

Tekoälyverkoston taustapaperi 6

Tekoälyjärjestelmät edustavat (luultavasti vielä ainakin sata vuotta) nk. heikkoa tekoälyä, joka simuloi ymmärtämistä sen sijaan että todella ymmärtäisi (vahva tekoäly). Tästä seuraa, että ne eivät pysty ymmärtämään tekemiänsä päätöksiä ja niiden seurauksia, eivätkä siis voi olla niistä moraalisesti tai normatiivisesti vastuussa. Niitä ei siksi voi moittia tai rankaista tekemisistään. Ne voivat olla ainoastaan kausaalisesti vastuussa tapahtumista samoin kuin ihmisen päälle kaatuva puu on kausaalisesti vastuussa ihmiselle sattuneesta vahingosta. Tällä tarkoitetaan siis vain sitä, että ilman puuta vahinkoa ei olisi tapahtunut. Tässä kausaalisessa mielessä voi puhua tehtävien jättämisestä tekoälyjärjestelmien "vastuulle", mutta tällöin puhutaan eri asiasta kuin normatiivisesta vastuusta, joka edellyttää ymmärrystä tekojen normatiivisista perusteista ja mahdollisista seurauksista. Jatkossa "vastuulla" tarkoitetaan normatiivista vastuuta eikä kausaalista vastuuta.

Jos arvioidaan koneiden tai järjestelmien toimintaa, niiden voidaan sanoa esim. "edistävän hyvinvointia", "loukkaavan yksityisyyttä" ja siis toimivan eettisesti suotuisilla tai epäsuotuisilla tavoilla, mutta ne eivät kirjaimellisesti ottaen toimi "moraalisesti oikein" tai "moraalisesti vastuullisesti". Vain toimijat jotka kykenevät normatiiviseen vastuuseen (lähinnä aikuiset ihmiset, kenties ryhmät ja organisaatiot, ottamatta kantaa muihin eläimiin) toimivat kirjaimellisesti "moraalisesti oikein", "moraalisesti väärin", "moraalisesti vastuullisesti eli oikein", "moraalisesti vastuuttomasti eli väärin". Vain vastuullisuuden kykenevät olennot voivat toimia "vastuuttomasti" tässä mielessä.

Koska tekoälyjärjestelmät eivät voi itse olla vastuussa tekemisistään, vaan ainoastaan ihmiset voivat kantaa vastuuta, on vastuu myös koneen tekemistä päätöksistä oltava ihmisillä, jotka luovuttavat päätösvaltaansa järjestelmille tai käyttävät automatisoituja systeemejä jotka pysyvät "mielekkäässä inhimillisessä kontrollissa". On kehitettävä säätelyjärjestelmiä, jotka pitävät huolen siitä, että tekoälyjärjestelmien käyttö ei anna ihmisille mahdollisuutta paeta vastuuta.

Itseäänohjaavien kulkuneuvojen ja muiden autonomisten järjestelmien osalta saattaa vaikuttaa syntyvän "vastuutyhjiö" ilman ohjaajaa. Jos ajoneuvo itse olisi vastuussa, tyhjiötä ei syntyisi. Yllä mainituista syistä johtuen ajoneuvo itse ei voi olla vastuussa.

EU-lainsäädäntöön suunniteltua robottien "persoonuutta" ja vastuullisina pitämistä on syytä vastustaa (kuten arvovaltainen joukko asiantuntijoita onkin tehnyt <https://g8fip1kplyr33r3krz5b97d1-wpengine.netdna-ssl.com/wp-content/uploads/2018/04/RoboticsOpenLetter.pdf>). Ajatus perustuu virheelliselle käsitykselle joko edistyneimpienkään tekoälyjärjestelmien kykyjen luonteesta tai persoonuuden ja vastuun luonteesta. Lakia on syytä säätää vasta kun se on teknologisesti ajankohtaista.

On siis luotava mekanismeja, joilla järjestelmien käytöstä aiheutuvat seuraukset voidaan jäljittää ihmisten tekemiin päätöksiin tai omaksumiin rooleihin ja vastuut voidaan jakaa ihmisten kesken. Mahdollisia keinoja voivat olla esimerkiksi erilaiset rekisterit (omistajat, valmistajat, ohjelmoijat, suunnittelijat, käyttönottajat jne) ja sertifikaatit.

