Ochrana dat

Umělá inteligence mění způsob, jakým pracujeme, komunikujeme i vyhledáváme informace. Nástroje generativní AI, jako jsou chatboti nebo systémy pro tvorbu textu a obrázků, přinášejí ohromné možnosti – ovšem jen tehdy, pokud s nimi zacházíme obezřetně.

Modely generativní AI mohou být trénovány mimo jiné na datech, která do nich vkládají jejich uživatelé. Proto nejsou zpravidla navrženy tak, aby zajišťovaly jejich důvěrnost nebo chránily informace, které jsou s chatboty a jinými nástroji sdíleny.
Myslete na to, že většina běžně dostupných chatbotů používá data, která do AI nástrojů vkládáte, k dalšímu trénování modelů umělé inteligence. Tato data zároveň opouští vaše zařízení a zpracovávají se na serverech provozovatele.

Do běžně dostupných chatbotů proto nevkládejte citlivá data, osobní údaje nebo interní data univerzity.

Tato chráněná a interní data můžete vkládat pouze do nástroje Microsoft Copilot pod univerzitní licencí nebo ve specifických případech do lokálně provozovaných modelů.

Chráněná, interní a veřejná data

Proč je třeba myslet na ochranu dat?

Z hlediska právní i etické odpovědnosti je nutné mít na paměti, že některá data podléhají zvláštní ochraně – zejména osobní a citlivé údaje podle legislativy (např. GDPR) by neměly být vkládány do nástrojů generativní AI.

Chatboti obvykle fungují jako cloudové služby, což znamená, že zadávané informace opouští zařízení uživatele a zpracovávají se na serverech provozovatele. Vkládání jakýchkoliv osobních údajů (např. jméno, adresa, hlasová nahrávka nebo IP adresa) – do těchto systémů bez výslovného souhlasu dotčených osob je v rozporu s GDPR.

Nejde jen o přímé vkládání dat do chatbotu, ale o jakoukoli formu zpřístupnění dat AI nástrojům. Pokud AI nástroji umožníme přístup k citlivým datům, je to z pohledu rizika v zásadě stejné, jako bychom tato data přímo vložili do chatu.

Udělování širokých oprávnění AI nástrojům (přístup k diskům, mailům, repozitářům)
je velmi rizikové a nedoporučované.

Většina veřejně dostupných chatbotů a nástrojů generativní AI navíc standardně používá vložená data pro další trénování modelu. To představuje významné riziko nejen z pohledu ochrany osobních údajů, ale také při práci s informacemi chráněnými autorským právem, obchodním tajemstvím či smluvní mlčenlivostí.

Pokud například do nástroje vložíme část textu z neveřejného výzkumu, studentské práce nebo licencovaného materiálu, systém si tyto informace může „zapamatovat“ a jejich fragmenty se mohou později objevit ve výstupech poskytovaných

jiným uživatelům. Navíc k těmto datům získává přístup samotný poskytovatel AI nástroje, který je může bez vašeho vědomí dále využívat.

Tím může dojít nejen k narušení důvěrnosti ke vkládaným datům, ale i k nechtěnému úniku chráněného obsahu. Proto je nezbytné pečlivě zvažovat, co do AI systémů vkládáme, a kdykoli je to možné, pracovat s anonymizovanými či veřejně dostupnými zdroji.

I v případě, kdy provozovatel AI nástroje nepoužívá data pro trénování nebo jiné vlastní účely, stále hrozí riziko ztráty dat či jejich zneužití v důsledku bezpečnostního incidentu (např. útoku hackera, chyby v konfiguraci, úniku záloh). Infrastruktura pro fungování AI nástrojů je zpravidla velmi rozsáhlá a zahrnuje mnoho subdodavatelů, díky čemuž významně roste i riziko takového úniku.

Při sdílení dat s AI nástroji musíme tedy vždy předpokládat, že k jejich úniku dojít může. Proto je důležité dobře zvážit, zda je pro nás takové riziko přijatelné.

Co jsou to chráněná data?
Osobní údaje
Jakákoliv informace, která umožňuje přímo či nepřímo ztotožnit konkrétní fyzickou osobu, je považována za osobní údaj podle obecného nařízení o ochraně osobních údajů (GDPR) a zákona č. 110/2019 Sb., o zpracování osobních údajů.

