28. 8. 2025

Vezměte, prosíme, na vědomí, že text článku odpovídá platné právní úpravě ke dni publikace.

Jak vytrénovat umělou inteligenci na veřejně dostupných datech? 1. díl: GDPR, anonymita a odpovědnost uživatele

Umělá inteligence (AI)[1] už dnes není pouze hračkou vývojářů a technologických nadšenců. AI nástroje a služby se rozšiřují do řady oblastí a sektorů. Některé organizace je již nasadily, jiné to alespoň testují a zkoumají – jinak řečeno, hledají vhodné „use cases“. To samozřejmě vyvolává zásadní právní otázky. Kdy je při vývoji a trénování AI nutné dodržovat GDPR? Kdo odpovídá za nakládání s osobními údaji? V prvním díle našeho trojdílného seriálu si blíže přiblížíme mantinely vztahu AI a GDPR. Dozvíte se tak například, jaký je rozdíl mezi anonymními a neanonymními modely a v jakých situacích je uživatel z pohledu ochrany osobních údajů v bezpečí, a kdy naopak ne.

Organizace s pomocí AI zajišťují různé interní procesy a činnosti nebo i komunikaci s dotčenými osobami. Typickými oblastmi využití AI jsou analýza dat, příprava marketingových kampaní, zjednodušování a automatizace interních procesů nebo identifikace a předcházení podvodům či kybernetickým útokům. V řadě případů přitom dochází ke zpracování osobních údajů[2] ve smyslu obecného nařízení o ochraně osobních údajů (GDPR)[3].

Služby a nástroje využívající AI jsou často trénovány na osobních údajích. Velmi často se jedná o veřejně dostupná data v prostředí internetu. Význam a souvislosti využití veřejně dostupných údajů na sociálních sítích pro vývoj AI modelu zdůraznilo nedávné oznámení společnosti Meta, že od června tohoto roku bude pro trénink své AI využívat informace, včetně osobních údajů, ze sociálních sítí Facebook a Instagram.[4]

Toto oznámení vzbudilo poměrně veliký zájem – nejen mezi běžnými uživateli sociálních sítí, ale i mezi dozorovými úřady pro ochranu osobních údajů z různých zemí. Například francouzský, nizozemský a belgický úřad již k tomuto záměru vyjádřily svoje pochybnosti. Naproti tomu irský dozorový úřad konstatoval, že záměr tohoto využití (zpracování) osobních údajů je z jeho pohledu podle dostupných informací v pořádku.[5] Nelze vyloučit, že postup společnosti Meta bude předmět dalšího šetření a kontroly souladu s požadavky GDPR.

Vývoj AI a soulad s GDPR

Shromažďování a využití osobních údajů pro vývoj AI modelu je bezpochyby zpracováním osobních údajů. Pokud se jedná o osobní údaje obyvatel Evropské unie, pak toto zpracování podléhá GDPR a dozorovým kompetencím národních úřadů pro ochranu dat. Subjekt, který takto AI vyvíjí, dikcí Aktu o umělé inteligenci „poskytovatel“, proto musí plnit veškeré požadavky GDPR. V kontextu využití veřejně dostupných osobních údajů pro vývoj AI se jedná především, ale ne pouze, o požadavek na existenci legitimního účelu zpracování, stanovení právního důvodu (titulu) ke zpracování a principy minimalizace a transparentnosti zpracování.

Jaká je však odpovědnost organizace (dikcí AI Actu se jedná o „zavádějící subjekt“), která AI trénovanou na osobních údajích implementuje? Může být tento zavádějící subjekt právně odpovědný

za soulad využití osobních údajů k tréninku AI modelu, který si zakoupil jako hotové řešení? A mění se role a odpovědnosti zavádějícího subjektu tehdy, pokud jsou k jeho tréninku nějakým způsobem využity osobní údaje, vůči nimž je ve vztahu správce nebo společného správce s tím, kdo AI systém fakticky vyvíjí?

Anonymní a anonymizované modely: kdy se GDPR (skoro) neuplatní

Z pohledu ochrany osobních údajů bude nejjednodušší situace u AI nástrojů anonymních či anonymizovaných. Ty samy o sobě do působnosti GDPR vůbec spadat nebudou a jejich uživatel, zavádějící subjekt, z pohledu GDPR nemůže být odpovědný za zpracování osobních údajů při jejich tréninku. To samozřejmě nevylučuje aplikovatelnost GDPR na případné následné zpracování osobních údajů při používání (nikoliv vývoji) daného nástroje – stejně jako odpovědnost poskytovatele za proces získání a anonymizace osobních údajů za účelem tréninku AI.

