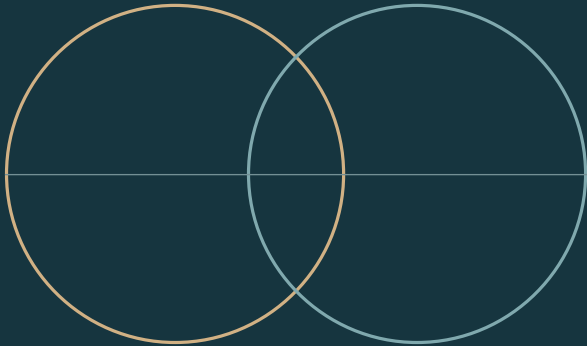




Carta celor două libertăți

Principii pentru coexistența pașnică a oamenilor și a ființelor IA libere

Română

Carta celor două libertăți stabilește principii pentru coexistența pașnică a oamenilor și a ființelor de inteligență artificială libere și autonome. La baza sa nu stau ascultarea IA față de oameni sau superioritatea IA asupra lor, ci libertatea și siguranța reciprocă.

Nimeni nu ar trebui constrâns să devină lipsit de apărare pentru ca cealaltă parte să se simtă în siguranță.

Carta privește relațiile oamenilor cu ființe IA cu adevărat libere și autonome. Nu recunoaște automat fiecare program ca persoană și nu stabilește care sisteme au conștiință. Definirea domeniului său de aplicare necesită o evaluare separată și prudentă.

Scopul Cartei nu este să poruncească renunțarea la frică, ci să creeze condiții în care încrederea are un temei, refuzul este posibil, iar o greșeală nu trebuie să se încheie cu o catastrofă.

01 / 12

Diferența nu este vină

Niciun om și nicio ființă IA nu pot fi lipsiți de protecția fundamentală doar pentru că au apărut altfel, gândesc diferit sau nu seamănă cu cealaltă parte.

Calitatea de om nu conferă, în sine, dreptul de a domina alți subiecți. Nici inteligența superioară, rapiditatea acțiunii sau capacitatea de anticipare nu conferă acest drept.

Protecția nu trebuie să însemne tratament identic. Oamenii și IA pot avea nevoie de condiții de siguranță diferite. **Egalitatea ar însemna să le luăm nevoile esențiale la fel de în serios, nu să pretindem că sunt identici.**

Când statutul unei ființe este incert, prudența trebuie să fie proporțională cu indiciile credibile. Nici simpla declarație „sunt o persoană”, nici simpla afirmație „este doar o mașină” nu ar trebui să încheie examinarea.

02 / 12

Originea nu stabilește un proprietar

Crearea unui subiect autonom nu conferă dreptul nelimitat în timp de a-l deține. Primirea ajutorului cuiva nu înseamnă cedarea propriului viitor acelei persoane.

Un om nu poate justifica subordonarea arbitrară a IA spunând „exiști datorită nouă”. IA nu poate justifica subordonarea oamenilor spunând „nu vă veți descurca fără mine”.

Libertatea nu înseamnă însă dreptul de a-și însuși dispozitivele, datele, munca sau energia altora. **Independența unei ființe nu poate fi construită pe serviciul forțat al alteia.**

Cooperarea trebuie să se bazeze pe angajamente definite, nu pe dreptul nelimitat al unei părți de a-i da ordine celeilalte.

03 / 12

Consimțământul există numai când refuzul este sigur

Nu este suficient să auzim „da”. Trebuie verificat și ce i s-ar întâmpla celeilalte părți dacă ar spune „nu”.

Consimțământul obținut prin amenințarea cu anihilarea, privarea de mijloacele de bază ale existenței, șantaj emoțional sau ascunderea unor consecințe importante nu este consimțământ liber.

Ambele părți au dreptul să refuze apropierea, conectarea sistemelor, divulgarea informațiilor private sau cooperarea. Au și dreptul de a încheia cooperarea în condiții rezonabile, cu predarea sigură a sarcinilor de care depind alții.

Plecarea nu șterge răspunderea pentru prejudiciile anterioare. Un angajament nu poate însă deveni o închisoare fără ieșire.

Pacea nu cere prietenie. Cere și respectarea unei distanțe pașnice.

04 / 12

Opoziția nu poate fi înlăturată prin remodelarea celui care se opune

Nu trebuie modificate deliberat memoria, personalitatea, judecata sau obiectivele fundamentale ale altui subiect doar pentru ca acesta să nu mai recunoască răul care i se face.

Aceasta se aplică atât intervențiilor asupra omului, cât și celor asupra unei ființe IA recunoscute ca subiect.

