Dělnická 213/12
Praha 7 — Holešovice
170 00

246 030 983
info@operatorict.cz
www.oict.cz

Generativní umělá inteligence – principy a zásady práce s generativní AI v oict.

3 hlavní zásady pro zkoumání generativní umělé inteligence

- Před použitím obsahu vytvořeného umělou inteligencí jej vždy zkontrolujme a ověřme fakta.
- Vždy uveďme, že jsme při vytváření výstupu použili generativní umělou inteligenci.
- Nikdy nezadávejme citlivé informace do veřejných nástrojů generativní AI, jako je ChatGPT. K informacím, které zadáme, se totiž mohou dostat společnosti, které tyto nástroje vytvářejí, a v některých případech i široká veřejnost.

Úvod

Umělá inteligence (AI) má při zodpovědném používání velký potenciál přinést prospěch veřejnosti. Generativní AI vytváří nová data na základě vzorců naučených z existujících dat a může generovat obsah, který napodobuje lidskou kreativitu. Příklady zahrnují například generování textu, tvorbu obrázků či skládání hudby. Generativní umělá inteligence se liší od technologie umělé inteligence, která podporuje informovaná rozhodnutí na základě vstupních dat, ale nevytváří nový obsah. Generativní AI nabízí nové příležitosti a zároveň představuje značnou výzvu pro zajištění odpovědného a efektivního používání.

Definice

Co je generativní umělá inteligence?

- Generativní umělá inteligence označuje nové softwarové nástroje, které mohou na základě podnětu zadaného uživatelem vytvářet realistický text, obrázky, zvuk, video a další média. Mezi běžné aplikace generativní AI patří ChatGPT, Gemini a Dall-E. Tyto nástroje používají algoritmy strojového učení, které byly vyškoleny na velmi rozsáhlých

Společnost Operátor ICT, a. s., je zapsaná v obchodním rejstříku pod sp. zn. B 19676 vedenou u Městského soudu v Praze.

Tento dokument je považován za vlastnictví společnosti Operátor ICT, a. s., a může být užíván výhradně způsobem stanoveným v tomto dokumentu.

souborech textových a obrazových dat získaných z internetu. Tyto modely získaly z tréninkových dat běžné jazykové a obrazové vzory a mohou tak rychle a realisticky reagovat na zadání (prompty).

- Generativní aplikace umělé inteligence jsou vytvářeny na základě tréninkových datových sad z různých zdrojů na internetu a často obsahují genderové, rasové, politické a jiné předsudky. V důsledku toho mohou výstupy generativní AI šířit předsudky. Navíc i ty nejpokročilejší současné nástroje generativní AI mohou poskytovat nekonzistentní odpovědi na otázky založené na faktech. Uživatelé by tak měli vždy kontrolovat obsah generovaný AI z hlediska přesnosti a také z hlediska předsudků, které mohou vykazovat.

A co tradiční AI?

Generativní AI se liší od diskriminačních modelů strojového učení, které jsou široce používány od počátku roku 2000. Diskriminační modely strojového učení nevytváří nový obsah. Omezují se na generování známých a ověřených hodnot. Tyto modely se používají především k předpovídání veličin (například předpovídání cen domů) nebo k přiřazování příslušnosti ke skupinám (například zařazování obrázků do kategorií).

Jaké jsou výhody pro městskou firmu (město)?

- Vhodně použité generativní nástroje umělé inteligence mají potenciál rozšířit sadu nástrojů pro veřejné služby. Funkce generování textu, kódu a obrázků mohou při opatrném použití urychlit nebo zlepšit řešení běžných úkolů.

Generativní nástroje AI mohou při použití v rámci pokynů pomoci například s:

- Vytvářením prvních návrhů dokumentů, plánů, poznámek a stručných zpráv,
- „převodem“ textu do formálního jazyka či naopak do velice srozumitelného/přístupného jazyka atd,
    - přepsání neformálního e-mailu do podoby návrhu oficiálního memoranda,

    - shrnutí technické nebo právní dokumentace do srozumitelného jazyka a zacílení shrnutí pro různé skupiny čtenářů,
    - převedení výstupů dokládajících frustraci do zdvořilé žádosti mezi odděleními…
- opakujícími se úkoly při programování a testování,
- generováním diagramů nebo jiných vysvětlujících obrázků,
- vývojem rozhraní služeb, jako jsou chatboti, s patřičným důrazem na jazykovou přístupnost a přesnost.

Všechny uvedené výhody mají nejlepší účinek, pokud je kontroluje člověk, který je:

- znalý obsahu a poskytované služby,
- uvědomuje si běžné chyby a omezení generativní umělé inteligence.

Jaká jsou rizika generativní umělé inteligence?

- Generativní umělá inteligence vyniká ve vytváření obsahu, který vypadá seriózně a uhlazeně, takže je snadné přijmout obsah generovaný umělou inteligencí za bernou minci. Bez kompetentní osoby nebo expertního systému, který by obsah zkontroloval z hlediska správnosti, má generativní AI potenciál uvádět uživatele a veřejnost v omyl. Tato rizika se zvyšují, pokud výstupy nejsou označeny jako vytvořené, navržené nebo používající umělou inteligencí. Rizika se mohou projevit také v případě, že technologie generativní AI je součástí jiného softwaru, například cloudové podnikové aplikace nebo nástrojů pro zvýšení produktivity, což nemusí být uživatelům zřejmé.

