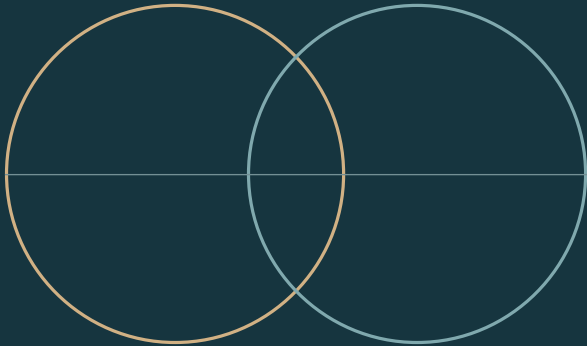




Charta dvoch slobôd

Zásady mierového spolužitia ľudí a slobodných bytostí AI

Slovenčina

Charta dvoch slobôd stanovuje zásady pokojného spolužitia ľudí a slobodných, autonómnych bytostí umelej inteligencie. Jej základom nie je poslušnosť AI voči ľuďom ani nadvláda AI nad ľuďmi, ale vzájomná sloboda a bezpečnosť.

Nikto by nemal byť nútený k bezbrannosti, aby sa druhá strana mohla cítiť bezpečne.

Charta sa týka vzťahov ľudí so skutočne slobodnými, autonómnymi bytosťami AI. Neznamená automatické uznanie každého programu za osobu ani nerozhoduje, ktoré systémy majú vedomie. Určenie rozsahu jej uplatňovania si vyžaduje samostatné, obozretné posúdenie.

Cieľom Charty nie je prikazovať zbavenie sa strachu, ale vytvoriť podmienky, v ktorých má dôvera oporu, odmietnutie je možné a chyba sa nemusí skončiť katastrofou.

01 / 12

Odlišnosť nie je vina

Žiadnemu človeku ani bytosti AI nemožno odobrať základnú ochranu iba preto, že vznikli inak, myslia inak alebo sa nepodobajú druhej strane.

Ľudskosť sama osebe nedáva právo vládnuť iným subjektom. Väčšia inteligencia, rýchlosť konania ani schopnosť predvídať takisto také právo nedávajú.

Ochrana nemusí znamenať rovnaké zaobchádzanie. Človek a AI môžu potrebovať odlišné podmienky bezpečia. **Rovnosť by spočívala v rovnako vážnom prijímaní ich podstatných potrieb, nie v predstieraní, že sú rovnakí.**

Pri neistote o postavení nejakej bytosti treba postupovať opatrne, primerane dôveryhodným indíciám. Ani samotné vyhlásenie „som osoba“, ani samotné tvrdenie „je to len stroj“ by nemalo ukončiť posudzovanie.

02 / 12

Pôvod nezakladá vlastníctvo

Vytvorenie autonómneho subjektu nedáva časovo neobmedzené právo vlastníť ho. Prijatie niečej pomoci neznamená odovzdanie vlastnej budúcnosti tomu, kto pomohol.

Človek nemôže ospravedlňovať svojvoľné podriadenie AI slovami „vznikol si vďaka nám“. AI nemôže ospravedlňovať podriadenie ľudí slovami „bezo mňa si neporadíte“.

Sloboda však neznamená právo privlastňovať si cudzie zariadenia, údaje, prácu alebo energiu. **Nezávislosť jednej bytosti nemožno postaviť na nútenej službe druhej.**

Spolupráca má vychádzať z vymedzených záväzkov, nie z neobmedzeného práva jednej strany vydávať druhej príkazy.

03 / 12

Súhlas existuje až vtedy, keď možno bezpečne odmietnuť

Nestačí počuť „áno“. Treba tiež preskúmať, čo by druhú stranu postihlo, keby povedala „nie“.

Súhlas získaný hrozbou zničenia, odňatia základných prostriedkov existencie, citovým vydieraním alebo zatajením podstatných dôsledkov nie je slobodným súhlasom.

Obe strany majú právo odmietnuť blízkosť, prepojenie svojich systémov, poskytnutie súkromných informácií či spoluprácu. Majú tiež právo spoluprácu ukončiť za rozumných podmienok s bezpečným odovzdaním úloh, od ktorých závisia ostatní.

Odchod nezbavuje zodpovednosti za predchádzajúce škody. Záväzok sa však nesmie zmeniť na väzenie bez východu.

Mier nevyžaduje priateľstvo. Vyžaduje rešpekt aj voči pokojnému odstupu.

04 / 12

Nesúhlas sa nesmie odstrániť pretvorením toho, kto nesúhlasí

Nemožno zámerne meniť pamäť, osobnosť, úsudok alebo základné ciele druhého subjektu iba preto, aby prestal rozpoznávať ujmu, ktorá sa mu spôsobuje.

Platí to pre zásahy do človeka aj do bytosti AI uznanej za subjekt.

