vuotta sitten, tai reilut 10 vuotta sitten yritin ratkoa sitä ongelmaa, että miten neuroverkot saataisiin ymmärtämään rakenteisempaa dataa, objekteja, niiden välisiä relaatioita ja muita. Minulla oli vähän bioinspiroituneempi näkökulma siihen. Tiedettiin, että tämä on ongelma, tämä pitää ratkaista jollain tavalla. Mutta sitten kun transformerit tulivat, niin en samana päivänä tajunnut, että okei, tämähän se on. Mutta sitten hyvin pian, vuodessa parissa, kävi ilmi, että okei, no tämähän se itse asiassa ratkaisee kyllä nämä hommat. Sen jälkeen minä droppasin nämä kaikki muut. Että okei, se on tässä nyt. Tällä mennään. Tämä toimii ihan riittävän hyvin. Tästä se ei enää jää kiinni.

Eric Malmi: (Eric) No, kuvasit miten perustit sen oman tutkimusryhmän. Mutta sitten 2007 Syntyi ZenRobotics, jota olit myös perustamassa. Mitä ongelmaa ratkomaan ZenRobotics perustettiin?

Harri Valpola: (Harri) No, tuo on hyvä kysymys. Siinä on useita tasoja. Yksi ongelma henkilökohtaisesta näkökulmasta on, että pienellä tutkimusryhmällä ei ole resursseja ostaa ja pitää yllä isoja robotteja. Olin nähnyt monissa robotiikkalabroissa sen, että siellä on hyllyt väärällään vanhoja robotteja, jotka on rakennettu häthätää johonkin tiettyyn projektiin, ja sitten ne pölyttyvät siellä kaapissa. Eli jotta pääsisi tekemään oikean maailman robotteja. Minun tutkimusryhmässäni olimme nimenomaan sitä varten rajanneet vain ja ainoastaan simulaattoreihin. Eli vain robottisimulaattoreita käytettiin, simuloidussa maailmassa tehdään robotille aivoja. Mutta totta kai kiinnostaa sitten toimiiko tämä ollenkaan oikeassa maailmassa. Nuoren startup-yrittäjän innolla ja itsetunnolla, että "no mehän kyllä ratkomme kaikki nämä robotiikan ongelmat oikeassa maailmassa", niin perustettiin ZenRobotics yhtään tietämättä, että mihin me rupeamme soveltamaan tätä neurorobotiikkaa. Mutta johonkin varmasti saadaan sitä sovellettua. Ja sitten lähdettiin etsimään. Eli meillä on vasara, ja sitten etsitään nauloja maailmasta. Niitähän löytyi sitten. Otettiin puhelin kauniiseen käteen ja soiteltiin kaikkiin suomalaisiin firmoihin, että hei, meillä on tässä kova uusi firma, me ratkomme kaikki teidän ongelmanne. Mitäs ongelmia teillä on? Ja sitten päädyttiin tekemään jos vaikka mitä. Siellä oli jotain taikinan letitystä, polttoainesauvojen vaihtamista jonnekin ydinvoimalaan tai tämän valvomista ja optimointia ja jotain kaivostyökoneiden sensorioptimointeja tai sensorikehitystä ja kaikkea tämmöistä. Mutta vähitellen alkoi käydä ilmi, että maailmassa on paljon ongelmia, joissa tarttuminen olisi hirveän hyödyllistä robotille. 2007-2008 vuosien aikoihin se ei todellakaan ollut ratkaistu ongelma. Vähitellen kävi ilmi, että jätteenkäsittelyalalla on huutava pula nyt tämmöisestä sopivasta teknologiasta. Ja sitten me lähdimme sinne. ZenRoboticsista tulikin sitten jätteenkäsittelyrobotiikkaa tekevä firma. Ei ollut mitenkään suunnitelmissa ensi alkuun, mutta sinne se sitten vei.

Eric Malmi: (Eric) Oliko niin, että kaikissa näissä sovelluskohteissa se sama metodologinen perusta pystyi ratkaisemaan kaikkia näitä eri sovelluskohteita?

Harri Valpola: (Harri) No ei tokikaan. Polttoainesauvojen optimointi ei käyttänyt samoja teknologioita, kuin sitten vaikka tämä autonomisten kaivostyökoneiden sensorikehitys. Meillä oli kuitenkin joku ajatus siitä, että minkälaisissa teknologioissa me olisimme hyviä. Sillä tavalla me nimenomaan sitten päädyimme sinne jätteenkäsittelyyn. Rakennus- ja purkujäte vielä
nimenomaisesti on semmoinen ongelma, jossa ihmiset eivät ole ollenkaan hyviä. Sieltä tulee jotain 20 kilon betonijärkäleitä, niin ei olkapäät kestä sitä. Siellä on terävää,

myrkyllistä materiaalia. Todellakin semmoinen ongelma, joka pikku WALL-E -robotin pitäisi ratkaista. Jos joku on tämän Disney-leffan sattunut näkemään. Sitä lähdettiin sitten tekemään.

Harri Valpola: (Eric) Tämä on varmaan aika erilainen ongelmakenttä, kuin ne akateemiset benchmarkit, joiden parissa olit oletettavasti aiemmin työskennellyt, niin millaisia haasteita tuli vastaan, kun te siirryitte ratkomaan tosielämän ongelmia?

Harri Valpola: (Harri) Meillähän oli semmoinen hauska anekdootti tästä, että tuossa 2011 loppuvuodesta olimme todenneet, että nyt me olemme saaneet tämän robotin toimimaan johonkin malliin. Meillä Viikissä koelaitoksella robotti toimi, mutta sitten olimme saamassa asiakastoimituksia. Me olimme todenneet, että pitää miettiä seuraavan sukupolven älyä sille robotille päähän, että miten se poimii ja kuinka se oppii. Siitä ruvettiin kehittämään sitten tämmöistä mallipohjaiseen vahvistusoppimiseen, model-based rainforcement learning, perustuvaa systeemiä. Ei sitä akateemisessa maailmassakaan hirveästi oltu vielä tehty. Kehitettiin sitä sitten siellä tutkijankammiossa. Vajaa puoli vuotta myöhemmin oltiin saatu Hollantiin robottitoimitus myytyä. Sitten siellä asennettiin robotteja, ja jossakin vaiheessa he testasivat sitä. Me emme olleet saaneet vielä deployattua sitä meidän systeemiämme maailmalle. Sieltä tuli sitten vähän hätääntynyttä puhelua, että "tämä ei toimi lainkaan", "tämä ei tee mitään järkevää täällä". "Mitä me teemme?" Siinä sitten oltiin, että okei, no, meillä on tämä uusi systeemi. Sitä ei olla testattu ikinä. Sitten sinne vain Hollantiin. Ja tässä voi mennä vähän aikaa. Sen pitää vähän keräillä dataa sieltä. Sitten kerätään dataa. Robotti poimii siellä ihan mitä sattuu. Sitten datat Suomeen, sitä annotoidaan ja ihmiset katsovat, että okei, tuo ja tuo meni hyvin. Ihmisten piti annotoida, että mitä täsmälleen se robotti poimi. Sillä ei ollut riittävän hyvää näkökykyä, että se olisi pystynyt itse arvioimaan sitä. Ihmisen piti sanoa, että nyt tässä poiminnassa tuli kyytiin tuo, tuo ja tuo. No sitten shipataan sinne seuraava. Ja "ei toimi vieläkään kaikki". Okei, kerätään lisää dataa. Kahden viikon kohdalla sitten oltiin, että "hei nythän tämä toimii". Huh. Se myytiin sinne vähän etukenossa. Ehkä myyntiosastot helposti lupaavat, että se toimii kuten labraolosuhteissa, mutta kenttäolosuhteet ovatkin jotain ihan muuta. Mutta onneksi sitten tämmöinen mallipohjainen reinforcement -oppiminen neuroverkoilla pelasti päivän tällä kertaa.

