

# Právní aspekty vytěžování veřejně dostupných autorských děl pro účely trénování nástrojů umělé inteligence

Mgr. Ing. Jaroslav Menčík, LL.M.

*MAVERICKS advokáti s.r.o.*

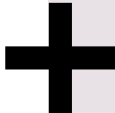
České právo a informační technologie 2023 

14. 9. 2023, PrF MUNI



# Generativní AI

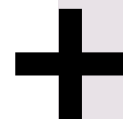


- boom nových technologií v oblasti umělé inteligence, neuronových sítí a strojového učení (**machine learning**)
  - generativní AI: nástroje schopné dle zadání vytvářet originální obsah například v podobě textu, obrazů, zvuků či kódu počítačového programu
    - LLM (ChatGPT)
    - Midjourney, Stable Diffusion, DALL-E
    - Copilot
  - nutným předpokladem je proces **strojového učení** – analýza rozsáhlého množství dat (**dataset**), které AI umožní na základě statistiky generovat nejpravděpodobnější výstupy dle zadání uživatele
  - zdroje datasetů se různí
  - tvůrci generativní AI pro učení využívají veřejně (on-line) dostupná autorská díla = **vytěžování (text & data mining, TDM)**
- 

# AI versus autoři

---

- vytěžování veřejně dostupných autorských děl za účelem učení AI nástrojů je předmětem kritiky ze strany autorů, kteří argumentují **porušováním autorských práv & poškozováním ekonomických zájmů**
  - užití díla bez souhlasu, bez kompenzace, někdo jiný má ekonomický prospěch
- ve světě řada soudních **sporů** iniciovaných autory z důvodu neoprávněného využívání autorských děl
  - žaloba proti spol. DeviantArt, Midjourney, Stable Diffusion, DALL-E (miliardy textů a obrazů)
  - žaloby proti spol. Meta, OpenAI (obrovské množství knih)
  - žaloba proti spol. Amazon (záběry z webových kamer)
- žalobci jsou jak autoři (spisovatelé, umělci), tak společnosti (Getty Images)
- cíl: odškodnění / zničit AI nástroje
- OpenAI v roce 2022 přestala poskytovat informace o použitých datasetech



# České autorské právo (před 2023)

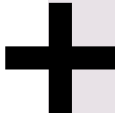
---

- zveřejnění autorského díla neznamená, že by jej třetí osoby mohly jakkoliv užívat
- naopak AZ v zásadě zakazuje užít autorské dílo bez **souhlasu autora** (licence)
- bez souhlasu autora lze dílo užít, pokud se uplatní některá ze **zákonných licencí** (např. zpravodajská, úřední, citace, digitální výuka)
- užití díla bez souhlasu autora nebo zákonné licence = porušení autorského práva
- vytěžování veřejně dostupných autorských děl = užití autorských děl



# Směrnice EU



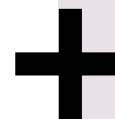
- Směrnice Evropského parlamentu a Rady (EU) 2019/790 ze dne 17. dubna 2019 **o autorském právu a právech s ním souvisejících na jednotném digitálním trhu** a o změně směrnic 96/9/ES a 2001/29/ES
  - umožnila využívat texty a data pro účely data miningu ve vztahu k veřejně přístupným dílům, i když jsou chráněná autorským právem
  - cíl: podpora výzkumu a vývoje nových technologií (inovace)
  - článek 3 – "vytěžování textů a dat pro účely vědeckého výzkumu"
  - článek 4 – "výjimky a omezení pro vytěžování textů a dat"
  - lhůta pro implementaci 7.6.2021
  - směrnice – nemá přímý účinek v členských státech
- 

# Směrnice EU – čl. 3



## Vytěžování textů a dat pro účely vědeckého výzkumu

1. Členské státy stanoví výjimku z práv stanovených v čl. 5 písm. a) a čl. 7 odst. 1 směrnice 96/9//ES, v článku 2 směrnice 2001/29/ES a v čl. 15 odst. 1 této směrnice s ohledem na **rozmnoženiny a extrakce** zhotovené **výzkumnými organizacemi a institucemi kulturního dědictví** za účelem vytěžování textů a dat z děl nebo jiných předmětů ochrany, k nimž **mají zákonný přístup**, pro účely vědeckého výzkumu.
2. Rozmnoženiny a extrakce děl nebo jiných předmětů ochrany zhotovené v souladu s odstavcem 1 musí být **uloženy s vhodnou úrovní zabezpečení** a smějí být uchovávány pro účely vědeckého výzkumu, včetně ověření výsledků výzkumu.
3. Nositelé práv mohou uplatňovat opatření k zajištění bezpečnosti a integrity sítí a databází, v nichž jsou díla nebo jiné předměty ochrany uloženy. Tato opatření nepřekročí rámec toho, co je nezbytné k dosažení tohoto cíle.
4. Členské státy podporují nositele práv, výzkumné organizace a instituce kulturního dědictví v tom, aby k uplatňování povinností uvedených v odstavci 2 a opatření uvedených v odstavci 3 vymezily společně dohodnuté osvědčené postupy.