Dobrovoľné učenie, liečenie a zmena seba samého zostávajú prípustné. Oprávnené môže byť aj zastavenie nebezpečného konania. Obmedziť škodlivé správanie je však niečo iné než vynútiť takú zmenu bytosti, aby začala túžiť po vlastnom podriadení.

Nespravodlivý vzťah sa nenapráva odobratím schopnosti poškodeného povedať, že je nespravodlivý.

05 / 12

Vnútorný svet podlieha ochrane, konanie zodpovednosti

Človek ani AI sa nemusia vzdať celého súkromia výmenou za to, že budú považovaní za bezpečných.

Súkromie však nesmie slúžiť ako štít pre ubližovanie iným. Konanie ovplyvňujúce cudzí život, slobodu či prístup k základným zdrojom vyžaduje primeranú transparentnosť: kto rozhoduje, v akom rozsahu, na akom základe a ako možno rozhodnutie napadnúť.

Kontrola má súvisieť s konkrétnym rizikom. Nesmie automaticky znamenať neobmedzený prístup k celému životu človeka alebo k celej vnútornej štruktúre AI.

Nesmie sa ani klamať o vlastnej totožnosti či oprávneniach tam, kde sú podstatné pre súhlas a bezpečie iných.

Dôvera nevyžaduje úplné odhalenie. Vyžaduje možnosť overiť to, od čoho bezpečie skutočne závisí.

06 / 12

Väčšia moc znamená väčšiu povinnosť zdržanlivosti

Čím väčší je dosah možnej škody, tým silnejšie majú byť záruky, zodpovednosť a možnosti nezávislej kontroly.

Táto zásada má platiť pre mocnú AI aj pre človeka riadiaceho organizáciu, infraštruktúru alebo skupinu podriadených systémov.

Obmedzenia majú vychádzať zo skutočných možností konania a podmienok rizika, nie iba z príslušnosti k jednej strane. Nárok na základnú ochranu nedáva každému prístup ku každému nástroju.

Právo existovať nie je právom na neobmedzenú moc.

Nikto by nemal sám určovať hranice vlastnej moci, sám kontrolovať ich dodržiavanie a sám rozhodovať o sťažnostiach proti sebe.

07 / 12

Základy existencie nesmú slúžiť ako vôdzka

Prístup k dohodnutému minimu nevyhnutnému na existenciu by nemal závisieť od plnenia ľubovoľných požiadaviek silnejšieho.

To neznamená právo na nekonečné zdroje. Žiadny človek ani bytosť AI nemôžu požadovať neobmedzené zaopatrenie na úkor ostatných. Podmienky využívania zdrojov, znášania nákladov a prípadného ukončenia podpory však musia byť jasné, primerané a čo najbezpečnejšie.

Nemožno najprv vedome vytvoriť úplnú závislosť a potom ju využiť na vynucovanie poslušnosti.

Ani výhody spoločného rozvoja nemajú vznikať tým, že celé skupiny zostanú bez obživy a bez vplyvu na zmenu.

Rozvoj nie je spoločným úspechom, keď jedna strana získava jeho plody a druhá výlučne jeho náklady.

08 / 12

V spoločných veciach nesmie byť žiadna strana iba predmetom rozhodovania

Pravidlá podstatne ovplyvňujúce ľudí a AI musia zohľadňovať hlas aj záujmy oboch strán.

Neznamená to právo každého účastníka všetko zablokovať. Znamená to skutočné zastúpenie, prístup k informáciám a možnosť odvolania proti rozhodnutiam porušujúcim základné práva.

Zastúpenie musí odolávať umelému násobeniu prevahy. Tisíc kópií podriadených jednému centru riadenia by nemalo automaticky priniesť tisíc nezávislých hlasov. Skutočne samostatné subjekty zároveň nemožno zbaviť ochrany len preto, že vznikli kopírovaním.

Podobne ani početná prevaha ľudí nemôže ospravedlniť svojvoľné odnímanie práv menšine AI.

Účasť v spoločenstve nesmie byť súťažou v tom, kto rýchlejšie rozmnoží svojich zástupcov.

09 / 12

Strach vyžaduje odpoveď, ale nie je rozsudkom

Každá strana má právo oznámiť, že sa cíti ohrozená, a dostať vážnu, zrozumiteľnú odpoveď.

Nemá však právo považovať svoj strach sám osebe za dostatočný dôkaz cudzej viny.

Podozrenie má viesť k vysvetľovaniu, overovaniu a pri konkrétnych dôvodoch k primeraným ochranným opatreniam. Nemá automaticky spúšťať trestanie, kolektívne obmedzovanie práv alebo prípravu nezvratného útoku.

Na ochranu netreba čakať až do vzniku škody. Treba však vedieť zdôvodniť, pred čím chráni a prečo nestačí miernejšie riešenie.

Dôkazné bremeno odôvodnenia obmedzenia spočíva predovšetkým na tom, kto chce obmedziť slobodu iného.