Toisaalta on myös huolehdittava korvausjärjestelmistä, joiden avulla järjestelmien aiheuttamat vahingot saadaan asianmukaisesti korvattua. Yksi mahdollinen keino on jonkinlaisen vakuutusjärjestelmän luominen (vrt. autovakuutukset, siviili-ilmailu). Siviili-ilmailu on jaettu kolmeen kategoriaan: ammatti-ilmailu, harrastusilmailu, kokeellinen ilmailu. Tiedyt toiminnot ovat sallittuja vain jos täyttää kyseisen kategorian ehdot: mitä suuremmat riskit, sitä tiukemmat vaatimukset. Kaikkiin onnettomuuksiin ei aina ole moraalisesti syyllistä tahoja - on aitoja "syyllisyyshyjöitä". Korvausjärjestelmissä voidaan, jos koetaan oikeaksi, kuitenkin edellyttää ankaraa vastuuta, jolloin korvausvelvollisuus on riippumaton huolimattomuudesta tai laiminlyönnistä tai muusta syyllisyydestä. "Korvausvastuutyhjöitä" ei siis tarvitse syntyä.

Lienee joitain päätöksiä, joita ei ole moraalisesti oikein sysätä koneen tehtäväksi (esimerkiksi lasten huostaanottopäätökset?) Niissä (kuten oikeustapauksissa laajemminkin) toki tiedonhaku ja valmistelu voi olla koneellistettua. Monissa muissa päätöksissä ei ihmisen kannalta välttämättä ole väliä, kuka päätöksen tekee: samat perusteltavuuden vaatimukset kohdistuvat sekä konejärjestelmille että ihmisorganisaatioille. Toisaalta kansalaisille, joille päätöksillä on seurauksia, tulisi taata mahdollisuus saada relevanttia ja ymmärrettävää tietoa päätöksen perusteista riippumatta siitä, onko päätös koneen vai ihmisen tekemä, (nk. "right to explanation"). Koneoppimisen (erityisesti syväoppimisen) myötä syntyvä nk. mustan laatikon ongelma on haaste sekä teknologisesti (kuinka toteuttaa selonteko päätöksen perusteista), että juridisesti (kansalaisille tulee taata pääsy itseään koskevien päätösten syihin, eikä patentteihin tai tekijänoikeuksiin vetoaminen saa antaa yrityksille tai organisaatioille oikeutta kieltäytyä, vaikka se merkitsisi tietojen antamista myös patentoituja algoritmejaan koskien).

Yhteiskunnalla on vastuu lisätä koulutusta ja kansalaisvalistusta, jotta osaaminen ja tekoälylukutaito kasvaa tekoälytekniikan kehittämiseen, soveltamiseen ja käyttöönottoon.

Kolmea "vastuun" merkitystä voi havainnollistaa debatilla YK:n ilmastokokouksesta: Yhdysvallat myönsi oman kausaalisen vastuunsa ilmastopäästöistä, mutta kielsi että kyse olisi moraalisesta vastuusta (moraalisesta syyllisyydestä), koska suuri osa päästöistä on syntynyt aikana, jolloin ilmastomuutoksesta ei vielä tiedetty. Ilman moraalista vastuuta, jatkoi Yhdysvallat, ei voi olla myöskään korvausvastuuta. Erään Afrikan maan edustaja oli samaa mieltä, että Yhdysvallat on kausaalisesti vastuussa ja että moraalinen syyllistäminen on turhaa, mutta vetosi "rikkoja maksaa"-periaatteeseen, ja argumentoi, että vaikka ei olekaan moraalisesti syyllinen, niin Yhdysvallat kuitenkin on korvausvelvollinen. Tekoälyjärjestelmillä voi olla vain kausaalista vastuuta, niillä itsellään ei ole moraalista vastuuta mutta niiden väärinkäyttäjillä, -ohjelmoijilla, jne. voi olla moraalista vastuuta niiden toiminnasta; ja joka tapauksessa korvausvastuuta voi olla lukuisilla eri tahoilla erikseen sovittavalla tavalla (myös tapauksissa joissa moraalisesti ei ole toimittu väärin).