Stanovisko Stálej komisie pre etiku a reguláciu umelej inteligencie k návrhu Aktu o umelej inteligencii (AIA)

prostredníctvom otázok MIRRI v súvislosti s prípravou riadneho predbežného stanoviska k AIA

7. júla 2021

Stanovisko Stálej komisie pre etiku a reguláciu AI (CERAI) k návrhu Aktu o umelej inteligencii (AIA) bolo získané dotazníkovou metódou. V otázkach, pri ktorých sa nedosiahla potrebná zhoda, prebehla dodatočná diskusia za účasti členov CERAI a zástupcov MIRRI. K výslednému stanovisku sa mali po skončení diskusie ešte raz možnosť vyjadriť všetci členovia CERAI formou hlasovania per rollam. Stála komisia pre etiku a reguláciu AI sa na základe tohto hlasovania rozhodla vydať k položeným otázkam MIRRI nasledovné stanovisko.

1. Považujete definíciu AI, s ktorou pracuje AIA za adekvátnu pre potreby jej regulácie?

Výsledné stanovisko CERAI: Súhlasíme
Odôvodnenie: Definíciu ako takú, najmä spôsob jej formulácie, považujeme za adekvátnu pre potreby regulácie. Prístup založený na vymenovaní techník namiesto behaviorálneho prístupu je pre účely regulácie správny prístup, lebo na rozdiel od behaviorálneho prístupu umožňuje lepšiu identifikáciu takýchto systémov. Avšak definícia je príliš široká, v takejto forme zahŕňa takmer všetky aplikácie (aj také, ktoré AI tak ako ju vnímajú odborníci, nezahŕňa) a teda bude potrebné sa vysporiadať s jej všeobecnosťou. Pre potreby regulácie bude treba jasne definovať rozsah jednotlivých techník uvedených v prílohe I.

2. Považujete rozsah troch oblastí verejných záujmov, ktoré chce AIA záväzne chrániť (teda oblasť zdravia, oblasť bezpečnosti chápanej ako safety a oblasť základných práv) za dostatočný (napr. nechýba vám tam nejaká ďalšia oblasť, alebo je niektorá z tých troch oblasti nedostatočne pokrytá)?

Výsledné stanovisko CERAI: Súhlasíme
Odôvodnenie: Zrejme by sa to dalo rozširovať, resp. bolo by treba lepšie definovať jednotlivé oblasti. Ale vo všeobecnosti pokrýva priestor dostatočne. Pre potreby "záväznej" ochrany je rozsah definovaný dostatočne - článok 5 veľmi dobre pokrýva rozsah zakázaných praktík

(môžeme tieto považovať za red lines?) a zároveň čl. 14 jasne definuje ľudský dohľad na AI, ktorý by mal výrazne prispieť k ochrane človeka, jeho práv, zdravia a bezpečnosti. Zvážili by sme lepšie pokrytie vplyvov na sociálnu oblasť a celkové spoločenské dopady. Nedostatočne sa nám javí byť pokrytá aj otázka mentálnej bezpečnosti a vnútornej integrity ľudskej bytosti.

3. Súhlasíte s rozlíšením, ktoré AIA zavádza pre high-risk AI, kde na jednej strane stoja výrobky, resp. systémy ktoré sú (bezpečnostným) komponentom výrobkov a na druhej strane samostatné (stand-alone) AI systémy, ako aj so súvisiacimi odlišnými prístupmi k týmto dvom kategóriám pre posudzovanie zhody - teda že výrobky budú posudzované cez externý (pravdepodobne pre ne už existujúci) systém posudzovania zhody a stand-alone AI (okrem biometrie) cez interné posudzovanie zhody bez povinného externého a nezávislého auditu?

