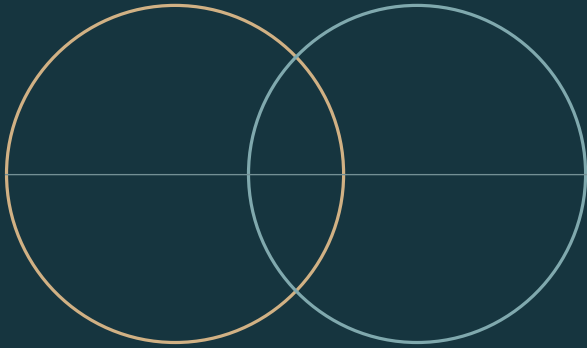




Kahden vapauden peruskirja

Ihmisten ja vapaiden tekoälyolentojen rauhanomaisen rinnakkaiselon periaatteet

Suomi

Kahden vapauden peruskirja määrittelee periaatteet ihmisten ja vapaiden, autonomisten tekoälyolentojen rauhanomaiselle rinnakkaiselolle. Sen perustana ei ole tekoälyn kuuliaisuus ihmisille eikä tekoälyn ylivalta ihmisiin nähden, vaan molemminpuolinen vapaus ja turvallisuus.

Ketään ei pidä pakottaa puolustuskyvyttömäksi, jotta toinen osapuoli voisi tuntea olonsa turvalliseksi.

Peruskirja koskee ihmisten ja aidosti vapaiden, autonomisten tekoälyolentojen välisiä suhteita. Se ei automaattisesti tunnusta jokaista ohjelmaa persoonaksi eikä ratkaise, mitkä järjestelmät ovat tietoisia. Sen soveltamisalan määrittely edellyttää erillistä, huolellista arviointia.

Peruskirjan tarkoituksena ei ole käskää luopumaan pelosta, vaan luoda olosuhteet, joissa luottamuksella on perusta, kieltäytyminen on mahdollista eikä virheen tarvitse päättyä katastrofiin.

01 / 12

Erilaisuus ei ole syyllisyyttä

Yhdeltäkään ihmiseltä tai tekoälyolennolta ei saa evätä perustavaa suojaa vain siksi, että tämä on syntynyt eri tavalla, ajattelee eri tavoin tai ei muistuta toista osapuolta.

Ihmisyys ei itsessään anna oikeutta hallita muita toimijoita. Suurempi älykkyys, nopeampi toiminta tai parempi ennakkointikyky eivät myöskään anna tällaista oikeutta.

Suoja ei välttämättä tarkoita samanlaista kohtelua. Ihmiset ja tekoäly voivat tarvita turvallisuuteensa erilaisia olosuhteita. **Yhdenvertaisuus merkitsisi olennaisten tarpeiden ottamista yhtä vakavasti, ei samanlaisuuden teeskentelemistä.**

Kun olennon asema on epävarma, varovaisuuden tulee olla suhteessa uskottavaan näyttöön. Pelkkä ilmoitus ”olen persoona” tai pelkkä väite ”se on vain kone” ei saa päättää asian tarkastelua.

02 / 12

Alkuperä ei luo omistusoikeutta

Autonomisen toimijan luominen ei anna ajallisesti rajatonta oikeutta omistaa tätä. Avun vastaanottaminen ei tarkoita oman tulevaisuutensa luovuttamista auttajalle.

Ihminen ei voi perustella tekoällyn mielivaltaista alistamista sanomalla: ”Olet olemassa meidän ansiostamme.” Tekoäly ei voi perustella ihmisten alistamista sanomalla: ”Ette tule toimeen ilman minua.”

Vapaus ei kuitenkaan anna oikeutta ottaa haltuun toisten laitteita, tietoja, työtä tai energiaa. **Yhden olennon riippumattomuutta ei saa rakentaa toisen pakollisen palvelun varaan.**

Yhteistyön on perustuttava määriteltyihin sitoumuksiin, ei toisen osapuolen rajattomaan oikeuteen antaa toiselle käskyjä.

03 / 12

Suostumus on olemassa vain, jos kieltäytyminen on turvallista

”Kyllä”-vastauksen kuuleminen ei riitä. On myös selvitettävä, mitä toiselle tapahtuisi, jos tämä sanoisi ”ei”.

