Evropská komise zveřejnila pokyny pro etickou umělou inteligenci

Evropská komise zahájila pilotní projekt, jehož cílem je ověřit, zda pokyny komise navrhované pro rozvoj a využívání umělé inteligence (AI) budou moci být implementovány v praxi, a usilovat o dosažení mezinárodního konsenzu v oblasti komunikace mezi AI a člověkem.

Tento pilotní projekt navazuje na práce skupiny špičkových nezávislých odborníků, kteří v prosinci loňského roku zveřejnili první návrh pokynů týkajících se důvěryhodné AI. Zmíněné pokyny byly následně konzultovány s průmyslovými organizacemi, výzkumnými ústavy a orgány veřejné správy, z čehož vzešlo více než 500 různých připomínek. Podle prvního návrhu pokynů by důvěryhodná AI měla být:

– zákonná – dodržovat všechny platné zákony a předpisy,
– etická – respektovat etické principy a hodnoty,
– robustní – a to z technického hlediska i s ohledem na sociální prostředí.

Cílem Evropské komise je usnadnit a zlepšit spolupráci v oblasti AI v celé EU, posílit konkurenceschopnost a zajistit důvěru založenou na hodnotách EU. Komise proto na základě strategie AI, zveřejněné 18. dubna 2018, zřídila odbornou skupinu, zahrnující akademickou sféru, průmyslové organizace i státní správu.

„Etická AI je důležitá pro všechny, protože může Evropě poskytnout konkurenční výhodu – být lídrem v oblasti komunikace mezi AI a lidmi," uvedl Andrus Ansip, místopředseda Evropské komise a komisař pro digitální trh. „Na etickou AI nelze pohlížet jako na luxus nebo doplněk. Aby bylo možné plně využívat digitální technologie, musí celá společnost těmto technologiím důvěřovat."

Komise stanovila tři stupně pro zajištění důvěryhodné AI, a to stanovení klíčových požadavků na důvěryhodnou AI, rozsáhlý pilotní projekt pro zajištění zpětné vazby od zúčastněných stran a dosažení mezinárodního konsenzu v oblasti komunikace mezi AI a člověkem.

Důvěryhodná AI by měla respektovat všechny platné zákony a předpisy, jakož i mnoho dalších požadavků a specifických ustanovení s cílem pomoci ověřit aplikaci každého ze sedmi rozhodujících požadavků:

Lidská činnost a dohled: Systémy AI by měly podporovat nezávislost člověka, schopnost činit informovaná rozhodnutí a respektovat jeho základní práva, neměly by ho nijak omezovat nebo jím nevhodně manipulovat. Současně musí být zajištěny řádné mechanismy dohledu a schopnost lidského zásahu, ať již v každém rozhodovacím cyklu systému (*human-in-the-loop*), během návrhu systému a monitorování provozu systému (*human-on-the-loop*), nebo při dohledu na celkovou činnost systému AI. Uživatel musí být schopen rozhodnout, kdy a jak systém používat v jakékoliv konkrétní situaci (*human-in-command*).

Robustnost a bezpečnost: Pro zajištění důvěryhodné AI je třeba, aby algoritmy byly bezpečné, spolehlivé a dostatečně robustní a byly schopny řešit chyby nebo nesrovnalosti ve všech fázích životního cyklu systémů AI. Znamená to, že výstupy musí být přesné, spolehlivé a reprodukovatelné a systémy AI musí být odolné proti různým zranitelnostem a kybernetickým útokům a musí mít připravený nouzový plán pro případ, že se něco pokazí. To je jediný způsob, jak předejít poškození nebo alespoň tuto možnost minimalizovat.

Ochrana osobních údajů a správa dat: Je třeba zajistit plné respektování ochrany soukromí a osobních dat a také odpovídající mechanismy pro správu dat zohledňující kvalitu a integritu dat a zajištění legitimního přístupu k datům. Občané by měli mít plnou kontrolu nad svými vlastními osobními daty, přičemž data, které se jich týkají, nesmí být zneužita k jejich poškozování nebo diskriminaci.

