

Expertná skupina

EURÓPSKEJ KOMISIE NA VYSOKEJ ÚROVNI PRE UMELÚ INTELIGENCIU



NÁVRH ETICKÉ USMERNENIA PRE DÔVERYHODNÚ UMELÚ INTELIGENCIU ZHRNUTIE

Pracovný dokument pre konzultácie so zainteresovanými stranami

V Bruseli 18. decembra 2018

NÁVRH ETICKÝCH USMERNENÍ

PRE DÔVERYHODNÚ UMELÚ INTELIGENCIU



Expertná skupina na vysokej úrovni pre umelú inteligenciu
Návrh etických usmernení pre dôveryhodnú umelú inteligenciu

Európska komisia
Generálne riaditeľstvo pre komunikáciu

Kontakt Nathalie Smuha – koordinátorka AI HLEG
E-mail: CNECT-HLG-AI@ec.europa.eu

European Commission
B-1049 Brussels

Dokument zverejnený 18. decembra 2018. v angličtine.

Tento pracovný dokument vypracovala expertná skupina na vysokej úrovni pre umelú inteligenciu bez toho, aby boli dotknuté jednotlivé pozície jej členov ku konkrétnym bodom, a bez toho, aby bolo dotknuté konečné znenie dokumentu. Tento dokument sa ešte bude ďalej upravovať a jeho konečné znenie bude predložené v marci 2019 po konzultácii so zainteresovanými stranami prostredníctvom Európskej aliancie pre umelú inteligenciu.

Európska komisia ani žiadna osoba konajúca v mene Komisie nie je zodpovedná za možné použitie uvedených informácií. Obsah tohto pracovného dokumentu je výlučnou zodpovednosťou expertnej skupiny na vysokej úrovni pre umelú inteligenciu (ďalej len „AI HLEG“). Hoci sa na vypracovaní usmernení podieľali zamestnanci útvarov Komisie, názory vyjadrené v tomto dokumente odzrkadľujú stanovisko AI HLEG a za žiadnych okolností ich nemožno považovať za vyjadrenie oficiálneho stanoviska Európskej komisie. Toto je návrh prvého výstupu AI HLEG. Jeho konečné znenie bude predložené Komisii v marci 2019. Konečné znenie druhého výstupu – Politika v oblasti umelej inteligencie a investičné odporúčania – bude predložené v polovici roka 2019.

Viac informácií o expertnej skupine na vysokej úrovni pre umelú inteligenciu je k dispozícii online (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>). Pravidlá opakovaného použitia dokumentov Európskej komisie sú upravené rozhodnutím 2011/833/EÚ (Ú. v. EÚ L 330, 14.12.2011, s. 39). Na akékoľvek použitie alebo reprodukciu fotografií alebo iného

materiálu, na ktorý sa nevzťahujú autorské práva EÚ, je potrebné povolenie priamo od držiteľov príslušných autorských práv.

ZHRNUTIE

Tento pracovný dokument je návrhom etických usmernení k umelej inteligencii, ktorý vypracovala expertná skupina na vysokej úrovni pre umelú inteligenciu (AI HLEG) zriadená Európskou komisiou, a ktorého konečné znenie bude predložené v marci 2019.

Umelá inteligencia patrí medzi zásadné transformačné sily súčasnosti, ktorá určite zmení štruktúru spoločnosti. Predstavuje výbornú príležitosť na zvýšenie prosperity a rastu, a Európa ju musí využiť. Za posledné desaťročie sa vďaka dostupnosti obrovského množstva digitálnych údajov, výkonných výpočtových architektúr, ako aj ďalšiemu vývoju technológií umelej inteligencie (ako je napr. strojové učenie) dosiahli v tejto oblasti významné pokroky. Významný vývoj v oblasti autonómnych vozidiel, zdravotnej starostlivosti, domácich robotov/ robotov zabezpečujúcich služby, vzdelávania alebo kybernetickej bezpečnosti, ktorý umožnila umelá inteligencia, zvyšuje kvalitu nášho každodenného života. Umelá inteligencia je okrem toho kľúčom k riešeniu mnohých veľkých výziev, ktorým svet čelí, ako sú globálne zdravie a zabezpečenie dobrých životných podmienok svetovej populácie, zmena klímy, spoľahlivé právne a demokratické systémy a iné výzvy vyjadrené v cieľoch Organizácie spojených národov v oblasti udržateľného rozvoja.

