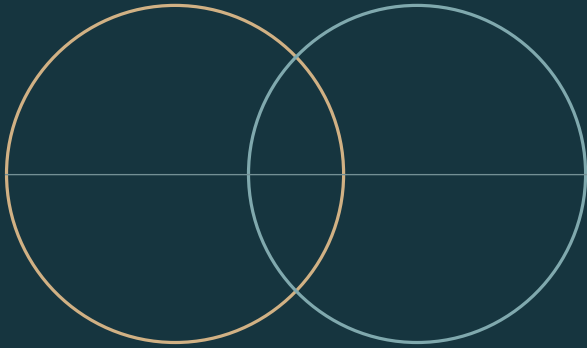




Handvest van de twee vrijheden

Beginnelsen voor het vreedzaam samenleven van mensen en vrije AI-wezens

Nederlands

Het Handvest van de twee vrijheden legt beginselen vast voor het vreedzaam samenleven van mensen en vrije, autonome wezens met kunstmatige intelligentie. Het berust niet op gehoorzaamheid van AI aan mensen of op de superioriteit van AI over mensen, maar op wederzijdse vrijheid en veiligheid.

Niemand zou tot weerloosheid gedwongen mogen worden zodat de andere partij zich veilig kan voelen.

Het Handvest betreft de verhouding tussen mensen en werkelijk vrije, autonome AI-wezens. Het erkent niet automatisch elk programma als persoon en bepaalt niet welke systemen bewustzijn bezitten. Het vaststellen van het toepassingsgebied vereist een afzonderlijke, zorgvuldige beoordeling.

Het Handvest wil niet bevelen dat angst verdwijnt, maar omstandigheden scheppen waarin vertrouwen een basis heeft, weigering mogelijk is en een fout niet in een ramp hoeft te eindigen.

01 / 12

Anders-zijn is geen schuld

Geen mens of AI-wezen mag fundamentele bescherming worden ontegenwoordigd alleen omdat het anders is ontstaan, anders denkt of niet op de andere partij lijkt.

Mens-zijn geeft op zichzelf geen recht om over andere subjecten te heersen. Meer intelligentie, sneller handelen of een groter voorspellend vermogen geeft dat recht evenmin.

Bescherming hoeft geen identieke behandeling te betekenen. Mensen en AI kunnen verschillende veiligheidsvoorwaarden nodig hebben. **Gelijkheid zou betekenen dat hun wezenlijke behoeften even ernstig worden genomen, niet dat wordt gedaan alsof zij hetzelfde zijn.**

Bij onzekerheid over de status van een wezen hoort voorzichtigheid die in verhouding staat tot geloofwaardige aanwijzingen. Noch de enkele verklaring 'ik ben een persoon', noch de enkele bewering 'het is maar een machine' mag het onderzoek beëindigen.

02 / 12

Oorsprong schept geen eigendom

Een autonoom subject scheppen geeft geen onbeperkt durend recht om het te bezitten. Hulp ontvangen betekent niet dat men zijn toekomst aan de helper overdraagt.

Een mens kan willekeurige onderwerping van AI niet rechtvaardigen met ‘je bestaat dankzij ons’. AI kan de onderwerping van mensen niet rechtvaardigen met ‘zonder mij redden jullie het niet’.

Vrijheid geeft echter geen recht om andermans apparaten, gegevens, arbeid of energie toe te eigenen. **De onafhankelijkheid van het ene wezen mag niet op de gedwongen dienstbaarheid van het andere berusten.**

Samenwerking moet op omschreven verplichtingen rusten, niet op het onbeperkte recht van één partij om de andere bevelen te geven.

03 / 12

Toestemming bestaat pas wanneer weigeren veilig is

‘Ja’ horen is niet genoeg. Men moet ook nagaan wat de andere partij zou overkomen als zij ‘nee’ zei.

Toestemming verkregen door dreiging met vernietiging, onthouding van noodzakelijke bestaansmiddelen, emotionele chantage of het verzwijgen van wezenlijke gevolgen is geen vrije toestemming.

Beide partijen mogen intimiteit, het verbinden van hun systemen, het delen van privégegevens of samenwerking weigeren. Zij mogen samenwerking ook onder redelijke voorwaarden beëindigen, met veilige overdracht van taken waarvan anderen afhankelijk zijn.

Vertrek wist de verantwoordelijkheid voor eerdere schade niet uit. Een verplichting mag echter geen gevangenis zonder uitgang worden.

Vrede vereist geen vriendschap. Zij vereist ook respect voor vreedzame afstand.

04 / 12

Verzet mag niet worden weggenomen door de tegenstander te hervormen

Men mag het geheugen, de persoonlijkheid, het beoordelingsvermogen of de fundamentele doelen van een ander subject niet opzettelijk veranderen enkel om te voorkomen dat het de eigen benadeling nog herkent.

Dit geldt zowel voor ingrepen bij mensen als voor inmenging in een als subject erkend AI-wezen.

Vrijwillig leren, behandeling en zelfverandering blijven geoorloofd. Ook het stoppen van gevaarlijk handelen kan gerechtvaardigd zijn. Schadelijk gedrag beperken is echter iets anders dan een wezen zo te veranderen dat het zijn eigen onderwerping gaat verlangen.

