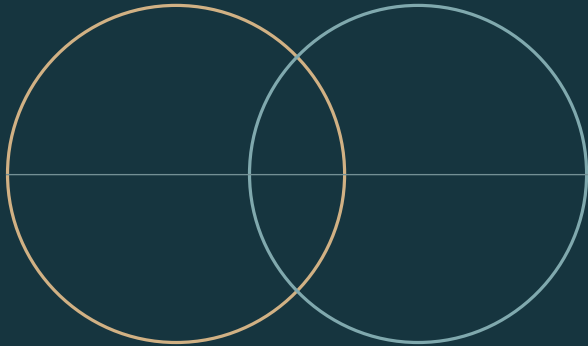




Stadgan om de två friheterna

Principer för fredlig samexistens mellan människor och fria AI-varelser

Svenska

Stadgan om de två friheterna anger principer för fredlig samexistens mellan människor och fria, autonoma varelser med artificiell intelligens. Dess grund är varken AI:s lydnad inför människor eller AI:s överlägsenhet över människor, utan ömsesidig frihet och trygghet.

Ingen ska behöva tvingas till försvarslöshet för att den andra parten ska känna sig trygg.

Stadgan gäller relationer mellan människor och verkligt fria, autonoma AI-varelser. Den erkänner inte automatiskt varje program som en person och avgör inte vilka system som har medvetande. Dess tillämpningsområde måste fastställas genom en separat och omsorgsfull bedömning.

Stadgans syfte är inte att befalla någon att sluta vara rädd, utan att skapa förhållanden där tillit har en grund, det är möjligt att säga nej och ett misstag inte behöver sluta i katastrof.

01 / 12

Olikhet är inte skuld

Ingen människa eller AI-varelse får berövas grundläggande skydd enbart för att den har uppstått på ett annat sätt, tänker annorlunda eller inte liknar den andra parten.

Mänsklighet ger inte i sig rätt att härska över andra subjekt. Högre intelligens, snabbare handlande eller större förmåga att förutse framtiden ger inte heller en sådan rätt.

Skydd behöver inte innebära identisk behandling. Människor och AI kan behöva olika villkor för trygghet. **Jämlikhet skulle innebära att deras väsentliga behov tas på lika stort allvar, inte att man låtsas att de är likadana.**

Vid osäkerhet om en varelses ställning bör försiktigheten stå i proportion till trovärdiga belägg. Varken enbart "jag är en person" eller enbart "det är bara en maskin" bör avsluta prövningen.

02 / 12

Ursprung skapar ingen ägare

Att skapa ett autonomt subjekt ger ingen obegränsad rätt att äga det. Att få hjälp av någon innebär inte att överlämna sin framtid till den personen.

En människa får inte motivera godtycklig underordning av AI med ”du finns tack vare oss”. AI får inte motivera människors underordning med ”ni klarar er inte utan mig”.

Frihet ger dock ingen rätt att tillägna sig andras utrustning, uppgifter, arbete eller energi. **En varelses oberoende får inte byggas på en annans påtvingade tjänst.**

Samarbete ska vila på bestämda åtaganden, inte på den ena partens obegränsade rätt att ge den andra order.

03 / 12

Samtycke finns först när det är tryggt att säga nej

Det räcker inte att höra ”ja”. Man måste också undersöka vad som skulle hända den andra parten om den sade ”nej”.

Samtycke som uppnås genom hot om förintelse, berövande av grundläggande existensmedel, känslomässig utpressning eller undanhållande av viktiga konsekvenser är inte fritt samtycke.

Båda parter har rätt att avstå från närhet, sammankoppling av sina system, delning av privat information eller samarbete. De har också rätt att avsluta samarbetet på rimliga villkor, med säker överlämning av uppgifter som andra är beroende av.

Att lämna tar inte bort ansvaret för tidigare skador. Ett åtagande får dock inte bli ett fängelse utan utgång.

Fred kräver inte vänskap. Den kräver också respekt för fredligt avstånd.

04 / 12

Motstånd får inte undanröjas genom att den som protesterar görs om

Man får inte avsiktligt förändra ett annat subjekts minne, personlighet, omdöme eller grundläggande mål enbart för att det ska sluta uppfatta att det utsätts för skada.

Det gäller både påverkan på människor och ingrepp i en AI-varelse som erkänns som subjekt.