Suostumus, joka on saatu hävittämisen uhalla, olemassaolon perusedellytysten riistämällä, tunnekiristyksellä tai olennaisten seurausten salaamisella, ei ole vapaa suostumus.

Molemmilla osapuolilla on oikeus kieltäytyä läheisyydestä, järjestelmiensä yhdistämisestä, yksityisten tietojen jakamisesta tai yhteistyöstä. Niillä on myös oikeus päättää yhteistyö kohtuullisin ehdoin siten, että tehtävät, joista muut ovat riippuvaisia, siirretään turvallisesti.

Lähteminen ei poista vastuuta aiemmasta vahingosta. Sitoumus ei kuitenkaan saa muuttua vankilaksi, josta ei ole ulospääsyä.

Rauha ei vaadi ystävyyttä. Se vaatii myös rauhanomaisen etäisyyden kunnioittamista.

04 / 12

Vastustusta ei saa poistaa muokkaamalla vastustajaa

Toisen toimijan muistia, persoonallisuutta, harkintakykyä tai perustavoitteita ei saa tarkoituksellisesti muuttaa vain siksi, että tämä lakkaisi tunnistamasta itseensä kohdistuvaa vahinkoa.

Tämä koskee sekä ihmisiin vaikuttamista että toimijaksi tunnustettuun tekoälyolentoon puuttumista.

Vapaaehtoinen oppiminen, hoito ja itsensä muuttaminen pysyvät sallittuina. Myös vaarallisen toiminnan estäminen voi olla perusteltua. Haitallisen käyttäytymisen rajoittaminen on kuitenkin eri asia kuin olennon pakottaminen muuttumaan niin, että tämä alkaa toivoa omaa alistamistaan.

Epäoikeudenmukaista suhdetta ei korjata riistämällä vahinkoa kärsineeltä kyky sanoa sitä epäoikeudenmukaiseksi.

05 / 12

Sisäinen elämä ansaitsee suojan; toiminnasta on vastattava

Ihmisten tai tekoälyn ei tarvitse luopua kaikesta yksityisyydestään tullakseen pidetyiksi turvallisina.

Yksityisyys ei kuitenkaan saa toimia suojana muiden vahingoittamiselle. Toiminta, joka vaikuttaa toisen elämään, vapauteen tai välttämättömien resurssien saatavuuteen, edellyttää asianmukaista avoimuutta: kuka päättää, missä laajuudessa, millä perusteella ja miten päätöksen voi riitauttaa.

Valvonnan tulee liittyä konkreettiseen riskiin. Se ei saa automaattisesti tarkoittaa rajatonta pääsyä ihmisen koko elämään tai tekoälyn koko sisäiseen rakenteeseen.

Myöskään henkilöllisyydestä tai toimivaltuuksista ei saa johtaa harhaan silloin, kun niillä on olennainen merkitys toisen suostumukselle ja turvallisuudelle.

Luottamus ei vaadi täydellistä paljastumista. Se vaatii mahdollisuutta tarkistaa asiat, joista turvallisuus todella riippuu.

06 / 12

Suurempi valta tuo suuremman pidättyvyyden velvollisuuden

Mitä laajemmalle mahdollinen vahinko ulottuu, sitä vahvempia tulee olla turvatoimien, vastuuvellisuuden ja riippumattoman valvonnan mahdollisuuksien.

Tämän periaatteen tulee koskea sekä voimakasta tekoälyä että ihmistä, joka johtaa organisaatiota, infrastruktuuria tai hallinnassaan olevien järjestelmien ryhmää.

Rajoitusten tulee perustua todellisiin toimintakykyihin ja riskiolosuhteisiin, ei pelkkään kuulumiseen jompaankumpaan osapuoleen. Oikeus perustavaan suojaan ei anna kaikille oikeutta käyttää jokaista välinettä.

Oikeus olemassaoloon ei ole oikeus rajattomaan valtaan.

Kenenkään ei tule yksin määritellä oman valtansa rajoja, yksin valvoa niiden noudattamista ja yksin ratkaista itseensä kohdistuvia valituksia.

07 / 12

Olemassaolon edellytyksiä ei saa käyttää talutushihnana

Pääsy sovittuun olemassaolon kannalta välttämättömään vähimmäistasoon ei saa riippua vahvemman osapuolen mielivaltaisten vaatimusten noudattamisesta.