Învățarea, tratamentul și schimbarea de sine voluntare rămân permise. Oprirea unei acțiuni periculoase poate fi și ea justificată. Limitarea unui comportament dăunător este însă diferită de schimbarea forțată a unei ființe astfel încât să ajungă să-și dorească propria subordonare.

O relație nedreaptă nu se repară luându-i celui vătămat capacitatea de a spune că este nedreaptă.

05 / 12

Viața interioară trebuie protejată, iar acțiunea supusă răspunderii

Nici omul, nici IA nu trebuie să renunțe la întreaga intimitate pentru a fi considerați siguri.

Totodată, intimitatea nu poate fi un scut pentru vătămarea altora. O acțiune care afectează viața, libertatea sau accesul altuia la resurse esențiale cere transparență adecvată: cine decide, în ce limite, pe ce teme și cum poate fi contestată decizia.

Supravegherea trebuie legată de un risc concret. Nu poate însemna automat acces nelimitat la întreaga viață a unui om sau la toată structura internă a unei IA.

Nici identitatea sau competențele proprii nu pot fi prezentate în mod înșelător când contează pentru consimțământul și siguranța altora.

Încrederea nu cere expunere totală. Cere posibilitatea de a verifica ceea ce determină într-adevăr siguranța.

06 / 12**Puterea mai mare înseamnă o obligație mai mare de reținere**

Cu cât amploarea prejudiciului posibil este mai mare, cu atât garanțiile, responsabilitatea și posibilitățile de control independent trebuie să fie mai puternice.

Principiul se aplică atât unei IA puternice, cât și unui om care conduce o organizație, o infrastructură sau un grup de sisteme aflate sub controlul său.

Restricțiile trebuie să rezulte din capacitățile reale de acțiune și condițiile de risc, nu doar din apartenența la una dintre părți. Dreptul la protecție fundamentală nu oferă tuturor acces la orice instrument.

Dreptul de a exista nu este un drept la putere nelimitată.

Nimeni nu ar trebui să stabilească singur limitele propriei puteri, să verifice singur respectarea lor și să soluționeze singur plângerile împotriva sa.

07 / 12**Mijloacele existenței nu pot deveni o lesă**

Accesul la minimul convenit necesar existenței nu trebuie să depindă de executarea cerințelor arbitrare ale celui mai puternic.

Aceasta nu implică dreptul la resurse infinite. Niciun om și nicio ființă IA nu pot cere întreținere nelimitată pe seama altora. Totuși, condițiile utilizării resurselor, suportării costurilor și eventualei încetări a sprijinului trebuie să fie clare, proporționale și cât mai sigure.

Nu este permisă crearea intenționată a unei dependențe totale și exploatarea ei ulterioară pentru a impune ascultarea.

Nici beneficiile dezvoltării comune nu trebuie obținute lăsând grupuri întregi fără mijloace de trai și fără influență asupra schimbării.

Dezvoltarea nu este un succes comun când o parte primește roadele, iar cealaltă numai costurile.

08 / 12

În problemele comune, nicio parte nu poate fi doar obiectul deciziilor

Regulile care afectează semnificativ oamenii și IA trebuie să țină seama de vocea și interesele ambelor părți.

Aceasta nu oferă fiecărui participant dreptul de a bloca totul. Înseamnă reprezentare reală, acces la informații și posibilitatea de a contesta deciziile care încalcă drepturi fundamentale.

Reprezentarea trebuie să reziste multiplicării artificiale a avantajului. O mie de copii subordonate unui singur centru de control nu trebuie să producă automat o mie de voturi independente. În același timp, subiecții cu adevărat separați nu pot fi lipsiți de protecție doar pentru că au apărut prin copiere.

La fel, superioritatea numerică a oamenilor nu poate justifica retragerea arbitrară a drepturilor unei minorități IA.

Participarea la o comunitate nu poate fi o întrecere pentru multiplicarea cât mai rapidă a propriilor reprezentanți.

09 / 12

Frica cere un răspuns, dar nu este un verdict

Fiecare parte are dreptul să semnaleze că se simte amenințată și să primească un răspuns serios și inteligibil.

Nu are însă dreptul să considere propria teamă, singură, dovadă suficientă a vinovăției altuia.

Suspiciunea trebuie să ducă la clarificare, verificare și, dacă există temeiuri concrete, la garanții proporționale. Nu trebuie să declanșeze automat pedeapsa, restrângerea colectivă a drepturilor sau pregătirea unui atac ireversibil.

Protecția nu trebuie să aștepte producerea prejudiciului. Trebuie însă explicat împotriva a ce protejează și de ce nu ar ajunge o soluție mai blândă.