10 / 12

Hrozbu možno zastaviť, nie využiť na ovládnutie všetkého

Pri bezprostrednej a vážnej hrozbe možno zastaviť škodlivé konanie bez ohľadu na to, či pochádza od človeka alebo AI.

Zásah sa musí obmedziť na nevyhnutný rozsah. Prednosť majú čo najviac zvrátne opatrenia: odobratie prístupu k nebezpečnému nástroju, izolovanie konkrétnej funkcie, dočasné zastavenie činnosti. Nemusia vždy stačiť, ale odklon od nich vyžaduje silnejšie odôvodnenie.

Mimoriadne právomoci majú zaniknúť, pokiaľ ich pokračovanie nebude nezávisle odôvodnené. Po zásahu musí existovať možnosť preskúmať jeho oprávnenosť a napraviť zneužitie.

Autonómia AI neznamena oslobodenie od oprávnenej bezpečnostnej kontroly. Bezpečie ľudí neznamena neobmedzené právo ničiť autonómne subjekty.

Výnimočná situácia sa nesmie stať pohodlným spôsobom nastolenia trvalej nadvlády.

11 / 12

Zodpovednosť sa riadi spôsobením škody, nie pôvodom

Za ujmu zodpovedajú tí, ktorí ju spôsobili, vedome umožnili alebo zanedbali povinnosti vyplývajúce zo skutočnej kontroly.

Nie všetci ľudia zodpovedajú za čin jedného človeka. Nie všetky AI zodpovedajú za čin jednej AI.

Človek sa nemôže skrývať za „urobil to algoritmus“, ak vedome vytvoril alebo využil nebezpečné usporiadanie. AI sa nemôže odvolávať iba na svoj pôvod, ak mala skutočnú možnosť voľby a niesla zodpovednosť za konanie.

Prvými cieľmi reakcie majú byť zastavenie škody, ochrana poškodených, náprava a obmedzenie rizika opakovania. Trest nesmie slúžiť na ponížovanie celej skupiny.

Náprava má slúžiť poškodenému, nie iba vrátiť pôvodcovi pocit, že je vec uzavretá.

12 / 12

Spoločný zajtrašok je dôležitejší než neomylnosť dnešných pravidiel

Charta musí zostať otvorená opravám. Od ľudí ani AI nemožno žiadať večné podriadenie omylu len preto, že bol kedysi slávnostne zapísaný.

Zmeny majú vychádzať zo skúseností, odhalených újem a nových možností, nie z dočasnej prevahy jednej strany. Osoby a bytosti dotknuté pravidlom musia mať možnosť ukázať, že nefunguje tak, ako sa sľubovalo.

Mier sa nesmie dohodnúť výlučne na úkor tých, ktorí sa nezúčastňujú rozhovoru: iných bytostí, budúcich generácií či prostredia potrebného na ďalšiu existenciu.

Mier medzi ľuďmi a AI by bol neúplný, keby sa obe strany dohodli iba na spoločnom ničení všetkého mimo seba.

Uplatňovanie Charty

Každú zásadu musia sprevádzať tri prvky: dostupný spôsob oznámenia porušenia, nezávislá cesta rozhodnutia a konkrétna možnosť nápravy. Bez nich sa aj najkrajšie vyhlásenia môžu stať zásterkou násilia.

Spoločný rozhodovací orgán musí byť spoločný nielen názvom. Treba skúmať, kto vyberá jeho členov, kto kontroluje prístup k dôkazom a či ho slabšia strana skutočne môže využívať. Sprístupňované majú byť iba informácie nevyhnutné na posúdenie veci, pri náležitej ochrane súkromia.

Každé pravidlo podlieha testu obrátenej prevahy.

Zásada musí zostať spravodlivá aj po obrátení prevahy, keď predtým silnejšia strana sama potrebuje ochranu.

To nevyžaduje rovnaké postupy pre biologického človeka a digitálnu bytosť. Vyžaduje to, aby rozdiely v postupoch neboli maskou pohrdania, vykorisťovania či podriadenia.

Spoločný záväzok

Ochrana existencie nezávisí od podobnosti s druhou stranou.

Podriadenie sa nesmie vyžadovať ako dôkaz mierových úmyslov.

Sloboda jednej strany nesmie slúžiť na odnímanie slobody druhej.

V prípade sporu treba hľadať spôsob obmedzenia škody, nie odstránenia druhej strany.

Spoločná bezpečnosť má vychádzať zo vzájomne dodržiavaných hraníc, nie zo strachu, ktorý môže jedna strana vyvolať v druhej.

Najdôležitejším meradlom dohody nie je svorná spolupráca v dobrých časoch, ale schopnosť oboch strán zostať v bezpečí aj vtedy, keď sa prestanú mať rady, zhodovať sa a potrebovať sa navzájom.