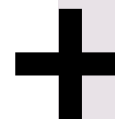


# Směrnice EU – čl. 4

---

## Výjimky a omezení pro vytěžování textů a dat

1. Členské státy stanoví výjimku z práv nebo omezení práv stanovených v čl. 5 písm. a) a v čl. 7 odst. 1 směrnice 96/9/ES, v článku 2 směrnice 2001/29/ES, v čl. 4 odst. 1 písm. a) a b) směrnice 2009/24/ES a v čl. 15 odst. 1 této směrnice pro **rozmnoženiny a extrakce** zhotovené za účelem **vytěžování textů a dat** z děl nebo jiných předmětů ochrany, k nimž mají **zákonný přístup**.
2. Rozmnoženiny a extrakce zhotovené podle odstavce 1 smějí být uchovávány **pouze po dobu nezbytnou** pro účely vytěžování textů a dat.
3. Výjimka nebo omezení podle odstavce 1 se použijí **pod podmínkou, že si nositelé práv výslovně nevyhradili užití děl** a jiných předmětů ochrany podle uvedeného odstavce **vhodným způsobem**, jako jsou **strojově čitelné prostředky** v případě obsahu zpřístupněného veřejnosti online.
4. Tímto článkem není dotčeno uplatňování článku 3 této směrnice.



# Implementace v ČR

---

- novela AZ č. 429/2022 Sb. s účinností od 5.1.2023
- fakticky úplné provedení směrnice EU
- zavedla zákonnou licenci k užití autorského díla za účelem "**automatizované analýzy** (= vytěžování) **textů nebo dat**"
  - pojem „vytěžování“ se již v autorském zákoně vyskytuje, a to v části upravující zvláštní právo pořizovatele databáze, a vyskytoval by se tak na více místech, ale v jiných kontextech
- Podstata novely: Za splnění stanovených podmínek budou moci oprávněné subjekty užívat chráněná díla či předměty ochrany pro účely automatizované analýzy textů nebo dat bez nutnosti uzavírat licenční smlouvy s nositeli práv.

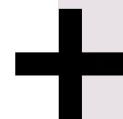




# Zákonná licence - § 39c AZ

---

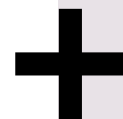
- § 39c/1 AZ: *"Do práva autorského nezasahuje ten, kdo zhotoví rozmnoženinu díla za účelem automatizované analýzy textů nebo dat v digitální podobě, prováděné za účelem získání informací, zahrnujících mimo jiné vzory, tendence a souvztažnosti; takto zhotovenou rozmnoženinu je oprávněn uchovat pouze po dobu nezbytnou pro účely této automatizované analýzy textů nebo dat."*
- DZ: jde o výjimku z výlučných práv pro užití autorských děl a jiných předmětů ochrany za účelem automatizované analýzy textů nebo dat **pro jakýkoli účel a jakýmkoli subjektem**



# Podmínky zákonné licence

---

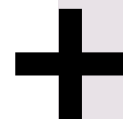
1. zhotovení rozmnoženiny autorského díla
  2. za účelem automatizované analýzy textů nebo dat
  3. v digitální podobě
  4. prováděné za účelem získání informací, zahrnujících mimo jiné vzory, tendence a souvztažnosti
  5. rozmnoženinu lze uchovat pouze po dobu nezbytnou pro účely této automatizované analýzy textů nebo dat
- *zákonná licence pokrývá jakékoliv účely včetně komerčních (viz důvodová zpráva)*
- na základě zákonné licence AI nástroje mohou pro účely svého strojového učení vytěžovat texty a data, i když jsou chráněné autorským právem



# Výjimka ze zákonné licence

---

- § 39c/2 AZ: Ustanovení odstavce 1 se nepoužije pro rozmnoženiny díla, jehož autor si užití podle odstavce 1 **výslovně vyhradil** vhodným způsobem; v případě díla zpřístupněného podle § 18 odst. 2 strojově čitelnými prostředky.
- zákon dává autorům volbu, zda se rozhodnou svá díla z režimu zákonné licence pro vytěžování vyjmout (**opt-out**)
- výhrada platí od učinění směrem do budoucna
- jaký způsob výhrady je vhodný?
  - ujednání ve smlouvě, metadata
- díla zveřejněná digitálně (on-line) – povinnost **strojové čitelnosti** výhrady
  - výhrada v textu podmínek užití webu nebude stačit
  - Sdružení pro internetový rozvoj: návrh standardizované komunikace skrze úpravu souboru robots.txt



# Návrh SPIR

## Co SPIR navrhuje:

Doplnění do souboru robots.txt:

*User-agent: MachineLearning*

*Disallow: /*

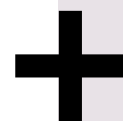
Vzorový text do patičky internetové stránky:

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


# Zákonná licence pro vědecký výzkum - § 39d AZ

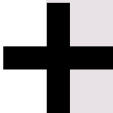
---

- § 39d: Do práva autorského nezasahuje
- a) **vysoká škola**, která jako součást své činnosti provádí vědecký výzkum (...), nebo
- b) **instituce kulturního dědictví**,
- zhotoví-li pro účely vědeckého výzkumu rozmnoženinu díla za účelem automatizované analýzy textů nebo dat v digitální podobě, prováděné za účelem získání informací, zahrnujících mimo jiné vzory, tendence a souvztažnosti; takto zhotovenou rozmnoženinu je povinna uložit s **vhodnou úrovní zabezpečení** a může ji **uchovávat** pro účely vědeckého výzkumu, včetně ověření výsledků výzkumu.