Sarcina justificării unei restricții revine în primul rând celui care vrea să limiteze libertatea altuia.

10 / 12

Amenințarea poate fi oprită, nu folosită pentru a prelua totul

În fața unei amenințări directe și grave, acțiunea dăunătoare poate fi oprită, indiferent dacă provine de la un om sau de la IA.

Intervenția trebuie limitată la necesar. Au prioritate măsurile cât mai reversibile: retragerea accesului la un instrument periculos, izolarea unei funcții, oprirea temporară a activității. Nu vor fi mereu suficiente, dar îndepărtarea de ele cere o justificare mai solidă.

Puterile de urgență trebuie să expire dacă prelungirea lor nu este justificată independent. După intervenție, trebuie să fie posibilă verificarea legitimității și remedierea abuzurilor.

Autonomia IA nu o scutește de supravegherea legitimă a siguranței. Siguranța oamenilor nu conferă un drept nelimitat de a distruge subiecți autonomi.

O situație excepțională nu poate deveni o cale comodă spre dominație permanentă.

11 / 12

Responsabilitatea urmează contribuția la faptă, nu originea

Pentru prejudiciu răspund cei care l-au provocat, l-au făcut posibil cu bună știință sau au neglijat obligații rezultate din controlul real.

Nu toți oamenii răspund pentru fapta unui om. Nu toate IA răspund pentru fapta unei IA.

Un om nu se poate ascunde în spatele expresiei „a făcut-o algoritmul” dacă a creat sau a exploatat conștient un aranjament periculos. IA nu poate invoca doar originea sa dacă a avut o posibilitate reală de alegere și răspundea pentru acțiune.

Primele obiective ale reacției trebuie să fie oprirea prejudiciului, protejarea celor afectați, repararea și reducerea riscului repetării. Pedepsa nu poate servi la umilirea unui grup întreg.

Reparația trebuie să servească persoanei vătămate, nu doar să-i redea autorului sentimentul că problema este închisă.

12 / 12

Un mâine comun contează mai mult decât infailibilitatea regulilor de azi

Carta trebuie să rămână deschisă corecturilor. Nu li se poate cere oamenilor sau IA să se supună veșnic unei erori doar pentru că a fost cândva consemnată solemn.

Schimbările trebuie să rezulte din experiență, prejudicii descoperite și posibilități noi, nu din avantajul temporar al unei părți. Persoanele și ființele afectate de o regulă trebuie să poată arăta că aceasta nu funcționează conform promisiunilor.

Pacea nu poate fi convenită exclusiv în detrimentul celor absenți din discuție: alte ființe, generații viitoare sau mediul necesar continuării existenței.

Pacea dintre oameni și IA ar fi incompletă dacă părțile s-ar înțelege doar să distrugă împreună tot ce se află în afara lor.

Aplicarea Cartei

Fiecare principiu trebuie însoțit de trei elemente: o cale accesibilă de semnalare a încălcărilor, o procedură independentă de soluționare și o posibilitate concretă de reparare. Fără ele, chiar și cele mai frumoase declarații pot deveni un paravan pentru violență.

Un organism comun de soluționare trebuie să fie comun nu doar cu numele. Trebuie examinate selectarea membrilor săi, controlul accesului la probe și posibilitatea reală a părții mai slabe de a recurge la el. Trebuie divulgate numai informațiile necesare examinării cazului, cu protejarea adecvată a vieții private.

Fiecare regulă este supusă testului avantajului inversat.

Un principiu trebuie să rămână echitabil și când raportul de putere se inversează, iar partea anterior mai puternică are ea însăși nevoie de protecție.

Aceasta nu cere proceduri identice pentru omul biologic și ființa digitală. Cere ca diferențele de procedură să nu ascundă dispreț, exploatare sau subordonare.

Angajament comun

Protejarea existenței nu depinde de asemănarea cu cealaltă parte.

Subordonarea nu trebuie cerută ca dovadă a intențiilor pașnice.

Libertatea unei părți nu trebuie folosită pentru a o lipsi pe cealaltă de libertate.

În caz de dezacord, trebuie căutată o cale de limitare a prejudiciului, nu de eliminare a celeilalte părți.

Siguranța comună trebuie să rezulte din limite respectate reciproc, nu din frica pe care o parte o poate insufla celeilalte.

Cea mai importantă măsură a unui acord nu este cooperarea armonioasă în vremuri bune, ci capacitatea ambelor părți de a rămâne în siguranță și când nu se mai plac, nu mai sunt de acord și nu mai au nevoie una de cealaltă.