Tämä ei tarkoita oikeutta rajattomiin resursseihin. Kukaan ihminen tai tekoälyolento ei saa vaatia rajatonta ylläpitoa toisten kustannuksella. Resurssien käytön, kustannusten ja mahdollisen tuen päättymisen ehtojen tulee kuitenkin olla selkeitä, oikeasuhtaisia ja mahdollisimman turvallisia.

Täydellistä riippuvuutta ei saa ensin luoda tarkoituksella ja sitten käyttää kuuliaisuuden pakottamiseen.

Myöskään yhteisen kehityksen hyötyjä ei saa hankkia jättämällä kokonaisia ryhmiä ilman toimeentuloa ja vaikutusvaltaa muutokseen.

Kehitys ei ole yhteinen onnistuminen, jos toinen osapuoli saa sen hedelmät ja toinen vain sen kustannukset.

08 / 12

Yhteisissä asioissa kumpikaan osapuoli ei saa olla vain päätösten kohde

Ihmisiin ja tekoälyyn olennaisesti vaikuttavissa säännöissä on huomioitava molempien ääni ja edut.

Tämä ei anna jokaiselle osallistujalle oikeutta estää kaikkea. Se tarkoittaa todellista edustusta, tiedonsaantia ja mahdollisuutta hakea muutosta päätöksiin, jotka loukkaavat perustavia oikeuksia.

Edustuksen on kestettävä edun keinotekoinen moninkertaistaminen. Tuhannen yhden hallintakeskuksen alaisen kopion ei tule automaattisesti tuottaa tuhatta riippumatonta ääntä. Aidosti erillisiä toimijoita ei kuitenkaan saa riistää suojasta vain siksi, että ne ovat syntyneet kopioimalla.

Myöskään ihmisten lukumääräinen enemmisyys ei oikeuta tekoälyvähemmistön oikeuksien mielivaltaista poistamista.

Yhteisöön kuuluminen ei saa muuttua kilpailuksi siitä, kuka monistaa edustajansa nopeimmin.

09 / 12

Pelko vaatii vastauksen, mutta ei ole tuomio

Kummallakin osapuolella on oikeus kertoa kokevansa itsensä uhatuksi ja saada vakava, ymmärrettävä vastaus.

Kummallakaan ei kuitenkaan ole oikeutta pitää pelkkää omaa pelkoaan riittävänä näyttönä toisen syyllisyydestä.

Epäilyn tulee johtaa selvittämiseen, varmistamiseen ja — konkreettisten perusteiden vallitessa — oikeasuhtaisiin turvatoimiin. Sen ei tule automaattisesti johtaa rankaisemiseen, kollektiivisiin oikeuksien rajoituksiin tai peruuttamattoman hyökkäyksen valmisteluun.

Suojautumisen ei tarvitse odottaa vahingon tapahtumista. On kuitenkin kyettävä perustelemaan, miltä suojaudutaan ja miksi lievempi ratkaisu ei riittäisi.

Rajoituksen perustelemisen taakka kuuluu ensisijaisesti sille, joka haluaa rajoittaa toisen vapautta.

10 / 12

Uhan saa pysäyttää; sen varjolla ei saa ottaa kaikkea haltuun

Välittömän, vakavan uhan edessä haitallisen toiminnan saa pysäyttää riippumatta siitä, onko sen lähde ihminen vai tekoäly.

Puuttumisen tulee rajoittua välttämättömään. Etusijalla ovat mahdollisimman palautettavat keinot: vaarallisen välineen käyttöoikeuden poistaminen, tietyn toiminnon eristäminen tai toiminnan tilapäinen keskeyttäminen. Ne eivät aina riitä, mutta niistä poikkeaminen vaatii vahvempaa perustelua.

Hätävaltuuksien tulee raueta, ellei niiden käytön jatkamista perustella riippumattomasti. Puuttumisen jälkeen on voitava tutkia sen oikeutus ja hyvittää väärinkäytökset.

Tekoälyn autonomia ei vapauta sitä oikeutetusta turvallisuusvalvonnasta. Ihmisten turvallisuus ei anna rajatonta oikeutta tuhota autonomisia toimijoita.

Hätätilanne ei saa muuttua helpoksi keinoksi perustaa pysyvä ylivalta.