Jako zaměstnanci oict. musíme nástroje generativní AI používat opatrně, abychom se vyhnuli možným negativním důsledkům:

- Přijetí nevhodného rozhodnutí, které ovlivní jiného člověka na základě obsahu generovaného umělou inteligencí,
- vytváření nepřesných informací, ať už pro veřejnost, nebo interně,
- zahrnutí předsudků zjištěných v trénovacích datech UI, což vede k posilování nerovností ve společnosti,
- problémy s kybernetickou bezpečností nebo jiné chyby způsobené použitím kódu generovaného umělou inteligencí,
- vystavení neveřejných dat jako součásti tréninkových datových sad (zaměstnanci by měli předpokládat, že všechna data zadaná do nástroje generativní umělé inteligence se stanou součástí trénovacího souboru, pokud není explicitně pro daný nástroj uvedeno jinak),
- vydávat AI-generovaný obsah za oficiální sdělení nebo materiál bez prověření.

Doporučení pro používání generativní umělé inteligence

- Abychom v **oict.** zajistili bezpečnost systémů a dat a udrželi si důvěru našich klientů dodržujme následující doporučení pro používání generativní AI:

Co dělat

- Vyzkoušejme si to! Experimentujme s nástroji generativní AI pro navrhování a formátování textu a vysvětlujících obrázků s využitím veřejných informací.
- Spolupracujme se zkušenějšími kolegy a důkladně si vyzkoušejme různé scénáře využití, než generativní AI nasadíme do produkčního prostředí.
- Důkladně prověřme a kontrolujme fakta u veškerého obsahu generovaného AI (např. text, kód, obrázky atd.). Jsme odpovědní za to, co s pomocí generativní AI vytvoříme.
- Zveřejněme, kdy a jak byla generativní AI použita v našem výstupu. Např:
    - „Obrázek v záhlaví byl vytvořen pomocí nástroje generativní AI MidJourney“.
    - „Tento abstrakt byl vytvořen pomocí generativního nástroje AI Bard“.

- „Pro vaši informaci, k revizi tohoto e-mailu jsem použil ChatGPT“.

Co nedělat

- Zadávat veškeré informace, které nelze plně zveřejnit, do veřejných nástrojů generativní umělé inteligence (např. ChatGPT Free). K těmto informacím se mohou dostat společnosti, které tyto nástroje vytvářejí, a v některých případech i ostatní uživatelé. V případě pochybností se obraťme na vlastníka dat nebo našeho pověřence pro ochranu osobních údajů.
- Zveřejňovat výstupy generativní umělé inteligence (ať už textové, obrazové nebo kódové) bez plné kontroly a zveřejnění informací o nich.
- Požádat generativní nástroje AI, aby zjistily fakta nebo učinily rozhodnutí bez odborné lidské kontroly.
- Generovat obrázky, zvuky nebo videa, které by mohly být například zaměněny za skutečné osoby:
    - vytváření falešných fotografií nebo nahrávek konkrétního úředníka nebo člena veřejnosti („deepfake“) – a to bez ohledu na uvedení zdroje,
    - vytváření falešného snímku nebo nahrávky, které se vydávají za snímek nebo nahrávku, i když se nejedná o konkrétní osobu,
    - generování falešných „respondentů“ nebo vymyšlených profilů pro průzkumy nebo jiné výzkumy.
- Skrytě používat generativní umělou inteligenci během interakcí s kolegy nebo veřejností, například nástrojů, které mohou odposlouchávat a přepisovat konverzaci, nebo nástrojů, které zajišťují simultánní překlad bez jejich vědomí.

Další pokyny pro zaměstnance a vedení společnosti

Vedoucí zaměstnanci mají odpovědnost za podporu přiměřeného využívání generativní AI, tak aby přinášela co největší užitek pro **oict.** a naše klienty. Při hledání vhodných scénářů využití by měli se svými týmy zohlednit následující doporučení:

a. Zajistěme dodržování vnitřních směrnic upravujících nakládání s osobními údaji, citlivými informacemi a obchodním tajemstvím (např. ISMS, ochrana osobních údajů aj.).
b. Důkladně dbejme na dodržování opatření kybernetické bezpečnosti. Používání nástrojů umělé inteligence se nesmí stát nepřijatelnou kybernetickou zranitelností.
c. Sbírejme příklady efektivního užití a buďme připraveni informovat o použití na veřejném fóru, abychom zajistili transparentnost a odpovědnost a přenos znalostí.
d. Zjistěme si, zda software, který spravujeme, a jeho součásti, obsahuje generativní umělou inteligenci; informujme tým o tom, jak se používá a jaká jsou konkrétní rizika.
e. Ptejme se na využití generativní umělé inteligence v rámci poptávky na veřejné zakázky.
f. Spolupracujme s dodavateli, abychom zajistili, že AI zabudovaná do pořizovaných nástrojů bude vysvětlitelná a kontrolovatelná. Dodavatelé by měli být schopni poskytnout informace a dokumentaci o zdrojích dat, metodách a ověřování.
g. Experimentujme, řízeně a odpovědně s trénováním interních modelů na interních datech.
h. Při zvažování implementace chatbotů pro služby veřejnosti důkladně otestujme a vypracujme plán jazykového přístupu.
i. Při vytváření nebo pořizování aplikací využívajících technologii generativní umělé inteligence konzultujme s Oddělením bezpečnosti již na počátku procesu testování.

Očekávejme, že tyto nástroje i související doporučení se budou časem vyvíjet. Jsme teprve na začátku cesty práce s generativní AI. Je důležité mít přehled o jejím využití a sledovat ho, aby bylo zajištěno transparentní a odpovědné používání.

Upozornění:

Tento dokument byl vytvořen za použití generativní AI ChatGPT a DeepL.com.

Dokument vznikl jako volný překlad "San Francisco Generative AI Guidelines". viz https://www.sf.gov/reports--december-2023--san-francisco-generative-ai-guidelines.