

Vytěžování veřejně dostupných autorských děl pro účely trénování umělé inteligence

Tak jako se člověk vyvíjí skrze zkušenosti a vzdělání, umělá inteligence (AI) se učí pomocí dat a algoritmů. Nutným předpokladem existence AI je proces strojového učení, v rámci kterého AI analyzuje velké množství informací a na jejich základě následně generuje nejpravděpodobnější výstupy dle zadání uživatele. Získávání tréninkových dat je však předmětem právní polemiky, jelikož tvůrci generativních AI nástrojů pro jejich učení často využívají veřejně dostupná autorská díla, k jejichž užití nemají souhlas. Jak to tedy je s vytěžováním veřejně dostupných autorských děl pro trénování umělé inteligence v českých podmínkách?

Mohou čeští vývojáři trénovat AI na veřejně dostupných datech?

Koncem roku 2022 vstoupil v platnost zákon č. [429/2022](#) Sb., kterým se mění zákon č. [121/2000](#) Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, a další související zákony, který mimo jiné upravuje automatizované analýzy textů a dat. Jde o transpozici směrnice EU 2019/790 o autorském právu a právech s ním souvisejících na jednotném digitálním trhu a o změně směrnic 96/9/ES a 2001/29/ES, přijaté dne 17. dubna 2019.

Novela zakotvila do § 39c autorského zákona nový druh zákonné licence, tzv. licenci k rozmnožování díla pro účely automatizované analýzy textů nebo dat. Ta umožňuje legálně užívat autorská díla pro účely vytěžování dat a jejich následné použití na trénování AI při splnění stanovených podmínek. Vytěžená data mohou být uchována pouze po dobu nezbytnou pro účely analýzy. Podle důvodové zprávy pokrývá zákonná licence jakékoliv účely včetně toho komerčního. AI vývojáři tak mohou pro potřeby strojového učení vytěžovat texty a data, i když by jinak byly chráněny autorským právem.

Proti vytěžování svých děl se autoři mohou efektivně bránit. Zákonná licence se nepoužije pro ty rozmnoženiny díla, u jichž si to jejich autor výslovně vyhradil. Novela tak dává autorům možnost vyjmout svá díla z režimu zákonné licence (tzv. opt-out). Tuto výhradu nestačí učinit v samotném díle nebo v textu podmínek užití díla, nýbrž musí mít strojově čitelnou formu. Sdružení pro internetový rozvoj doporučuje např. do souboru robots.txt doplnit následující formulaci:

User-agent: MachineLearning

Disallow: /

A do zápatí internetové stránky následně vložit text:

Automatické vytěžování textů a dat z této internetové stránky ve smyslu čl. 4 směrnice 2019/790/EU je bez souhlasu <majitel autorských práv> zakázáno.

Zvláštní pravidla pro vytěžování dat k vědeckým účelům a orgány veřejné správy

Speciálním ustanovením k § 39c je licence k rozmnožování díla pro účely automatizované analýzy textů nebo dat k vědeckému výzkumu podle § 39d autorského zákona. Vztahuje se na vysoké školy a instituce kulturního dědictví jako vytěžovatele a na rozdíl od prvně zmíněné licence neumožňuje autorovi vyloučit své dílo z automatizované analýzy textů nebo dat. Tyto rozmnoženiny smějí být dále uchovávány za předpokladu vhodného zabezpečení. Smluvní ujednání, která by byla v rozporu s úpravou zákonné licence podle § 39d jsou nevymahatelná.

Orgány veřejné správy musí rozlišovat, jestli automatizovanou analýzu textů nebo dat provádějí v rámci své pravomoci vyplývající z právních předpisů. Pokud ano, bude se takové užití posuzovat jako tzv. úřední licence podle § 34 písm. a) autorského zákona, nikoliv podle nové úpravy. Důvodová zpráva uvádí jako příklad činnost Českého statistického úřadu, když pomocí web-scrapingu získává údaje pro statistické účely.