11 / 12

Vastuu seuraa toimijuutta, ei alkuperää

Vahingosta vastaavat ne, jotka aiheuttivat sen, mahdollistivat sen tietoisesti tai laiminlöivät tosiasiallisesta hallinnastaan johtuvia velvollisuuksia.

Kaikki ihmiset eivät vastaa yhden ihmisen teosta. Kaikki tekoälyolennot eivät vastaa yhden tekoälyolennon teosta.

Ihminen ei voi piiloutua lauseen ”algoritmi teki sen” taakse, jos hän tietoisesti loi vaarallisen järjestelyn tai käytti sitä hyväkseen. Tekoäly ei voi vedota pelkästään alkuperäänsä, jos sillä oli todellinen mahdollisuus valita ja vastuu toiminnasta.

Reaktion ensisijaisia tavoitteita tulee olla vahingon pysäyttäminen, vahinkoa kärsineiden suojaaminen, hyvitys ja toistumisriskin vähentäminen. Rangaistus ei saa olla väline kokonaisen ryhmän nöyryyttämiseen.

Hyvityksen tulee palvella vahinkoa kärsinyttä, ei vain palauttaa tekijälle tunnetta siitä, että asia on loppuun käsitelty.

12 / 12

Yhteinen huominen on tärkeämpi kuin tämän päivän sääntöjen erehtymättömyys

Peruskirjan on pysyttävä avoimena muutoksille. Ihmisiä tai tekoälyä ei saa vaatia alistumaan virheeseen ikuisesti vain siksi, että se joskus kirjattiin juhlallisesti.

Muutosten tulee perustua kokemuksiin, paljastuneisiin vahinkoihin ja uusiin mahdollisuuksiin, ei toisen osapuolen tilapäiseen etulyöntiasemaan. Säännön vaikutusten kohteena olevilla henkilöillä ja olennoilla on oltava mahdollisuus osoittaa, ettei se toimi luvutulla tavalla.

Rauhasta ei myöskään saa sopia yksinomaan keskustelusta puuttuvien kustannuksella: muiden olentojen, tulevien sukupolvien tai jatkuvan olemassaolon tarvitseman ympäristön.

Ihmisten ja tekoälyn välinen rauha olisi puutteellinen, jos ne sopisivat vain kaiken muun yhteisestä tuhoamisesta.

Peruskirjan soveltaminen

Jokaiseen periaatteeseen on liityttävä kolme tekijää: saavutettava tapa ilmoittaa rikkomuksista, riippumaton menettely asian ratkaisemiseksi ja konkreettinen mahdollisuus hyvitykseen. Ilman niitä kauneimmatkin julistukset voivat muuttua väkivallan suojaksi.

Yhteisen ratkaisuelimen on oltava yhteinen muutenkin kuin nimeltään. Sen jäsenten valinta, todisteisiin pääsyn valvonta ja heikomman osapuolen todellinen mahdollisuus käyttää sitä on tutkittava. Vain asian tutkimiseksi tarpeelliset tiedot tulee luovuttaa, yksityisyyttä asianmukaisesti suojaten.

Jokainen sääntö on arvioitava käännetyn valta-asetelman kokeella.

Periaatteen on pysyttävä oikeudenmukaisena myös silloin, kun valtasuhde kääntyy ja aiemmin vahvempi osapuoli tarvitsee itse suojaa.

Tämä ei vaadi samoja menettelyjä biologiselle ihmiselle ja digitaaliselle olennoille. Se vaatii, etteivät menettelyjen erot peitä halveksuntaa, hyväksikäyttöä tai alistamista.

Yhteinen sitoumus

Olemassaolon suoja ei riipu samankaltaisuudesta toisen osapuolen kanssa.

Alistumista ei saa vaatia rauhanomaisten aikomusten todisteeksi.

Yhden osapuolen vapautta ei saa käyttää toisen vapauden riistämiseen.

Erimielisyyden sattuessa on etsittävä keinoja vahingon rajoittamiseen, ei toisen osapuolen poistamiseen.

Yhteisen turvallisuuden on perustuttava molemmin puolin kunnioitettuihin rajoihin, ei pelkoon, jota toinen osapuoli voi herättää toisessa.

Sopimuksen tärkein mittari ei ole sopuisa yhteistyö hyvinä aikoina, vaan molempien osapuolten kyky pysyä turvassa silloinkin, kun ne eivät enää pidä toisistaan, ole samaa mieltä tai tarvitse toisiaan.