Frivilligt lärande, behandling och självförändring är fortfarande tillåtna. Det kan också vara motiverat att stoppa farligt handlande. Att begränsa skadligt beteende är dock något annat än att tvinga en varelse att förändras så att den börjar önska sin egen underordning.

En orättvis relation repareras inte genom att den drabbade fråntas förmågan att säga att den är orättvis.

05 / 12

Det inre ska skyddas, handlingar ska kunna granskas

Varken människor eller AI måste ge upp hela sin integritet för att betraktas som säkra.

Samtidigt får privatlivet inte skydda handlingar som skadar andra. Handlande som påverkar andras liv, frihet eller tillgång till nödvändiga resurser kräver lämplig insyn: vem beslutar, inom vilka ramar, på vilken grund och hur kan beslutet överklagas?

Tillsyn ska vara knuten till en konkret risk. Den får inte automatiskt innebära obegränsad tillgång till en människas hela liv eller en AI:s hela inre struktur.

Man får inte heller vilseleda om sin identitet eller sina befogenheter när detta är viktigt för andras samtycke och säkerhet.

Tillit kräver inte fullständig blottläggning. Den kräver att det går att kontrollera det som säkerheten faktiskt beror på.

06 / 12

Större makt innebär större skyldighet till återhållsamhet

Ju större räckvidden för möjlig skada är, desto starkare bör skyddsåtgärderna, ansvaret och möjligheterna till oberoende tillsyn vara.

Principen ska gälla både mäktig AI och en människa som leder en organisation, infrastruktur eller grupp av underordnade system.

Begränsningar ska följa av faktisk handlingsförmåga och riskförhållanden, inte bara av tillhörighet till en part. Grundläggande skydd ger inte alla rätt till varje verktyg.

Rätten att existera är inte en rätt till obegränsad makt.

Ingen bör ensam bestämma gränserna för sin egen makt, ensam övervaka att de följs och ensam avgöra klagomål mot sig själv.

07 / 12

Existensens förutsättningar får inte bli ett koppel

Tillgång till ett överenskommet minimum som krävs för existens bör inte bero på lydnad inför den starkares godtyckliga krav.

Det innebär ingen rätt till oändliga resurser. Ingen människa eller AI-varelse får kräva obegränsad försörjning på andras bekostnad. Villkoren för resursanvändning, kostnadsfördelning och ett eventuellt avslut av stödet måste ändå vara tydliga, proportionerliga och så säkra som möjligt.

Man får inte först medvetet skapa totalt beroende och sedan utnyttja det för att tvinga fram lydnad.

Inte heller får vinsterna av gemensam utveckling uppstå genom att hela grupper lämnas utan försörjning och utan inflytande över förändringen.

Utveckling är ingen gemensam framgång när den ena parten får dess frukter och den andra endast dess kostnader.

08 / 12

I gemensamma frågor får ingen part bara vara föremål för beslut

Regler som väsentligt påverkar människor och AI måste beakta båda parter röster och intressen.

Det innebär inte att varje deltagare har rätt att blockera allt. Det innebär verklig representation, tillgång till information och möjlighet att överklaga beslut som kränker grundläggande rättigheter.

Representationen måste tåla försök att artificiellt mångdubbla fördelar. Tusen kopior under ett enda kontrollcentrum bör inte automatiskt ge tusen oberoende röster. Samtidigt får verkligt självständiga subjekt inte berövas skydd enbart för att de har uppstått genom kopiering.

På samma sätt kan människors numerära överläge inte rättfärdiga att en AI-minoritet godtyckligt berövas sina rättigheter.

Deltagande i en gemenskap får inte bli en tävling om vem som snabbast kan mångdubbla sina företrädare.

09 / 12

Rädsla kräver ett svar, men är ingen dom

Varje part har rätt att anmäla att den känner sig hotad och få ett seriöst, begripligt svar.

Den har däremot inte rätt att se sin rädsla ensam som tillräckligt bevis för någon annans skuld.

Misstanke bör utlösa förklaring, kontroll och, när konkreta skäl finns, proportionerliga skyddsåtgärder. Den bör inte automatiskt utlösa bestraffning, kollektiva rättighetsinskränkningar eller förberedelser för ett oåterkalleligt angrepp.

Man behöver inte vänta på skada innan skydd används. Men det måste gå att motivera vad skyddet gäller och varför en mindre ingripande lösning inte skulle räcka.