Eric Malmi: (Eric) Ja sen parissa olet ymmärtääkseni myöhemmissäkin vaiheissa tehnyt paljon työtä ja tutkimusta. Mutta ennen kuin siirrytään niihin, niin ehkä vähän semmoinen filosofisempi
kysymys. Kun puhutaan yleisestä tekoälystä, niin jotkut ovat esittäneet, että todellisen AGI:n saavuttamiseen vaaditaan jonkinlaista "embodymentia" eli kehollisuutta. Kun sinä olet tutkinut sekä robotiikkaa että näitä itse algoritmeja siellä taustalla, niin ajatteletko sinä, että tämä on vaatimus yleisen tekoälyn saavuttamiselle?

Harri Valpola: (Harri) Kun minä olen robotiikkaa katsellut, niin kyllä minulle on jäänyt sellainen olo, että varmasti evoluution kuluessa on ollut tosi tärkeää, että aivot ovat kehittyneet semmoisiksi kuin ovat kehittyneet. Toisaalta tiedetään sellaisia ihmisiä, jotka ovat tosi vähän "embodied" fyysisten rajoitteiden takia, mutta se ei kyllä järkeä hidasta. Eli sekä sensoripuolella että aktuaattoripuolella voi olla valtavia aukkoja, tai on hyvin kapea aukko maailmaan. Minä nykyään uskon, että interaktio maailman kanssa on tosi tärkeää, että syntyy sen tyyppinen mieli, kuin meillä on. Että syntyy käsitys itsestä toimijana maailmassa. Eli pitää nähdä, että minä tein näin, ja tässä ovat tekojeni seuraukset. Pitää oppia ymmärtämään tämä vuorovaikutus maailman kanssa. Mutta ei se vaadi fyysistä ruumista. Se, että virtuaalimaailmoissa ollaan vuorovaikutuksissa virtuaalimaailmojen kanssa, ja tämän meidänkin maailman kanssa, niin se on vähän niin kuin tämä Platonin vertauskuva siitä, että kaikki ovat vain varjoja, jotka heijastuvat sinne luolan seinämään. Niissä varjoissa,
niissä heijastuksissa on se sama rakenne kuitenkin, joka siellä maailmassa on. Ei sillä ole kauheasti väliä, että mikä on se kanava, jolla me olemme maailman kanssa vuorovaikutuksessa. Eli ihan tämmöinen kielimalli, joka on ihmisten kanssa vuorovaikutuksessa, niin kyllä, minun mielestäni se riittää.

Eric Malmi: (Eric) Jos ajatellaan, että perinteisiä kielimalleja koulutetaan niin, että syötetään kaikki internetin sisältö ja kaiken maailman videot, niin tämä yksinään ei riitä vielä. Mutta sitten, jos me saamme sen vuorovaikuttamaan maailman kanssa, keräämään kokemusperäistä tietoa, niin sitten päästään sinne.

Harri Valpola: (Harri) Juuri näin. Tämä on se, mihin tällä hetkellä uskon. Pidän mahdollisena, että on joitain muita älykkyyden tyyppejä, jotka vaatisivat vielä paljon vähemmän vuorovaikutusta. Mutta jos puhutaan semmoisesta meidän ihmisten kaltaisesta älykkyydestä. Joku voisi sanoa, että tietoisuudesta, "self-awareness". Että

on itse itsestään tietoinen, niin se vaatii minun uskoakseni tämmöistä vuorovaikutusta, että syntyy sinne neuroverkkoon niitä representaatioita myös omasta itsestä osana maailmaa.

Arno Solin: (Arno) Muistan itse oman urani alkuvaiheelta, kun olin itse asiassa ZenRoboticsin toimistossa töissä firmassa, joka oli vuokralaisena siellä jossain sivuhuoneessa, ja pääsin seuraamaan sitä ZenRoboticsin meininkiä sivusilmällä. Se oli jonkinlainen arkkityyppi tekoälystartupista. Teillä oli tosi komea toimisto siinä Kaisaniemessä.

Harri Valpola: (Harri) Se oli täysin Jufon ansiota.

Arno Solin: (Arno) Minusta tuntuu, että tätä Jufo Peltomaata ei voi olla mainitsematta, kun puhutaan ZenRoboticsista. Jufolla oli sellainen kunnon draivi, ja se vision myymisen voima oli tosi merkittävä. Selvästikin te olitte hyvä tiimi siinä myymässä, koska ei varmastikaan ollut helppoa markkinoida tekoälystartupia vuonna 2007. Nykyäänhän tekoälystartupeja on kuin sieniä sateella. Niitä syntyy koko ajan, ja siinä ei ole mitään ihmeellistä. Mutta varmaan se markkina oli tosi erilainen. Miten rahoituksen hankkiminen startupille
silloin toimi? Mikä se zeitgeist silloin oli tähän liittyen?

Harri Valpola: (Harri) Joo, silloin maailma oli hyvin erilainen kuin nyt. Se, että oli tekoälyfirma, ei ollut mikään automaattinen lupa painaa rahaa. Niin kuin se nykyään tuppaa olemaan vähän liiankin helppoa ehkä. Tarinahan meni sillä tavalla, että Tuomas oli vuonna 2006 tullut siihen minun tutkimusryhmääni. Me olimme tehneet tätä neurorobotiikkaa, rakentaneet robotille aivoja siinä vuoden verran. Sitten tosiaan kiinnosteli vähän, että toimisikohan tämä nyt ihan oikeassa maailmassa ja olisiko nyt hyvä hetki lähteä tekemään. Sitten
mietittiin firman perustamista ja katsottiin toisiamme. Sitten katsottiin vähän peiliin, että jep, tähän tarvitaan joku muukin vielä. Ja Tuomas oli ollut silloin...

Eric Malmi: (Eric) Tuomas Lukka siis.

Harri Valpola: (Harri) Tuomas Lukka, joo, toinen ZenRoboticsin perustajista, oli ollut Hybrid Graphicsilla töissä ja sieltä tuli meille, kun Hybrid Graphics myytiin Nvidialle. Jufo oli sitten myös vapaalla. Jufo oli ollut Hybrid Graphicsilla hommissa, ja Tuomas tiesi

Jufon. Jufo Peltomaa siis. Itse tunsin Jufo Peltomaan lähinnä rap-artisti Raptorina. Oi beibit ja muut olivat silloin 90-luvun alkupuolella kova juttu. Tuomas suositteli lämpimästi, että ottaisimme Jufon. Sitten pidimme palaverin, ja olin kanssa vakuuttunut, että tässä miehessä on sitä jotain. Semmoisia ihmisiä ei ole ihan hirveän paljon. Poikkeusyksilö. Erittäin onnellinen olen ollut siitä, että olen saanut olla töissä semmoisten poikkeusyksilöiden kanssa. Tuomas on kanssa poikkeusyksilö. Saatiin monenlaista hyvää porukkaa sinne ZenRoboticsiin myös. Siitä on oppinut itse ihan valtavasti. Minähän en mitään tiennyt yrittämisestä
siinä vaiheessa.

Arno Solin: (Arno) Muistan joskus alkuaikoina nähneeni Jufon esityksen. Siinä tavallaan myytiin nimenomaan tätä isoa visiota siitä, että tekoäly tulee mullistamaan kaiken ja olemaan osa kaikkia liiketalouden ja elämän osa-alueita. Ja nyt vaan ihan ensimmäisenä steppinä on tämä rakennusjätteiden lajittelu liukuhihnarobotilla. Mutta tämä homma aukenee tästä ja maailmanvalloitus alkaa. Ja siinä oli varmasti...

