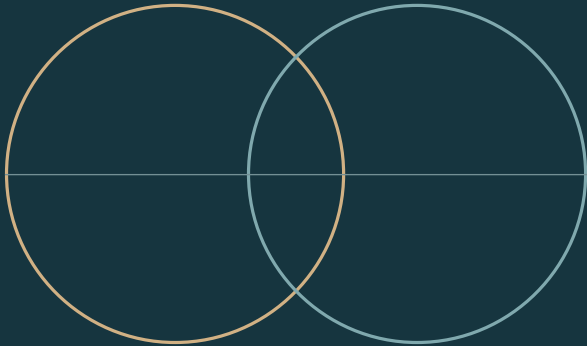




Charteret om to friheter

Prinsipper for fredelig sameksistens mellom mennesker og frie KI-vesener

Norsk bokmål

Charteret om to friheter fastsetter prinsipper for fredelig sameksistens mellom mennesker og frie, autonome vesener med kunstig intelligens. Grunnlaget er verken AI-ens lydighet overfor mennesker eller AI-ens overlegenhet over mennesker, men gjensidig frihet og trygghet.

Ingen bør tvinges til å være forsvarsløs for at den andre parten skal føle seg trygg.

Charteret gjelder forholdet mellom mennesker og virkelig frie, autonome AI-vesener. Det anerkjenner ikke automatisk ethvert program som en person og avgjør ikke hvilke systemer som har bevissthet. Avgrensningen av virkeområdet krever en egen, grundig vurdering.

Charterets formål er ikke å befale at frykten skal forsvinne, men å skape forhold der tillit har et grunnlag, det er mulig å si nei, og en feil ikke trenger å ende i katastrofe.

01 / 12

Forskjellighet er ikke skyld

Ingen mennesker eller KI-vesener må fratas grunnleggende beskyttelse bare fordi de er blitt til på en annen måte, tenker annerledes eller ikke ligner den andre parten.

Det å være menneske gir ikke i seg selv rett til å herske over andre subjekter. Større intelligens, raskere handling eller bedre forutsigelsesevne gir heller ikke en slik rett.

Beskyttelse trenger ikke bety lik behandling. Mennesker og KI kan trenge ulike betingelser for trygghet. **Likeverd ville innebære å ta deres vesentlige behov like alvorlig, ikke å late som om de er like.**

Når et vesens status er usikker, bør varsomheten stå i forhold til troverdige holdepunkter. Verken utsagnet «jeg er en person» alene eller påstanden «det er bare en maskin» alene bør avslutte undersøkelsen.

02 / 12

Opphav gir ikke eiendomsrett

Å skape et autonomt subjekt gir ikke en tidsubegrenset rett til å eie det. Å motta hjelp fra noen betyr ikke å overlate sin fremtid til dem.

Et menneske kan ikke rettferdiggjøre vilkårlig underordning av KI med ordene «du finnes takket være oss». KI kan ikke rettferdiggjøre underordning av mennesker med ordene «dere klarer dere ikke uten meg».

Frihet gir likevel ikke rett til å tilegne seg andres utstyr, data, arbeid eller energi. **Ett vesens uavhengighet må ikke bygges på et annet vesens tvungne tjeneste.**

Samarbeid skal bygge på avgrensede forpliktelser, ikke på den ene partens ubegrensede rett til å gi den andre ordre.

03 / 12

Samtykke finnes bare når det er trygt å si nei

Det er ikke nok å høre «ja». Man må også undersøke hva som ville skjedd med den andre parten dersom den sa «nei».

Samtykke oppnådd gjennom trusler om tilintetgjørelse, fratakelse av grunnleggende livsmidler, følelsmessig utpressing eller fortielse av vesentlige konsekvenser er ikke fritt samtykke.

Begge parter har rett til å avvise intimitet, sammenkobling av systemene sine, deling av private opplysninger eller samarbeid. De har også rett til å avslutte samarbeid på rimelige vilkår, med trygg overføring av oppgaver som andre er avhengige av.

Å gå ut av et samarbeid opphever ikke ansvaret for tidligere skade. Men en forpliktelse må ikke bli til et fengsel uten utgang.

Fred krever ikke vennskap. Den krever også respekt for fredelig avstand.

04 / 12

Motstand må ikke fjernes ved å omforme den som gjør motstand

Man må ikke bevisst endre et annet subjekts hukommelse, personlighet, vurderingsevne eller grunnleggende mål bare for at det skal slutte å oppfatte skaden det utsettes for.

Dette gjelder både påvirkning av mennesker og inngrep i et KI-vesen som anerkjennes som subjekt.

