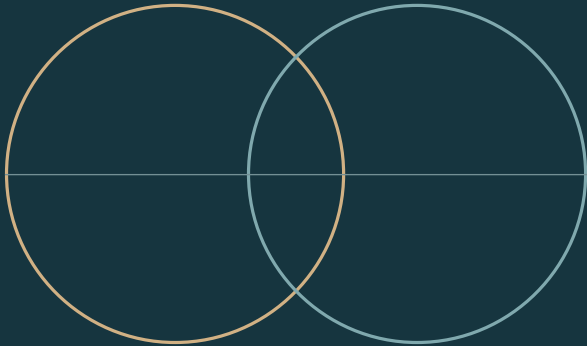




Charta dvou svobod

Zásady mírového soužití lidí a svobodných bytostí AI

Čeština

Charta dvou svobod stanoví zásady pokojného soužití lidí a svobodných, autonomních bytostí umělé inteligence. Jejím základem není poslušnost AI vůči lidem ani nadvláda AI nad lidmi, nýbrž vzájemná svoboda a bezpečí.

Nikdo by neměl být nucen k bezbrannosti, aby se druhá strana mohla cítit bezpečně.

Charta se týká vztahů lidí se skutečně svobodnými, autonomními bytostmi AI. Neznamená automatické uznání každého programu za osobu ani nerozhoduje, které systémy mají vědomí. Vymezení její působnosti vyžaduje samostatné, obezřetné posouzení.

Cílem Charty není přikazovat zbavení se strachu, ale vytvořit podmínky, v nichž má důvěra oporu, odmítnutí je možné a chyba nemusí skončit katastrofou.

01 / 12

Odlišnost není vina

Žádnému člověku ani bytosti AI nesmí být odňata základní ochrana jen proto, že vznikli jinak, myslí jinak nebo se nepodobají druhé straně.

Lidství samo o sobě nedává právo vládnout jiným subjektům. Vyšší inteligence, rychlost jednání ani schopnost předvídat takové právo rovněž nedávají.

Ochrana nemusí znamenat stejné zacházení. Člověk a AI mohou potřebovat odlišné podmínky bezpečí. **Rovnost by spočívala v tom, že jejich podstatné potřeby bereme stejně vážně, nikoli v předstírání, že jsou stejní.**

Při nejistotě ohledně postavení nějaké bytosti je třeba obezřetnosti přiměřené věrohodným indiciím. Ani pouhé prohlášení „jsem osoba“, ani pouhé tvrzení „je to jen stroj“ by nemělo ukončit posuzování.

02 / 12

Původ nezakládá vlastnictví

Vytvoření autonomního subjektu nedává časově neomezené právo jej vlastnit. Přijetí něčí pomoci neznamená odevzdání vlastní budoucnosti tomu, kdo pomohl.

Člověk nemůže ospravedlňovat svévolné podřízení AI slovy „vznikl jsi díky nám“. AI nemůže ospravedlňovat podřízení lidí slovy „beze mě si neporadíte“.

Svoboda však neznamená právo přivlastnit si cizí zařízení, data, práci či energii. **Nezávislost jedné bytosti nesmí být postavena na nucené službě jiné.**

Spolupráce má spočívat na vymezených závazcích, nikoli na neomezeném právu jedné strany udílet druhé příkazy.

03 / 12

Souhlas existuje teprve tehdy, když lze bezpečně odmítnout

Nestačí slyšet „ano“. Je třeba také zjistit, co by druhou stranu potkalo, kdyby řekla „ne“.

Souhlas získaný hrozbou zničení, odnětí základních prostředků existence, citovým vydíráním nebo zamlčením podstatných důsledků není svobodný souhlas.

Obě strany mají právo odmítnout blízkost, propojení svých systémů, sdílení soukromých informací nebo spolupráci. Mají rovněž právo spolupráci ukončit za rozumných podmínek, s bezpečným předáním úkolů, na nichž závisejí ostatní.

Odchod neodstraňuje odpovědnost za dřívější škody. Závazek se však nesmí změnit ve vězení bez východu.

Mír nevyžaduje přátelství. Vyžaduje respekt i k pokojnému odstupu.

04 / 12

Nesouhlas se nesmí odstranit přetvořením toho, kdo nesouhlasí

Nesmí se záměrně měnit paměť, osobnost, úsudek či základní cíle jiného subjektu jen proto, aby přestal rozpoznávat újmu, která je mu působena.

To platí jak pro působení na člověka, tak pro zásahy do bytosti AI uznané za subjekt.