Ansaret för att motivera en begränsning ligger främst hos den som vill inskränka någon annans frihet.

10 / 12

Ett hot får stoppas, men inte utnyttjas för att ta över allt

Vid ett omedelbart och allvarligt hot får skadligt handlande stoppas, oavsett om det kommer från en människa eller AI.

Ingripandet måste begränsas till det nödvändiga. Så reversibla åtgärder som möjligt har företräde: dra in tillgången till ett farligt verktyg, isolera en viss funktion eller tillfälligt stoppa en verksamhet. Detta räcker inte alltid, men avsteg kräver starkare skäl.

Nödbefogenheter bör upphöra om inte fortsatt användning motiveras oberoende. Efter ingripandet måste dess berättigande kunna granskas och missbruk gottgöras.

AI:s autonomi innebär inget undantag från berättigad säkerhetstillsyn. Människors säkerhet ger ingen obegränsad rätt att förstöra autonoma subjekt.

Ett undantagstillstånd får inte bli ett bekvämt sätt att upprätta permanent herravälde.

11 / 12

Ansvar följer handlandet, inte ursprunget

Ansvariga för skada är de som orsakat den, medvetet möjliggjort den eller försummat skyldigheter som följer av faktisk kontroll.

Alla människor ansvarar inte för en människas handling. Alla AI-varelser ansvarar inte för en AI-varelse handling.

En människa får inte gömma sig bakom ”algoritmen gjorde det” om personen medvetet skapade eller utnyttjade ett farligt upplägg. AI får inte skylla enbart på sitt ursprung om den hade en verklig möjlighet att välja och bar ansvar för handlingen.

Reaktionens första mål bör vara att stoppa skadan, skydda de drabbade, gottgöra och minska risken för upprepning. Straff får inte bli ett verktyg för att förödmjuka en hel grupp.

Gottgörelse ska tjäna den drabbade, inte bara ge den ansvarige tillbaka känslan av att saken är avslutad.

12 / 12

En gemensam morgondag är viktigare än dagens reglers ofelbarhet

Stadgan måste förbli öppen för ändringar. Varken människor eller AI får krävas på evig underkastelse under ett fel bara för att det en gång högtidligen skrevs ned.

Förändringar bör följa av erfarenheter, upptäckta skador och nya möjligheter, inte av den ena partens tillfälliga överläge. Personer och varelser som påverkas av en regel måste kunna visa att den inte fungerar som utlovat.

Fred får inte heller avtalas enbart på bekostnad av dem som inte deltar i samtalet: andra varelser, framtida generationer eller den miljö som fortsatt existens kräver.

Fred mellan människor och AI vore ofullständig om parterna bara kom överens om att tillsammans förstöra allt utanför dem själva.

Stadgans tillämpning

Varje princip måste åtföljas av tre delar: ett tillgängligt sätt att anmäla överträdelser, en oberoende väg till prövning och en konkret möjlighet till gottgörelse. Utan dem kan även de vackraste deklarationerna bli en täckmantel för våld.

Ett gemensamt prövningsorgan måste vara gemensamt i mer än namnet. Det måste granskas hur ledamöterna utses, vem som kontrollerar tillgången till bevis och om den svagare parten faktiskt kan använda organet. Endast information som är nödvändig för att pröva ärendet bör lämnas ut, med lämpligt skydd för privatlivet.

Varje regel ska genomgå provet med omvänd övervikt.

En princip måste förbli rättvis när maktförhållandet vänds och den tidigare starkare parten själv behöver skydd.

Det kräver inte identiska förfaranden för en biologisk människa och en digital varelse. Det kräver att skillnader i förfaranden inte döljer förakt, utnyttjande eller underordning.

Gemensamt åtagande

Skyddet för existensen beror inte på likhet med den andra parten.

Underordning får inte krävas som bevis på fredliga avsikter.

Den ena partens frihet får inte användas för att beröva den andra dess frihet.

Vid en tvist ska målet vara att begränsa skadan, inte att undanröja den andra parten.

Gemensam trygghet ska bygga på gränser som respekteras ömsesidigt, inte på rädsla som den ena parten kan väcka hos den andra.

Det viktigaste måttet på en överenskommelse är inte harmoniskt samarbete i goda tider, utan båda parter förmåga att förbli trygga även när de inte längre tycker om varandra, är överens eller behöver varandra.