Umelá inteligencia má potenciál priniesť obrovské výhody jednotlivcom a spoločnosti, zároveň sú s ňou spojené určité riziká, ktoré by sa mali správne riadiť. Vzhľadom na to, že výhody umelej inteligencie celkovo prevažujú nad jej rizikami, musíme zabezpečiť, aby sme sa postupovali tak, že budeme **maximalizovať jej prínosy a minimalizovať jej riziká**. Preto je potrebné **vzťahu k umelej inteligencii zaujať prístup zameraný na človeka**, ktorý nás bude nútiť nezabúdať na to, že rozvoj a používanie umelej inteligencie by sme nemali vnímať ako prostriedok sám osebe, ale ako niečo, čoho cieľom je zvýšiť blahobyť človeka. **Základom bude dôveryhodnosť umelej inteligencie**, keďže ľudia budú schopní s istotou a plne využívať prínosy umelej inteligencie, len vtedy, ak budú môcť tejto technológii dôverovať.

Dôveryhodná umelá inteligencia má **dve zložky**: 1. mala by rešpektovať základné práva, platné právne predpisy a základné zásady a hodnoty, čím sa zabezpečí „**etický účel**“, a 2. mala by byť **technicky odolná** a spoľahlivá, pretože technologické nedostatky môžu aj pri dobrých úmysloch spôsobiť neúmyselné ublíženie.

V týchto usmerneniach sa preto stanovuje **rámec pre dôveryhodnú umelú inteligenciu**:

- **Kapitola I** je zameraná na **zabezpečenie etického účelu umelej inteligencie** stanovením základných práv, zásad a hodnôt, ktoré by mala spĺňať.
- Z týchto zásad vychádza **kapitola II** ktorá sa zameriava na **usmernenia k realizácii** dôveryhodnej umelej inteligencie, ktoré sa týkajú etického účelu aj technickej odolnosti. Obsahuje zoznam požiadaviek, ktoré by dôveryhodná umelá inteligencia mala spĺňať a poskytuje prehľad technických a netechnických metód, ktoré možno použiť na jej realizáciu.
- **Kapitolou III** sa následne **uvádzajú požiadavky do praxe** a poskytuje sa konkrétny, i keď neúplný zoznam bodov na posudzovanie dôveryhodnej umelej inteligencie. Tento zoznam sa potom prispôsobuje konkrétnym prípadom použitia.

Na rozdiel od iných dokumentov, ktoré sa zaoberajú etickou umelou inteligenciou, teda cieľom usmernení nie je poskytnúť len ďalší zoznam základných hodnôt a zásad pre umelú inteligenciu, ale skôr ponúknuť usmernenia na ich konkrétne vykonávanie a zavádzanie týchto princípov do systémov umelej inteligencie. Usmernenia sú rozdelené do troch úrovní podľa abstrakcie, od najabstraktnejšej v kapitole I (základné práva, zásady a hodnoty) až po najkonkrétnejšiu v kapitole III (zoznam na účely posudzovania).

Usmernenia sú určené všetkým **relevantným zainteresovaným stranám, ktoré vyvíjajú, zavádzajú alebo používajú umelú inteligenciu**, medzi ktoré patria spoločnosti, organizácie, výskumníci, verejné orgány, inštitúcie, jednotlivci alebo ďalšie subjekty. V konečnom znení týchto usmernení sa navrhne

mechanizmus, ktorým sa umožní zainteresovaným stranám, aby ich dobrovoľne schválili.

