

NADĚJE
VNITŘNÍ PŘEDPIS Č. 315
ZE DNE 18. 3. 2024



ETICKÝ KODEX A PRAVIDLA PRO VYUŽITÍ NÁSTROJŮ UMĚLÉ INTELIGENCE

I. OBECNĚ

1. Tento vnitřní předpis stanovuje etické principy a pravidla pro využití nástrojů umělé inteligence (dále jen AI) v NADĚJI.
2. Zodpovědnost za využití nástrojů AI a jejich produktů má člověk, konkrétní zaměstnanci a jejich nadřízení.
3. Nástroje AI využíváme transparentně. Jejich použití musí být transparentní vůči příjemcům pomoci, donorům, partnerům, komunikačním protějškům a dalším zúčastněným stranám. Nikdy například zpětně neupravujeme existující fotografie, videa či audia nebo je negenerujeme tak, aby u nich mohlo dojít k záměně s médiem fotografie, videozáznamu nebo záznamu reálného zvuku, krom výjimečných případů, viz níže.
4. Využití nástrojů AI je odůvodněné zkvalitněním naší práce. Nechceme generovat vyšší množství obsahů na úkor kvality.
5. Jejich využití musíme umět vysvětlit, abychom zajistili kontrolu nad vlastním rozhodováním a uměli interpretovat jejich výsledky.
6. Nástroje AI využíváme v souladu s Všeobecnou deklarací lidských práv a svobod, posláním NADĚJE, hodnotami NAĚJE a etickým kodexem NADĚJE. S ohledem na možné zesílení předsudků a diskriminačních zkreslení v použitých modelech a nástrojích AI dbáme na to, abychom vyloučili diskriminaci skupin a jednotlivců na základě rasy, původu, pohlaví, náboženství, politických názorů apod.
7. Dbáme na zajištění rovnosti a psychické integrity, aby při využití nástrojů AI nedocházelo k zneužití lidských a psychických slabin jednotlivců či skupin a jiné informační a mocenské nerovnováze vůči zúčastněným stranám, abychom zachovali rovnost a lidskou důstojnost.
8. Zohledňujeme dopad na společnost a životní prostředí. Zkoumáme a vyhodnocujeme, zda nástroje, prostředky a procesy, které využíváme, nemají širší negativní dopad na společnost a životní prostředí.
9. Usilujeme o to, aby přístup k AI, informace o přínosech i rizicích AI a know-how ve využití jejich nástrojů bylo dostupné nejzranitelnějším skupinám, aby z nich měla prospěch celá společnost.
10. Nástroje AI využíváme pro rozšíření přístupnosti našich služeb a informací lidem se sluchovým, zrakovým, mentálním a jiným postižením.
11. Chráníme osobní údaje a další důvěrné informace, aby používání nástrojů AI bylo v souladu s GDPR a příslušnými zákony a předpisy o ochraně údajů a bezpečnosti a nemohlo být zneužité třetí stranou.

II. PRAKTICKÁ PRAVIDLA PRO VYUŽITÍ NÁSTROJŮ AI

PRÁCE S JAZYKOVÝMI MODELY (LLM)

12. Veřejné nástroje jako ChatGPT, Bing, Bard atd. můžeme využít pro inspiraci, návrhy textů či úpravu stylistiky. Dají se rovněž využít na překlady, sumarizace, hledání synonym a odborných termínů, úpravu tónu, zkracování, návrhy titulků a další práci s obsahem. Některé specializované LLM je možné využít k řešení matematických a logických úloh, při tvorbě, úpravě a doplňování tabulek, při kategorizaci, zjišťování excelových funkcí a podobně.
13. Na tyto nástroje nespolehneme při zjišťování či ověřování faktických informací. Všechny texty a informace generované AI zpětně kontrolujeme.
14. Při využití veřejně dostupných jazykových modelů dbáme na ochranu osobních údajů a citlivých interních informací. Nikdy do nich nezadáme osobní údaje a jakékoli další interní informace, jako jsou vnitřní a metodické předpisy, metodiky, písemně stanovené pracovní postupy apod.