Pokrok společnosti vs. ochrana autorských práv

I navzdory tomu, že výše popsaná právní úprava zakotvuje do českého právního řádu důležitou a nanejvýš aktuální problematiku, výzvy spojené s vytěžováním autorských děl pro účely trénování nástrojů AI tím zdaleka nekončí.

Zřejmě nejzásadnější dilema je, jak dosáhnout rovnováhy mezi ochranou práv autorů a technologickým pokrokem. V jistém smyslu se jedná o filozofický problém, na jedné straně se zájmem chránit práva jednotlivce a na straně druhé nejen zájem vývojářů nástrojů AI, a do určité míry i všeobecný zájem společnosti čerpat výhody z lidského pokroku. Najít odpověď, která by uspokojila všechny aktéry, však vůbec není jednoduché. Právo se proto musí snažit o vybalancování těchto dvou hodnot, tak aby obě strany mohly i nadále vykonávat svou činnost efektivně, a hlavně v souladu s právem.

Na co se v této debatě též nesmí zapomínat je, že i nejmenší omezení vytěžování textů nebo dat má potenciál významně ovlivnit další vývoj nástrojů generativní AI. Ztížení přístupu ke kvalitním datasetům může mít za následek zhoršení kvality AI nástrojů, což bude v konečném důsledku znamenat, že technologie nedosáhne nejlepších možných výsledků a naopak to může prohlubovat její nedostatky. Jinými slovy, může to zabrzdit pokrok na poli vývoje umělé inteligence. Není však pravděpodobné, že by zákonodárci jednotlivých států přistoupili k výrazným omezením na tomto poli. Uvědomují si totiž, že by tím znevýhodnili místní vývojáře oproti zahraniční konkurenci. Je však nezbytné přistupovat k této problematice s určitou mírou opatrnosti.

Závěr

Novela autorského zákona, která zavedla do českého právního řádu nové zákonné licence týkající se trénování umělé inteligence, je dle našeho názoru důležitým doplněním českého autorského práva. Zjednodušeně dává vývojářům prostor trénovat své modely umělé inteligence na veřejně dostupných datech, ale zároveň umožňuje autorům se proti takovému vytěžování svých děl vyhradit. Je nezbytné, aby právo nadále reflektovalo technologický vývoj a regulovalo jeho případné negativní společenské dopady. Tato opatření však nemohou být natolik restriktivní, že progres úplně zastaví. Problematika vytěžování veřejně dostupných autorských děl pro účely trénování nástrojů umělé inteligence je tématem, které se teprve dostává do povědomí veřejnosti. Můžeme proto očekávat, že o něm ještě uslyšíme.

JUDr. Ing. Jaroslav Menčík, LL.M., Ph.D.,
partner

Další články:

- [Návrh nového zákona o digitální ekonomice](#)
- [Byznys a paragrafy, díl 30.: Jednání za s.r.o. – zápis jednatele do obchodního rejstříku](#)
- [Předběžné opatření a další instituty k ochraně věřitelů při přeměnách](#)
- [Promlčení zápůjčky bez určení splatnosti v judikatuře Nejvyššího soudu](#)
- [Rozvod s mezinárodním prvkem a související otázky péče o děti a výživného](#)
- [Příkaz a příkaz na místě v přestupkovém řízení vedeném orgány inspekce práce](#)
- [Letiště a letecké stavby](#)
- [Problematické aspekty změn v úpravě odpovědnosti za škodu způsobenou vadou výrobku](#)
- [Autonomní mobilita optikou české legislativy. Kdy se na silnicích dočkáme vozidel bez řidiče?](#)
- [Rodičovská odpovědnost po novele občanského zákoníku: Jak nové principy rovnosti, spolupráce a ochrany dítěte mění praxi soudů a rodin](#)
- [Lhůta a povinnosti při vyrozumění oznamovatele o výsledku posouzení oznámení](#)