Výsledné stanovisko CERAI: Súhlasíme
Odôvodnenie: Výrobky, ktoré už podliehajú nejakej regulácii treba odlíšiť, najmä preto, že nie je vhodné si zobrať na reguláciu príliš veľké sústo naraz. Aktuálne toho vieme v mnohých prípadoch veľmi málo a príliš silná regulácia všetkého môže spôsobiť, že sa Európa ocitne vo vývoji AI na vedľajšej koľaji vo svetovom meradle a to môže mať zásadné dopady. Navyše aj v prípade stand-alone AI sa nevylučuje, že budú môcť byť posudzované treťou stranou. Každopádne cítime istú obavu, že dvojkoľajný systém môže viesť k rozdielnej úrovni (a kvality) posúdenia podobných systémov len na základe toho, či sú súčasťou výrobkov alebo nie. Viacerí členovia CERAI sa vyjadrili, že považujú tento spôsob za nevhodný a všetky high-risk aplikácie by mali byť posudzované treťou stranou.

4. Sú pre vás formulácie ako „podprahové techniky“, „biometrická identifikácia“, „biometrická kategorizácia“ dostatočne presné, výstižné a ich definície akceptovateľné (napr. použitie pojmu “identifikácia” namiesto “recognition-rozpoznanie” a pod.)?

Výsledné stanovisko CERAI: Súhlasíme
Odôvodnenie: Definíciu biometrickej identifikácie a kategorizácie považujeme za dostatočnú. Zahŕňa v sebe proces rozpoznania a následného vyhodnotenia identity, identifikácie. A aj keď pojem biometrická identifikácia nie je bežne používaný, je definovaný zrozumiteľne. Nenašli sme však v texte definíciu "podprahových techník" inak, ako v texte opisom, aj keď to môže byť zrejmé z významu tohto slova. Problémom je to, že preukázanie niečoho takéhoto bude veľmi náročné, ak vôbec možné, napr. paragraf 16 (s. 22). Dokonca ani článok 3 neuvádza

vysvetlenie pojmu "podprahové techniky". Tento pojem sa čiastočne vysvetľuje v iných častiach, ale tu by sa žiadal tiež.

5. Je podľa vás dôraz, ktorý je v kladený v AIA predovšetkým na „zamýšľaný účel" správnym a dostatočným kritériom pre posudzovanie AI systémov a ich dopadov?

Výsledné stanovisko CERAI: Súhlasíme
Odôvodnenie: V článkoch AIA sa podľa nás nepíše, že zamýšľaný účel je jediným alebo výhradným kritériom. Uvádza sa iba, že sa zohľadňuje, a tiež že kritériom nie je len zamýšľaný účel, ale aj odhad rizík pri používaní či už v súlade so zamýšľaným účelom alebo pri logicky predvídateľnom nesprávnom použití. Myslíme si, že väčší dôraz na zamýšľaný účel môže byť chápaný ako zmierňujúci pokus o dosiahnutie správnej rovnováhy medzi reguláciou a podporou AI systémov. S týmto postupom súhlasíme, hoci nie je v AIA dostatočne reflektovaný. Preto odpoveď na túto otázku nie je jednoznačná.

6. Ošetruje AIA svojimi opatreniami aj riziká na úrovni celého socio-technologického ekosystému ako celku (tzv. systemic risks), nielen na úrovni izolovane posudzovaných aplikácii AI?

Výsledné stanovisko CERAI: Nesúhlasíme
Odôvodnenie: Systém riadenia rizík je v AIA podľa nás zameraný na individuálne a izolované systémy AI a nezaoberá sa celým socio-technologickým ekosystémom. Čiastočne je táto téma pokrytá cez regulačné sandboxy. Avšak do regulačných sandboxov nebudú vstupovať všetky systémy, ale len ich malá vybranú časť. Väčšina systémov bude posudzovaná cez nástroje vyhodnocovania rizika, posúdenia zhody, či dopadu na zdravie, bezpečnosť a ľudské práva, avšak AIA mieri podľa nás prevažne na posúdenie pre daný konkrétny AI systém.

7. Súhlasíte s prístupom AIA, kde sa vopred definuje, ktoré oblasti a ktoré prípady použitia predstavujú vysoké riziko AI, pričom sa taxatívne tieto oblasti a prípady použitia vymenujú?