15. Negenerujeme a nepoužíváme texty vytvořené AI mechanicky a bez důvodu. Nechceme zvyšovat produkci na úkor kvality nebo nahradit osobitost průměrem. Nástroje AI mají napomáhat tomu, aby naše texty byly ve výsledku lepší, než kdybychom je psali sami.
16. Za texty vytvářené s pomocí AI ručí vždy příslušný zaměstnanec a podepisuje je vlastním jménem.
17. Speciálně dbáme na bezpečnost dětí v našich službách. Podle [pravidel ChatGPT](#) její aplikaci nemají samostatně používat děti do 13 let, ani kdyby měly souhlas zákonného zástupce. Děti od 13 let mohou ChatGPT se souhlasem zákonného zástupce využívat, i tady se ale doporučuje dohled.

VYUŽITÍ GENERÁTORŮ OBRAZŮ, VIDEA A AUDIA

18. Nikdy negenerujeme obrazy, videa či audia tak, aby mohly být chápány jako dokumentace naší činnosti a vytvářely falešný dojem o reálných událostech nebo o naší práci. Audiovizuální materiál v našich výstupech není pouhou ilustrací aktivit, které děláme, ale dodává jim věrohodnost, kterou nesmíme ohrozit.
19. Nikdy zpětně neupravujeme existující fotografie, videa či audia s využitím nástrojů AI, vyjma technických a stylistických úprav bez zásahu do obsahu nebo v odůvodněných případech pro anonymizaci osob či míst kvůli zajištění bezpečnosti. V takovém případě transparentně informaci o jejich využití uvádíme spolu s obrazem, videem či audiem.
20. Nikdy negenerujeme obrazy, videa či audia, u kterých by mohlo dojít k záměně s médii fotografie, videozáznamu nebo záznamu reálného zvuku. Přípustnou výjimkou jsou situace, kdy pro vzdělávací účely chceme ilustrovat schopnosti nástrojů AI – v takovém případě transparentně uvádíme informaci o jejich využití.
21. Generátory obrazů, videí či zvuku můžeme využít pro vytváření kreseb a ilustrací, zvláště v tvorbě napodobujících ruční či digitální umělecké styly a v grafickém designu. V takovém případě zjišťujeme, že jsme oprávněni je daným způsobem použít a transparentně informaci o jejich využití uvádíme spolu se jménem autora.

VYUŽITÍ DALŠÍCH NÁSTROJŮ AI

22. Při dodržení tohoto vnitřního předpisu a podmínek výše můžeme využívat mnoho dalších nástrojů AI pro zkvalitnění a zefektivnění naší práce. Nabízí se například nástroje pro tvorbu prezentací, přepisovače mluveného slova na text, pomocné nástroje pro evaluaci, automatizaci, programování, práci s databázemi, personalizaci nebo boj s dezinformacemi. Ve všech případech využití nástrojů AI zvažujeme účelnost a rizika v souladu s hodnotami a etickým kodexem NADĚJE a celkovým přínosem.

OCHRANA DAT

23. Do veřejně přístupných AI nástrojů zadáváme jen takový obsah, který bychom se nebáli zveřejnit kdekoli na internetu, tedy u kterých nehrozí únik citlivých informací. Platí to i pro aplikace nebo služby, ke kterým se připojujeme svým soukromým účtem.

III. ZÁVĚREČNÁ USTANOVENÍ

24. Oblastní ředitelé ve lhůtě do 30 dnů od nabytí účinnosti tohoto vnitřního předpisu seznámí s jeho obsahem podřízené pracovníky.
25. Oblastní ředitelé nesou odpovědnost za dodržování ustanovení tohoto vnitřního předpisu v jimi řízených oblastech a mohou oblastním vnitřním předpisem stanovit nebo vymezit podrobnější pravidla, pověřené osoby pro práci s AI apod.
26. Tento vnitřní předpis nabývá účinnosti dnem 1. dubna 2024.

Mgr. Jan Vaněček
Ústřední ředitel