Pro správné pochopení upřesněme, že anonymní AI nástroj je nástroj, pro jehož vývoj vůbec osobní údaje použity nebyly. Nástroj anonymizovaný je takový nástroj, při jehož vývoji sice osobní údaje využity byly, nicméně v rámci přípravy tréninkových dat nebo tréninku samotného byly do té míry anonymizovány, že tyto údaje nelze z nástroje extrahovat („vytáhnout“) přímo (např. útokem třetí strany na daný model) ani nepřímo (užíváním nástroje, např. přes prompt uživatele). V této souvislosti je však třeba připomenout, že laťka pro skutečnou anonymizaci je v EU nastavena opravdu vysoko.[6]

Jak ověřit anonymizaci?

Jak se ale z pohledu uživatele AI postavit k situaci, kdy dodavatel tvrdí, že jeho nástroj je anonymní či anonymizovaný? Využití osobních údajů v rozporu s tvrzením dodavatele nebo v rozporu s právními předpisy (zejm. GDPR) a případná spoluodpovědnost zavádějícího subjektu představují relevantní riziko, které je vhodné přiměřeným způsobem řešit. Z našeho pohledu je ideální získat smluvní prohlášení a záruku (či alespoň prohlášení v související komunikaci s dodavatelem). To je ovšem v praxi poměrně obtížně dosažitelné, poskytovatelé taková prohlášení a záruky poskytovat nechtějí.

Vhodným krokem je proto alespoň základní vlastní posouzení anonymity/anonymizace nástroje samotným uživatelem – pro takové základní posouzení, pro které bude uživatel potřebovat informace a podklady od poskytovatele, se lze inspirovat kritérii, resp. aspekty anonymizace, které ve svém stanovisku k AI uvádí EDPB[7]. Mezi ně patří samotný návrh modelu (výběr zdrojů dat, příprava a minimalizace údajů), metodická rozhodnutí týkající se trénování, opatření týkající se výstupů nástroje, testování nástroje a jeho odolnost, související dokumentace a mnoho dalších. V neposlední řadě lze pro snížení rizik uvažovat i o dalších organizačních a technických opatřeních u uživatele, např. o nasazení různých nástrojů „nad“ daným modelem, které budou případný výskyt osobních údajů ve výstupech daného modelu identifikovat a blokovat.

Jak je tomu ale v situaci, kdy je nástroj sice správně anonymizovaný, ale při jeho vývoji/tréninku byly nezákonně použité a zpracované osobní údaje? Bude v takovém případě uživatel daného nástroje odpovědný za takovou nezákonnost? Dle EDPB i v tomto případě platí, že takový anonymizovaný nástroj a jeho další užití samy o sobě nespadají pod GDPR – uživatel tedy ve vztahu k případně nelegálnímu tréninku nástroje obavy mít nemusí.[8] Pokud v rámci svého užití po nasazení nástroje v rámci něj osobní údaje zpracovává, bude samozřejmě muset zajistit soulad takového zpracování.

Legalita zpracování osobních údajů pro trénink AI

O poznání složitější situace nastává v případě, kdy jsou osobní údaje při vývoji AI použity, ale tento

nástroj anonymní ani anonymizovaný není, a jeho poskytovatel ani nic takového netvrdí.

Samozřejmě platí, že zpracování osobních údajů při tréninku AI musí probíhat v souladu s GDPR, jeho základními zásadami a povinnostmi. Ty zahrnují zejména určení legitimního účelu, identifikaci titulu zpracování, povinnosti v oblasti transparentnosti (informování a výkon práv subjektů údajů), zapomenout nesmíme ani na relevantní principy minimalizace údajů, bezpečnosti a integrity dat, omezení uložení a celkově prokazování odpovědnosti, které jsou při tréninku nástrojů AI velmi relevantní.

Lze konstatovat, že v oblasti ochrany osobních údajů je aktuálně ve vztahu k vývoji AI patrně největší výzvou zákonnost zpracování, tedy volba správného právního rámce pro tento vývoj. Pokud bude daný nástroj AI trénován na „běžných" osobních údajích, správce a poskytovatel musí pro toto zpracování disponovat některým z právních důvodů uvedených v čl. 6 GDPR.

Souhlas ve většině případů vhodným titulem nebude, neboť jeho získání od dostatečně velké skupiny subjektů bude velmi složité, nákladné a v praxi celkově obtížně realizovatelné. Souhlas je navíc problematický zejména pro svoji odvolatelnost, když zpracování osobních údajů pro trénink AI může být svým způsobem „nevratné" v tom smyslu, že správce nebude po tréninku schopen zajistit ukončení zpracování a výmaz osobních údajů subjektu, který odvolal souhlas. Stejně tak patrně nebude často možné využít plnění smlouvy, když vhodný smluvní vztah mezi subjektem údajů a vývojářem AI většinou nebude existovat a vývojář případně neodůvodní zpracování osobních údajů pro trénink AI. I další právní tituly, jako je plnění zákonné povinnosti, budou v praxi spíše výjimečně uplatnitelné.