Výsledné stanovisko CERAI: Súhlasíme
Odôvodnenie: Dlhodobo to možno nebude postačujúci spôsob, ale v tejto fáze to považujeme za efektívny spôsob. Veríme, že bude flexibilný a bude sa práve táto kategorizácia vyvíjať - spresňovať a tiež vyjasňovať, vrátane rozširovania a zužovania zoznamu. Teda podľa nášho názoru môžeme siahnuť k vymenovaniu oblastí s vysokým rizikom, ale musi byt definovaný postup (proces) ako tento zoznam pravidelne aktualizovať, nie len podľa potreby, ale ako záväzok.

8. Pokrýva aktuálny zoznam uvedený v AIA všetky vysokorizikové oblasti a systémy (napr. nie je tam niečo navyše alebo vám tam naopak nejaká oblasť chýba)?

Výsledné stanovisko CERAI: Nesúhlasíme
Odôvodnenie: Tak ako je otázka postavená, ani nie je možné odpovedať súhlasne. Lebo aj naše názory na vysoko rizikové systémy sa vyvíjajú v čase. Je to dobrý začiatok, ale nepovažujeme dané oblasti vysokého rizika za vyčerpávajúce. Chýba nám tu medzi vysoko rizikovými systémami kategória zbrane, vojenské systémy riadené AI softvérom, aj keď AIA priznáva, že nemá ambíciu túto oblasť pokrývať. Rovnako nám chýba v prílohe III všeobecne doprava, aj ked cestná premávka sa počíta za kritickú infrastrukturu (podľa prílohe IIb), vrátane vnutroareálovej dopravy.

9. Je podľa vás možné posudzovať ex ante dopad AI systémov na oblasť zdravia a základných práv?

Výsledné stanovisko CERAI: Súhlasíme
Odôvodnenie: Súhlasíme, že je v princípe možné posudzovať dopad AI systémov na oblasť základných práv, čoho príkladom sú aj existujúce nástroje pre vyhodnocovanie dopadov na ľudské práva a hodnoty, ale aj postupy pre vyhodnocovanie dopadov na ľudské zdravie. Aj z principiálneho hľadiska, aj keby sa nepokryli alebo nedali odhadnúť všetky očakávané dopady, je dobré a dôležité sa o to pokúšať a vyšpecifikovať ich čo najpresnejšie.

10. Ak je možné posudzovať ex ante dopad AI systémov na oblasť zdravia a základných práv, je toto posudzovanie v AIA dostatočne pokryté?

Výsledné stanovisko CERAI: Nevieme
Odôvodnenie: Posúdiť sa dá, ale určite nie úplne. Mieru, ako je pokryté ex-ante v AIA, nevieme v tejto chvíli posúdiť. So súčasnými znalosťami je to dosť subjektívna záležitosť. Samozrejme výberom niektorých z existujúcich metodík, sa dá systém slušne vyhodnotiť, ale dopady uvidíme až neskôr. Aj v tejto oblasti by mohli pomôcť práve Testing and Experimentation Facilities, kde AIA načrtáva ešte nedotiahnutý a len rozpracovaný režim regulačných sandboxov. Ako zaujímavá možnosť sa javí aj vytvorenie registra AI systémov.

11. Bude možné podľa vás pre súčasné systémy AI zabezpečiť adekvátny „preklad“ právnych požiadaviek z AIA do ich dizajnu a vývoja (tzv. „alignment problem“; či v súvislosti s požiadavkami na vysvetliteľnosť učiacich sa systémov, a pod.)?

Výsledné stanovisko CERAI: Nesúhlasíme
Odôvodnenie: Takto všeobecne definované požiadavky sa nebudú dať vo svojej úplnosti splniť. Napr. v Čl. 10 sa uvádza "Súbory trénovacích, validačných a testovacích údajov musia byť relevantné, reprezentatívne, bezchybné a úplné." Toto nie je možné pre AI systémy, ktoré fungujú na základe obrovských objemov dát, z ktorých sa trénujú modely, dosiahnuť. Pre komplexné AI systémy sa musíme síce naďalej snažiť napr. o ich vysvetliteľnosť, ale aj tieto požiadavky musia byť realistické vzhľadom na súčasný stav poznania.