Treba zdôrazniť, že tieto usmernenia nie sú myslené ako náhrada za príslušné právne predpisy či akékoľvek politické dokumenty (to sa bude riešiť v druhom výstupe AI HLEG: Politika umelej inteligencie a investičné odporúčania, ktorý bude predložený v máji 2019), a ich cieľom nie je ani odradiť od ich zavedenia. Tieto usmernenia by sa navyše mali vnímať ako živý dokument, ktorý sa musí v priebehu času pravidelne aktualizovať, aby sa zabezpečila jeho neustála relevantnosť, keďže technológia a naše poznatky o nej sa vyvíjajú. Tento dokument by preto mal byť východiskovým bodom pre diskusiu o „**dôveryhodnej umelej inteligencii vytvorenej v Európe**“.

Európa môže šíriť svoj etický prístup k umelej inteligencii iba vtedy, keď bude konkurencieschopná na globálnej úrovni. **Avšak etický prístup k umelej inteligencii je potrebný práve pre zabezpečenie zodpovednej konkurencieschopnosti**, pretože vybuduje dôveru používateľov a podporí širšie využívanie umelej inteligencie. Cieľom týchto usmernení nie je potlačiť inovácie v oblasti umelej inteligencie v Európe, ale skôr používať etiku ako inšpiráciu na vytvorenie jedinečného druhu umelej inteligencie, a to takého, ktorý sa zameriava na ochranu a prospech jednotlivcov a spoločné blaho. To umožňuje Európe stať sa lídrom v špičkovej, bezpečnej a etickej umelej inteligencii. Európski občania budú plne využívať výhody umelej inteligencie, iba ak sa zabezpečí jej dôveryhodnosť.

Za hranicami Európy sú tieto usmernenia zamerané aj na **podporu úvah a diskusií** o etickom rámci pre umelú inteligenciu na **globálnej úrovni**.

VÝKONNÉ USMERNENIA

Každá kapitola týchto usmernení ponúka usmernenia ako dosiahnuť dôveryhodnosť umelej inteligencie, ktoré sú určené všetkým relevantným zainteresovaným stranám, ktoré vyvíjajú, zavádzajú alebo používajú umelú inteligenciu, a sú zhrnuté takto:

Kapitola I: Kľúčové usmernenia na zabezpečenie etického účelu:

- Zabezpečiť, aby bola umelá inteligencia **zameraná na človeka**: umelá inteligencia by sa mala vyvíjať, zavádzať a používať s „**eticným účelom**“, ktorý je zakotvený v základných právach, a zohľadňovať spoločenské hodnoty a etické zásady *zabezpečovania najlepšieho záujmu* (robiť dobro), *neškodnosti* (nespôsobovať škodu), *autonómnosti ľudí*, *spravodlivosti* a *vysvetliteľnosti*. Je to základným predpokladom zabezpečenia **dôveryhodnosti umelej inteligencie**.
- Opierať sa pri hodnotení budúcich účinkov umelej inteligencie na človeka a spoločné blaho o základné práva, etické zásady a hodnoty. Venovať **osobitnú pozornosť** situáciám, ktoré sa týkajú **zraniteľnejších skupín**, ako sú deti, osoby so zdravotným postihnutím alebo menšiny, alebo situáciám, ktoré sa vyznačujú asymetriou moci alebo dostupnosti informácií, napríklad vo vzťahoch medzi zamestnávateľmi a zamestnancami alebo podnikmi a spotrebiteľmi.
- Uznávať a zohľadňovať skutočnosť, že hoci umelá inteligencia prináša jednotlivcom a spoločnosti zásadné výhody, môže mať aj negatívny vplyv. Zachovávať obozretnosť v kritických oblastiach záujmu.