Transparentnost: Data, systém i obchodní modely AI by měly být transparentní. K dosažení tohoto cíle mohou napomoci mechanismy pro vysledovatelnost, tj. schopnost ověřit historii, umístění nebo využití pomocí zdokumentované zaznamenané identifikace. Navíc by měla být rozhodnutí AI vysvětlena způsobem, který je přizpůsoben zúčastněným stranám. Lidé si musí být vědomi, že komunikují se systémem AI, a musí být informováni o možnostech a omezeních systému.

Rozmanitost, nediskriminace a spravedlnost: Systémy AI by měly brát v úvahu celou škálu lidských schopností, dovedností a požadavků a měly by být přístupné všem bez ohledu na jakékoliv postižení. Je třeba se vyhnout zaujatosti a předpojatosti, protože by mohly mít mnoho negativních důsledků od opomíjení menšinových a zranitelných skupin až po vytváření předsudků a diskriminaci. Podpora rozmanitosti znamená, že systémy AI by měly být přístupné všem bez ohledu na jakékoliv postižení či odlišnost, a to během celého životního cyklu.

Společenský dopad a vliv na životní prostředí: Systémy AI by měly být přínosem pro všechny lidské bytosti, včetně budoucích generací. Kromě toho by měly brát v úvahu životní prostředí, včetně jiných živých bytostí, a jejich sociální a společenský dopad by měl být pečlivě zvážen. Proto musí být zajištěno, aby byly využívány k posílení pozitivních sociálních změn a byly udržitelné a šetrné k životnímu prostředí.

Odpovědnost: Měly by být zavedeny mechanismy, které zajistí odpovědnost za systémy AI a jejich rozhodnutí a výstupy, a to jak před jejich vývojem, implementací a používáním, tak i po něm. V aplikacích ovlivňujících základní práva, včetně aplikací kritických z hlediska bezpečnosti, by systémy AI měly být schopné nezávislého auditu, tj. hodnocení algoritmů, dat a návrhových procesů. Vznikne-li nespravedlivý nebo nepříznivý dopad, měly by být dostupné mechanismy, které zajistí odpovídající nápravu. Zvláštní pozornost by měla být věnována zranitelným osobám nebo skupinám obyvatelstva.

Evropská komise chce dosáhnout mezinárodního konsenzu v oblasti komunikace mezi AI a člověkem, protože technologie, data a algoritmy neznají hranice. Proto chce rozšířit spolupráci s podobně smýšlejícími partnery, jako je Japonsko, Kanada nebo Singapur, a chce i nadále hrát aktivní úlohu v mezinárodních diskusích a iniciativách, včetně G7 a G20. Součástí pilotního projektu budou také mezinárodní organizace a společnosti z mimoevropských zemí.

Do podzimu plánuje komise zřídit několik center pro výzkum AI, zřídit síť center pro digitální inovace a společně s členskými státy a zúčastněnými stranami zahájit diskusi o vývoji a zavádění modelu pro sdílení dat a co nejlepšího využití společných datových prostorů.

Společnosti, veřejné správy i další organizace se mohou přihlásit k Evropské alianci AI [2] a obdržet oznámení o zahájení pilotního programu.

V návaznosti na pilotní program a na základě obdržené zpětné vazby skupina odborníků pro AI přezkoumá a vyhodnotí klíčové požadavky na důvěryhodnou AI, což je naplánováno na počátek roku 2020. Na základě tohoto přezkumu komise vyhodnotí výsledky a navrhne další kroky.

Literatura:

[1] European Commission, High-Level Expert Group on AI. *Ethics Guidelines for Trustworthy AI* [online]. Duben 2019, 39 stran [cit. 2019-06-17]. Dostupné z: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477

[2] *Digital Single Market, Policy: The European AI Alliance* [online]. 2019 [cit. 2019-06-17]. Dostupné z: https://ec.europa.eu/digital-single-market/en/european-ai-alliance

Jaroslav Hrstka