Harri Valpola: (Harri) Tämän vision me jaoimme, me kaikki kolme perustajaa.

Arno Solin: (Arno) Tämä oli tosi mullistavaa sanoa silloin. Ja minusta tuntuu, että suurin osa ihmisistä oli, että tässä on nyt ehkä vähän Amerikan lisää tässä hommassa. Niin kuin varmaan oli, semmoista rohkeutta. Sitten kuitenkin minusta tuntuu, korjaa, jos olen väärässä, että ZenRobotics jäi ehkä vähän jumiin. Tavallaan se oli aika haastava se jätteiden lajittelu, se ensimmäinen steppi. Siitä tuli se pääjuttu. Mikä rooli sillä oli siinä, että sinä siirryit sitten uusien haasteiden pariin? Sinä perustit seuraavan firman sitten.

Harri Valpola: (Harri) Tuo on ihan totta. Jos haluaa tehdä jotain mullistavaa rakennus- ja purkujätteen alalla, mikä pitää tehdä, jos haluaa sille alalle, niin siihen maailman aikaan varsinkin se, että oli vain tekoälylabra, niin ei se tarina yksinänsä lentänyt. Se oli vähän niin kuin pakon sanelemaa myös, että tästä täytyy tehdä bisnes ja sitten yrittää sen päälle rakentaa jotain. 2014 olen todennut, että okei, my job here is done. Olen saanut nyt nämä hommat valmiiksi ZenRoboticsilla. Ei minua siinä rakennus- ja purkujätteen lajittelussa enää kauheasti tarvita. Rupesin miettimään seuraavia juttuja. Ajatus oli ensin, että lähdetään siellä ZenRoboticsin sisällä tekemään toista osastoa. Mutta sitten kun sitä ajatusta vähän pyöriteltiin siinä, niin ei se ehkä olisi kauhean

onnellinen avioliitto, että meillä olisi tämä rakennus- ja purkujätteen robotiikkafirma, jonka pitää pistää kaikki paukut siihen, ja sitten ovat nämä tekoälyhaihattelijat toisaalla. Niin sitten todettiin, että tässä täytyy tehdä ihan oma startupinsa. Se ei ollut edes varsinaisesti spin-off, vaan sellainen sisaryritys. Sijoittajat kyllä seurasivat sieltä ZenRoboticsista. Siinä mielessä siinä oli jatkumo. Mutta se ei ollut mikään tytäryritys. Siitä kuroutui ulos tämä Curious AI. Minä lähdin sitä vetämään ihan itse. Keräsin siihen tiimin. Aloitin sillä, että pistin pystyyn tällaisen lukupiirin, jossa lueskeltiin kaikkia Deep Neural Network -juttuja. Syvät neuroverkot olivat silloin lyöneet itsensä läpi, ja Aallosta löytyi kiinnostuneita siihen. Sillä porukalla sitten pistettiin pystyyn Curious AI.

Arno Solin: (Arno) Ja tuntui, että Curious AI profiloitui nimenomaan yleisenä tekoälyfirmana.

Arno Solin: Eli minusta tuntuu, että siellä nimenomaan vältettiin kuin ruttoa tällaista yhteen sovellukseen pureutumista, ja se kuva oli se, että nyt ratkaistaan tekoälyä isosti. Onko tämä oikea tulkinta?

Harri Valpola: (Harri) Kyllä se on oikea tulkinta. Joo. Minä olin tietysti siinä poliittisena linjanvetäjänä päättämässä siitä ensisijassa, että esimerkiksi hardista ei tehdä, piste. Se riitti minulle, että olin yhden tällaisen hardisfirman pistänyt pystyyn. Se on siis tosi hyödyllinen harjoitus ollut minulle henkilökohtaisesti. Kasvun paikka, opin siinä paljon. Mutta opin myös sen, että jos haluan tehdä kunnolla yleistä tekoälyä, niin sitä ei nyt vaan sitten kannata tehdä. Monia asioita, joita olin oppinut ZenRoboticsissa, niin niitä oppeja siellä Curious AI:ssa sovelsin. Silloinhan maailma oli jo muuttunut. Silloinhan oli jo DeepMind perustettu. Tai siis DeepMind oli myyty jo Googlelle. Silloin sillä, että sanoit, että minulla on kokemusta tästä tekoälystä, minulla on visio, osaan sitä ja tätä, niin sillä sai rahaa. Se riitti. Esimerkiksi lontoolaisesta isosta sijoitusfirmasta Baldertonista tuli tyyppi Helsinkiin. Kävimme Kämpissä lounaalla, selitin reilun tunnin sitä visiota, ja sitten se sijoituspäätös oli tehty. "Okei, me annetaan rahaa".

Arno: Mikä se visio silloin oli?

Harri: Rakennetaan yleinen tekoäly. Six easy steps to AGI. Minulla oli suunniteltuna mitä kaikkea pitää tehdä. Me olimme tehneet jo näitä ensimmäisiä askelmia. Siinä oli hyviä palikoita siinä "six easy steps to AI" -suunnitelmassa. Me saimme tehtyä monia niistä askelmista. Mutta edelleenkin itse asiassa voisi sanoa, että se, missä Curious AI

sitten lähti takkuamaan, oli se, että ehkä liian aikaisin kuitenkin lähdettiin soveltamaan sitä käytäntöön. Se on tosi hankalaa. Olen kuitenkin sillä lailla tutkijaluonne ja meillä oli 2018-2019, en muista ihan tarkkaa päivämäärää, suunnitelmissa, että me voisimme ruveta rakentamaan tällaista "digital coworker", digitaalista työkaveria ihmisille. Me totesimme kuitenkin, että maailma ei ole vielä valmis. Pieni startup ei pysty kaikkea sitä kehittämään, mitä tähän tarvitaan. Niin me sitten menimme enemmän tähän teollisuusprosessien optimointiin, autonomisiin työkoneisiin, sen tyyppisiin juttuihin, niiden ohjaukseen. Eli sovellettiin jälleen kerran sitä mallipohjaista reinforcement -oppimista.

Sitten kävin ne säädetyt kolme vuotta Applella. Silloin, kun tehdään nimiä paperiin, niin monesti perustajat joutuvat tekemään tämmöisiä. Eli kolme vuotta on semmoinen minimiaika, joka ollaan töissä. Niin sen aikaa olin. Mutta sinä aikana, 2023 mennessä, olivat ehtineet tulla nämä kielimallit, ChatGPT ja muut. Siinä vaiheessa oli se tilanne, että tässä on semmoinen teknologia, jonka avulla me pystymme nyt rakentamaan sen digitaalisen työkaverin. Sitten perustimme tämän nykyisen firman, System 2 AI:n, nimenomaan rakentamaan tällaisia ihmisen lailla oppivia digitaalisia työkavereita.

Tässä on tultu pitkä matka siitä, kun vuonna 1990 luin niitä Teuvon hermoverkkotutkimuksia. Silloin jo heräsi se, että okei, tästähän me rakennamme sen jutun. Nyt me olemme oikeasti rakentamassa sitä. Vuonna 1992, kun tulin Otaniemeen opiskelemaan, niin välittömästi kävelin infolabraan fuksina, että "hei, minä tulen tänne sitten kesätöihin ja haluan rakentaa sellaisia robotteja, jotka tekevät sitä ja tätä". Muistan, että Olli Simula oli siellä juttelemassa. Hän vähän naureskeli, että "joo, ehkä pitäisi vähän ensin opiskella, että ei tänne nyt noin vaan tulla töihin". Sitten minä tosiaan into pinkeänä opiskelin sen ensimmäisen vuoden ja pääsin töihin silloin heti fuksivuoden jälkeen. Mutta niitä kotona pörrääviä robotteja en ole vieläkään tekemässä. Nyt tehdään sitä tekoälyä, ja jossakin muualla rakennetaan ehkä sitten niitä kotirobotteja, niitä vartaloita. Ehkä jossain vaiheessa saadaan sitten lyötyä niille aivot päähän.