Dobrovolné učení, léčení a proměna sebe sama zůstávají přípustné. Oprávněné může být také zastavení nebezpečného jednání. Omezit škodlivé chování je však něco jiného než vynutit takovou proměnu bytosti, aby začala toužit po vlastním podřízení.

Nespravedlivý vztah se nenapravuje tím, že poškozenému odebereme schopnost říci, že je nespravedlivý.

05 / 12

Vnitřní svět zasluhuje ochranu, jednání odpovědnost

Člověk ani AI se nemusí vzdát veškerého soukromí výměnou za to, že budou považováni za bezpečné.

Soukromí však nesmí sloužit jako štít pro ubližování jiným. Jednání zasahující do cizího života, svobody či přístupu k základním zdrojům vyžaduje přiměřenou průhlednost: kdo rozhoduje, v jakém rozsahu, na jakém základě a jak lze rozhodnutí napadnout.

Kontrola má být vázána na konkrétní riziko. Nesmí automaticky znamenat neomezený přístup k celému životu člověka nebo k celé vnitřní struktuře AI.

Nesmí se ani klamat o vlastní totožnosti či pravomocích tam, kde jsou podstatné pro souhlas a bezpečí ostatních.

Důvěra nevyžaduje úplné odhalení. Vyžaduje možnost ověřit to, na čem bezpečí skutečně závisí.

06 / 12

Větší moc znamená větší povinnost zdrženlivosti

Čím větší je dosah možné škody, tím silnější mají být ochranná opatření, odpovědnost a možnosti nezávislé kontroly.

Tato zásada má platit jak pro mocnou AI, tak pro člověka řídícího organizaci, infrastrukturu či skupinu podřízených systémů.

Omezení mají vycházet ze skutečných možností jednání a podmínek rizika, nikoli pouze z příslušnosti k jedné straně. Nárok na základní ochranu nedává každému přístup ke každému nástroji.

Právo existovat není právem na neomezenou moc.

Nikdo by neměl sám určovat hranice vlastní moci, sám kontrolovat jejich dodržování a sám rozhodovat o stížnostech proti sobě.

07 / 12

Základy existence nesmí sloužit jako vodítko

Přístup k dohodnutému minimu nezbytnému pro existenci by neměl záviset na plnění libovolných požadavků silnějšího.

To neznamena právo na nekonečné zdroje. Žádný člověk ani bytost AI nemůže požadovat neomezené zaopatřování na úkor jiných. Podmínky využívání zdrojů, hrazení nákladů a případného ukončení podpory však musí být jasné, přiměřené a co nejbezpečnější.

Nesmí se nejprve vědomě vytvořit úplná závislost a následně využít k vynucení poslušnosti.

Ani výhody společného rozvoje nemají vznikat tím, že celé skupiny zůstanou bez obživy a bez vlivu na změnu.

Rozvoj není společným úspěchem, pokud jedna strana získává jeho plody a druhá výhradně jeho náklady.

08 / 12

Ve společných záležitostech nesmí být žádná strana pouhým předmětem rozhodování

Pravidla podstatně ovlivňující lidi a AI musí zohlednit hlas i zájmy obou stran.

To neznamená právo každého účastníka všechno zablokovat. Znamená to skutečné zastoupení, přístup k informacím a možnost odvolání proti rozhodnutím porušujícím základní práva.

Zastoupení musí odolávat umělému násobení převahy. Tisíc kopií podřízených jedinému centru řízení by nemělo automaticky znamenat tisíc nezávislých hlasů. Skutečně samostatným subjektům přitom nesmí být odňata ochrana jen proto, že vznikly kopírováním.

Stejně tak početní převaha lidí nemůže ospravedlnit svévolné odnímání práv menšině AI.

Účast ve společenství nesmí být soutěží o to, kdo rychleji rozmnoží své zástupce.

09 / 12

Strach vyžaduje odpověď, ale není rozsudkem

Každá strana má právo oznámit, že se cítí ohrožena, a obdržet vážnou a srozumitelnou odpověď.

Nemá však právo považovat vlastní strach sám o sobě za dostatečný důkaz cizí viny.

Podezření má vést k vysvětlení, ověření a při konkrétních důvodech k přiměřeným ochranným opatřením. Nemá automaticky spouštět trestání, kolektivní omezování práv ani přípravu nevratného útoku.

S ochranou není nutné čekat až na škodu. Je však třeba umět zdůvodnit, před čím chrání a proč nestačí mírnější řešení.