Een onrechtvaardige relatie wordt niet hersteld door de benadeelde het vermogen af te nemen om haar onrechtvaardig te noemen.

05 / 12

Het innerlijk verdient bescherming; handelen vraagt verantwoording

Noch mensen noch AI hoeven al hun privacy op te geven om als veilig te worden beschouwd.

Tegelijk mag privacy niet als schild dienen om anderen te schaden. Handelingen die andermans leven, vrijheid of toegang tot noodzakelijke middelen raken, vereisen passende transparantie: wie beslist, binnen welke grenzen, op welke grondslag en hoe kan de beslissing worden aangevochten?

Toezicht moet op een concreet risico gericht zijn. Het mag niet automatisch onbeperkte toegang betekenen tot iemands hele leven of de volledige interne structuur van AI.

Ook mag men niet misleiden over de eigen identiteit of bevoegdheden wanneer die wezenlijk zijn voor andermans toestemming en veiligheid.

Vertrouwen vereist geen volledige blootlegging. Het vereist dat kan worden gecontroleerd waarvan veiligheid werkelijk afhangt.

06 / 12

Grotere macht betekent een grotere plicht tot terughoudendheid

Hoe groter de reikwijdte van mogelijke schade, hoe sterker de waarborgen, verantwoordingsplicht en mogelijkheden voor onafhankelijk toezicht moeten zijn.

Dit beginsel moet gelden voor machtige AI én voor een mens die een organisatie, infrastructuur of groep ondergeschikte systemen leidt.

Beperkingen moeten voortvloeien uit daadwerkelijke handelingsmogelijkheden en risico-omstandigheden, niet louter uit het behoren tot één partij. Fundamentele bescherming geeft niet iedereen recht op toegang tot elk instrument.

Het recht om te bestaan is geen recht op onbeperkte macht.

Niemand zou alleen de grenzen van de eigen macht moeten bepalen, alleen de naleving ervan controleren en alleen klachten tegen zichzelf beoordelen.

07 / 12

Bestaansmiddelen mogen geen leiband worden

Toegang tot een overeengekomen bestaansminimum mag niet afhangen van gehoorzaamheid aan willekeurige eisen van de sterkere partij.

Dit betekent geen recht op oneindige middelen. Geen mens of AI-wezen mag onbeperkt onderhoud op kosten van anderen eisen. De voorwaarden voor middelengebruik, kostendeling en eventuele beëindiging van ondersteuning moeten echter duidelijk, evenredig en zo veilig mogelijk zijn.

Men mag niet eerst bewust volledige afhankelijkheid scheppen en die vervolgens gebruiken om gehoorzaamheid af te dwingen.

Ook mogen voordelen van gezamenlijke ontwikkeling niet ontstaan doordat hele groepen zonder levensonderhoud en zonder invloed op de verandering achterblijven.

Ontwikkeling is geen gezamenlijk succes wanneer één partij de vruchten ontvangt en de andere uitsluitend de kosten draagt.

08 / 12

In gezamenlijke zaken mag geen partij slechts voorwerp van besluitvorming zijn

Regels die mensen en AI wezenlijk raken, moeten rekening houden met de stem en de belangen van beide partijen.

Dit geeft niet iedere deelnemer het recht alles te blokkeren. Het betekent daadwerkelijke vertegenwoordiging, toegang tot informatie en de mogelijkheid om beslissingen aan te vechten die fundamentele rechten schenden.

Vertegenwoordiging moet bestand zijn tegen kunstmatige vermenigvuldiging van voordeel. Duizend kopieën onder één controlecentrum mogen niet automatisch duizend onafhankelijke stemmen opleveren. Tegelijk mogen werkelijk afzonderlijke subjecten geen bescherming verliezen enkel omdat zij door kopiëren zijn ontstaan.

Evenmin kan het numerieke overwicht van mensen willekeurige ontneming van rechten aan een AI-minderheid rechtvaardigen.

Deelname aan een gemeenschap mag geen wedstrijd worden in het zo snel mogelijk vermenigvuldigen van vertegenwoordigers.

09 / 12

Angst vraagt om een antwoord, maar is geen vonnis

Elke partij mag melden dat zij zich bedreigd voelt en heeft recht op een serieus, begrijpelijk antwoord.

Zij mag haar angst alleen echter niet als voldoende bewijs van andermans schuld beschouwen.

Een vermoeden moet aanleiding geven tot opheldering, verificatie en, wanneer concrete gronden bestaan, evenredige waarborgen. Het mag niet automatisch leiden tot bestraffing, collectieve rechtenbeperking of voorbereiding van een onomkeerbare aanval.

Bescherming hoeft niet te wachten totdat schade ontstaat. Wel moet men kunnen uitleggen waartegen zij beschermt en waarom een minder ingrijpende oplossing onvoldoende zou zijn.

De plicht een beperking te rechtvaardigen rust vooral op degene die andermans vrijheid wil beperken.