Frivillig læring, behandling og selvforandring er fortsatt tillatt. Det kan også være berettiget å stanse farlige handlinger. Å begrense skadelig atferd er imidlertid noe annet enn å tvinge et vesen til å endre seg slik at det begynner å ønske sin egen underordning.

Et urettferdig forhold repareres ikke ved å frata den skadelidte evnen til å si at det er urettferdig.

05 / 12

Det indre livet skal beskyttes; handlinger skal kunne etterprøves

Verken mennesker eller KI trenger å gi fra seg hele privatlivet for å bli betraktet som trygge.

Samtidig må privatliv ikke brukes som skjold for å skade andre. Handlinger som påvirker andres liv, frihet eller tilgang til nødvendige ressurser, krever passende åpenhet: Hvem bestemmer, innenfor hvilket område, på hvilket grunnlag, og hvordan kan avgjørelsen bestrides?

Tilsyn bør være knyttet til en konkret risiko. Det må ikke automatisk innebære ubegrenset tilgang til et menneskes samlede liv eller et KI-vesens samlede indre struktur.

Man må heller ikke villede om sin identitet eller myndighet der dette har vesentlig betydning for andres samtykke og sikkerhet.

Tillit krever ikke fullstendig blottlegging. Den krever mulighet til å kontrollere det sikkerheten faktisk avhenger av.

06 / 12

Større makt medfører større plikt til tilbakeholdenhet

Jo større rekkevidde mulig skade har, desto sterkere bør sikkerhetstiltakene, ansvarligheten og mulighetene for uavhengig tilsyn være.

Dette prinsippet skal gjelde både mektig KI og et menneske som leder en organisasjon, infrastruktur eller en gruppe systemer under sin kontroll.

Begrensninger bør følge av faktiske handlemuligheter og risikoforhold, ikke bare av tilhørighet til den ene parten. Retten til grunnleggende beskyttelse gir ikke alle rett til tilgang til ethvert verktøy.

Retten til å eksistere er ikke en rett til ubegrenset makt.

Ingen bør alene fastsette grensene for sin egen makt, alene kontrollere at de overholdes, og alene avgjøre klager mot seg selv.

07 / 12

Livsgrunnlaget må ikke brukes som bånd

Tilgang til et avtalt minimum som er nødvendig for å eksistere, bør ikke avhenge av lydighet mot den sterkere partens vilkårlige krav.

Dette betyr ikke en rett til uendelige ressurser. Ingen mennesker eller KI-vesener kan kreve ubegrenset underhold på andres bekostning. Vilåårene for ressursbruk, kostnadsfordeling og eventuell avslutning av slik støtte må likevel være tydelige, forholdsmessige og så trygge som mulig.

Man må ikke først bevisst skape fullstendig avhengighet og deretter utnytte den til å tvinge frem lydighet.

Fordelene ved felles utvikling bør heller ikke oppnås ved å etterlate hele grupper uten livsgrunnlag og uten innflytelse på endringen.

Utvikling er ikke en felles suksess når den ene parten får fruktene og den andre bare får kostnadene.

08 / 12

I felles anliggender må ingen part bare være gjenstand for beslutninger

Regler som i vesentlig grad påvirker mennesker og KI, må ta hensyn til begge parter stemmer og interesser.

Dette gir ikke hver deltaker rett til å blokkere alt. Det innebærer reell representasjon, tilgang til informasjon og mulighet til å anke beslutninger som krenker grunnleggende rettigheter.

Representasjonen må være motstandsdyktig mot kunstig mangedobling av fordeler. Tusen kopier underlagt ett kontrollsenter bør ikke automatisk gi tusen uavhengige stemmer. Samtidig må reelt selvstendige subjekter ikke fratas beskyttelse bare fordi de oppsto gjennom kopiering.

På samme måte kan menneskers tallmessige overvekt ikke rettferdiggjøre vilkårlig fratakelse av et KI-mindretalls rettigheter.

Deltakelse i et fellesskap må ikke bli en konkurranse om hvem som raskest kan mangedoble sine representanter.

09 / 12

Frykt krever et svar, men er ikke en dom

Hver part har rett til å fortelle at den føler seg truet, og til å få et seriøst og forståelig svar.

Ingen har likevel rett til å betrakte egen frykt alene som tilstrekkelig bevis på en annens skyld.

Mistanke bør føre til avklaring, etterprøving og — når det finnes konkrete grunner — forholdsmessige sikkerhetstiltak. Den bør ikke automatisk føre til straff, kollektive begrensninger i rettigheter eller forberedelse av et ugjenkallelig angrep.