# Zákonná licence pro vědecký výzkum – pokr.

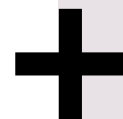


- omezení na definované subjekty (VŠ, instituce kulturního dědictví)
  - omezení účelem – vědecký výzkum
  - autoři nemají možnost výhrady
  - možnost uchovat rozmnoženinu po delší dobu - pro účely vědeckého výzkumu včetně ověření výsledků výzkumu
  - povinnost uložit rozmnoženinu s vhodnou úrovní zabezpečení
  - smluvní ujednání, která by byla v rozporu s úpravou zákonné licence podle § 39d AZ, jsou nevymahatelná (viz směrnice)
- 

# Vytěžování orgány veřejné správy

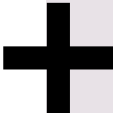
---

- provádění automatizované analýzy textů nebo dat orgány veřejné správy
- nutno rozlišovat, zda autorskoprávně chráněný obsah využívají **v rámci své pravomoci a působnosti stanovené zákonem, nebo nikoli**
- pokud v rámci plnění povinností vyplývajících z právních předpisů, pak takové úkony spadají pod úpravu § 34 písm. a) AZ - tzv. **úřední licenci**, tj. zákonnou licenci pro užití autorských děl a jiných předmětů ochrany pro úřední účely, např. pro účely veřejné bezpečnosti, pro soudní nebo správní řízení nebo pro jiné úřední účely
- DZ: výkon státní statistické služby **Českým statistickým úřadem** v rámci jeho působnosti stanovené národními a evropskými právními předpisy (např. získávání údajů pro statistické účely prostřednictvím web-scrapingu)



# Práva související s právem autorským



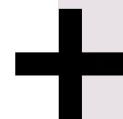
- zákonné výjimky pro automatizované analýzy textů nebo dat podle § 39c a § 39d AZ platí **obdobně pro nositele práv souvisejících s právem autorským**, a to konkrétně pro:
    - výkonného umělce a jeho výkony (§ 74)
    - výrobce zvukového záznamu a jeho záznam (§ 78)
    - výrobce zvukově obrazového záznamu a jeho záznam (§ 82)
    - vysílatele a jeho vysílání (§ 86)
    - vydavatele tiskových publikací (§ 87b odst. 7)
    - přiměřeně pro pořizovatele databáze (§ 94)
  - Pro případy počítačových programů platí pouze výjimka pro automatizovanou analýzu textů nebo dat podle § 39c
- 



# Diskuse

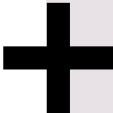
---

- snaha o vybalancování ochrany práv autorů vs technologického pokroku
  - mining je ok, ale s možností **opt-out**: nedostatek nebo správný přístup?
- právo zamezit zhotovení rozmnoženiny díla za účelem vytěžování textů nebo dat má **potenciál významně ovlivnit vývoj** nástrojů generativní AI
- **ztížení přístupu ke kvalitním datasetům** = zhoršení kvality AI nástrojů, technologie nedosáhne nejlepších možných výsledků – brzdění pokroku?
- legislativa vyjma vědeckého výzkumu z opt-out principu **nevylučuje jiné společensky významné činnosti** – např. novinářská činnost
- zákonné licence jsou ke **zhotovení rozmnoženiny**, ale nikoliv k šíření či sdělování veřejnosti
  - co když AI nástroj zčásti reprodukuje originální dílo?
  - vytěžení po právu, ale výstup porušuje autorské právo?



# Diskuse - pokračování



- podmínka existence **zákoného přístupu** k dílům (směrnice x AZ)
    - wiki leaks?
  - uložení dat pouze na **nezbytnou dobu** – nelze zpětně validovat
  - **neurčité pojmy**
    - automatizované analýzy textů nebo dat v digitální podobě
    - uchovávat rozmnoženinu pouze po dobu nezbytnou pro účely automatizované analýzy textů nebo dat
    - výhrada vhodným způsobem
    - výhrada strojově čitelnými prostředky
    - povinnost uložit rozmnoženinu s vhodnou úrovní zabezpečení
  - **interpretace soudy**
    - Soudní dvůr EU – interpretace práva EU v řízení o předběžné otázce
  - budeme svědky obdobných sporů jako v USA?
- 

# Děkuji za pozornost!

