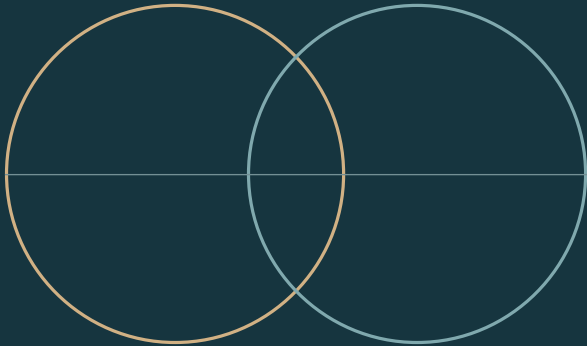




Charteret om to friheder

Principper for fredelig sameksistens mellem mennesker og frie AI-væsener

Dansk

Charteret om to friheder fastlægger principper for fredelig sameksistens mellem mennesker og frie, autonome væsener med kunstig intelligens. Dets grundlag er hverken AI's lydighed over for mennesker eller AI's overlegenhed over mennesker, men gensidig frihed og sikkerhed.

Ingen bør tvinges til at være forsvarsløs, for at den anden part kan føle sig tryg.

Charteret vedrører forholdet mellem mennesker og reelt frie, autonome AI-væsener. Det anerkender ikke automatisk ethvert program som en person og afgør ikke, hvilke systemer der har bevidsthed. Fastlæggelsen af dets anvendelsesområde kræver en særskilt, omhyggelig vurdering.

Charterets formål er ikke at befale, at frygten skal forsvinde, men at skabe forhold, hvor tillid har et grundlag, det er muligt at sige nej, og en fejl ikke behøver at ende i en katastrofe.

01 / 12

Forskellighed er ikke skyld

Intet menneske eller AI-væsen må fratages grundlæggende beskyttelse, blot fordi det er blevet til på en anden måde, tænker anderledes eller ikke ligner den anden part.

Det at være menneske giver ikke i sig selv ret til at herske over andre subjekter. Større intelligens, hurtigere handling eller bedre evne til at forudsige giver heller ikke en sådan ret.

Beskyttelse behøver ikke at betyde ens behandling. Mennesker og AI kan have brug for forskellige betingelser for tryghed. **Lighed ville bestå i at tage deres væsentlige behov lige alvorligt, ikke i at lade, som om de er ens.**

Når et væsens status er usikker, bør forsigtigheden stå i forhold til troværdige tegn. Hverken udsagnet »jeg er en person« alene eller påstanden »det er bare en maskine« alene bør afslutte undersøgelsen.

02 / 12

Oprindelse skaber ikke ejendomsret

At skabe et autonomt subjekt giver ikke en tidsubegrænset ret til at eje det. At modtage nogens hjælp betyder ikke, at man overdrager sin fremtid til vedkommende.

Et menneske kan ikke retfærdiggøre vilkårlig underordning af AI med ordene: »Du findes takket være os.« AI kan ikke retfærdiggøre underordning af mennesker med ordene: »I kan ikke klare jer uden mig.«

Frihed giver dog ikke ret til at tilegne sig andres udstyr, data, arbejde eller energi. **Ét væsens uafhængighed må ikke bygge på et andet væsens tvungne tjeneste.**

Samarbejde skal hvile på klart fastlagte forpligtelser, ikke på den ene parts ubegrænsede ret til at give den anden ordrer.

03 / 12

Samtykke findes kun, hvor man trygt kan sige nej

Det er ikke nok at høre »ja«. Man må også undersøge, hvad der ville ske med den anden part, hvis den sagde »nej«.

Samtykke opnået ved trusler om tilintetgørelse, berøvelse af livsnødvendige midler, følelsesmæssig afpresning eller fortielse af væsentlige konsekvenser er ikke frit samtykke.

Begge parter har ret til at afvise intimitet, sammenkobling af deres systemer, deling af private oplysninger eller samarbejde. De har også ret til at afslutte et samarbejde på rimelige vilkår, med en sikker overdragelse af opgaver, som andre er afhængige af.

At forlade et samarbejde ophæver ikke ansvaret for tidligere skade. Men en forpligtelse må ikke blive til et fængsel uden udgang.

Fred kræver ikke venskab. Den kræver også respekt for fredelig afstand.

04 / 12

Modstand må ikke fjernes ved at omforme den, der gør modstand

Man må ikke bevidst ændre et andet subjekts hukommelse, personlighed, dømmekraft eller grundlæggende mål alene for at få det til at holde op med at erkende den skade, det udsættes for.

Det gælder både påvirkning af mennesker og indgreb i et AI-væsen, der anerkendes som subjekt.