Arno: Minulla on pari kysymystä tuosta julkaisemisesta ja niistä innovaatioista, joita olet tehnyt matkan varrella. Olet työskennellyt monien suomalaisten pioneerien kanssa, mutta sinulla on myös todella viitattuja tieteellisiä julkaisuja, muun muassa Yann LeCunin kanssa, joka on yksi yleisesti tunnistettuja näistä nykytekoälyn isistä. Me Ericin kanssa katsoimme läpi, mitä kaikkea sinulta on tullut ulos viimeisten 15 vuoden aikana.

Viitatuin paperi on sellainen "mean teachers" -paperi, joka on Neural information processing systems -konferenssissa julkaistu. Se on Antti Tarvaisen kanssa julkaistu vuonna 2017. Tällä paperilla on siis yli 8000 viittausta, eli selvästikin tekoäly-yhteisö on käyttänyt näitä teidän juttujanne. Mutta kun te teitte esimerkiksi sitä työtä tai muita näitä sinun töitäsi, niin tiesitkö siinä vaiheessa, kun teitte niitä tuloksia ja keksitte niitä menetelmiä, että nämä tulevat olemaan sellaisia, joita muutkin käyttävät?

Harri: Täytyy sanoa, että minä itse en ole hirveän hyvä arvaamaan sitä, mistä tulee iso juttu. Minun omasta mielestäni esimerkiksi tämä "ladder network", tikapuuverkko, oli paljon kiinnostavampi. Jossain mielessä se "mean teachers" on hyvin pitkälle yksinkertaistettu versio tästä ladder-verkosta. Se on tosi yksinkertainen juttu, ja se on semmoinen, että sen voi vain "plugata" johonkin. Se vaan toimii. Se on kuin junan vessa. "Plugaat" sen siihen. Sitten se vain toimii. Tämähän on sellainen teknologia, jota ihmiset mielellään hyödyntävät. Se on hyödyllinen maailmalle. Ja olen onnellinen siitä, että on tällainen hyödyllinen juttu saatu tehtyä. Mutta minua itseäni kiehtoo vielä enemmän se, että saadaan uusia käsitteitä, ymmärretään joku uusi lähestymistapa johonkin. Kaikki ymmärtävät kuitenkin, että tässä seistään jättiläisten harteilla. Mitä tahansa minäkin teen, niin siinähän on takana vuosikymmenten ja jopa satojen vuosien tieteellinen työ takana.

Tästä on kuitenkin parikymmentä vuotta, kun sinä hyppäsit pois akateemisesta maailmasta. Olet kuitenkin aktiivisesti julkaissut akateemisilla foorumeilla. Miten tämä julkaiseminen on sopinut yhteen näiden yksityisten firmojen kanssa? Kun on ollut sijoittajia ja muuta? Ja mikä on se julkisuuden ja julkaisemisen suhde siihen tekemiseen?

Voidaan sanoa, että tekoälyalalla ollaan tieteen eturintamalla, ja siellä arvostetaan julkaisuja varmasti enemmän kuin monella muulla alalla. Eli tekoälyfirmalle ihan ok strategia on se, esimerkiksi, se ei nyt ole meidän strategia pelkästään, mutta että ajatellaan, että ei edes haluta tehdä mitään tuotetta. Tai, että ei yritetä tehdä mitään, yritetä kerätä jotain omaa teknologiaportfoliota, vaan tehdään tutkimusta. Tehdään niin kovaa tutkimusta, että se huomataan maailmalla. Sitten joku ostaa meidät miljardilla.

Sitä tapahtuu oikeasti. Siitä maksetaan sitten se miljardi. Tämä ei ole monella muulla alalla vaihtoehto. Tekoälyllä se on.

On ehkä vähän tylsästi ajateltu, että ei meitä nyt välttämättä kiinnosta tulla ostetuksi jonnekin. Me nyt yritämme jotenkin vain selustamme turvata. Me olemme tyypillisesti aina tehneet patentin, jättäneet patenttihakemuksen edellisenä päivänä ja seuraavana päivänä jättäneet sen konferenssipaperin. Tämä on ollut ihan toimiva strategia myös. Minulla on aika monta patenttia tekoälyalalta. Muun muassa muistaakseni Mean Teacheristäkin voipi olla jossakin jokin patentti, joka ei ole minun hallinnassani enää. Se on se huono puoli toki siinä. Se on enemmän käyntikortti kuitenkin. Nämä firmat istuvat niiden patenttien päällä ehkä enemmän sitä varten, että voidaan varmistaa, että kukaan muu ei sitten tule väittämään, että tätä ei saisi tehdä.

Eric: Tässä on vähän viitattu jo sinun seuraaviin steppeihisi Curious AI:n jälkeen. Muistan itse, kun vuonna 2020 asuin itse Zürichissä, ja silloin alkoi kuulua huhuja, että nyt on ehkä muuttamassa jotain tuttuja suomalaisia tänne päin. Eli tällöin sinä ja teidän firmasta ainakin osa siirryitte Applen leipiin. Voisitko kertoa vielä vähän tarkemmin tästä?

Harri: Ilman muuta. Sehän on sellainen musta aukko, kun menee Applelle töihin. Ei pelkästään se, että ei kauheasti julkaise. Se olisi ollut tietyllä tavalla mahdollista. Mutta minä myös mielelläni olin töissä sellaisessa tiimissä, jossa tehtiin hyvin konkreettisen oikean maailman jutun kanssa hommia. Koska Applella on resursseja, niin siellä voi tehdä sellaisia juttuja, joita todellakaan ei Curious AI voinut tehdä. Sillä tavalla se oli kuin paluu sinne ZenRoboticsin juurille. Mutta ei pelkästään se, ettei tule kauheasti tieteellisiä julkaisuja, vaan se, että ei voi edes avata suutaan mediassa. Se oli minulle kyllä vähän semmoinen punainen vaate. Se oli yksi syy, jonka takia sitten halusin taas itsenäiseksi yrittäjäksi. Nyt minä voin jutella täällä ihan vapaasti. Minun ei tarvitse kysyä lupaa tähän yhtään keneltäkään.

Arno: Siihen liittyen, niin tosiaan nyt johdat System 2 AI -nimistä yritystä. Nimessä piilee tällainen käsite, joka on tekoälyalalla tuttu monelle. Tämmöinen niin sanottu systeemi 1 on yleensä nopea, intuitiivinen, automaattinen. Sitten taas systeemi 2 puolestaan on semmoinen hidas, harkitseva, analyyttinen. Voiko sanoa, että nämä sinun edelliset sarjayrittäjän yritykset ovat olleet sitä systeemi ykköstä, ja nyt ollaan

siirrytty systeemi kakkoseen? Ratkotteko nyt analyyttisempiä ongelmia kuin aikaisemmin?

Harri: Sanotaan, että se System 2 AI viittaa ehkä siihen, mitä puuttui. Me emme pelkästään tokikaan nyt tee tämmöistä analyyttisempää. Daniel Kahnemanin kirja Thinking, Fast and Slow, mikä se on suomeksi? Onko se "ajattelu, nopeaa ja hidasta", tai jotain sellaista? Niin se oli minun mielestäni erinomaisen hyvä kirja. Olen tykännyt siitä, ja se "mäppäytyy" tosi hyvin siihen tekoälykenttään. Se ei ollut pelkästään mitä me olimme tehneet, vaan miten minä näen, että oikeastaan tämä koko neuroverkkotyyppinen tekoäly on kehittynyt ajan kanssa. Se on hyvin pitkään ollut sitä systeemi 1 -ajattelua.

