Stanovisko CERAI k vybraným otázkam MIRRI k návrhu AIA

Bratislava, 16.5.2022

Ministerstvo investícií, regionálneho rozvoja a informatizácie Slovenskej republiky (MIRRI) požiadalo Stálu komisiu pre etiku a reguláciu umelej inteligencie (CERAI) o vyjadrenie k AIA. Na 6. riadnom zasadnutí CERAI konanom dňa 26.4.2022 členovia CERAI diskutovali dve otázky z oblasti testovania AI systémov. Na základe materiálu predloženého vybranými členmi CERAI a následnej diskusie pripravila Stála komisia spoločné stanovisko k položeným otázkam. Toto stanovisko bolo následne schválené per rollam. Na základe výsledkov hlasovania sa Stála komisia rozhodla vydať nasledujúce stanovisko.

Otázka 1: Je možné a má zmysel testovať AI systémy v reálnych podmienkach („real-world conditions“) spôsobom, že musí byť vopred zadefinovaný ich zamýšľaný účel? Nestráca sa tým pridaná hodnota experimentovania a testov?

Stanovisko:

Komisia má za to, že každý systém má mať definovaný zamýšľaný účel, ktorý sa zohľadňuje pri testovaní AI systémov. Pre systémy so všeobecným účelom (“general-purpose”) sa za účel dá považovať jeho funkcionalita využiteľná v rôznych (často vopred nedefinovateľných) kontextoch AI systémov (napr. klasifikácia textu, obrazu).

Testovanie nenasadeného AI systému v “reálnych podmienkach” je pre veľké množstvo AI systémov prakticky nemožné, nakoľko sa dopredu nedajú získať testovacie dáta a zabezpečiť také prostredie, aby bolo možné vyhodnotiť všetky jeho vlastnosti. Existujú metódy, ktoré v rôznych oblastiach umožňujú testovanie blízke reálnym podmienkam, ale okrem toho, že to je oblasť stále aktívna aj vo výskume, výsledky sú rôzne pre rôzne domény. Podobne to platí aj pri testovaní testovacou skupinou používateľov.

Je bežné, že po uvedení AI systému do používania sa AI technológia (napr. modely nasadené v tomto AI systéme) ďalej vylepšuje, napr. na základe nových dát z používania.

Pri používaní AI systému sa môže vyskytnúť/objaviť nový účel tohto AI systému, resp. jeho častí. Takáto zmena účelu by sa mala posudzovať samostatne.

Otázka 2: Ak bude celkom vyňatý „development“ z pôsobnosti AIA, bude možné dostatočne priebežne kontrolovať systémy strojového učenia? Ak nie, ako je systémovo/legislatívne možné zabezpečiť dostatočnú kontrolu ich priebežného vývoja?

Poznamka: Ide o “priebežný vývoj” - je myslený ako dotrénovanie, ladenie systemu po jeho spustení.

Stanovisko:

AI systémy (podobne ako aj iné softvérovo intenzívne systémy) sa po ich nasadení spravidla ďalej vyvíjajú. Je to spôsobené/zapríčinené najmä potrebou prispôsobovania sa reálnemu a meniacemu sa prostrediu (vo všetkých jeho aspektoch, vrátane ekonomického) a potrebám užívateľa (pri zachovaní účelu).

Zároveň pre AI systémy musíme zobrať do úvahy ich nedeterministickú povahu a v súčasných architektúrach závislosť od dát, na ktorých sa trénujú modely. Datasety na trénovanie modelov môžu zahŕňať dáta vytvorené inými systémami alebo priamo človekom (vrátane značkovania dát) a môžu sa kontinuálne vylepšovať (samotná kvalita dát, biasy, ale aj rozsah). Rovnako to platí aj pre nastavovanie parametrov modelov pre ich trénovanie alebo dotrénovanie. Kontinuálne vylepšovanie charakteristík AI systému, resp. modelov, s ktorými pracuje, je základnou vlastnosťou AI systémov.

Vývoj AI systémov, ktoré sú určené na uvedenie na trh, je nutné zabezpečiť nastavením procesov, ktorými sa zabezpečí tak transparentnosť dát, ako aj algoritmov, a priebežným vyhodnocovaním jednotlivých aspektov, ktoré primerane dokumentujú jednotlivé kroky a samotné dáta spolu so zabezpečením transparentnosti a posudzovaním po uplynutí stanoveného času alebo pri rozsahu zmien spôsobených vývojom, ktoré prekročia stanovené hranice (ak to je možné definovať, mal by sa brať do úvahy zamýšľaný účel a jeho zmeny).