V rámci studia mohou často studující pracovat s osobními údaji v souvislosti s dotazníkovými šetřeními, rozhovory, nahrávkami či při práci se zdravotnickou dokumentací. Kromě typických příkladů osobních údajů, jako jsou identifikační údaje člověka (např. rodná čísla/jiná identifikační čísla u cizinců, jméno a příjmení, datum narození, adresa), to mohou být i údaje, u kterých to na první pohled nemusí být tak zjevné, např. nahrávka hlasu, rentgenový snímek nebo IP adresa.

Materiály obsahující osobní údaje je možné před vložením do nástroje generativní AI anonymizovat či přeformulovat. V praxi to znamená například vymazání jmen, jakýchkoliv identifikačních čísel, adres apod.

Zvlášť chráněnou kategorii pak představují tzv. zvláštní kategorie osobních údajů, mezi které jsou řazeny informace vypovídající o rasovém či etnickém původu, politických názorech, náboženském vyznání či filozofickém přesvědčení nebo členství v odborech, genetické a biometrické údaje a údaje o zdravotním stavu či o sexuálním životě nebo sexuální orientaci.

Základní pravidlo tedy zní, že uživatelé v zásadě nesmí do chatbotů a dalších nástrojů AI vkládat osobní údaje, pokud k tomu nemají výslovný souhlas od osob, ke kterým se osobní údaje vztahují. Hrozí nejen riziko zásahu do soukromého života osob, ale také možné zneužití osobních údajů ze strany poskytovatele nástroje. Současně se také jedná o zpracování osobních údajů, které musí být v souladu s nařízením GDPR.

Některé externí nástroje generativní AI – zejména v placených verzích – nabízejí uživatelům možnost vypnout využití zadaných dat pro další trénování modelu. Ačkoli tuto funkci doporučujeme vždy aktivovat, pokud je dostupná, je potřeba si uvědomit, že i v takovém případě data opouštějí počítač uživatele a zpracovávají se na serverech poskytovatele dané služby.

Z tohoto důvodu nedoporučujeme vkládat citlivé informace nebo osobní údaje ani do chabotů, kde je toto trénování vypnuto, pokud není ochrana dat ošetřena smluvně.

Data chráněná autorským zákonem
Data, která jsou jedinečným výsledkem tvůrčí činnosti fyzické osoby (tedy člověka, nikoli právnické osoby nebo nástroje generativní AI), jsou chráněna autorským právem jako autorská díla podle zákona č. 121/2000 Sb. (autorský zákon). Autorských děl může vznikat široká škála; může jít například o dílo slovesné, fotografické, výtvarné nebo kartografické. Autorským právem je chráněn i software, resp. jeho zdrojový kód.

Je třeba mít na paměti, že již samotné vložení autorského díla do nástroje generativní AI je užitím autorského díla. K takovému užití díla musíme mít licenci (oprávnění od autora či držitele autorských práv), nebo nám musí svědčit některá z výjimek autorského zákona, např. volné užití díla pro osobní potřebu. Mimo to je však nutné také věnovat pozornost podmínkám užití daného nástroje AI včetně skutečnosti, zdali se nástroj generativní AI trénuje na uživatelských promptech.

V případě výstupů generativní AI si uživatelé musí dát pozor na to, že výstupy mohou obsahovat části autorských děl třetích osob, jejichž zveřejnění prostřednictvím akademických výstupů by mohlo představovat užití díla, ke kterému je potřeba získat od autora licenci. To platí také pro situace, kdy dochází k využití nástrojů generativní AI k úpravě díla, například obrázku, neboť dochází ke vzniku tzv. odvozeného díla.

V případě, že pak dochází k užití díla bez licence a současně takovému užití nesvědčí žádná z výjimek autorského zákona (např. citační licence), tak se jedná o zásah do práv autora původního díla. Takovému autorovi následně vzniká právo domáhat se odstranění následků a poskytnutí přiměřeného zadostiučinění za způsobenou újmu.

Jiná citlivá data
V případě, že při tvorbě akademických výstupů na univerzitě má studující vztah k nějakému subjektu (např. pracovněprávní vztah k?obchodní společnosti či státní instituci) a dochází k využití dat tohoto subjektu, tak je také nutné myslet na to, že tato data mohou být pro tento subjekt velmi cenná.

Typicky by mohlo jít o data, na která se vztahuje povinnost mlčenlivosti upravená ve smlouvě (v pracovní smlouvě, dohodě o práci konané mimo pracovní poměr, ve smlouvě o spolupráci apod.), data představující obchodní tajemství nebo data mající povahu utajované informace dle zákona č. 412/2005 Sb., o ochraně utajovaných informací a o bezpečnostní způsobilosti.