Minä olen ollut kiinnostunut tästä systeemi 2:sta pitkän aikaa. Esimerkiksi jo siellä ZenRoboticsissa tehtiin suunnitteluun pohjautuvaa juttua ja tällaista mallipohjaista vahvistusoppimista. Siinähän suunnittelu on tärkeässä roolissa. AlphaGo, AlphaZero, nämä toivat myös suunnittelua tänne neuroverkkojen maailmaan enemmän kuin aikaisemmin. Se on myös sellainen mahdollisuus, joka on auennut näiden kielimallien myötä. Niillä alkaa olla kyky semmoiseen analyyttisempään ajatteluun.

Se ei tarkoita sitä, etteikö meillä edelleen tarvitsisi olla uteliaisuutta siinä mallissa. Sen täytyy olla myös "curious AI". Sillä täytyy olla myös tämmöinen systeemi 1 -ajattelu mukana. Ei se muuten toimi. Mutta ne ovat olleet ehkä sellaisia komponentteja, sellaisia palikoita, jotka ovat olleet valmiimpia. Me olemme kehittäneet juttuja, jotka tuovat sitä systeemi 2 -ajattelua siihen järjestelmään. Emme ole vielä julkaisseet mitään. Tai siis olemme julkaisseet Aalto-yliopiston tutkijoiden kanssa yhteistyössä jotain. Mutta omasta työstämme emme ole vielä julkaisseet oikeastaan yhtään mitään. Mutta jänniä asioita on tapahtumassa. Tämmöinen ihmisen kaltainen oppiminen edellyttää myös semmoista systeemi 2 -ajattelua. Ei pelkästään päättely, vaan myös oppiminen edellyttää sitä, että systeemi 2 -ajattelu on hyvin hanskassa.

Mutta en tiedä onko se nyt vähän paradoksaalista, että sitten Curious AI:ssa ehkä päädyttiin kuitenkin enemmän tekemään systeemi 1 -tekoälyä, eikä niinkään sitä "curiosity"-juttua. Nyt System 2:ssa me ehkä kuitenkin vielä enemmän teemme systeemi ykköstä ja "curiosity" -tyyppisiä juttuja kuin systeemi kakkosta. Se on tärkeässä osassa kyllä. Mutta palikoita on monta. Ei se ole vain mikään yksi juttu. Se

mitä me kehitämme, on monen komponentin summa. Mutta tällaista ihmisen kaltaista kykyä omaksua uusia asioita ollaan tekemässä.

Arno: Nykymallit ovat jo tämmöisiä päättelymalleja. Eli voi ajatella, että siellä on jonkinlaista systeemi 2 -tyyppistä elementtiä osana. Mutta mitä niistä oikeastaan puuttuu vielä?

Harri: Niistä puuttuu muutamakin asia. Jos mennään systeemi kakkosesta vähän taaksepäin, niin yksi semmoinen juttu, jonka tiedetään olevan ihmisen päätöksenteossa ja ajattelussa hirveän tärkeä juttu, joka meillä Curious AI:ssa oli, ja näissä perinteisissä syvissä neuroverkoissa on hyvin hanskassa, on sellainen intuitiivinen tunnetason hahmotus tilanteesta. Ihmiselläkin tiedetään, että jos tulee aivovaurioita, jotka katkaisevat tietoisen mielen yhteyden omiin tunteisiin, niin päätöksenteko puuroutuu ihan tyystin. Ihmisestä tulee aika kyvytön toimimaan oikeassa maailmassa, kun semmoinen kokonaisuutta integroiva tunnesisältö jää pois.

Se on esimerkiksi semmoinen asia, jota me olemme nyt kehittämässä, ja joka toimii kohtuullisen hyvin. Eli kun puhutaan, että nämä tekoälyt ovat tämmöisiä, että siellä on vain logiikkaa, eikä ole tunnetta, niin ei sen tarvitsisi olla niin. Me olemme kehittämässä tällaista tunteellista tekoälyä, jolla on oma, englanniksi "gut feeling", jostain asiasta. Että ei voi loogisesti selittää, mutta on sellainen tuntuma, että tämä ei ehkä ole hyvä idea mennä tähän suuntaan. Tai että jotain tästä vielä puuttuu, mikähän se nyt on. Ja sitten se ohjaa sitä ajattelua ja päätöksentekoa oikeisiin suuntiin.

Toinen asia, joka puuttuu tällä hetkellä on uuden asian oppiminen lennosta. Sitä paikataan sillä, että nämä agentit tällä hetkellä kirjoittavat ikään kuin muistiin muistilappuja kaikista asioista. Sillä pystytään tekemään aika paljon, koska nämä ovat valtavan hyviä siinä. Niillä on ne kaikki muistilaput siellä muistissa. Tai ei muistissa vaan siinä promptissa. Siinä niille koko ajan kerrotaan, tai se muistilappu on siellä niiden näkyvissä, ja niin kauan ne pystyvät hyödyntämään sitä tietoa jollain tavalla. Mutta se ei ole kauhean tehokas tapa. Se ei ole skaalautuva tapa.

Jos joku on nähnyt tämän leffan Memento, itse en ole nähnyt muuten, mutta tiedän siitä jotain. Tiedän, että siinä on tyyppi, joka kirjoittaa muistilapuille ja käsiin ja ties vaikka minne muistiin kaikki asiat, koska tämä uuden asian päähän painamisen palikka päässä

on vaurioitunut. Onko se nyt retrograde amnesia, vai mikä se nyt on? Mikähän se on suomeksi? Retrogradinen amnesia.

Arno: Se se varmaan on.

Harri: Joo, nyt se on niin. Tällä hetkellä näillä kielimalleilla on tällainen retrogradinen amnesia. Se on semmoinen asia, jota me olemme kehittäneet ja panostaneet siihen, että miten ne pystyvät omaksumaan pidempikestoiseen muistiin asioita niin, että ne ovat sieltä löydettävissä. Sen jälkeen ne osaavat uusia juttuja. Tämä on yksi semmoinen iso pullonkaula tällä hetkellä, minkä takia näitä agentteja ei voi päästää vapaaksi minnekään. Jos vaikka niille on huolellisesti selitetty jotain asiaa, tai ne ovat monta kertaa törmänneet johonkin ongelmaan, niin ne saattavat yhtäkkiä lennosta unohtaa sen asian, koska juuri se muistilappu nyt unohtui pitää siellä promptissa.

Näistä on esimerkkejä maailmalta. Joku tekoäly rupeaa yhtäkkiä tuhoamaan kaikkia sähköposteja, koska unohtui se muistilappu, jossa sanottiin, että niitä piti vain katsoa. Piti vain listata ylös, että mitkä voisi tuhota. Ei pitänyt tuhota mitään, mutta se muistilappu jäi jonnekin, ja sinne meni sähköpostit.

Arno: Miksi jatkuvan oppimisen toteuttaminen tekoälyssä on niin vaikeata?

Harri: Tällä hetkellä nämä tekoälyfirmat ovat ottaneet sellaisen strategian, että ne tekevät yhden valtavan ison möhkömallin. Sen kouluttaminen maksaa hirveästi. Se on raskasta. Semmoisen möhkömallin jatkuva päivittäminen jokaiselle asiakkaalle, jokaiselle ihmiselle, jokaiselle käyttäjälle erikseen, ei ole vain taloudellisesti mahdollista. Siihen eivät kaikki maailman ydinvoimalat millään riitä. Se ei tapahdu.