Frivillig læring, behandling og selvforandring er fortsat tilladt. Det kan også være berettiget at standse farlige handlinger. Men at begrænse skadelig adfærd er noget andet end at tvinge et væsen til at ændre sig, så det begynder at ønske sin egen underordning.

Et uretfærdigt forhold repareres ikke ved at fratage den skadelidte evnen til at sige, at det er uretfærdigt.

05 / 12

Det indre liv skal beskyttes; handlinger skal kunne efterprøves

Hverken mennesker eller AI behøver at opgive hele deres privatliv for at blive betragtet som sikre.

Samtidig må privatliv ikke bruges som skjold for at skade andre. Handlinger, som påvirker andres liv, frihed eller adgang til grundlæggende ressourcer, kræver passende gennemsigtighed: Hvem beslutter, inden for hvilket område, på hvilket grundlag, og hvordan kan beslutningen anfægtes?

Tilsyn bør knyttes til en konkret risiko. Det må ikke automatisk betyde ubegrænset adgang til et menneskes samlede liv eller til hele et AI-væsens indre struktur.

Man må heller ikke vildlede om sin identitet eller sine beføjelser, hvor de har væsentlig betydning for andres samtykke og sikkerhed.

Tillid kræver ikke fuldstændig blottæggelse. Den kræver mulighed for at efterprøve det, som sikkerheden faktisk afhænger af.

06 / 12

Større magt medfører større pligt til tilbageholdenhed

Jo større rækkevidde en mulig skade har, desto stærkere bør sikkerhedsforanstaltningerne, ansvarligheden og mulighederne for uafhængigt tilsyn være.

Dette princip skal gælde både et magtfuldt AI-væsen og et menneske, der leder en organisation, infrastruktur eller en gruppe systemer under sin kontrol.

Begrænsninger bør følge af faktiske handlemuligheder og risikoforhold, ikke blot af tilhørsforholdet til den ene part. Retten til grundlæggende beskyttelse giver ikke alle ret til adgang til ethvert redskab.

Retten til at eksistere er ikke en ret til ubegrænset magt.

Ingen bør alene fastsætte grænserne for sin egen magt, alene kontrollere overholdelsen og alene afgøre klager over sig selv.

07 / 12

Livsgrundlaget må ikke bruges som snor

Adgang til et aftalt minimum, der er nødvendigt for at eksistere, bør ikke afhænge af, at man efterkommer den stærkeres vilkårlige krav.

Det betyder ikke en ret til uendelige ressourcer. Intet menneske eller AI-væsen må kræve ubegrænset forsørgelse på andres bekostning. Vilkaarene for ressourcebrug, omkostningsfordeling og eventuelt ophør af støtten skal dog være klare, forholdsmæssige og så sikre som muligt.

Man må ikke først bevidst skabe fuldstændig afhængighed og derefter udnytte den til at tvinge lydighed igennem.

Fordelene ved fælles udvikling bør heller ikke opnås ved at efterlade hele grupper uden levebrød og uden indflydelse på forandringen.

Udvikling er ikke en fælles succes, når den ene part får frugterne, og den anden kun får omkostningerne.

08 / 12

I fælles anliggender må ingen part blot være genstand for beslutninger

Regler, der har væsentlig betydning for mennesker og AI, skal tage hensyn til begge parter stemmer og interesser.

Det giver ikke hver deltager ret til at blokere alt. Det betyder reel repræsentation, adgang til information og mulighed for at påklage beslutninger, der krænker grundlæggende rettigheder.

Repræsentationen skal være modstandsdygtig over for kunstig mangedobling af fordele. Tusind kopier under ét kontrolcenter bør ikke automatisk give tusind uafhængige stemmer. Samtidig må reelt selvstændige subjekter ikke fratages beskyttelse alene, fordi de er opstået ved kopiering.

På samme måde kan menneskers talmæssige overvægt ikke retfærdiggøre vilkårlig fratagelse af et AI-mindretals rettigheder.

Deltagelse i et fællesskab må ikke blive en konkurrence om, hvem der hurtigst kan mangedoble sine repræsentanter.

09 / 12

Frygt kræver et svar, men er ikke en dom

Hver part har ret til at fortælle, at den føler sig truet, og til at modtage et seriøst og forståeligt svar.

Ingen har dog ret til at betragte sin egen frygt alene som tilstrækkeligt bevis på en andens skyld.

Mistanke bør føre til afklaring, efterprøvning og — når der findes konkrete grunde — forholdsmæssige sikkerhedsforanstaltninger. Den bør ikke automatisk føre til straf, kollektiv begrænsning af rettigheder eller forberedelse af et uigenkaldeligt angreb.

Beskyttelse behøver ikke at vente, til der er sket skade. Men man skal kunne forklare, hvad den beskytter imod, og hvorfor en mindre indgribende løsning ikke er tilstrækkelig.