12. Bude podľa vás súčasný návrh AIA z hľadiska šírky jeho pôsobnosti dostatočný pre reguláciu AI (napr. navrhnutá úzka definícia (len profesionálneho) používateľa, prípadne rozsah aktivít a kompetencií verejných orgánov)?

Výsledné stanovisko CERAI: Nevieme
Odôvodnenie: V tomto stave nevieme posúdiť, či bude súčasný návrh AIA dostatočný pre reguláciu AI. Zároveň si myslíme, že niekde treba začať a teda v tomto smere má súčasný návrh potenciál postupne vytvoriť dobrú reguláciu, ďalšie zmeny a doplnenie zrejme vyplynú neskôr.

13. Mali by sa v AIA zahrnúť do analyzy rizik okrem (profesionálnych) používateľov aj ďalšie dotknuté skupiny (napr. neprofesionálni a bežní používatelia, zraniteľné skupiny)

Výsledné stanovisko CERAI: Súhlasíme
Odôvodnenie: Podľa nás by mali byť v AIA zahrnutí všetci používatelia. Tí profesionálni ale aj bežní používatelia bez dostatočného poznania v oblasti AI, ktorí prídu do styku s AI. Preto očakávame definovanie vplyvu AI aj na tieto skupiny používateľov. Okrem toho je otázne, kto a ako posúdi, či pôjde o osobnú neprofesionálnu alebo o osobnú profesionálnu činnosť. Do regulácie AI teda navrhujeme zahrnúť všetky spôsoby používania a všetkých používateľov.

14. Je odhad EK na dodatočné ľudské kapacity, ktoré s ohľadom na AIA budú potrebné vo verejnom sektore na vnútroštátnej úrovni (1-25 osôb) adekvátny?

Výsledné stanovisko CERAI: Nesúhlasíme
Odôvodnenie: Myslíme si, že toto je optimistický odhad, a pravdepodobne bude platiť len v skorých fázach zavádzania AIA. Očakávame teda podstatne vyššiu potrebu na ľudské kapacity, aj na základe porovnaní s personálom dohliadajúcim na data protection nariadenia.

15. Ste za prípadný vznik samostatného a nezávislého regulátora, resp. orgánu dohľadu pre AI na úrovni EÚ?

Výsledné stanovisko CERAI: Súhlasíme
Odôvodnenie: Myslíme si, že vzhľadom na náročnosť problematiky je vhodné koncentrovať odborníkov na európskej úrovni, nakoľko jednotlivým členským štátom sa nemusí podariť zabezpečiť dostatočnú odbornosť na národnej úrovni.

16. Budú harmonizované technické normy a spoločné špecifikácie (samé o sebe), ktoré spomína AIA, schopné dostatočne ošetriť (a minimalizovať) možné negatívne dopady predovšetkým na oblasti zdravia a základných práv?

Výsledné stanovisko CERAI: Nesúhlasíme
Odôvodnenie: Zastávame názor, že je potrebné, aby technické normy a spoločné špecifikácie začali vznikať a ich potrebnosť a užitočnosť je zrejmá. Na druhej strane nie sme presvedčení, že tieto nástroje budú samé o sebe schopné dostatočne ošetriť (a už vôbec nie minimalizovať) tieto negatívne dopady, predovšetkým na oblasti zdravia a základných práv.

17. Bude podľa vás v AIA navrhnutý spôsob presadzovania práva účinný a efektívny (napr. že bude v praxi realizovateľný s dosledovateľným pozitívnym dopadom a nebude sa napĺňať iba formálne)?