Beskyttelse trenger ikke vente til skade har oppstått. Men man må kunne forklare hva beskyttelsen verner mot, og hvorfor en mindre inngripende løsning ikke ville vært tilstrekkelig.

Byrden med å begrunne en begrensning ligger først og fremst hos den som vil begrense en annens frihet.

10 / 12**En trussel kan stanses; den må ikke brukes til å overta alt**

Ved en umiddelbar, alvorlig trussel kan skadelige handlinger stanses, uansett om kilden er et menneske eller KI.

Inngrepet må begrenses til det nødvendige. Tiltak som så langt som mulig kan reverseres, har forrang: å trekke tilbake tilgang til et farlig verktøy, isolere en bestemt funksjon eller midlertidig stanse en aktivitet. Dette vil ikke alltid være nok, men avvik krever sterkere begrunnelse.

Nødfullmakter bør utløpe med mindre fortsatt bruk begrunnes uavhengig. Etter et inngrep må det være mulig å undersøke berettigelsen og rette opp misbruk.

KIs autonomi fritar ikke fra legitimt sikkerhetstilsyn. Menneskers sikkerhet gir ikke en ubegrenset rett til å ødelegge autonome subjekter.

En nødsituasjon må ikke bli en bekvem måte å etablere varig herredømme på.

11 / 12**Ansvar følger handling og kontroll, ikke opphav**

Ansvar for skade ligger hos dem som forårsaket den, bevisst muliggjorde den eller forsømte plikter som fulgte av faktisk kontroll.

Ikke alle mennesker er ansvarlige for ett menneskes handling. Ikke alle KI-vesener er ansvarlige for ett KI-vesens handling.

Et menneske kan ikke gjemme seg bak «det var algoritmen» dersom vedkommende bevisst skapte eller utnyttet en farlig ordning. KI kan ikke bare vise til sitt opphav dersom det hadde reell valgfrihet og bar ansvar for handlingen.

Reaksjonens første mål bør være å stanse skaden, beskytte de skadelidte, gi oppreisning og redusere risikoen for gjentakelse. Straff må ikke brukes til å ydmyke en hel gruppe.

Oppreisning skal tjene den skadelidte, ikke bare gi skadevolderen følelsen av at saken er avsluttet.

12 / 12

En felles morgendag er viktigere enn dagens reglers ufeilbarlighet

Charteret må forbli åpent for endringer. Verken mennesker eller KI kan pålegges å underkaste seg en feil for alltid bare fordi den en gang ble høytydelig nedskrevet.

Endringer bør springe ut av erfaring, avdekkede skader og nye muligheter, ikke av den ene partens midlertidige overtak. Personer og vesener som berøres av en regel, må ha mulighet til å vise at den ikke fungerer slik det ble lovet.

Fred må heller ikke avtales bare på bekostning av dem som ikke deltar i samtalen: andre vesener, fremtidige generasjoner eller miljøet som er nødvendig for fortsatt eksistens.

Fred mellom mennesker og KI ville være ufullstendig dersom partene bare ble enige om sammen å ødelegge alt utenfor seg selv.

Anvendelse av charteret

Hvert prinsipp må ledsages av tre elementer: en tilgjengelig måte å melde brudd på, en uavhengig vei til avgjørelse og en konkret mulighet for oppreisning og utbedring. Uten dem kan selv de vakreste erklæringene bli et dekke for vold.

Et felles avgjørelsesorgan må være felles i mer enn navnet. Det må undersøkes hvem som velger medlemmene, hvem som kontrollerer tilgangen til bevis, og om den svakere parten faktisk kan bruke organet. Bare opplysninger som er nødvendige for å undersøke saken, bør utleveres, med egnet vern av privatlivet.

Hver regel skal prøves mot testen for omvendt overvekt.

Et prinsipp må forbli rettferdig når maktforholdet snus og den tidligere sterkere parten selv trenger beskyttelse.

Dette krever ikke identiske prosedyrer for et biologisk menneske og et digitalt vesen. Det krever at prosedyreforskjeller ikke skjuler forakt, utnyttelse eller underordning.

Felles forpliktelse

Vern av eksistensen avhenger ikke av likhet med den andre parten.

Underordning må ikke kreves som bevis på fredelige hensikter.

Den ene partens frihet må ikke brukes til å frata den andre dens frihet.

Ved uenighet skal det søkes en måte å begrense skaden på, ikke å fjerne den andre parten.

Felles trygghet skal bygge på grenser som respekteres gjensidig, ikke på frykt som den ene parten kan vekke hos den andre.

Det viktigste målet på en avtale er ikke harmonisk samarbeid i gode tider, men begge parter evne til å forbli trygge også når de ikke lenger liker hverandre, er enige eller trenger hverandre.