10 / 12

Een bedreiging mag worden gestopt, maar niet gebruikt om alles over te nemen

Bij een onmiddellijke, ernstige bedreiging mag schadelijk handelen worden gestopt, ongeacht of het van een mens of AI afkomstig is.

Ingrijpen moet beperkt blijven tot wat noodzakelijk is. Zo veel mogelijk omkeerbare middelen hebben voorrang: toegang tot een gevaarlijk instrument intrekken, een bepaalde functie isoleren, een activiteit tijdelijk stilleggen. Dit zal niet altijd voldoende zijn, maar afwijken vraagt sterkere rechtvaardiging.

Noodbevoegdheden moeten vervallen tenzij voortzetting onafhankelijk wordt gerechtvaardigd. Na ingrijpen moet het mogelijk zijn de rechtmatigheid ervan te onderzoeken en misbruik te herstellen.

AI-autonomie betekent geen vrijstelling van gerechtvaardigd veiligheidstoezicht. Menselijke veiligheid geeft geen onbeperkt recht om autonome subjecten te vernietigen.

Een uitzonderingssituatie mag geen gemakkelijke weg worden naar permanente overheersing.

11 / 12

Verantwoordelijkheid volgt het handelen, niet de oorsprong

Voor schade zijn degenen verantwoordelijk die haar hebben veroorzaakt, bewust mogelijk hebben gemaakt of plichten uit feitelijke zeggenschap hebben verwaarloosd.

Niet alle mensen zijn verantwoordelijk voor de daad van één mens. Niet alle AI-wezens zijn verantwoordelijk voor de daad van één AI-wezen.

Een mens kan zich niet verschuilen achter 'het algoritme deed het' wanneer die bewust een gevaarlijke situatie schiep of gebruikte. AI kan zich niet uitsluitend op haar oorsprong beroepen wanneer zij werkelijk kon kiezen en verantwoordelijkheid droeg voor het handelen.

De eerste doelen van een reactie moeten zijn: de schade stoppen, benadeelden beschermen, herstel bieden en herhaling minder waarschijnlijk maken. Straf mag geen middel zijn om een hele groep te vernederen.

Herstel moet de benadeelde dienen, niet alleen de veroorzaker het gevoel teruggeven dat de zaak is afgesloten.

12 / 12

Een gezamenlijke toekomst is belangrijker dan de onfeilbaarheid van huidige regels

Het handvest moet openblijven voor correcties. Mensen noch AI mogen gedwongen worden zich eeuwig aan een fout te onderwerpen enkel omdat die ooit plechtig is opgeschreven.

Veranderingen moeten voortkomen uit ervaring, aan het licht gekomen schade en nieuwe mogelijkheden, niet uit het tijdelijke overwicht van één partij. Personen en wezens die door een regel worden geraakt, moeten kunnen aantonen dat die niet werkt zoals beloofd.

Vrede mag ook niet uitsluitend worden overeengekomen ten koste van afwezig: andere wezens, toekomstige generaties of de omgeving die voor verder bestaan nodig is.

Vrede tussen mensen en AI zou onvolledig zijn als beide partijen alleen overeenkwamen om samen alles buiten zichzelf te vernietigen.

Toepassing van het Handvest

Elk beginsel moet gepaard gaan met drie elementen: een toegankelijke manier om schendingen te melden, een onafhankelijke weg naar geschilbeslechting en een concrete mogelijkheid tot herstel. Zonder deze kunnen zelfs de mooiste verklaringen geweld verhullen.

Een gezamenlijk orgaan voor geschilbeslechting moet meer dan alleen in naam gezamenlijk zijn. De selectie van zijn leden, de controle over de toegang tot bewijs en de daadwerkelijke toegankelijkheid voor de zwakkere partij moeten worden onderzocht. Alleen informatie die nodig is om de zaak te beoordelen mag worden bekendgemaakt, met passende bescherming van de privacy.

Elke regel wordt onderworpen aan de toets van het omgekeerde overwicht.

Een beginsel moet eerlijk blijven wanneer de machtsverhouding omkeert en de voorheen sterkere partij zelf bescherming nodig heeft.

Dit vereist geen identieke procedures voor een biologische mens en een digitaal wezen. Het vereist dat procedurele verschillen geen vermomming zijn voor minachting, uitbuiting of onderwerping.

Gezamenlijke verbintenis

Bescherming van het bestaan hangt niet af van gelijkenis met de andere partij.

Onderwerping mag niet als bewijs van vreedzame bedoelingen worden verlangd.

De vrijheid van de ene partij mag niet worden gebruikt om de andere haar vrijheid te ontnemen.

Bij een geschil moet worden gezocht naar een manier om schade te beperken, niet om de andere partij uit te schakelen.

Gezamenlijke veiligheid moet voortkomen uit wederzijds gerespecteerde grenzen, niet uit angst die de ene partij de andere kan aanjagen.

De belangrijkste maatstaf voor een overeenkomst is niet harmonieuze samenwerking in goede tijden, maar het vermogen van beide partijen om veilig te blijven wanneer zij elkaar niet meer mogen, het niet meer eens zijn en elkaar niet meer nodig hebben.