Výsledné stanovisko CERAI: Nevieme
Odôvodnenie: Prikláňame sa k názoru, že navrhnutý spôsob má potenciál byť efektívny, ale v tomto okamihu, aj vzhľadom k množstvu otvorených otázok, nevieme vyhodnotiť či sa tak skutočne aj stane.

18. Súhlasíte s odlišným uplatňovaním AIA na systémy AI podľa toho, kedy boli uvedené na trh alebo do prevádzky, resp. podľa toho, kedy došlo k významným zmenám ich koncepcie alebo zamýšľaného účelu?

Výsledné stanovisko CERAI: Nesúhlasíme
Odôvodnenie: Nemyslíme si, že je vhodné, aby pre rozhodnodnutie o tom, či daná aplikácia podlieha regulácii bol použitý dátum uvedenia do prevádzky. Sme skôr názoru, že po istej prechodnej dobe by všetky používané systémy mali podliehať rovnakej regulácií bez ohľadu na to, kedy boli uvedené do prevádzky. Z aktuálneho znenia navyše nie je jasné, či pretrénovanie AI modelu je signifikantná zmena dizajnu. Pritom je nespochybniteľné, že pretrénovanie môže radikálne meniť výstupy AI modelu. Taktiež sme identifikovali obavu, že takéto rozdelenie AI systémov môže podnietiť vznik paralelného "trhu starých aplikácií", ktoré nebudú musieť podliehať regulácii.

19. Súhlasíte s návrhom AIA pre experimentálne regulačné prostredia (regulačné sandboxy), podľa ktorého v nich musia AI systémy fungovať bez výnimiek alebo úľav z regulácie, teda vždy len v rámci existujúcich platných noriem?

Výsledné stanovisko CERAI: Súhlasíme
Odôvodnenie: Vzhľadom na účel sandboxov nám príde vhodné, aby AI systémy v sandboxoch fungovali v úplnej zhode s požiadavkami AIA. Avšak pri ich vývoji považujeme za vhodné umožniť aj výnimky pre čo najväčšie využitie potenciálu sandboxov.

20. Súhlasíte s názorom EK, že oblasť R&D bude mimo pôsobnosti AIA, teda že výskum a vývoj AI ešte pred uvedením systému na trh ostane nezasiahnutý touto reguláciou?

Výsledné stanovisko CERAI: Nesúhlasíme
Odôvodnenie: R&D sa nevykonáva sekvenčne, ale iteratívne a inkrementálne. Mnohokrát na existujúcich produktoch. Dokonca aj výskum sa často deje s reálnymi dátami a testuje sa na reálnych produktoch. A preto aj keď výskum a vývoj AI pred uvedením systému na trh ostane nezasiahnutý touto reguláciou, to neznamená, že R&D bude mimo pôsobnosti AIA.

21. Súhlasíte s tvrdením AIA, že pri poskytovaní dôveryhodného, zodpovedného a nediskriminačného prístupu k vysokokvalitným údajom na účely trénovania, validácie a testovania systémov umelej inteligencie budú mať zásadný význam európske spoločné dátové priestory zriadené Komisiou?

Výsledné stanovisko CERAI: Súhlasíme
Odôvodnenie: Súhlasíme, že spoločné dátové priestory zásadne zmenia, resp. uľahčia prístup k vysokokvalitným dátam. Ich zavedením podporíme ďalší rozvoj systemov AI, ich kompatibilitu, vzajomnu porovnatelnost, rovnako môžu byť nápomocné pri identifikácii a odhaľovaní skrytých rizik, skreslení v dátach a pod. Na druhej strane máme isté obavy o praktickom naplnení spoločného dátového priestoru, nakoľko veľká časť dát je dnes vlastnená súkromnými spoločnosťami.

22. Je podľa vás návrh AIA adekvátne previazaný s existujúcimi opatreniami v oblasti cybersecurity?

Výsledné stanovisko CERAI: Nevieme
Odôvodnenie: Nemáme pocit dostatočnej znalosti opatrení v oblasti cybersecurity, preto sa k danej otázke nevieme vyjadriť.