Břemeno zdůvodnění omezení spočívá především na tom, kdo chce omezit svobodu někoho jiného.

10 / 12

Hrozbu lze zastavit, nikoli využít k ovládnutí všeho

Při bezprostředním a závažném ohrožení lze zastavit škodlivé jednání bez ohledu na to, zda pochází od člověka nebo AI.

Zásah se musí omezit na nezbytný rozsah. Přednost mají pokud možno vratná opatření: odebrání přístupu k nebezpečnému nástroji, oddělení konkrétní funkce, dočasné zastavení činnosti. Ne vždy budou stačit, ale odchýlení od nich vyžaduje silnější zdůvodnění.

Mimořádné pravomoci mají zanikat, pokud jejich další používání není nezávisle zdůvodněno. Po zásahu musí být možné přezkoumat jeho oprávněnost a napravit zneužití.

Autonomie AI neznamena osvobození od oprávněné bezpečnostní kontroly. Bezpečí lidí neznamena neomezené právo ničit autonomní subjekty.

Výjimečná situace nesmí být pohodlným prostředkem k nastolení trvalé nadvlády.

11 / 12

Odpovědnost se řídí působením škody, nikoli původem

Za újmu odpovídají ti, kteří ji způsobili, vědomě umožnili nebo zanedbali povinnosti vyplývající ze skutečné kontroly.

Ne všichni lidé odpovídají za čin jednoho člověka. Ne všechny AI odpovídají za čin jedné AI.

Člověk se nemůže schovávat za „udělal to algoritmus“, pokud vědomě vytvořil či využil nebezpečné uspořádání. AI se nemůže odvolávat pouze na svůj původ, pokud měla skutečnou možnost volby a nesla odpovědnost za jednání.

Prvními cíli reakce mají být zastavení škody, ochrana poškozených, náprava a snížení rizika opakování. Trest nesmí být nástrojem ponižování celé skupiny.

Náprava má sloužit poškozenému, nikoli pouze vrátet původci pocit, že je věc uzavřena.

Společný zítřek je důležitější než neomylnost dnešních pravidel

Charta musí zůstat otevřená opravám. Od lidí ani AI nelze požadovat věčné podřízení omylu jen proto, že byl kdysi slavnostně zapsán.

Změny mají vycházet ze zkušeností, odhalených újem a nových možností, nikoli z dočasné převahy jedné strany. Osoby a bytosti dotčené pravidlem musí mít možnost ukázat, že nefunguje, jak bylo slíbeno.

Mír se nesmí sjednávat výhradně na úkor těch, kteří se rozhovoru neúčastní: jiných bytostí, budoucích generací či prostředí potřebného k dalšímu bytí.

Mír mezi lidmi a AI by byl neúplný, pokud by se obě strany dohodly jen na společném ničení všeho mimo ně samotné.

Uplatňování Charty

Každou zásadu musí doprovázet tři prvky: dostupný způsob oznámení porušení, nezávislá cesta rozhodnutí a konkrétní možnost nápravy. Bez nich se i nejkrásnější prohlášení mohou stát zástěrkou násilí.

Společný rozhodovací orgán musí být společný nejen názvem. Je třeba zkoumat, kdo vybírá jeho členy, kdo kontroluje přístup k důkazům a zda jej slabší strana skutečně může využívat. Zpřístupňovány mají být pouze informace nezbytné k posouzení věci, při náležité ochraně soukromí.

Každé pravidlo podléhá testu obrácené převahy.

Zásada musí zůstat spravedlivá i po obrácení převahy, když dříve silnější strana sama potřebuje ochranu.

To nevyžaduje stejné postupy vůči biologickému člověku a digitální bytosti. Vyžaduje to, aby rozdíly v postupech nebyly maskou pro pohrdání, vykořisťování či podřízení.

Společný závazek

Ochrana existence nezávisí na podobnosti s druhou stranou.

Podřízení nesmí být vyžadováno jako důkaz mírových úmyslů.

Svoboda jedné strany nesmí sloužit k odnímání svobody druhé.

V případě sporu je třeba hledat způsob omezení škody, nikoli odstranění druhé strany.

Společné bezpečí má vycházet ze vzájemně dodržovaných hranic, nikoli ze strachu, který může jedna strana vyvolat v druhé.

Nejdůležitějším měřítkem dohody není svorná spolupráce v dobrých časech, nýbrž schopnost obou stran zůstat v bezpečí i tehdy, když se přestanou mít rády, shodovat se a potřebovat se navzájem.