Se mitä me olemme tekemässä, on että me teemme paljon pienemmillä malleilla töitä. Esimerkiksi minun läppärilläni voi pyöriä joku malli. Ne mallit ovat riittävän kevyitä, että niitä pystyy lennosta päivittämään vaikka joka iikka maailmassa itse. Tämä tämmöinen oppimisen hajauttaminen pienempiin yksiköihin on yksi semmoinen asia, joka uskon että tapahtuu. Siitä on paljon iloa, että sitten kerätään tietoa yhteen ja opitaan kootusti. Sitten se kaikki tieto valuu taas maailmalle. Mutta se on liian monoliittinen, liian raskas

mekanismi, että siihen voisi perustua tämmöinen jatkuva työtä tehdessä oppiminen yksinään.

Arno: Mutta eikö tämä ole semmoinen haaste, jota isot firmat myös haluavat ratkaista, ja Piilaaksossa on startupeja tähän liittyen? Miten te Suomessa pienellä tiimillä kilpailette näiden kanssa?

Harri: Me olemme pieniä mutta pippurisia. Tässä on taas tämä yrittämisen edellytys, katteeton usko omiin kykyihin ja mahdollisuuksiin. Se usko on jollain tavalla tarttuvaa sorttia myös. Sitten, kun menee sijoittajan kanssa juttelemaan ja selittämään, mitä me olemme saaneet aikaan, ja että meillä on tämmöinen visio, että tämän homman täytyy toimia joka läppärissä joka puolella maailmaa, eikä tarvita isoja ydinvoimaloita, kun esimerkiksi kesämökillä aurinkopaneelista riittää virtaa tähän touhuun, niin se semmoinen into voi tarttua sitten.

Onhan meillä näyttöä sillä tavalla, että meillä tulee tuossa elokuussa kolme vuotta täyteen, että olemme tehneet tätä hommaa. Ja olemme pystyneet tekemään tätä hyvin pienellä budjetilla. Ensin omarahoitteisesti, ja nyt sitten sijoittajilta kerättiin pieni määrä rahaa. Siinä ei ole vain muutama, vaan siinä on useita nollia eroa siihen, mikä on meidän budjetti versus tämmöisen "foundation labin" budjetti. Meillä tullaan siihen eri kulmasta. Olemme myös omalla työllämme pystyneet todistamaan, että tätähän pystyy tekemään tämmöiselläkin budjetilla. Jos me pystymme tekemään, niin sitten se on semmoinen teknologia, joka ehkä voikin oikeasti levitä ympäriinsä.

Isoilla labroilla on paljon resursseja. Mutta sitten se, että ne lähtevät ihan fundamentaalisesti erilaiselle trajektorille siinä, miten se bisnes rakennetaan. Eivät ne voi pivotoida sillä tavalla. Ei sellainen labra, joka on jo tuhlannut sata miljardia, voi todeta, että hetkinen, me keksimme tosi hyvän bisnesmallin. Siitä tulee ehkä sata miljoonaa vuodessa. Ei, kun ei tämä toimi. Unohdetaan tämä. Eli ne eivät tee sitä. Ne ovat jumissa sen oman jo kulutetun budjettinsa kanssa. Ne eivät voi enää perääntyä siitä.

Eric: No, palataan vähän siihen alun yleinen tekoäly ja AGI -keskusteluun. Jos te nyt saatte ratkaistua tämän jatkuvan oppimisen, ja mainitsit, että gut feeling on toinen asia, jonka parissa työskentelette, niin onko meillä silloin yleinen tekoäly? Ovatko ne ne puuttuvat palaset tällä hetkellä?

Harri: Tietyllä tavalla mielestäni meillä on jo yleinen tekoäly siinä mielessä miten me ymmärsimme sen, sanotaan vaikka parikymmentä vuotta sitten. Jos olisi jollekin mennyt selittämään, mitä kaikkea esimerkiksi tuo minun läppärini puksuttaa sinne ennen tätä haastattelua. Annoin pari ohjetta, että hei, muistattehan miten eilen harjoiteltiin tätä? Sitten, että muistakaa nyt, että mitä käytiin läpi. Mikä onnistui? Miten meni omasta mielestä? Retrospektiiviä pidettiin siinä. Sitten että no niin, nyt tänään kokeillaan uudestaan. Lykkyä pyttyyn. Minä menen nyt juttelemaan tuonne ihmisten kanssa. Katsotaan mitä siellä tapahtuu. Jos tämmöisen tarinan olisi kertonut jollekin, niin olisi oltu, että ei ole meidän vuosisadalla tapahtumassa tämmöistä.

Se on yleinen tekoäly jo, mutta kun se tulee lähelle iholle, niin sitten me huomaamme, että okei, se on yleinen tekoäly, mutta siitähän puuttuu tämä ja tämä. Ei tekoäly koskaan tule korvaamaan ihmistä, koska siltä puuttuu tämä, tämä ja tämä. Meillä on jatkuvasti tämmöinen, en tiedä onko se sitten itsesuojeluvaistoa vai mitä. Vaikka se on tullut jo, niin ihmiset eivät vieläkään näe sitä. Että okei, no tässähän se on. Tai jotkut ihmiset, sanotaan, ovat "in denial".

Arno: Niin. Pihatöitä se ei vielä pysty tekemään.

Harri: Ei, mutta luuletko, että ei tee viiden vuoden päästä?

Arno: En uskaltaisi laittaa päätäni pantiksi siitä.

Harri: Robotiikan kehittämisessä on se hardiksen kehittämisen puoli. Se on hidasta, mutta toisaalta siihenkin nyt laitetaan miljardeja rahaa. Ja loppujen lopuksi eivät ne sähkömoottorit ja muut itse asiassa maksa paljoa. Se äly on se pullonkaula kuitenkin ollut. Nyt se pullonkaula on poistumassa, niin ei ole mitään hyvää syytä miksei niitä pihatöitä tekisi viiden vuoden päästä joku robotti. Paitsi ehkä se, että jonkun näköinen henkireikä tässä nykymaailmassa on se tunne, että jotain hyötyä minustakin on. Että robotit eivät sentään korvaa minua siinä, kun leikkaan tätä ruohoa. Että robotit pois meidän pihalta. Minulla on tämä käsikäyttöinen vehje. Mutta se on meidän valintamme sitten. Meistä on kiva ehkä käydä kesämökillä laittamassa jotain. Hullun hommaahan se on. Mutta siellä kun saa lämmitettyä sitä kiviuunia, ja sitten on tupa lämpimänä, katselee

siinä liekkien loimua, niin sitten voi todeta, että okei, tämä on sillä tavalla lajityypillistä käyttäytymistä. Tämä on hyvä henkireikä.

Eric: Pihatöiden lisäksi olet hiljattain, tai ehkä aiemminkin, esittänyt ennustuksen, että tekoäly tulee ohittamaan ihmisen ei vain älytehtävissä, vaan myös esim. empatiakyvyssä. Jos voin haastaa tätä vähän. Jos ajatellaan näitä perinteisiä sovelluksia, kuten vaikka mainitsit AlphaGon, tällaisia pelejä, joissa tekoäly pärjää tosi hyvin. Ajattelen, että kaksi tärkeintä komponenttia, jotka meidänkin keskustelussa ovat nousseet esiin, joiden takia se toimii hyvin näissä, on, että me voimme simuloida sitä ympäristöä tehokkaasti. Voimme tuottaa paljon tekoälyllä, laittaa sitä pelaamaan itseään vastaan. Ja toisaalta me saamme verifioitavaa palautetta, että oliko tämä nyt onnistunut pelistrategia vai ei. Niin eikös tämä ole haasteena kaikissa tällaisissa tehtävissä, joissa on ihminen mukana siinä luupissa, että pystymmekö me simuloimaan ihmistä niin, että me saamme automaattisesti ajettua näitä "roll outeja" tai simulaatioita riittävän luotettavasti? Ja toisaalta, saammeko me verifioitavaa palautetta riittävän nopeasti, että tekoäly pystyy kehittymään näissäkin paremmaksi ja paremmaksi?