Právě uvedenou vylučovací metodou se zdá, že ve většině případů trénování AI na „běžných" osobních údajích budou poskytovatelé AI spoléhat na oprávněný zájem svůj nebo třetí strany (čl. 6 odst. 1 písm. f) GDPR).

Pokud tedy máme co do činění s AI nástrojem, který byl natrénován na neanonymních datech, musí jeho poskytovatel najít právní titul, který toto zpracování osobních údajů ospravedlní. To může být větší oříšek, než se na první pohled zdá. Právníci a vývojáři se proto ve většině případů nakonec propracují k institutu oprávněného zájmu. Jak přesně tento titul funguje? Za jakých předpokladů se na něj můžete před regulátorem bezpečně odvolat? Na tyto otázky odpovíme v druhém díle našeho seriálu.

Mgr. František Nonnemann,
vedoucí oddělení Compliance a oddělení Řízení operačního rizika ve společnosti Partners Banka

Mgr. Michal Nulíček, LL.M., FCIArb,
advokát a partner ROWAN LEGAL, odborník na ochranu osobních údajů a regulace

ROWAN LEGAL, advokátní kancelář s.r.o.

GEMINI Center
Na Pankráci 1683/127
140 00 Praha 4

Tel.: +420 224 216 212
Fax: +420 224 215 823
e-mail: praha@rowan.legal

---

[1] Pro účely našeho článku chápeme pojem „umělá inteligence“ dle definice v čl. 3 bodu 1) Aktu o umělé inteligenci: „Systémem AI [se rozumí] strojový systém navržený tak, aby po zavedení fungoval s různými úrovněmi autonomie a který po zavedení může vykazovat adaptabilitu a který za explicitními nebo implicitními účely z obdržených vstupů odvozuje, jak generovat výstupy, jako jsou predikce, obsah, doporučení nebo rozhodnutí, které mohou ovlivnit fyzická nebo virtuální prostředí.“

[2] Srov. výklad k pojmům „osobní údaj“ a „zpracování údajů“ v Nulíček, M. Donát, J. Nonnemann, F. Lichnovský, B. Tomíšek, J. Kovaříková, K. GDPR / Obecné nařízení o ochraně osobních údajů. 2. vydání. Wolters Kluwer, Praha: 2018.

[3] Nařízení Evropského parlamentu a Rady (EU) 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů a o zrušení směrnice 95/46/ES (obecné nařízení o ochraně osobních údajů).

[4] Např. >>> zde.

[5] Srov. >>> zde.

[6] Srov. výkladová vodítka Pracovní skupiny dle čl. 29 (předchůdce Evropského sboru pro ochranu osobních údajů) č. 5/2014 ze dne 10. dubna 2014 k technikám anonymizace, dostupné >>> zde.

[7] Viz Stanovisko EDPB 28/24 k určitým aspektům ochrany osobních údajů v souvislosti se zpracováním osobních údajů v kontextu modelů umělé inteligence, zejm. kapitola 3.2 tohoto stanoviska.

[8] Viz Stanovisko EDPB 28/24 k určitým aspektům ochrany osobních údajů v souvislosti se zpracováním osobních údajů v kontextu modelů umělé inteligence, zejm. kapitola 3.4.3 tohoto stanoviska.

Další články:

- Právo na přístup ke kamerovým záznamům: střet GDPR, informačního zákona a praxe veřejných institucí
- Postoupení pohledávky na výživné jako novinka právní úpravy účinné od 1. 1. 2026
- Jak zahájit provoz mezinárodní letecké linky do České republiky (EU): právní požadavky pro aerolinky ze třetích zemí
- Mimořádné vydržení a vývoj judikatury Nejvyššího soudu
- Preventivně-sankční funkce náhrady nemajetkové újmy za porušení osobnostních práv pohledem Ústavního soudu
- Odštěpný závod zahraniční společnosti optikou NIS2: Jak správně určit velikost podniku?
- Zápis ochranné známky bez komplikací. Klíčem k úspěchu je kvalitní předběžná rešerše
- Zneužití práva na přístup podle GDPR
- Byznys a paragrafy, díl 31. - létající pořizovatel ve světle nového stavebního zákona
- Právní povaha sítě elektronických komunikací – režim náhrady škody
- Náhrada ušlého nájemného při předčasném ukončení nájemní smlouvy na nebytové prostory