Vložením takových dat do nástrojů generativní AI by mohlo dojít k jejich vyzrazení, což by mohlo mít pro subjekt závažné důsledky. Proto v případě, kdy používáme taková data, je vždy v první řadě vhodné konzultovat zpracování dat prostřednictvím nástrojů generativní AI s tímto subjektem.

Microsoft Copilot
Microsoft Copilot je nástroj umělé inteligence, který je v podobě webového chatbota dostupný všem studujícím a zaměstnancům UK.

Přihlášení uživatelé Microsoft Copilot mají několik výhod. Jde o přístup k výkonějšímu modelu a možnost delších konverzací.

Zásadní výhoda však spočívá ve zvýšené ochraně dat. Když se přihlásíte pomocí studentského nebo zaměstnaneckého účtu, Copilot poskytuje komerční ochranu dat - to znamená, že vaše data jsou lépe a smluvně chráněna.

Jako každý nástroj, ani MS Copilot nelze považovat za stoprocentně bezpečný. Možnosti kontroly nad tím, co se s daty v celé infrastruktuře skutečně děje, jsou velmi omezené a i tato infrastruktura může být napadena a data odcizena. Na rozdíl od běžně a zdarma dostupných chatbotů u MS Copilotu ovšem ochrana dat existuje a data jsou smluvně chráněna.

Zvýšená ochrana dat se vztahuje pouze na verzi Copilotu používanou pod univerzitním účtem. Nesmí se zaměňovat s běžnou veřejně dostupnou verzí Copilotu, která je zdarma. Do MS Copilot Chatu se vždy přihlašujte univerzitním účtem.

Přihlaste se zde.

Jak se přihlásit a více informací najdete zde.

Lokální modely

Lokální modely generativní AI jsou modely, které běží přímo na zařízení uživatele – například na jeho osobním počítači nebo na zabezpečeném serveru dané instituce – a nekomunikují s externími cloudovými službami. Díky tomu nedochází k odesílání ani ukládání dat mimo prostředí uživatele, což významně snižuje riziko úniku nebo zneužití informací.

Právě z tohoto důvodu lze do těchto modelů, při dodržení odpovídajících bezpečnostních opatření, vkládat i citlivá nebo interní data (např. výsledky výzkumu, studentské práce nebo osobní údaje), což je u veřejně dostupných online nástrojů zpravidla nežádoucí či nepřípustné.

Vyhodnocení, zda je prostředí, ve kterém je provozován lokální model, skutečně bezpečné, vyžaduje odborné znalosti. Doporučeno tak je používat jen takové lokální modely a prostředí, které jsou oficiálně prověřeny a schváleny.

Pět zásad práce s nástroji AI

Do běžných chatbotů nevkládajte osobní, citlivá, interní nebo jinak chráněná data.

Vyhnete se tím riziku, že vaše data budou použita na trénování dalších modelů anebo že dojde k úniku nebo zneužití vašich dat.

Používejte nástroje Generativní AI doporučené univerzitou nebo fakultou.

Tyto nástroje mají ochranu dat smluvně ošetřenou.

Když to jde, anonymizujte svá data.

Při práci s jakýmkoliv chatbotem je nejbezpečnější data, která do nich vkládáte, anonymizovat. V praxi to znamená například smazat jména, osobní čísla, adresy.

Sdílejte s AI nástroji vždy jen nezbytné minimum dat pro daný úkol.

Je rozdíl svěřit nástroji jeden konkrétní soubor a nebo mu otevřít přístup k celému disku či úložišti.

Nesvěřujte AI nástrojům vysoce citlivá data, u kterých si v žádném případě nemůžete dovolit jejich případný únik nebo zneužití.

Žádný AI nástroj není zcela bez rizika.

Teď, když znáte základní zásady ochrany dat, se můžete směle pustit do promptování a experimentování s umělou inteligencí.

Univerzita Karlova podporuje využívání nástrojů generativní umělé inteligence studujícími za předpokladu, že je využití AI transparentní a v souladu s právním řádem ČR, předpisy Univerzity Karlovy či jejích součástí.

Podívejte se na nástroje dostupné na univerzitě, přečtěte si první rady o promptování nebo nás kontaktujte, pokud máte další otázky.

Nástroje doporučené univerzitou Rychlý start v promptování

Doporučení vedení Univerzity Karlovy k bezpečnému užívání nástrojů generativní umělé inteligence

Kontaktujte nás