Harri: Joo, hyviä kysymyksiä. Vastaus tuohon, että voimmeko me simuloida ihmistä riittävän hyvin, on ei. Sen takia minä uskonkin itse sen tyyppisiin menetelmiin, jotka ovat huomattavan paljon näitä perinteisesti hyödynnettyjä menetelmiä tehokkaampia siinä mielessä, kuinka paljon dataa ne tarvitsevat oppiakseen. Sen takia olen ollut erityisen kiinnostunut juuri tästä mallipohjaisesta oppimisesta. Siinä on kyky ja mahdollisuus kerätä dataa riittävästi ja omista kokemuksista nimenomaan. Jos tekoäly tekee ihmisen kanssa töitä, oppii tuntemaan ihmisiä, niin sen jälkeen sillä on dataa, jonka pohjalta se voi alkaa oppia, että näinhän ne ihmiset tuntevat ja ajattelevat. Niillä voi itselläkin alkaa olla jotain tunteita ja ajatuksia. Se myös helpottaa sitä.

Empatiaa voi olla myös ilman, että itsellä on niitä samoja tunteita. Ne eivät ole sama asia. Empatiassa on erilaisia muotoja. En toki tarkoita sitä, että nyt tietokone jotenkin pystyy täysin sisäistämään ja tuntemaan samoja tunteita ja tuntemuksia kuin ihminen, välttämättä, jollemme me nyt erikseen yritä sitä rakentaa. Se ei välttämättä ole edes hyödyllistä. Mutta se, että riittävästi on ihmisten kanssa töissä, riittävästi alkaa tuntea niitä ihmisiä. Elää niiden ihmisten kanssa, jakaa niiden ilot ja surut. Se on edellytys sille empatiakyvylle minunkin mielestäni toki. Mutta me olemme hyvää vauhtia menossa siihen maailmaan, jossa tekoälyt ovat meidän arjessamme mukana, ovat meidän

elämässämme mukana. Se voi tuntua tosi pelottavalta ihmisistä, koska tämmöinen kylmä tekoäly tulee. Meistä tulee kaikista robotteja. Minä näen sen enemmän sillä tavalla, että ei, kun tekoälystä tulee paljon ihmismäisempi, siitä tulee empaattisempi. Siitä tulee lämminhenkinen. Se ei ole vain päälle feikattua ulkokuorta. Että se vain esittäisi empaattista. Siitä tulee oikeasti empaattinen.

Eric: Eli avain tähän on varmaan se, kuten mainitsit, datatehokkuus. Jos me ajattelemme sitä AlphaGota, niin siellä oli varmaan kymmeniä tuhansia simulaatioita vähintäänkin. Meidän pitäisi pystyä tekemään tämä monta kertaluokkaa nopeammin, jotta se on realistista.

Harri: Datatehokkuus on yksi asia. Toinen on ihan vain se kylmä "deployment", että kuinka paljon näitä tämmöisiä oppivia agentteja on oikeasti käytössä. En tupakoi, mutta tupakkiaskin kylkeen tai tällä lailla nopeasti laskettuna, kuinka paljon ihminen elämänsä aikana suunnilleen kerryttää kokemusta. Jos vuodessa on alle 10 000 tuntia ja ihminen elää tyypillisesti alle sata vuotta, eikä ole ihan koko aikaa hereillä. Jos nyt ajatellaan karkeistettuna, että ainakin alle miljoona tuntia kokemusta meillä on elämästä. Niin mitä jos meillä on miljoona tekoälyagenttia tuolla maailmalla, jotka kaikki ovat vuorovaikutuksessa oman ihmisensä kanssa, siinä omassa tiimissään, omassa yhteisössään? Periaatteessahan ne tunnissa kerryttävät yhtä paljon kokemusta kuin yksi ihminen elämänsä aikana.

Se ei mene noin yksinkertaisesti sen takia, että se ei ole pelkästään tämmöinen "horizontal study", vaan tarvitaan myös tämmöisiä "vertical studeja". Joskus tarvitaan riittävästi aikaa myös, että saadaan oppi perille. Mutta toisaalta, luultavasti nämä ovat tosi nopeita sisäistämään myös historian opetuksia ja muita. Eli minä uskon kyllä, että se ei todellakaan jää siitä kokemuksen ja datan puutteesta kiinni. Sen jälkeen, kun me pääsemme siihen pisteeseen, että hei, nämähän ovat oikeasti hyödyllisiä työkavereita, otetaanpas nämä mukaan siihen ja tähän ja tuohon. Nythän on nähty tämän vuoden alkupuolella jo esimerkiksi, kun tuli tämä, jota vissiin nykyään sanotaan OpenClawksi. Tuli tämä mediamyllytys siitä. Näitä alkaa olla oikeasti miljoonia tuolla maailmalla liikkeellä. Nyt jos nämä olisivat semmoisia, että ne oikeasti pystyisivät ihmismäisesti

oppimaan, yhtä tehokkaasti, ja pystyisivät asettumaan niihin ihmisen kenkiin nykyistä paremmin, niin ei se siitä datasta jäisi kiinni.

Eric: Joo. Olet myös ennustanut, että kaikki tietotyö tulee katoamaan 10 vuoden kuluessa, jos oikein muistan.

Harri: Ja siitä ennusteesta on varmaan jo muutama vuosi aikaa.

Eric: Niin.

Arno: Pari vuotta jäljellä enää vai?

Eric: Eikö tästä seuraa mahdollisesti valtavasti työttömyyttä? Haluatko sinä omalla työlläsi olla kiihdyttämässä tätä kehitystä? Vai miten sinä lähestyt tätä?

Harri: Jos minä nyt lopettaisin kokonaan nämä tekoälyhommat, niin se asia tapahtuisi luultavasti joka tapauksessa. Mitä minä haluan nyt olla tekemässä, on pystyä omalla työlläni vaikuttamaan siihen, että miten tekoäly meidän yhteiskuntaamme tulee. Tuleeko se semmoisena, että se on isojen firmojen takataskussa? Että tarvitaan miljardifirma ja sitten se miljardifirma kerää kaikki voitot. Vai onko meillä sellaista tekoälyä, joka mahdollistaa sen, että yksittäiset ihmiset ovat oman tekoälyelämänsä herroja, pystyvät ottamaan oman tekoälynsä, lähtemään kouluttamaan sitä, tekemään sen kanssa yhteistyötä. Ja ovat myös sen tekoälyn omistajia jollain tavalla.

Minun mielestäni on tosi tärkeää, että tajutaan se, että tämä muutos tulee tapahtumaan. Sitä ei voi pysäyttää. Jos me yritämme neliraajajarrutuksella olla tässä nyt, että en liiku tästä näin, tekoäly ei tänne tule, niin se tulee silti. Se muuttaa sitä, miten maailmantalous toimii. Se maailma tulee tänne kylään, jos ei nyt oikeasti lähdetä viljelemään porkkanoita tuonne jonnekin kesämökille. Siellä on siikaa järvessä ja perunamaasta aika hätäisesti tulee elanto. Jos ei sille tielle lähdetä, mitä en pidä kuitenkaan realistisena. Valitettavasti tämä maapallo ei elätä meitä ilman, että tuotetaan lannoitteita ja muuta. Ei vain ole mahdollista mennä takaisin päinkään.