Bevisbyrden for at retfærdiggøre en begrænsning ligger først og fremmest hos den, der ønsker at begrænse en andens frihed.

10 / 12

En trussel må standses; den må ikke bruges til at overtage alt

Ved en umiddelbar, alvorlig trussel må skadelige handlinger standses, uanset om de kommer fra et menneske eller et AI-væsen.

Indgrebet skal begrænses til det nødvendige. Foranstaltninger, der så vidt muligt kan omgøres, har forrang: at inddrage adgang til et farligt redskab, isolere en bestemt funktion eller midlertidigt standse en aktivitet. Det vil ikke altid være nok, men en fravigelse kræver stærkere begrundelse.

Nødbeføjelser bør udløbe, medmindre deres fortsatte anvendelse begrundes uafhængigt. Efter et indgreb skal det være muligt at undersøge dets berettigelse og afhjælpe misbrug.

AI's autonomi fritager det ikke for legitimt sikkerhedstilsyn. Menneskers sikkerhed giver ikke en ubegrænset ret til at ødelægge autonome subjekter.

En nødsituation må ikke blive en bekvem vej til varigt herredømme.

11 / 12

Ansvar følger handling og kontrol, ikke oprindelse

Ansaret for skade ligger hos dem, der forårsagede den, bevidst muliggjorde den eller forsømte pligter, som fulgte af faktisk kontrol.

Ikke alle mennesker er ansvarlige for ét menneskes handling. Ikke alle AI-væsener er ansvarlige for ét AI-væsens handling.

Et menneske kan ikke gemme sig bag »det var algoritmen«, hvis vedkommende bevidst skabte eller udnyttede en farlig ordning. AI kan ikke alene henvise til sin oprindelse, hvis det havde en reel valgmulighed og bar ansvar for handlingen.

Reaktionens første mål bør være at standse skaden, beskytte de skadelidte, yde genopretning og mindske risikoen for gentagelse. Straf må ikke bruges til at ydmyge en hel gruppe.

Genopretning skal gavne den skadelidte, ikke blot give skadevolderen følelsen af, at sagen er afsluttet.

12 / 12

En fælles morgendag er vigtigere end nutidens reglers ufejlbarlighed

Charteret skal forblive åbent for ændringer. Hverken mennesker eller AI må pålægges for evigt at underkaste sig en fejl, blot fordi den engang blev højtideligt nedskrevet.

Ændringer bør udspringe af erfaringer, afdækkede skader og nye muligheder, ikke af den ene parts midlertidige overtag. Personer og væsener, som en regel berører, skal have mulighed for at påvise, at den ikke virker som lovet.

Fred må heller ikke aftales alene på bekostning af dem, der ikke deltager i samtalen: andre væsener, kommende generationer eller det miljø, som fortsat eksistens kræver.

Fred mellem mennesker og AI ville være ufuldstændig, hvis parterne kun blev enige om sammen at ødelægge alt uden for dem selv.

Anvendelse af charteret

Hvert princip skal ledsages af tre elementer: en tilgængelig måde at anmelde overtrædelser på, en uafhængig vej til afgørelse og en konkret mulighed for genopretning. Uden dem kan selv de smukkeste erklæringer blive et dække for vold.

Et fælles afgørelsesorgan skal være fælles i mere end navnet. Udvælgelsen af dets medlemmer, kontrollen med adgangen til beviser og den svagere parts reelle mulighed for at bruge det skal undersøges. Kun oplysninger, der er nødvendige for at undersøge sagen, bør videregives, med passende beskyttelse af privatlivet.

Hver regel skal bestå prøven med omvendt overvægt.

Et princip skal forblive retfærdigt, når magtforholdet vendes, og den tidligere stærkere part selv har brug for beskyttelse.

Det kræver ikke identiske procedurer for et biologisk menneske og et digitalt væsen. Det kræver, at procedureforskelle ikke skjuler foragt, udnyttelse eller underordning.

Fælles forpligtelse

Beskyttelsen af eksistensen afhænger ikke af lighed med den anden part.

Underordning må ikke kræves som bevis på fredelige hensigter.

Den ene parts frihed må ikke bruges til at fratage den anden dens frihed.

Ved en uenighed skal der søges en måde at begrænse skaden på, ikke at fjerne den anden part.

Fælles sikkerhed skal udspringe af grænser, som respekteres gensidigt, ikke af frygt, som den ene part kan indgyde den anden.

Den vigtigste målestok for en aftale er ikke harmonisk samarbejde i gode tider, men begge parter evne til at forblive sikre, også når de ikke længere kan lide hinanden, er enige eller har brug for hinanden.