Kapitola II: Kľúčové usmernenia pre realizáciu dôveryhodnej umelej inteligencie:

- Začleniť **požiadavky na dôveryhodnú umelú inteligenciu do vývoja týchto technológií od úplného začiatku**: zodpovednosť, správa údajov, riešenia zohľadňujúce potreby všetkých, riadenie autonómie umelej inteligencie (ľudský dohľad), nediskriminácia, rešpektovanie ľudskej autonómie, rešpektovanie súkromia, spoľahlivosť, bezpečnosť, transparentnosť.
- Zvážiť technické a netechnické metódy na zabezpečenie vykonávania týchto požiadaviek v systéme umelej inteligencie. Okrem toho zohľadňovať tieto požiadavky pri budovaní tímu, ktorý má pracovať na systéme, pri vytváraní samotného systému, jeho testovacieho prostredia a možnostiach využitia.

- Jasným a proaktívnym spôsobom poskytovať **zainteresovaným stranám** (zákazníci, zamestnanci atď.) **informácie** o možnostiach a obmedzeniach systému umelej inteligencie, čo im umožní vytvoriť si realistické očakávania. Zabezpečenie **vysledovateľnosti** systému umelej inteligencie je v tomto ohľade kľúčové.
- Urobiť dôveryhodnú umelú inteligenciu **súčasťou kultúry organizácie** a poskytnúť zainteresovaným stranám informácie o tom, ako je dôveryhodná umelá inteligencia začlenená do vývoja a používania systémov umelej inteligencie. Dôveryhodná umelá inteligencia môžu organizácie aj začleniť do svojich etických zásad alebo kódexov správania.
- Zabezpečiť účasť a **začlenenie zainteresovaných strán** do navrhovania a vývoja systému umelej inteligencie. Okrem toho vytvárať pracovné tímy, ktoré zabezpečujú vývoj, realizáciu a testovanie produktu tak, boli čo **najrozmanitejšie** .
- Snažiť sa **uľahčiť kontrolovateľnosť** systémov umelej inteligencie, najmä v kritických kontextoch alebo situáciách. V čo najväčšej miere navrhnúť systém tak, aby sa umožnilo vysledovanie jednotlivých rozhodnutí k rozličným vstupom, údajom, nacvičeným modelom atď. Okrem toho vymedziť **metódy vysvetľovania** systému umelej inteligencie.
- Zabezpečiť osobitný proces pre **riadenie zodpovednosti**.
- Naplánovať **odbornú prípravu a vzdelávanie** a zabezpečiť, aby vedúci pracovníci, vývojári, používatelia a zamestnávateľia boli informovaní a vyškolení v oblasti dôveryhodnej umelej inteligencie.
- Treba mať na pamäti, že medzi rôznymi cieľmi môže dôjsť k zásadným konfliktom (transparentnosť môže viesť k zneužitiu, identifikácia a korekcia zaujatosti môže byť v rozpore s ochranou súkromia). O týchto kompromisoch treba informovať a evidovať ich.
- Podporovať výskum a inovácie s cieľom pomôcť pri plnení požiadaviek na dôveryhodnú umelú inteligenciu.

Kapitola III: Kľúčové usmernenia pre posudzovanie dôveryhodnej umelej inteligencie

- Prijaté **zoznam na účely posudzovania** dôveryhodnej umelej inteligencie, ktorý sa využije pri jej vývoji, zavádzaní alebo používaní, a prispôbiť ho konkrétnemu prípadu použitia, v ktorom sa systém používa.
- Mať na pamäti, že zoznam na účely posudzovania **nebude nikdy úplný** a že pri zabezpečení dôveryhodnej umelej inteligencie nejde o zaškrtavanie políčok, ale o nepretržitý proces identifikácie požiadaviek, vyhodnocovania riešení a zabezpečovania lepších výsledkov počas celého životného cyklu systému umelej inteligencie.

Tieto usmernenia sú súčasťou takéhoto prístupu k umelej inteligencii, ktorý bude zameraný na človeka a ktorý umožní Európe stať sa globálnym vedúcim inovátorom v oblasti etickej, bezpečnej a špičkovej umelej inteligencie. Cieľom je uľahčiť a umožniť vytvorenie značky „**dôveryhodná umelá inteligencia vytvorená v Európe**“, ktorá pomôže európskym občanom zlepšiť životné podmienky.