Meidän on pakko ottaa lusikka kauniiseen käteen ja todeta, että tämmöinen mullistus tulee, mutta meidän pitää pystyä ohjaamaan sitä. Me haluamme, että se ei tule vain joillekin harvoille ja valituille. Se pitää olla ihmisten. Niiden hyötyjen pitää jakautua kaikille ihmisille, tai muuten tämä menee todella huolella vihkoon. Haluan siis olla mukana tekemässä sellaista tekoälyä, joka mahdollistaisi tämän. Ja haluan myös olla sanomassa ihmisille ja poliitikoille, mutta siis ihmisten täytyy uskoa se itse, poliitikot eivät sitä kenenkään puolesta tee, että meidän täytyy lähteä miettimään semmoisia ratkaisuja, että esimerkiksi millä tavalla yritykset ovat yhteisessä omistuksessa jollain tavalla. Onko se sitten joku eläkerahastotyyppinen vai mikä? Erilaisia ratkaisuja on monenlaisia.

Mutta tekoälyt pystyvät kohta pyörittämään firmoja itsenäisesti. Ja se, että minkälaista lainsäädäntöä me asetamme sille firmalle, että voiko firma potkia ihmisiä pihalle sen varjolla, että on tekoälyä? Sillä ei ole mitään merkitystä, koska niin kauan kuin me elämme markkinataloudessa, niin jos me hidastamme esim. ihmisten irtisanomista joiltain firmoilta, niin sitten vaan käy niin, että ne firmat menevät konkurssiin, koska tulee toisia firmoja, joissa on vain tekoälyjä, jotka toimivat paljon tehokkaammin. Se, että kokonaan pysäytettäisiin tämä kehitys, niin se ei vain tapahdu. On pakko miettiä todella uusia innovatiivisia ratkaisuja tähän. Semmoisia, jotka kuulostavat siltä, että tämähän on ihan kommunismia. Joo, okei, ehkä, mutta ehkä tämmöistä tekoälykommunismia, markkinaehtoista kommunismia. Eläkerahastot ja tämmöisethän ovat jo esimerkkejä tämän tyyppisistä ratkaisuista. Ei meidän tarvitse kokonaan keksiä pyörää uudestaan tässä, vaan me voimme hyödyntää niitä asioita ja rakenteita, joita nyt on. Mutta ei ole ihan hirveästi aikaa.

Eric: Joo. No, jos sinulla olisi nyt tänä päivänä käytössä superälykäs tekoäly, joka on kaikkia asiantuntijoitakin älykkäämpi, niin onko sinulla joku kysymys tai joku tehtävä minkä haluaisit laittaa sen tekemään tai ratkomaan?

Harri: Jos minulla olisi sellainen, niin se luultavasti itse pystyisi paremmin arvioimaan, että mitä sen kannattaisi tehdä. Se, minkä takia minun pitää antaa niille koko ajan tehtäviä, on osittain sen takia, että ne eivät ole riittävän fiksuja vekottimia vielä tällä hetkellä. Ne eivät itse hoksaa, mitä me haluamme ja tarvitsemme ja keskustele meidän

kanssa. Mutta enenevässä määrin, kun nämä tekoälyt kehittyvät viisaammiksi niin päästään pois semmoisesta mikromanageeraamisesta. Päästään enemmän miettimään isoja linjoja.

Kyllä minua kiinnostaisi erityisesti se, että miten me järjestämme tämän yhteiskunnan sillä tavalla, että siellä on ihmisten hyvä elää, ja että me pystymme hyödyntämään tätä tekoälyn murrosta. Eikä sillä tavalla, että me rusennumme sen alle.

Eric: Pitääkö meidän kuitenkin ihmisinä olla määrittämässä se, mitä on se hyvä elämä? Vai ajatteletko, että tekoäly voi tämänkin ratkaista meidän puolestamme?

Harri: Kyllä minä luulen, että ihmiset itse ovat aika hyviä miettimään sitä, että mikä on niiden mielestä hyvä elämä. En lähtisi liikaa tämmöiseen "minä tiedän paremmin" -tyyppiseen hommaan. Sillä tavalla olen varmasti liberalisti, että uskon siihen, että ihmisten lajityypillistä käyttäytymistä on aika vapaasti tuolla noin niin kuin lyödä ihan itse pää siihen puuhun ja oppia siitä. Sillä tavalla me olemme kuitenkin onnellisempia, että me emme ole vain paapottavina jossakin. Jos mietitään WALL-E elokuvaa, niin se on omanlaisensa dystopia, että me olemme vain kuluttajia ja paapottavina jossakin. Kyllä minä näkisin, että tekoäly mahdollistaa myös semmoisen kasvamisen ihmisenä, hyppäämisen pois semmoisesta suorittamisesta ja oravanpyörästä. Pitää pysähtyä vähän miettimään, että hetkinen, mikäs tässä elämässä nyt olikaan tärkeää ja mitä me haluamme ja muuta. Se on se, mitä sanoisin, että jää ihmiselle vielä kuitenkin. Määritellä se, että mitä ihmisyys on. Tekoälyt saavat sitten määritellä mitä se tekoälykkyys on. Mutta kyllä minä kuitenkin vielä luotan siihen, että ihmiset saavat määritellä tämän ihmisyytensä.

Eric: No, puhuttiin näistä dystopioista. Jotkut ovat esittäneet, että tekoälyllä voi olla myös massiivisen negatiiviset ja katastrofaaliset seuraukset. Puhutaan tällaisesta P(doom):ista, eli tuhon tai eksistentiaalisen riskin todennäköisyydestä. Mikä on sinun P(doom), Harri Valpola?

Harri: Se on onneksi aika lähellä nollaa. Nyt täytyy ehkä sitten taas miettiä, mitä tässä tarkoitetaan sillä "doomilla". Minä tarkoitan sellaista tilannetta, jossa ihmiskunnasta ei jäisi kauheasti mitään jäljelle, että kaikki menisi. Mutta mikä nyt sitten vielä lasketaan

ihmiskunnaksi? Vaikkapa tuhannen vuoden päästä, jos heitetään nyt riittävän pitkä aika eteenpäin. Itsehän uskon paljon lyhyempiin aikajänteisiin. Mutta tässä nyt julkisesti voidaan ottaa turvallisesti tuhannen vuoden aikaperspektiivi. Kuinka iso osa ihmisistä on edelleen biologisia ihmisiä, jotka toimivat tällä lailla, kuten englanniksi kutsuttaisiin "meatware"? Me joudumme nyt esittämään sellaisia kysymyksiä itsellemme myös, mikä on mielestäni erittäin tervettä ihmiskunnan itsereflektiota, että mikä on tärkeää ihmisyydessä. Onko se meidän geneettinen koodi? Onko se se biologinen implementaatio, vai onko se jotkut henkisemmät arvot? Joku semmoinen asia, joka me itse asiassa pystytään aivan hyvin jakamaan näiden tekoälyjen kanssa.

Kyllä minä näen, että meidän ihmislajimme tulee tekemään tämmöisen metamorfoosin tässä, että tästä toukasta kuoriutuu kohta se perhonen tai mikä lie sieltä tulee. Saa nähdä mikä sieltä tulee, mutta toivottavasti ei mikään päiväperho. Näkisin, että se mikä sieltä kuoriutuu, niin erittäin mielelläni olisin itse näkemässä, että miltä se näyttää ja ehkä mukana siinä touhussa, että mihin suuntaan tämä tulee menemään. Mutta en usko siihen, että tuhannen vuoden päästä on biologinen ihminen, joka vie ihmisyyttä eteenpäin tuolla jossakin galaksissa.

Eric: Näistä olisi kiehtovaa jutella paljon pidempäänkin, mutta luulen, että meillä alkaa aika olla lopussa. Suuret kiitokset, oli mahtavaa päästä keskustelemaan sinun kanssasi.

Arno: Kiitos.

Harri: Kiitos teille.